UNDERSTANDING NUAGE-MITOCHONDRIAL COUPLING IN *DROSOPHILA* PIRNA BIOGENESIS

A Dissertation Presented

By

TIANFANG GE

Submitted to the Faculty of the
University of Massachusetts Graduate School of Biomedical Sciences, Worcester
In partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 15, 2018

INTERDISCIPLINARY GRADUATE PROGRAM

DEDICATION

To the apple of my eye—

Crystal

for being heirs with me

of the gracious gift of life

ACKNOWLEDGEMENTS

My gratitude goes first to my advisor Phil Zamore, for taking me into his lab after I had a “faulty start”, for providing me with unlimited freedom to pursue scientific questions that interest me, for supporting me and my family through big life events with zero stress or pressure—the birth and raising of our two kids. Most importantly, for his professionalism as a scientist and his diligent teaching that “know the reason why each component is in your buffer,” “controls can’t be too many,” and “it’s never too late to start the right experiment”. It is the training I got from Phil’s lab that gives me the confidence to be a scientist. Two things Phil said to me that I will not forget, “I want you to do good science and have a balanced life”, and “I support you all the same no matter which career path you take.”

I would like to thank Wei Wang, my fellow graduate student and collaborator on the Armi project, for being very kind and helpful since my first day in the lab. Being my teacher in bioinformatics, she answered all my naïve questions with patience, and challenged me with questions like “what would this analysis tell you?” Wei’s hardworking spirit and constant strive to become better always inspire me. I want to thank her particularly for continually helping me with computational analysis even after her graduation in 2015. For almost two years, while working full-time for a private company, she often made time out of her busy schedule to work on my problems and answer my questions. This project

would not be possible without her commitment to the collaboration from start to end. Wei, Thank you!

I would like to thank Zhiping Weng, a supportive “big sister” that I am fortunate to get acquainted with. While being the head of a big research group and the mom of two growing kids, she took her precious evening time every other Thursday to accommodate Wei’s working schedule and brainstormed with us on the project—not just once or twice, but for well over a year! I still don’t fully understand why she did that, but I guess it is because of her passion for science. Her commitment to the success of her trainees—“If I can make it, you can too.”—will certainly leave a long-lasting influence on me.

I would like to thank Cindy Tipping, for being a wonderful helper on our CRISPR project, trusting my leadership and faithfully performed our screening experiments over and over: putting hundreds of flies into tubes individually, making lysate and setting up PCR for each of them, running large-scale agarose gels…just to list a few. Our paper would not be possible without her patiently and meticulously working out the details. When I’m tired of the tedious procedures and tried to cut corners, Cindy will remind me gently that “we should do it the right way, it will save us time in the end,” although it would be her who end up doing the work! I really enjoyed working with Cindy for the past few years and learned a lot from her.

I would like to thank my TRAC for encouraging and supporting me over the years. They are the scientists that I look up to, but they are all amiable, respectful to my ideas, and willing to help. Thanks to Dave Lambright and Bill Theurkauf for many good ideas, to Allan Jacobson for pushing me to think and present my data as a story rather than many loose ends, and to my Chair Melissa Moore, for sticking to her role even after joining Moderna and could not be at the school all the time. She would always schedule one-on-one meetings with me when there is a need to. Thanks to my external examiner Julius Brennecke for being here for my defense, and all the exciting and thought-provoking discussions on piRNAs that we had.

My thanks go to all the Zamore lab members for making the lab a pleasant and fun place to work. I would like to thank our “lab mom”, Gwen Farley, for running the lab smoothly, and together with Phil, “enforcing” a lab culture that is fair, constructive, and altruistic. To our “lab secretary”, Tiffanie Gardner, for scheduling all my TRAC meetings and booking my travels for more than six years! I will miss her neatly prepared “travel folder” waiting for me in the Zamore lab mailbox. To Alicia Boucher, for wearing many hats and doing lab jobs that are often unnoticed or unappreciated. Thank you, Alicia! To Elif Cenik, Cha San Koh, Bo Han and Liang Meng Wee for helping me with experiments or analyses. Tracey Lincoln, Xin Li and Chengjian Li for teaching me how to think. Tim Chang, Sam Jolly, Amena Arif, Yongjin Lee and Cansu Colpan for sharing the struggles

of a graduate student. The post-doc team, Xuan Wu, Paul Albosta, Deniz Ozata and Ildar Gainetdinov, for being good friends and great resources. Kaycee Quarles, Jen Broderick and Wes Salomon for advices on career development.

Outside of the lab, I would like to thank Donggi Paik (Neal Silverman lab) and Kate Koles (Vivian Budnik lab) for teaching me fly genetics in my early graduate school years. Thanks to Vivian Budnik who rescued me from giving up on basic science research. Thanks to Siu Kai Kong and Yuan-Yuan Ho, my master's degree and undergraduate degree mentors, respectively, without whose help and guidance I would not be able to get into a graduate school in the US. I would also like to thank my three Christian friends at UMass, Huaiqing Chen, Cheng Chang and Haiwei Mou, for their spiritual support and scientific insights. I will truly miss our lunch time sharing. Thanks to all our brothers and sisters at the Chinese Gospel Church for their love, support, guidance and advice.

My gratitude also goes to my ever-supportive parents and my four 90-year-old grandparents, who are always understanding and always have confidence in me; in particular to my mom, for all the difficulties she went through in raising this thin and weak kid. I remember the days of my elementary school when she biked me across the city every weekend to treat my increasingly heavy nearsightedness. Her efforts eventually paid off: my prescription has not changed since the 9th grade. Now whenever I work long hours in front of the computer and feel strained in my eyes, I know that this would have been so much worse

without mom's perseverance. I would like to thank my mother-in-law for coming here several times to help with child-rearing. It is not fun to take a 15-hour flight followed by 4 hours of road trip, but she never hesitated when we needed her help the most.

When we entered the graduate school, Dean Tony Carruthers said, "We don't give out PhDs, you have to earn it." After nine years, I can truly attest to that. However, I could not have earned it without my dear wife Crystal. She keeps me going not only with every meal she cooks and every chore she does around the house, but also with advices, corrections and perspectives that I sometimes disagree with in the beginning, but turns out to be extremely useful. Leaving her big family and the warm weather in Hong Kong and spending eight "unbearable" winters in Worcester, she is always ready to sacrifice for me to get this PhD. In my first few *hard-working* months in the Zamore lab, she would cook a nutritious meal after work, drive back to school and have dinner with me; in my last few *hard-working* months here, she took the kids back China so that I could focus on writing this thesis. Last but not least, I thank her for bearing, nurturing and raising our two young children, Samuel and Rachel, and taking such good care of them. She is the best mom ever.

Lastly, to my two little angels, Samuel and Rachel, for putting a smile on my face every day. You are truly gifts from above and I'm eternally grateful for that!

ABSTRACT

In the *Drosophila* ovary, PIWI-interacting RNAs (piRNAs) suppress transposon expression, ensuring female fertility. PIWI proteins Aub and Ago3, loaded with ping-pong piRNAs and reside in perinuclear nuage granules, engage in reciprocal transposon transcript cleavage termed the ping-pong cycle. The other PIWI protein Piwi, loaded with phased piRNAs and resides in the nucleus, silences transposon transcriptionally. Ping-pong piRNAs are made through the ping-pong amplification loop by Aub and Ago3, whereas phased piRNAs are made through consecutive endonucleolytic cleavages that spread in 5′-to-3′ direction, presumably by Zucchini (Zuc), an endonuclease resides on the surface of mitochondria. The ping-pong and phasing biogenesis pathways are coupled genetically and molecularly. However, it is not known how such coupling is achieved at the mechanistic level.

We found that nuage and mitochondria are physically separate under the confocal and electron microscopy. Zuc interacts with other known phasing factors on the mitochondrial surface, including an RNA-binding ATPase Armitage (Armi). Relying on its ATPase activity, Armi avoids binding to genic mRNAs, instead binds to piRNA precursors engaged in ping-pong or phasing, and localizes to both nuage and mitochondria. Armi localization is dynamically regulated by the ping-pong and phasing pathways. In *armi* loss-of-function mutants, ping-pong still operates, but phasing is disrupted. Therefore, the coupling between ping-pong

and phasing pathways can be explained by Armi shuttling between nuage and mitochondria. An Armi ATPase mutant retains the interactions with piRNA biogenesis factors and piRNA precursors, but is insufficient to support phasing, suggesting an additional role of the Armi ATPase activity in ribonucleoprotein complex (RNP) remodeling.

Our study suggests that the dynamic distribution of an RNA-binding ATPase serves to transfer piRNA precursors between distinct subcellular compartments. It furthers our understanding of the complex coordination between piRNA biogenesis pathways and may serve to guide future studies on the mitochondrial phase of piRNA biogenesis. A few important questions remain to be answered: what interactions or conformational changes need to happen on Armi for it to anchor at nuage or mitochondria? How does Armi remodel the phasing RNP? Why are ping-pong and phasing machineries separated, and why does phasing happen on mitochondria?

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

COPYRIGHT NOTICE

ABBREVIATIONS

ANP: adenosine-5′-(β,γ-imido)triphosphate

Armi: Armitage

ATP: adenosine triphosphate

ATP5A: ATP synthase complex V alpha subunit

Aub: Aubergine

C. elegans: *Caenorhabditis elegans* (round worm)

CRISPR: clustered regular interspersed short palindromic repeat

D. melanogaster: *Drosophila melanogaster* (fruit fly)

DAPI: 4′,6-diamidino-2-phenylindole

DSB: double-stranded DNA break

DTME: dithiobismaleimidoethane

FISH: fluorescent *in situ* hybridization

GOI: gene of interest

GPAT: glycerol-3-phosphate *O*-acetyltransferase

H3K9me3: trimethylated histone H3 lysine 9

HR: homologous recombination

iBAQ: Intensity Based Absolute Quantification

IP: immunoprecipitation

kbp: kilobase pairs

l-Deg: long (regular) degradome-seq

Lig4: ligase 4

LINE: long interspersed nuclear element

LTR: long terminal repeat

miRNA: microRNA

MMEJ: microhomology-dependent end joining

MS: mass spectrometry

NHEJ: non-homologous end joining

NMD: nonsense mediated decay

nt: nucleotide

OSC: ovarian somatic cell

PA: phosphatidic acid

PAR-CLIP: Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation

PFA: paraformaldehyde

piRNA: PIWI-interacting RNA

PLD: phospholipase D

ppm: parts per million genome mappers

RITS: RNA-induced transcriptional gene silencing

RNAi: RNA interference

RNP: ribonucleoprotein particle

s-Deg: short degradome-seq

sgRNA: single-guide RNA

SINE: short interspersed nuclear element

siRNA: small interfering RNA

ssRNA: single-strand RNA

TE: transposable element

TEM: transmission electron microscopy

UTR: untranslated region

w: the *white* gene of *Drosophila melanogaster*

Zuc: Zucchini

CHAPTER I: INTRODUCTION

TRANSPOSABLE ELEMENTS AND HOST CONTROL MECHANISMS

Approximately 50%–60% of the human genome is composed of interspersed repeats. In the much smaller fly genome, transposable elements (TE) take about 4–9% of the space (Huang et al., 2012). While most TEs in the genome are truncated or decayed and no longer active, in the human genome about 0.01% of all LINE insertions (~100 copies) are still active (Friedli and Trono, 2015). Transposons use virus-like replication strategies without the extracellular phase of the viral life cycle. DNA transposons use a cut-and-paste mechanism that takes advantage of the host DNA replication or DNA repair mechanisms to increase copy number (Feschotte and Pritham, 2007). In contrast, retrotransposons use a copy-and-paste mechanism to propagate in the host genome. Retrotransposons can be further categorized into LTR (long terminal repeats) and non-LTR retrotransposons, based on the presence or absence of long terminal repeats at the retrotransposon genome ends. LTR retrotransposons are also called endogenous retroviruses (ERV), which carry prototypical *gag*, *pol* and *env* coding sequences in between the LTRs. LINE and SINE are two major groups of non-LTR retrotransposons. In *Drosophila*, the majority of TE sequences are derived from LTR retrotransposons (Kaminker et al., 2002).

Because of the wide spread of TEs in host genomes, it is no surprise that some of them can affect host gene expression. They can alter gene transcription by serving as promoters or enhancers, change mRNA splicing or poly-

adenylation, nucleate repressive heterochromatin, or alter coding sequences by disrupting exons (Friedli and Trono, 2015). Activated TEs are also associated with erratic homologous recombination sites (Zamudio et al., 2015). While some controlled TE activation is likely beneficial in host adaptation by generating genetic diversity, uncontrolled TE activity is detrimental to the animal germline: in flies, embryos that cannot silence *I-element* or *P-element* TEs develop into sterile adults with under-developed gonads, a phenomenon termed hybrid dysgenesis (Khurana et al., 2011; Brennecke et al., 2008).

In order to keep TEs in check, hosts have developed two main TE control strategies: transcriptional silencing and post-transcriptional silencing. Transcription silencing through heterochromatin formation at TE loci can be accomplished through DNA methylation or histone modification. This is achieved in fungi, worms, plants and animals, in part through small RNA-mediated processes. In fungi and plants, small interfering RNAs (siRNAs) act via the RNA interference (RNAi) pathway to silence transposons and other types of repetitive DNA. A representative example is the RNAi-mediated packaging of repetitive sequences into heterochromatin in the fission yeast, *Schizosaccharomyces pombe*. siRNAs bound to *S. pombe* Ago1 guide the “RITS” complex to nascent transcripts from transposon-like repeats near the centromere, where the complex recruits proteins to establish repressive heterochromatin (Castel and Martienssen, 2013). In animals, particularly in the germline, piRNAs replace

siRNAs to fulfill TE transcription repression, which is described in detail below. In tetrapod animals, the KRAB-zinc finger family of proteins with hundreds of members, use the sequence-specific zinc finger domain to target specific TEs, while the KRAB domain serves as a scaffold to assemble heterochromatin-inducing complexes, such as histone methyltransferase and DNA methyltransferase (Ecco et al., 2017).

Post-transcriptional silencing of TE transcripts or cDNA intermediates may involve siRNA- (Sigova et al., 2004) or piRNA-mediated target cleavage (see below), or small RNA-independent mechanisms, such as transposition repression by the APOBEC cytidine deaminase, the 3′-repair exonuclease 1 (Trex1), or the DNA repair machinery (Friedli and Trono, 2015; Levin and Moran, 2011).

piRNA IN TRANSPOSON SILENCING

Animals use PIWI-interacting RNAs (piRNAs), a class of small silencing RNAs distinct from siRNAs, to silence germline transposons and ensure fertility (Han and Zamore, 2014). Like siRNAs and the mRNA-regulating microRNAs (miRNAs), piRNAs direct Argonaute proteins to silence complementary nucleic acid targets. Unlike siRNAs and miRNAs, piRNAs guide a specialized sub-class of Argonautes, the PIWI proteins, which are found exclusively in animals and nearly always in the germline or germline-related cells.

In *Drosophila*, 23–29 nt long piRNAs bind three different PIWI proteins: P-element-induced wimpy testes (Piwi), Aubergine (Aub) and Argonaute3 (Ago3). Aub and Ago3 localize to cytoplasmic “nuage” granules (see below) and act to silence transposons by destroying their RNA transcripts. Aub, guided by antisense piRNAs, targets and cleaves sense TE transcripts. The cleavage products mature into sense piRNAs that load into Ago3, which in turn cleave antisense piRNA precursor transcripts to produce antisense, Aub-bound piRNAs that are identical to the ones that started the cycle. These reciprocal cleavage events, cleaving TE transcripts while generating new piRNAs, is termed “ping-pong”.

In contrast, Piwi resides in the nucleus, where it represses transposon transcription by recruiting the zinc finger protein Asterix (Muerdter et al., 2013)/Gtsf1 (Jin et al., 2013) and the scaffolding protein Panoramix (Yu et al., 2015)/Silencio (Sienski et al., 2015), which in turn recruit the histone methyltransferase Eggless/SetDB1 to install H3K9me3 (trimethylation of the lysine 9 of histone 3) marks on TE loci (McCue and Slotkin, 2012; Le Thomas et al., 2013; Rozhkov et al., 2013). These H3K9me3 marks are recognized by heterochromatin protein 1 (HP1, officially named Su(var)205), generating chromatin that is refractory to transcription. It is thought that Piwi uses antisense piRNA guides to recognize nascent sense transposon transcripts. Supporting

that, RNA is found to be required for Piwi to co-immunoprecipitate with proteins known to bind nascent RNAs (Le Thomas et al., 2013).

In flies, in the absence of DNA methylation, Piwi-mediated histone modification becomes the major TE silencing mechanism in the germline (Wang et al., 2015; Senti et al., 2015). In mice, DNA methylation silences transposons in addition to histone modification, which also depends, at least in part, on piRNA (Aravin et al., 2008; Carmell et al., 2007). Based on the observations in flies (*Drosophila melanogaster*) and mice, it was thought that piRNA function is restricted to the germline (Girard and Hannon, 2008). Only recently it was found that piRNAs exist and likely repress transposons in somatic cells of most surveyed arthropods, including *Drosophila virilis*, leaving *D. melanogaster* an evolutionary outlier (Lewis et al., 2018).

CYTOPLASMIC PIRNA BIOGENESIS PATHWAYS

This thesis focuses on the cytoplasmic processing of piRNA precursor transcripts into piRNAs in the *Drosophila* germline, for which two coupled pathways are involved: the upstream ping-pong and the downstream phasing (Han et al., 2015a; Mohn et al., 2015).

Ping-pong

In the ping-pong cycle, reciprocal cleavages of sense transposon message or antisense piRNA cluster transcripts by Aub or Ago3 generates new piRNAs for each other (Brennecke et al., 2007; Gunawardane et al., 2007) (Figure 1.1). The DEAD-box protein Vasa delivers the Aub cleavage product to Ago3 (Xiol et al., 2014), and the cochaperone Shutdown presumably helps to load the ping-pong cleavage products into Aub or Ago3 (Olivieri et al., 2012; Preall et al., 2012). Tudor-domain-containing proteins bind symmetric di-methylated Arginine residues (sDMAs) on Aub or Ago3 to assemble a scaffold for efficient ping-pong amplification (Nishida et al., 2009). Aub, Vas, Shutdown and the RNA-binding ATPase Spn-E are indispensable for the ping-pong machinery to operate (as defined by a statistically significant enrichment of piRNAs on opposite genomic strands with 10 nt overlap between their 5′-ends) (Han et al., 2015a). Many other factors (Ago3, Krimp, Qin, Tudor, BoYb, Tejas, Tapas and Vret) either boost the efficiency of the cycle or ensure an antisense bias of the produced piRNAs (Figure 1.1). By converting the sense transposon transcript into a piRNA that directs the production of its cognate antisense “silencing” piRNA, the pong-pong cycle amplifies the piRNA pool targeting active transposons. However, since every 5′-end of a piRNA is specified by a pre-existing piRNA, ping-pong offers limited ability to make new piRNAs.

Figure 1.1

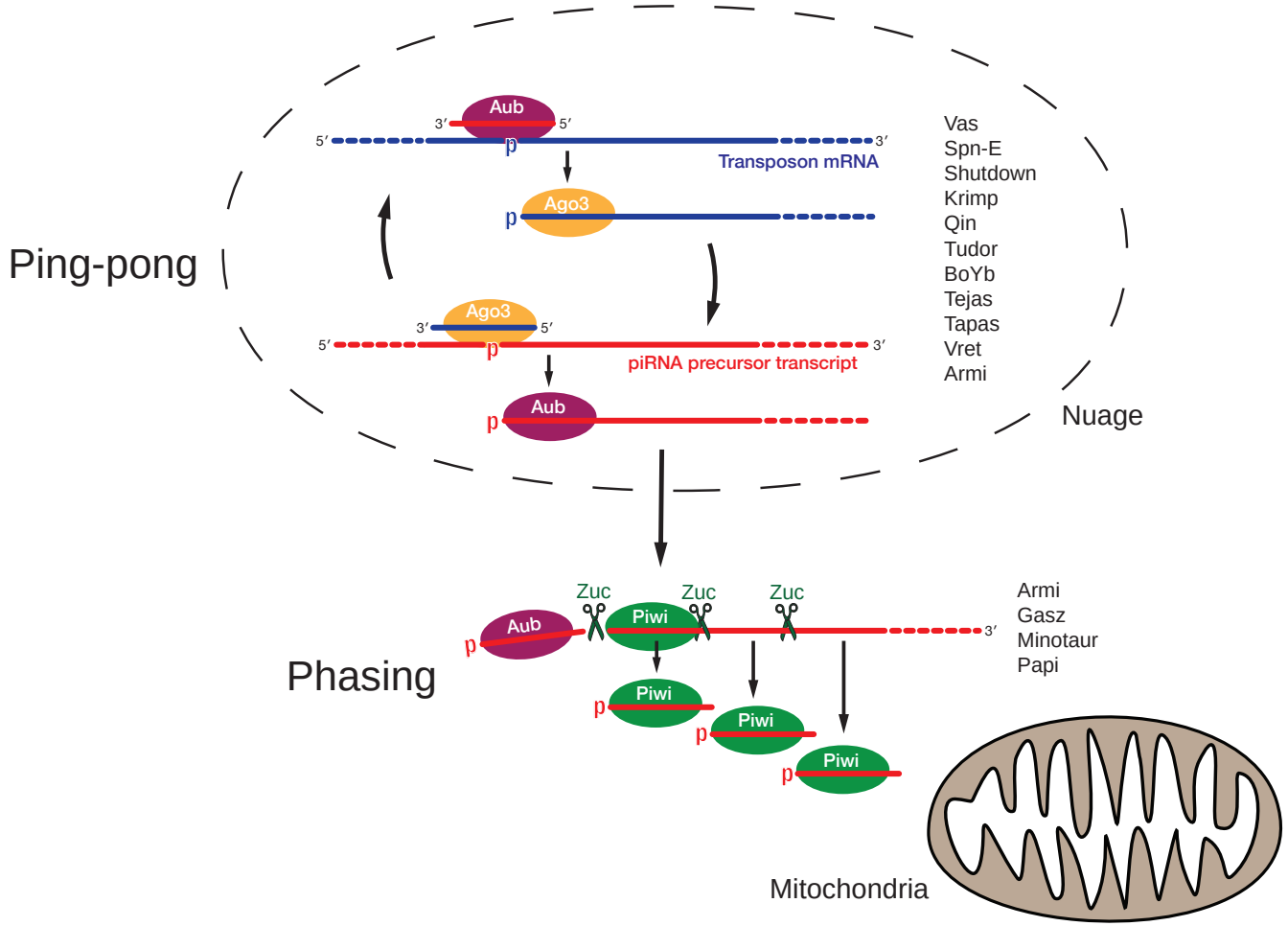

Ping-pong
Aub
3′
5′
p
Transposon mRNA
Ago3
Ago3
piRNA precursor transcript
Aub
Vas
Spn-E
Shutdown
Krimp
Qin
Tudor
BoYb
Tejas
Tapas
Vret
Armi
Nuage
Phasing
Zuc
Zuc
Zuc
Aub
Piwi
Piwi
Piwi
Piwi
Armi
Gasz
Minotaur
Papi
Mitochondria

Figure 1.1: The current model of piRNA biogenesis in the *Drosophila* germ cell cytoplasm

In fly ovaries, Aub, Ago3, Piwi, Qin and Squash co-immunoprecipitate with Spn-E (Andress et al., 2016). In the silk moth ovarian BmN4 cells, Siwi (Aub homolog), Ago3 and Qin co-immunoprecipitate with a catalytically inactive Vasa (Xiol et al., 2014). These results suggest that ping-pong factors function in a complex to efficiently amplify piRNAs.

The ping-pong cycle is believed to take place in the nuage (French for “cloud”), electron-dense perinuclear granules, based on the observations that most factors participating in the ping-pong pathway concentrate in the nuage. Examples include the PIWI-clade Argonaute proteins Aub and Ago3 (Brennecke et al., 2007), RNA-binding ATPases Vasa (Liang et al., 1994) and Spn-E (Andress et al., 2016), Tudor domain-containing proteins Krimp (Lim and Kai, 2007), Qin (Zhang et al., 2011; White-Cooper, 2012), Tudor (Nishida et al., 2009), BoYb (Handler et al., 2011) , Tejas (Patil and Kai, 2010), Tapas (Patil et al., 2014), and the cochaperone Shutdown (Olivieri et al., 2012). In addition, piRNA precursor transcripts were found to colocalize with Vasa in the nuage (Mohn et al., 2014). Furthermore, it is shown that Vasa accumulates in perinuclear foci (nuage) that are juxtaposed to the sites of piRNA precursor transcription (marked by Rhino) on the other side of the nuclear envelop (Zhang et al., 2012), suggesting the channeling of piRNA precursor transcripts directly into the nuage after nuclear export.

Nuage is a collective term for electron-dense, non-membrane-bound granulo-fibrous bodies found in immature and differentiating germ cells, originally called “fibrous bodies” (Mahowald, 1971). Under the electron microscope, bodies of similar appearance or morphology, and sometimes even with the same protein constituents, have been described in the germ cells of various animals, from worms to mammals (Eddy, 1975). Examples include those found in oocytes (Balbiani body, sponge body, polar granules), spermatocytes (intermitochondrial cement), and spermatids (chromatoid body) (Voronina et al., 2011). In a broader sense, nuage has been used to describe most of the bodies mentioned above (Jaglarz et al., 2011; Eddy, 1975). However, for the sake of simplicity, throughout this thesis a narrower definition of nuage is used, referring specifically to perinuclear nuage found in immature and differentiating germ cells.

Nuage is enriched in RNA-protein complexes (Voronina et al., 2011) and can quickly exchange components with the surrounding cytoplasm (Snee and Macdonald, 2004; Andress et al., 2016; Webster et al., 2015). In *C. elegans*, nuage has been shown to behave like liquid droplets (Brangwynne et al., 2009). Interestingly, during early *Drosophila* oogenesis, the post-meiotic pro-oocyte gradually loses nuage on its way to become the oocyte, while the other 15 nurse cells retain nuage (Mahowald, 1970; Mahowald, 1971). The loss of nuage coincides with the oocyte nucleus condensing into the karyosome and becoming

transcriptionally inactive (Bastock and St Johnston, 2008). Therefore, the structural maintenance of nuage may require active transcription.

Nuage is long known to be a conserved feature of germ cells, so is Vasa, a DEAD-box ATPase and a well-established marker for the nuage (Hay et al., 1988). The function of Vasa in transposon silencing is revealed after the discovery of piRNAs and many other nuage factors that similarly participate in piRNA biogenesis (Lim and Kai, 2007). Since a *vasa* mutation only affects transposon silencing, but not protein-coding gene expression (Zhang et al., 2012), and the nuage granules disappear in *vasa* mutants under the electron microscope (Liang et al., 1994), it is thought that the main function of the nuage is to produce piRNAs and silence transposons.

Phasing

The phasing pathway (traditionally called “primary biogenesis”) processes long piRNA precursor transcripts into head-to-tail linked strings of piRNAs. An endonuclease (presumably Zucchini, see below) is thought to simultaneously generate the 3′-end of the preceding piRNA and the 5′-end of the trailing piRNA (Figure 1.1). Five factors, Zuc, Armi, Gasz, Minotaur and Piwi are necessary for the phasing machinery to operate (as defined by a statistically significant enrichment of head-to-tail-linked piRNAs on the same genomic strand) (Han et al., 2015a). In contrast to ping-pong, the phasing pathway diversifies the piRNA

pool, because once endonucleolytic cleavages begin on a piRNA precursor, piRNAs are made downstream from the entry site in a length-dependent, but sequence-independent manner.

Phasing is thought to happen on the mitochondrial surface, because all factors implicated in the phasing pathway (Zuc, Armi, Gasz, Minotaur and Piwi) localize to mitochondria, at least partially (in the case of Armi and Minotaur (Vagin et al., 2013)) or transiently (in the case of Piwi, (Olivieri et al., 2012; Olivieri et al., 2010)). Armi, Zuc, Ago3, to a lesser degree Aub, but not Tudor, have been shown to associate with fractionated mitochondria in flies (Huang et al., 2014). In the silk moth ovarian cell line BmN4, Zuc, Papi, Siwi (Aub homolog), Ago3 and Trimmer (PNLDC1) ((Izumi et al., 2016), not conserved in flies) are detected in fractionated mitochondria, albeit in different sucrose gradient fractions (Nishida et al., 2018). However, it is unclear what relationships exist between mitochondria and piRNAs. Both *zuc* and *gasz* mutants are defective at mitochondrial fusion (Handler et al., 2013; Huang et al., 2011; Watanabe et al., 2011; Choi et al., 2006). In addition, both Zuc and Minotaur/GPAT2 have been implicated in the biosynthesis of phosphatidic acid, an important signaling molecule mediating membrane fusion (Kameoka et al., 2018). However, knocking down mitochondrial fusion factors such as Mitofusin has no impact on transposon silencing (Handler et al., 2013; Baena-Lopez et al., 2013; Muerdter et al., 2013).

PROTEIN FACTORS REQUIRED FOR PHASING

Zucchini

Zucchini (Zuc) is a piRNA biogenesis factor identified through a genetic screen for female sterility mutants (Pane et al., 2007). Its mammalian homolog, MitoPLD, is also required for piRNA biogenesis and male fertility in mice (Huang et al., 2011; Watanabe et al., 2011). Zuc and MitoPLD both harbor N-terminal mitochondrial localization sequences and reside at the surface of mitochondria (Choi et al., 2006; Saito et al., 2010; Handler et al., 2013). Zuc belongs to the phospholipase D (PLD) superfamily whose members harbor an HKD catalytic domain that hydrolyzes phosphodiester bonds in phospholipids or nucleic acids (Selvy et al., 2011). The crystal structure of Zuc shows extensive similarity to Nuc, a bacterial PLD-family nuclease: two Zuc monomers form a homodimer to constitute the HKD active site, and a positively charged groove lying across the active site potentially accommodates the RNA substrate (Nishimasu et al., 2012; Ipsaro et al., 2012; Voigt et al., 2012). Consistently, Zuc cleaves single-stranded RNA in vitro, albeit the cleavage only occurs at non-physiological salt concentrations, and shows no preference for uridine nucleotides (Nishimasu et al., 2012; Ipsaro et al., 2012; Nishida et al., 2018), a feature expected for the endonuclease generating phased piRNAs (Han et al., 2015a; Mohn et al., 2015). Nonetheless, Zuc is thought to be the phasing endonuclease because no better candidates have been identified, even after several genome-wide RNAi screens

for piRNA factors (Handler et al., 2013; Baena-Lopez et al., 2013; Muerdter et al., 2013). Alternatively, it has been suggested that Zuc/MitoPLD promote piRNA biogenesis indirectly through the generation of the signaling lipid phosphatidic acid at mitochondrial surface (Huang et al., 2011; Watanabe et al., 2011). MitoPLD catalytic activity induces mitochondrial fusion (Choi et al., 2006), a process likely mediated by MitoPLD-dependent production of phosphatidic acid (Huang et al., 2011). Recombinant mitoPLD has been shown to hydrolyze cardiolipin into phosphatidic acid (Choi et al., 2006), though it was not reproduced in a later study (Ipsaro et al., 2012).

Armitage

Armitage (Armi) was identified through a genetic screen for female fertility and egg patterning defects (Cook et al., 2004). Armi mutants cause premature translation of *oskar* mRNA and microtubule depolarization in early oogenesis. During mid-oogenesis, *armi* mutants disrupt *oskar* or *gurken* mRNA localization and dorsal-ventral patterning of the oocyte (Cook et al., 2004). Later, *armi* mutation was found to affect the accumulation of *Suppressor of Stellate* rasiRNA (a class of piRNA) in *Drosophila* testes (Vagin et al., 2006). The persistence of DNA double-strand break marker gamma-H2Av past region 2b of the germarium in *armi* mutants suggests unrepaired DNA damage (Klattenhoff et al., 2007), which is linked to global transposon activation (Klattenhoff et al., 2009) and loss

of piRNAs (Malone et al., 2009). Using the somatic follicle cells of the ovary or follicle cell-derived, immortalized cell line (OSS or OSC cells) as model systems, Armi is found to be an essential factor responsible for the production of primary piRNAs (renamed to phased piRNAs in this thesis), together with Zuc and Yb (Saito et al., 2010; Haase et al., 2010; Olivieri et al., 2010). Armi is dispensable for the ping-pong cycle (traditionally called “secondary piRNA biogenesis”) (Handler et al., 2011; Malone et al., 2009). More recently, it was shown that tethering Armi artificially to a transcript is sufficient to trigger the production of piRNA-length small RNAs from the tethered transcript (Rogers et al., 2017; Pandey et al., 2017). Purified, recombinant Armi can unwind RNA duplexes in a 5′-to-3′ direction in an ATP-dependent manner in vitro (Pandey et al., 2017). However, the mechanistic role of Armi in vivo remains largely unknown.

Armi subcellular localization is not clear-cut. It was suggested to be a nuage protein (Cook et al., 2004), but was later shown to form cytoplasmic “clouds” (Olivieri et al., 2010; Saito et al., 2010) that colocalize with mitochondria (Handler et al., 2013). Both nuage and cytosolic localization of Armi leads to the hypothesis that Armi is a shuttle between nuage and mitochondria (Huang et al., 2014). Interestingly, while in wild-type ovaries Armi mainly associates with the mitochondrial fraction of Ago3, in flies where the catalytic residue of Ago3 is mutated (DDH → AAH) and ping-pong is inhibited, Armi is found to associate more tightly with Ago3AAH in the nuage (Huang et al., 2014). At the time, it was

thought that the primary pathway (phasing) is upstream of the secondary pathway (ping-pong). It is therefore not understood why a block in the downstream pathway (ping-pong) would affect the behavior of an upstream pathway component (Armi). This observation made sense in light of the later finding that ping-pong works upstream of phasing (Han et al., 2015a; Mohn et al., 2015).

The mammalian homologs of Armi, Mov10l1 (germline-specific) and Mov10 (ubiquitous) both have RNA-binding properties (Vourekas et al., 2015; Gregersen et al., 2014). Mov10l1 functions in primary piRNA biogenesis just like Armi (Frost et al., 2010; Zheng et al., 2010), but whether it associates with bona fide piRNA precursors is not conclusive: RNAs that UV-crosslink to Mov10l1 in mouse testicular cells do not have a first-uridine nucleotide (1U) signature, a hallmark for primary piRNAs, and do not share 5′-ends with mature piRNAs (Vourekas et al., 2015). Like flies, mouse piRNA precursors undergoing phasing are expected to carry both features (Gainetdinov et al., 2018).

Gasz, Minotaur and Papi

Gasz is another factor required for phased piRNA biogenesis (Handler et al., 2013; Baena-Lopez et al., 2013). It is an ankyrin repeat-containing protein localized to mitochondria through its C-terminal peptide. Gasz colocalizes with Zuc and is required for Armi localization (Handler et al., 2013; Baena-Lopez et

al., 2013). In mice, Gasz is required to stabilize multiple components of the intermitochondrial cement, and the cementing material itself (Ma et al., 2009).

Minotaur belongs to the glycerol-3-phosphate *O*-acetyltransferase (GPAT) family, an enzyme with the potential to function in phosphatidic acid biosynthesis. *minotaur* mutant phenocopies *zuc* in piRNA loss (Vagin et al., 2013; Han et al., 2015a). However, the predicted GPAT active site is dispensable for transposon silencing in both Minotaur and its mammalian homolog GPAT2 (Fu et al., 2013; Vagin et al., 2013). In flies, Minotaur localizes to both mitochondria and the endoplasmic reticulum (Vagin et al., 2013).

Papi is a Tudor domain-containing protein best studied in the silk moth ovarian BmN4 cells. BmPapi is required to both maintain piRNA levels (Izumi et al., 2016; Nishida et al., 2018) and promote piRNA 3′-trimming by recruiting the Trimmer enzyme PNLDC1 (Izumi et al., 2016; Honda et al., 2013). BmPapi directly interacts with PIWI protein (Zhang et al., 2018; Zhang et al., 2017), associates with piRNA precursors (ping-pong intermediates (Nishida et al., 2018)), and localizes to the mitochondrial outer membrane (Honda et al., 2013). In *Drosophila*, Papi has been shown to interact with Ago3 (Liu et al., 2011), but the loss of Papi in flies has a much milder impact than in BmN4 cells: only Piwi-bound piRNAs are extended, on average, ~0.5 nt (Hayashi et al., 2016), transposons are largely repressed and female fertility is close to normal (Zhang

et al., 2018). It is possible that other factors that share similar domains with Papi, such as Yu/Spoon (Handler et al., 2011), serve redundant functions in flies.

NUAGE-MITOCHONDRIAL COUPLING

Nuage-mitochondrial association

Nuage often, but not always, associates with mitochondria (Kloc et al., 2014; Eddy, 1975). Particularly in mammalian early germ cells, “cementing material” in the interstices of mitochondrial clusters (Eddy, 1974) has been linked to piRNA biogenesis (Aravin et al., 2009). In the absence of MitoPLD in mice prospermatogonia, mitochondria show perinuclear polarized clustering, and the “cementing material” is lost (Watanabe et al., 2011; Huang et al., 2011). The loss of intermitochondrial cement is unlikely the result of piRNA loss in *mitoPLD* mutants, because in *miwi2* mutants, where piRNAs are also significantly reduced, intermitochondrial cement is unaffected (Kuramochi-Miyagawa et al., 2010). An intriguing finding is that the loss of Lipin 1 (fld, fatty liver dystrophy), a mitochondrial surface enzyme converting phosphatidic acid to diacylglycerol, increases the length and density of intermitochondrial cement, supporting the notion that phosphatidic acid signals nuage-mitochondrial association in mammalian male germ cells (Huang et al., 2011). It is not known if piRNA production is affected by Lipin 1, but the fact that Lipin 1 mutant mice are sterile (Huang et al., 2011) suggests such a possibility.

Nuage-mitochondrial association is, however, not universal. A survey of the literature on electron micrographs of *Drosophila* nurse cells reveals a general lack of nuage-mitochondria association (Jaglarz et al., 2011; Wilsch-Bräuninger

et al., 1997; Mahowald, 1971; Mahowald, 1970; Liang et al., 1994; Dapples and King, 1970), except one report of nuage-mitochondria attachment in stage 1 nurse cells, but the lack of such association in later stages of oogenesis (Mahowald, 1971). Indeed, the association between mitochondria and nuage-like granules seems to vary significantly between species. For example, polar granules are often found to attach to mitochondria in late stage oocytes of *D. melanogaster*, *D. willistoni* and *D. immigrans*, but not *D. hydei* (Mahowald, 1971; Mahowald, 1962). Therefore, the knowledge about nuage-mitochondrial association in mammalian germ cells may not be directly transferrable to flies.

Ping-pong-phasing coupling

The ping-pong machinery appears to operate independently of phasing: the 10-nt-overlap “ping-pong signature” between piRNA 5′-ends is unaffected by all the known phasing mutants (Han et al., 2015a). However, phasing depends on ping-pong in the fly germline (Wang et al., 2015; Senti et al., 2015), although it functions without ping-pong in the somatic follicle cells of the fly ovary.

In 2015, it was found that in germline cells, ping-pong piRNA-directed cleavage by Ago3 or Aub specifies the entry site of phased piRNA production (Han et al., 2015a; Mohn et al., 2015) (Figure 1.1). The 3′-cleavage product, which bears a 5′-monophosphate, is first loaded into Aub or Ago3 (termed “ping-pong intermediate” in this thesis), which subsequently becomes the substrate for

phasing. Zuc cleaves in front of the first uridine unprotected by Aub or Ago3 binding, to form the 3′-end of the new ping-pong piRNA (Mohn et al., 2015; Gainetdinov et al., 2018). The remainder of the 5′-monophosphorylated precursor is made into multiple consecutive piRNAs in a similar fashion: the 5′ end of precursor is loaded into Piwi (and to a lesser degree, Aub, but not Ago3) before Zuc cleaves at the first available downstream uridine to make a new phased piRNA. The loading-cleavage cycle can happen multiple times on the same precursor, yielding head-to-tail-linked phased piRNAs. The relative positions of the phased piRNAs are in phases downstream of the first ping-pong piRNA (multiples of mature piRNA length, i.e., 27 nt, 55 nt, etc.). Therefore, the process is termed phased piRNA biogenesis (Han et al., 2015a).

STUDY AIMS

Deep sequencing of piRNAs and their precursor molecules revealed ping-pong-initiated phased piRNA production (Han et al., 2015a; Mohn et al., 2015), and studies of genetic mutants revealed a strong interaction between the ping-pong and phasing pathways (Wang et al., 2015; Senti et al., 2015). However, there is a lack of mechanistic understanding of how such coupling is achieved. For example, what relationships exist between nuage and mitochondria, e.g., do they come in contact? Are phasing factors exclusively localized to the mitochondria, or are they also present in the nuage? Do phasing factors function together as a complex or separately in different steps? If the ping-pong and phasing machineries are physically close, what governs their sequential processing of a piRNA precursor? On the other hand, if the ping-pong and phasing machineries are physically separate, how is a common piRNA precursor first processed by ping-pong in the nuage, then transferred to mitochondria for phased processing? In this thesis I will present my findings toward these ends.

CHAPTER II: THE RNA-BINDING ATPASE ARMITAGE COUPLES PIRNA AMPLIFICATION IN NUAGE TO PHASED PIRNA PRODUCTION ON MITOCHONDRIA

PREFACE

The work presented in this chapter was a collaborative effort: Wei Wang did most of the bioinformatic analyses. Cindy Tipping helped with fly ovary dissections. The UMass Proteomics and Electron Microscopy Core Facilities performed mass spectrometry and transmission electron microcopy studies, respectively. Phil Zamore and I designed the experiments with inputs from Wei Wang and Zhiping Weng. I performed the rest of the experiments and analyzed some of the sequencing data.

SUMMARY

In the *Drosophila* ovary, PIWI-interacting RNAs (piRNAs) suppress transposon expression, ensuring female fertility. Germline piRNAs are made through two coupled pathways, the upstream ping-pong amplification loop and the downstream phased piRNA production, presumed to take place in the perinuclear nuage and the mitochondrial surface, respectively. We found that nuage and mitochondria are physically separate in *Drosophila* nurse cells. Both Zuc (endonuclease) and Armi (ATPase) are required for the downstream phasing, but not the upstream ping-pong pathway. While Zuc localizes exclusively to mitochondria, Armi localizes to both nuage and mitochondria. A block of phasing traps Armi on mitochondria, suggesting a dynamic distribution of Armi between the two compartments. Armi binds to protein factors and RNA precursors that participate in ping-pong or phasing, and the ATPase activity is required for it to selectively bind piRNA precursors, rather than genic mRNAs. The Armi ATPase mutant retains binding to RNA and other piRNA factors, but is dispersed in the cytoplasm and fails to support phased piRNA production. We propose a model that Armi uses ATP to quickly dissociate from genic mRNAs, associates with ping-pong-cleavage products in the nuage and anchors them to the mitochondrial surface, where Armi uses ATP to remodel the phasing machinery to activate Zuc cleavages. Armi therefore enables phased piRNA

production at two levels: providing the correct RNA substrate, and remodeling the catalytic complex.

RESULTS

Nuage and mitochondria are physically separate in nurse cells

Physical association of nuage and mitochondria might explain the molecular coupling between ping-pong and phasing. However, a survey of the literature on electron micrographs of *Drosophila* nurse cells reveals a general lack of nuage-mitochondria association. To further investigate this finding, we examined the cytologic relationship between ATP synthase complex V alpha subunit (ATP5A), an inner mitochondrial membrane protein, and Vasa, a nuage protein. We focused on stage 3 egg chambers, where the 15 nurse cell nuclei are sufficiently separated to allow unambiguous detection of cytoplasmic proteins, and where the piRNA biogenesis factors are expressed at relatively high levels (Figure 2.1). In wild-type ovaries, nuage shows perinuclear punctate staining in germline nurse cells. Mitochondria are more evenly distributed in the cytoplasm, with a tendency to clump around the nucleus (Figure 2.2). Quantification of the fluorescence signal showed that only 18% ± 9% (mean ± standard deviation) of the Vas signal overlapped with ATP5A, while 88% ± 4% of the Vas signal overlapped with Aub, another nuage protein. Because of the diffraction barrier, the resolution of optical microscopy is limited to approximately half the wavelength of the laser light (~250 nm), which is not significantly smaller than the typical size of nuage (Jaglarz et al., 2011). Therefore, the observed low level of

nuage/mitochondria colocalization may be caused by diffracted/diffused light, rather than true physical proximity.

Figure 2.1

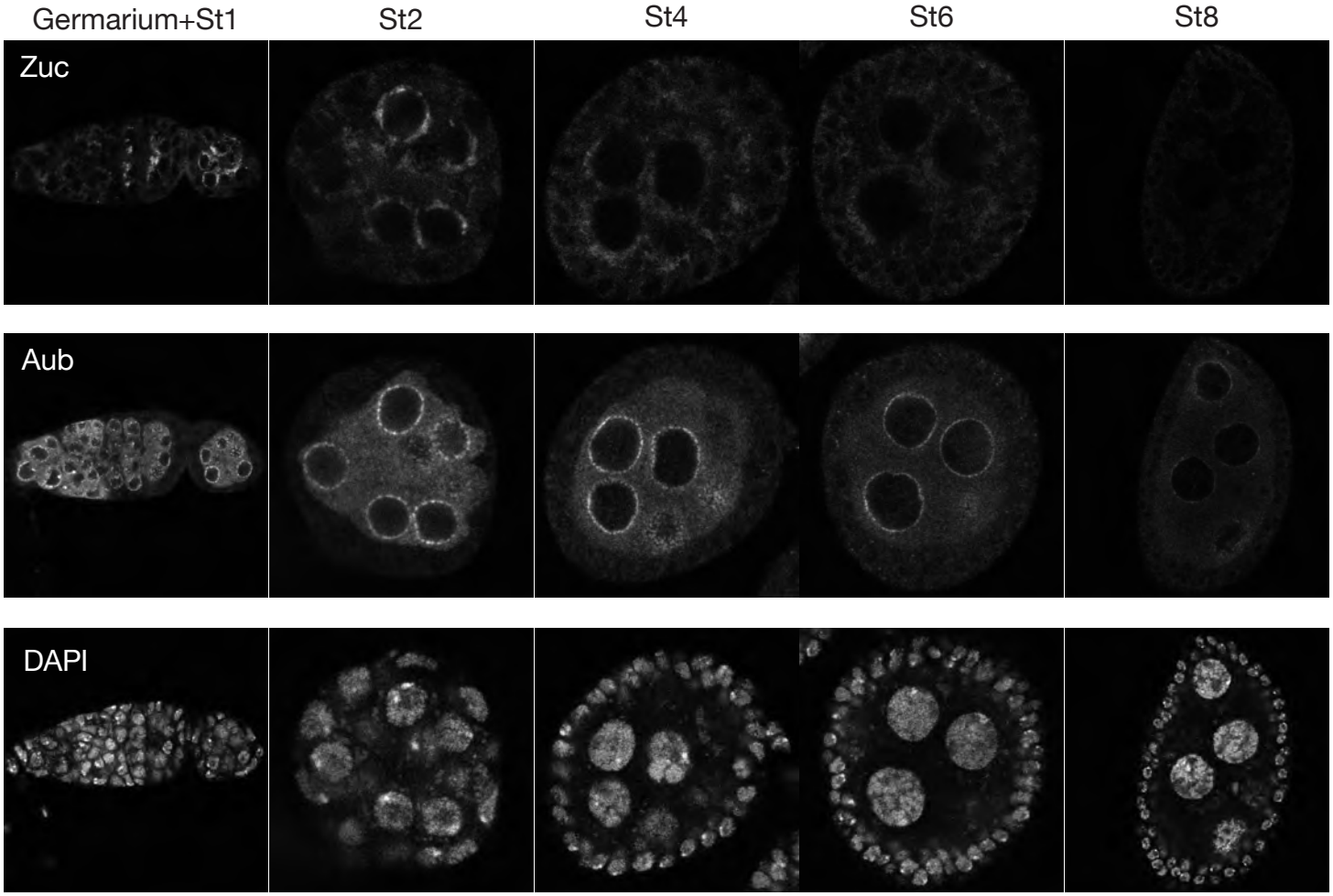


Figure 2.1: Zuc and Aub expression levels through early and mid oogenesis

Immunofluorescence detection of Zuc-3×FLAG (using anti-FLAG), Aub and nucleic acids (DAPI) in the germarium and stage 2, 4, 6 and 8 egg chambers. Both Zuc and Aub are expressed higher and show more distinct localization in earlier stages.

Figure 2.2

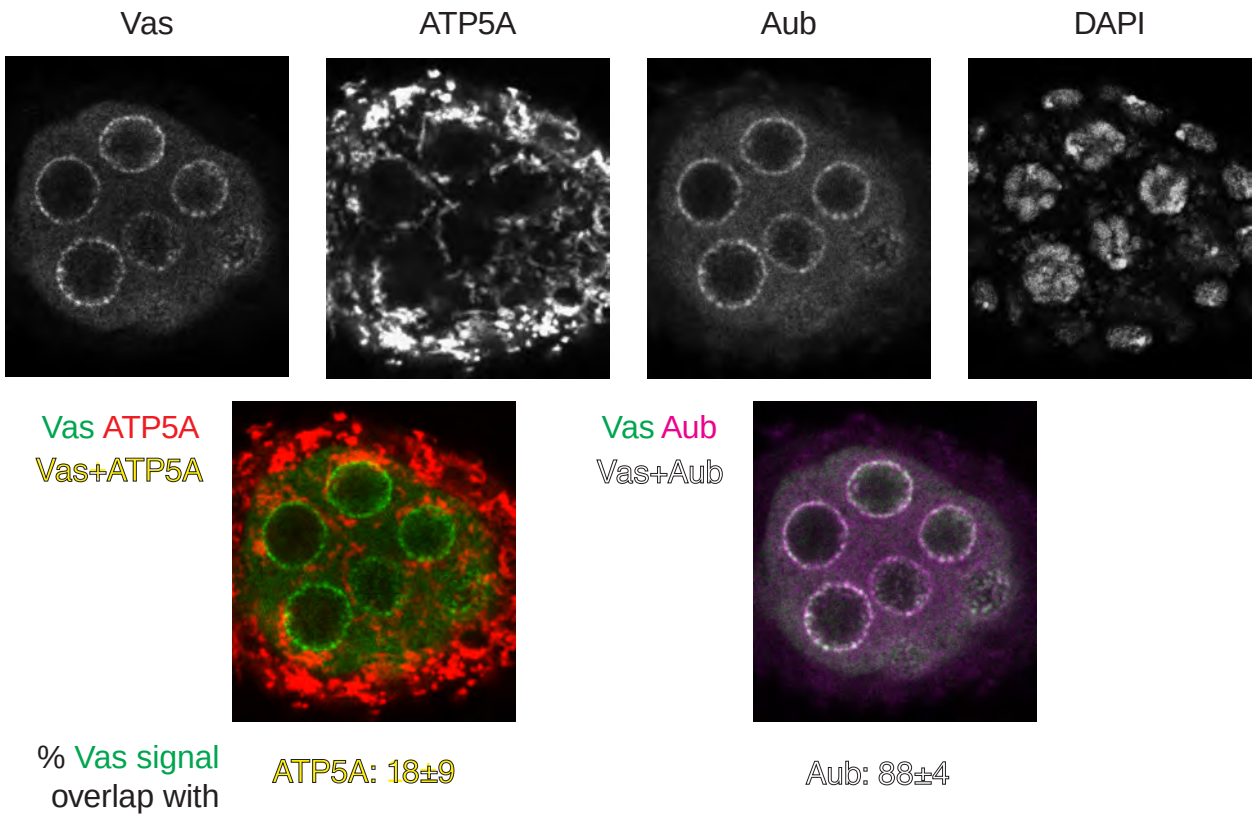

Vas
ATP5A
Aub
DAPI
Vas ATP5A
Vas+ATP5A
Vas Aub
Vas+Aub
% Vas signal
overlap with
ATP5A: 18±9
Aub: 88±4

Figure 2.2: Vas and ATP5A shows minimal overlap in immunofluorescence

Immunofluorescence detection of Vas, ATP5A, Aub and nucleic acids (DAPI) in wild-type stage 3 egg chambers. Each channel and the overlapping signal are color-coded. The percentage of Vas signal overlapping with either ATP5A or Aub was quantified using a CellProfiler custom script, and 54 serial z scan images at 1 µm interval were quantified. Numbers represent mean ± S.D. from three egg chambers.

To discern whether nuage and mitochondria touch each other, we used transmission electron microscopy (TEM), a technique that unequivocally recognizes both nuage (electron dense fibrous granules not bound by membrane) and mitochondria (double-membrane-bound organelles with internal cristae). Despite of our best efforts, nuage and mitochondria were never observed to contact each other in nurse cells (Figure 2.3), consistent with the literature (Jaglarz et al., 2011; Wilsch-Bräuninger et al., 1997; Mahowald, 1971; Mahowald, 1970; Liang et al., 1994; Dapples and King, 1970). We conclude that nuage and mitochondria are physically separate compartments in germline nurse cells.

Figure 2.3

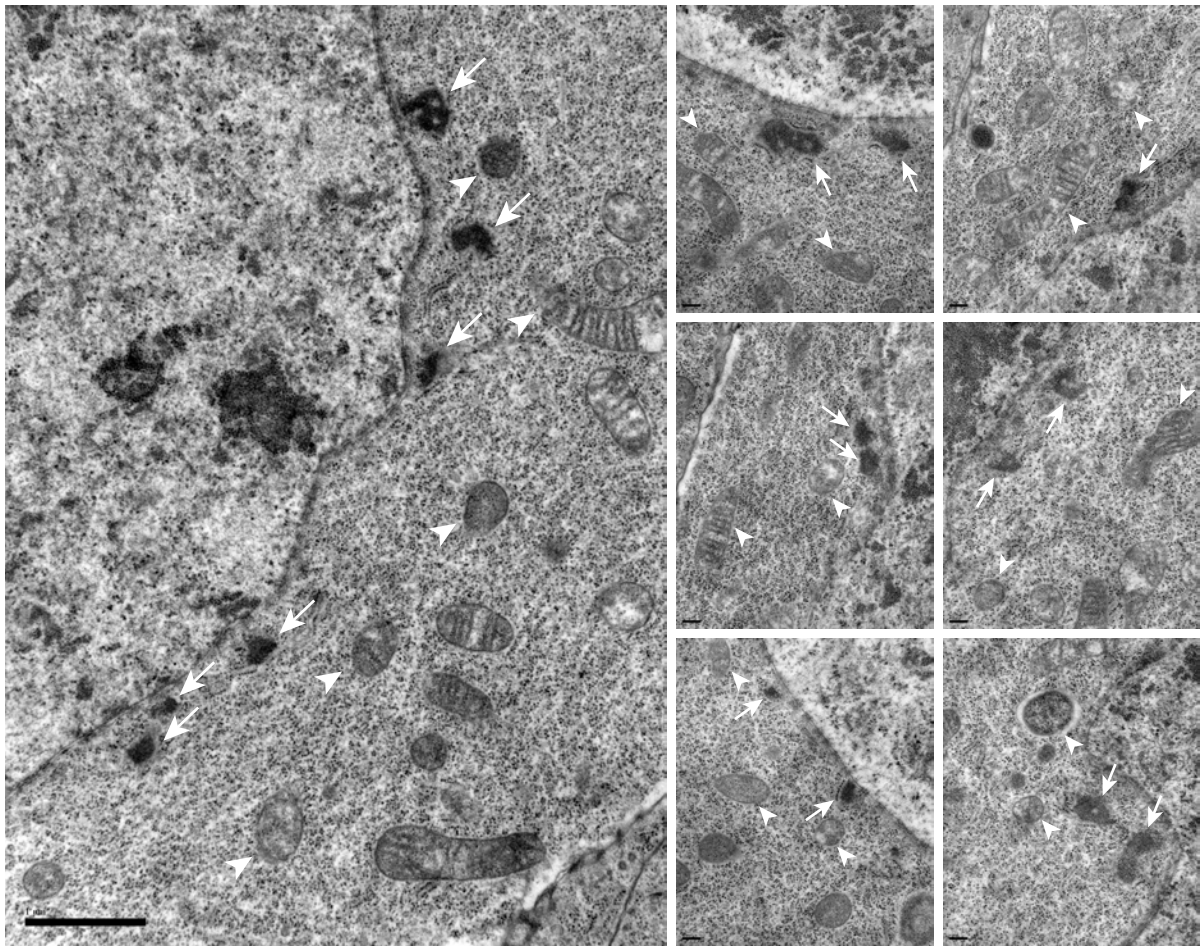


Figure 2.3: Lack of nuage-mitochondria contact in nurse cells

Transmission electron microscopy on wild-type stage 3 egg chambers. Ultrathin sections were stained with uranyl acetate and lead citrate. Scale bar, 1 μm for the image at left and 0.2 μm for the images at right. Arrow, nuage; arrowhead, mitochondrion.

Zuc localizes to mitochondria, but not nuage

Tagged, overexpressed Zuc has been shown to localize to mitochondria through the N-terminal anchor in OSC cells or in fly ovaries (Saito et al., 2010; Handler et al., 2013). However, it is not known if the endogenous Zuc behaves similar to the overexpressed protein. It is possible that a proportion of the endogenous Zuc can reside in the nuage (e.g., by losing the mitochondrial membrane anchor) to participate in piRNA biogenesis. Since there is no Zuc antibody available, we engineered a fly strain where the endogenous Zuc is tagged with 3×FLAG at the C-terminus. The tag is not expected to be affected by the presence or absence of the N-terminal anchor. Flies homozygous for *zuc-3×FLAG* are viable and fertile. The amount of piRNAs and normal comparing to w^{1118}, a laboratory “wild-type” strain (Figure 2.4A). Consistently, RNA expression of most of the transposon families are comparable between *zuc-3×FLAG* and w^{1118}, except for a few that changes in either direction, likely due to background variations (Figure 2.4B). Therefore, the tagged Zuc is functionally indistinguishable from the untagged version.

Figure 2.4

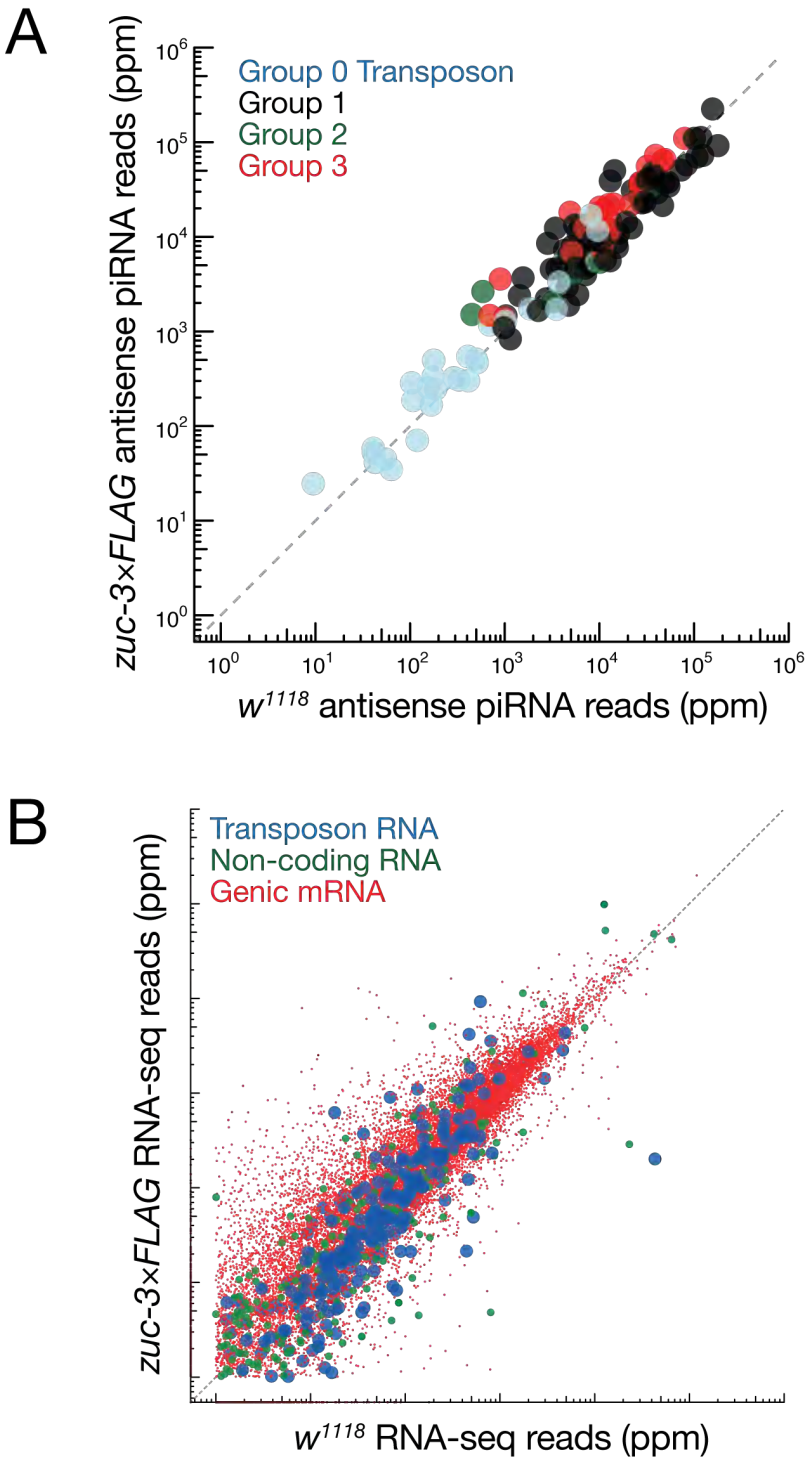

A
Group 0 Transposon
Group 1
Group 2
Group 3
zuc-3×FLAG antisense piRNA reads (ppm)
w1118 antisense piRNA reads (ppm)
B
Transposon RNA
Non-coding RNA
Genic mRNA
zuc-3×FLAG RNA-seq reads (ppm)
w1118 RNA-seq reads (ppm)

Figure 2.4: *zuc-3×FLAG* flies have normal piRNA expression and transposon silencing

(A) Scatterplot showing the level of piRNAs antisense to transposons in *zuc-3×FLAG* ovaries versus that of w^{1118} (Bo Han and Phil Zamore, unpublished). Each dot represents one transposon family, classified into Groups 0–3 according to (Li et al., 2009). (B) Scatterplot showing the level of long RNA in *zuc-3×FLAG* ovaries versus that of w^{1118} ((Zhang et al., 2011)). Each dot represents a transposon family (a total of 238 families shown), a non-coding RNA, or a genic mRNA. All genome mappers are shown and displayed as ppm (parts per million genome mappers).

We then performed immunofluorescence on *zuc-3×FLAG* ovaries using anti-FLAG antibody to examine Zuc localization relative to nuage or mitochondria. While Zuc almost perfectly overlaps with ATP5A, it has minimal overlap with the nuage marker Vas (Figure 2.5). Therefore, endogenous Zuc appears to be predominantly mitochondrial.

Figure 2.5

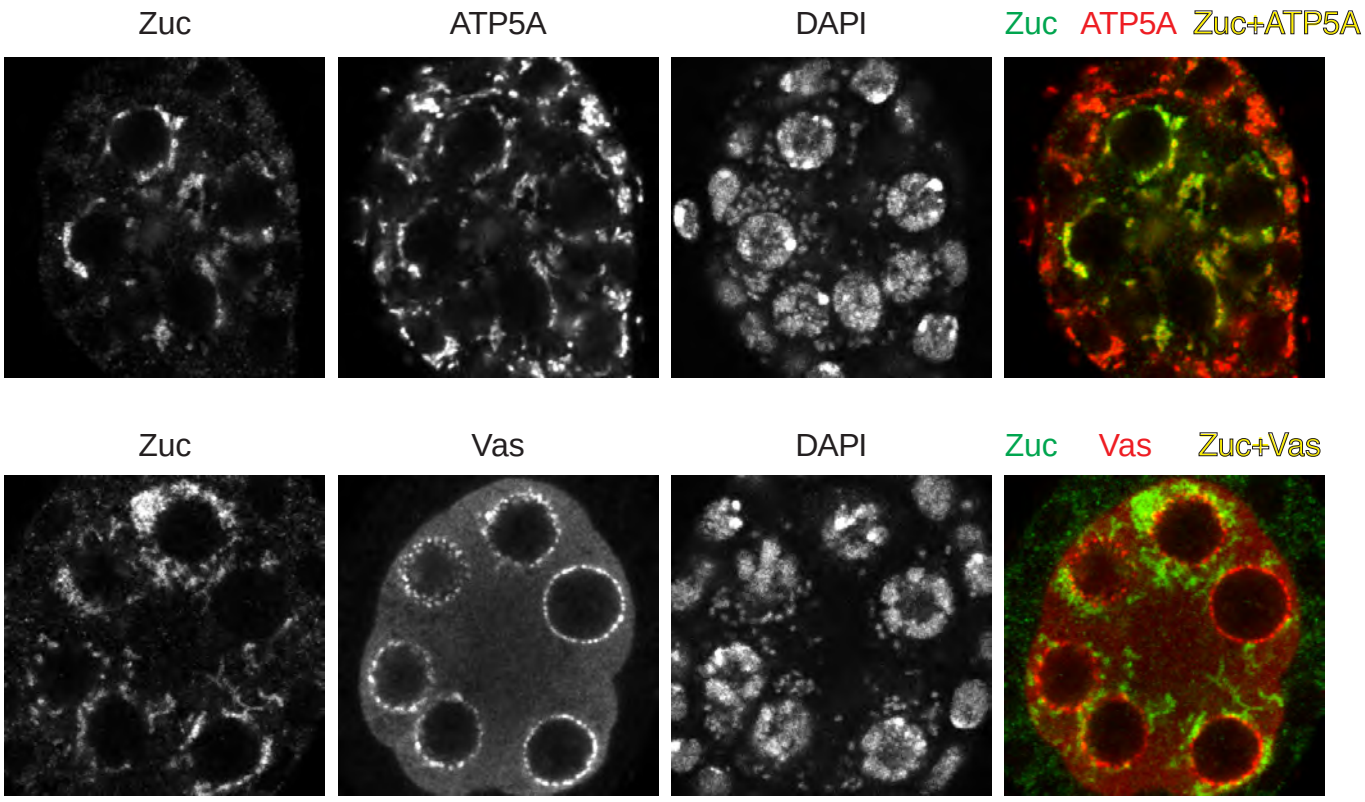

Zuc
ATP5A
DAPI
Zuc ATP5A Zuc+ATP5A
Zuc
Vas
DAPI
Zuc Vas Zuc+Vas

Figure 2.5: Zuc colocalizes with mitochondria, but not nuage

Immunofluorescence detection of Zuc-3×FLAG, ATP5A, Vas and nucleic acids (DAPI) in stage 3 egg chambers of *zuc-3×FLAG* female flies. Each channel and the overlapping signal are color-coded.

Zuc interacts with mitochondrial, but not nuage proteins

A small amount of Zuc beyond the detection limit of immunofluorescence may present in nuage and contribute to piRNA biogenesis. To exclude that possibility, we took another approach to test if Zuc comes in contact with nuage proteins. We reasoned that by immunoprecipitating Zuc and examining its interacting proteins using mass spectrometry, its interaction partners may be detected, even at low stoichiometry. Using wild-type ovaries that don't express any FLAG-tagged protein as negative control, FLAG IP from *zuc-3×FLAG* lysate does not enrich for any known piRNA factors under native conditions (data not shown), even though 12 peptides of Zuc were detected in the experimental IP and none in the control IP. We therefore tested a series of membrane-permeable, reversible chemical crosslinkers that can potentially stabilize protein-protein interactions. Armi is predicted to interact with Zuc because the mouse Armi homolog Mov10l1 co-immunoprecipitates with mouse Zuc homolog mitoPLD, when the two proteins are co-expressed in mammalian 293T cells that do not normally make piRNAs (Vourekas et al., 2015). The following crosslinkers were compared for efficiencies to crosslink Zuc to Armi in intact ovaries: paraformaldehyde (PFA), which crosslinks primary amine groups in proteins to a neighboring nitrogen atom in proteins or nucleic acids, with a spacer arm of 2.5 Å, reversible by heating above 65˚C in the presence of water; DTME (dithiobismaleimidoethane), a sulfhydryl-to-sulfhydryl crosslinker with a 13.3 Å spacer arm, reversible by reducing agent at

temperatures above 37˚C; DST, EGS and DSP, amine-to-amine crosslinkers with spacer arms of 6 Å, 16 Å, and 12 Å, and reversible by periodate oxidation, hydroxylamine and reducing agent, respectively. Among them, DTME best stabilizes the interaction between Zuc and Armi (Figure 2.6).

Figure 2.6

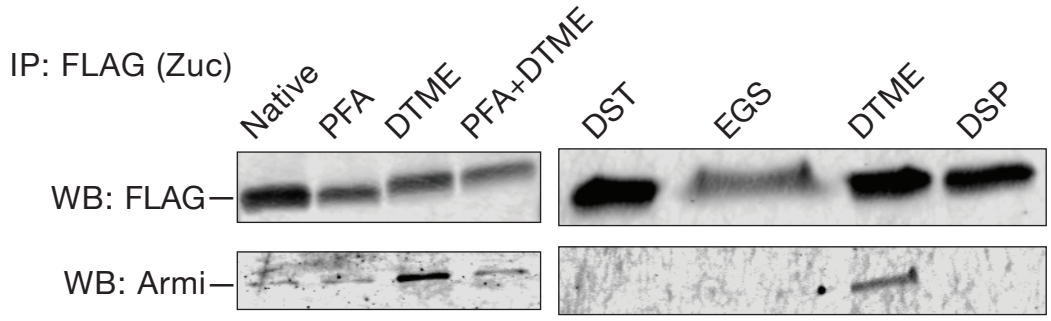

IP: FLAG (Zuc)
Native
PFA
DTME
PFA+DTME
DST
EGS
DTME
DSP
WB: FLAG
WB: Armi

Figure 2.6: DTME best stabilizes Zuc-Armi interaction

Zuc-3×FLAG fly ovaries were crosslinked with the indicated chemical crosslinkers before tissue lysis. Eluate from the FLAG IP was subjected to Western blot using anti-FLAG or anti-Armi antibodies.

DTME-crosslinked FLAG IP eluate from *zuc-3×FLAG* or Oregon R ovaries were subjected to mass spectrometry, and the enriched proteins from three biological replicates were identified by Fisher's exact test using weighted spectra and a threshold of Benjamini-Hochberg multiple test-corrected $p < 0.05$ (Table 2.1). Among the 40 enriched proteins, we found all four piRNA phasing factors: Armi, Gasz, Minotaur and Papi. Another piRNA biogenesis factors SoYb was also enriched. Although not well characterized, GFP-SoYb colocalizes with Armi in cytoplasmic clouds that we believe to be mitochondria (see below) (Handler et al., 2011). Importantly, among the 40 enriched proteins, only Armi is a known component of the nuage, in contrast to more than half (23 of 40) being known mitochondrial proteins (GO cellular component analysis on either the fly protein or its mammalian homolog). It is worth noting that without DTME crosslinking, Zuc IP does not enrich for any of the 40 proteins other than CG7461, which precludes the possibility that stable Zuc-mitochondrial protein interactions form after cell lysis.

Taken together, nuage and mitochondria are physically separate in *Drosophila* nurse cells, and Zuc resides on the mitochondrial surface with Armi, Gasz, Minotaur, Papi and SoYb. How can the same piRNA precursor be processed first by the ping-pong machinery in nuage, then by the phasing machinery on the mitochondrial surface (Figure 1.1)?

Table 2.1

Fly name	Human name	Mean fold enrichment
Zuc (bait)	Pld6	34767
Gasz	Asz1	2569
CG12360	Trabd	1152
CG10880	–	929
nmd	Atad1	867
Papi	Tdrkh	694
Armi	Mov10l1	472
Aldh-III	Aldh3b1	432
Miro	Rhot1	384
CG7705	–	372
Miga	Miga2	315
Faf	Faf2	278
CG9855	March5	268
Minotaur	GPAM	256
Spoon/Yu	Akap1	222
Pex3	Pex3	170
Marf	Mfn2	144
CG2316	Abcd2	144
Usp30/CG3016	Usp30	134
Myt1	PKMYT1	130
Men	Me1	126
CG1291	Alg2	113
Mtch	Mtch2	102
Ptp61F	Ptpn1	100
CG7639	SAMM50	98
CG1665	Marc1	93
CG8735	LNP	82
CG6550	CDKAL1	75
Sec63	Sec63	74
CG9853	Get4	69
CG6744	Exd2	64
CG7461	Acadvl	58
CG7546	Bag6	56
Tom70	Tomm70a	52
CG3394	SLC27A1	47
Pmp70	Abcd3	46
CG6089	Tmem214	18
iPLA2-VIA	Pla2g6	14
SoYb	Tdrd12	13
Acsl	Acsl3	7

Table 2.1: Proteins co-immunoprecipitate with Zuc-3×FLAG

Zuc-3×FLAG ovaries were crosslinked with DTME before cell lysis, followed by FLAG IP and mass spectrometry. Only proteins passing the Fisher's exact test using weighted spectra and a threshold of Benjamini-Hochberg multiple test corrected $p < 0.05$ are shown. Each *Drosophila* protein was searched in the Alliance of Genome Resources database (https://www.alliancegenome.org/) to obtain its human ortholog. Fold enrichment was calculated by dividing normalized iBAQ quantification score in the experimental *zuc-3×FLAG* IP by that in the Oregon R control IP. A pseudo-count equals to the average of the lowest 10 iBAQ values in each sample was added to all proteins in that sample to eliminate zeros. Shown is the mean fold enrichment from three biological replicates. Known mitochondrial proteins are shown in red.

Armi localizes to both nuage and mitochondria

Without physical proximity, one possible way to couple the nuage and mitochondrial phases of piRNA biogenesis is by transferring the piRNA precursor from nuage to mitochondria. Among the four identified factors required for phasing (Zuc, Armi, Gasz and Minotaur), the RNA-binding ATPase Armi is the most likely candidate. RNA-binding ATPases utilize the energy of ATP hydrolysis to perform an expanding repertoire of functions including RNA-protein interaction remodeling and RNA duplex unwinding (Pyle, 2011). Armi localization in nuage (Pandey et al., 2017; Huang et al., 2014) or mitochondria (Handler et al., 2013; Huang et al., 2014) has been observed in nurse cells at various stages of oogenesis, but it is not clear whether Armi localizes to different subcellular compartments in different developmental stages, or to both nuage and mitochondria in the same cell. To answer this question, we triple-stained Armi, mitochondria (ATP5A) and nuage (Vas) in the same stage 3 egg chamber. This shows that Armi colocalizes with both ATP5A and Vas at the same time (Figure 2.7). These results are consistent with the hypothesis that Armi moves between nuage and mitochondria.

Figure 2.7

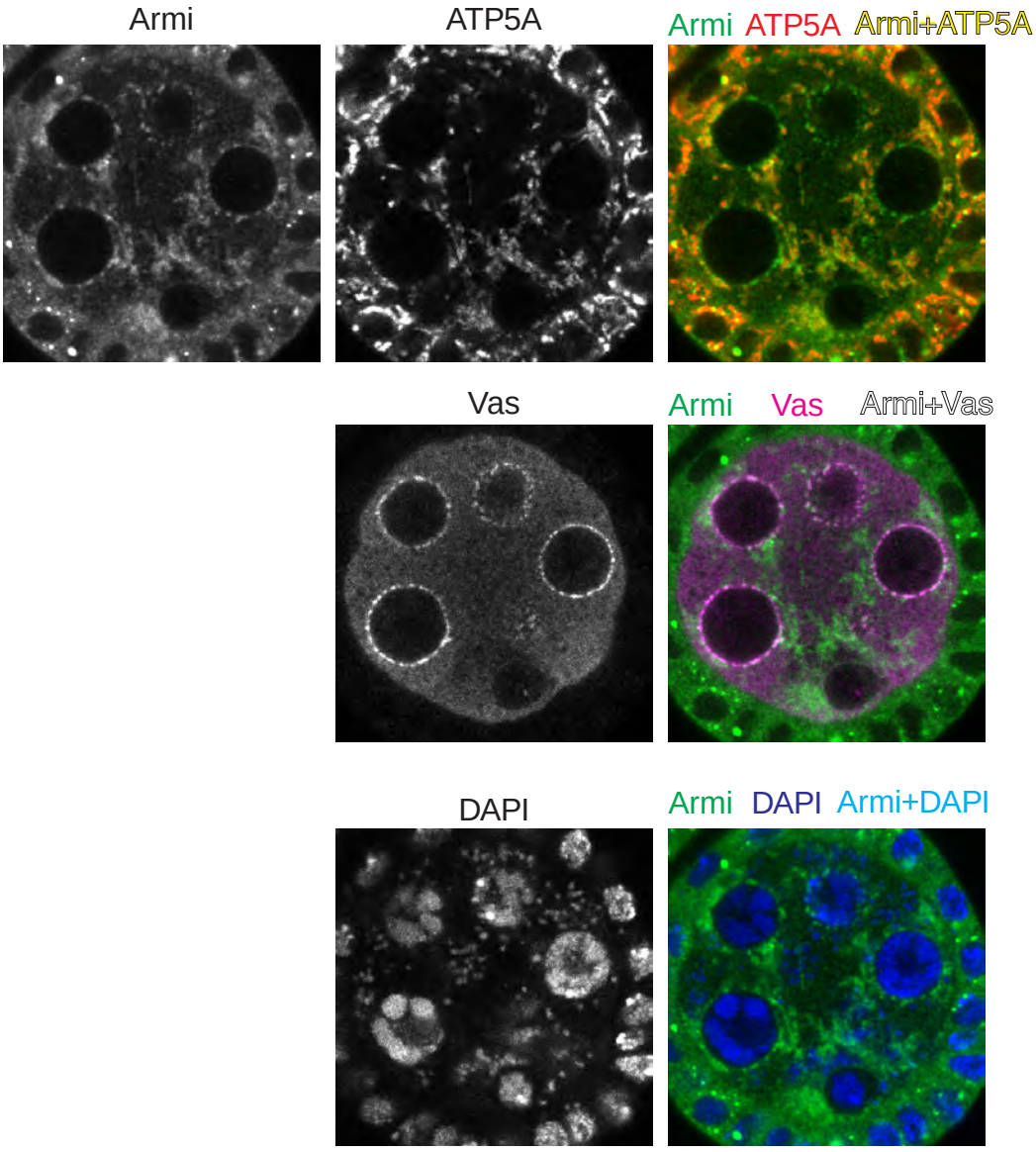

Armi
ATP5A
Armi ATP5A Armi+ATP5A
Vas
Armi Vas Armi+Vas
DAPI
Armi DAPI Armi+DAPI

Figure 2.7: Armi localizes to both nuage and mitochondria

Immunofluorescence detection of Armi, ATP5A, Vas and nucleic acids (DAPI) in the same stage 3 egg chamber of wild-type flies. Each channel and the overlapping signals are color-coded.

Armi interacts with both nuage and mitochondrial piRNA factors

We then asked what proteins interact with Armi in nurse cells. Ovaries from transgenic flies overexpressing germline-specific N-terminal 3×FLAG-6xMyc tagged Armi (FM-Armi) were crosslinked with DTME before cell lysis, followed by FLAG IP and mass spectrometry. Flies of the same genetic background, but not carrying the UAS-FM-Armi transgene serve as the negative control. Of the 95 significantly enriched proteins identified using Fisher's exact test from three biological replicates, 16 are known piRNA factors (Table 2.2). Consistent with the immunofluorescence result, Armi interacts with both nuage piRNA factors (Shu, Ago3, Spn-E, Tapas, Aub, BoYb, Qin and Vas) and mitochondrial piRNA factors (Gasz, Minotaur, SoYb, Papi). In addition, Vret, a tudor-domain containing protein required for the proper localization of all three PIWI proteins (i.e., Piwi, Aub and Ago3) also interacts with Armi, reproducing previous reports (Zamparini et al., 2011; Handler et al., 2011). Interestingly, GFP-Vret localization highly resembles that of GFP-Armi (Handler et al., 2011), suggesting that they may colocalize. Piwi also interacts with Armi, consistent with previous reports (Olivieri et al., 2010; Saito et al., 2010; Haase et al., 2010). Zuc was not detected in Armi IP-mass spec, likely because Zuc is a small protein with only 253 amino acids, and its expression level in total ovary lysate is beyond the detection limit of anti-FLAG Western blot (see *Discussion*).

Table 2.2

Fly name	Human name	Mean fold enrichment
Armi (bait)	Mov10l1	12900
Gasz	Asz1	238
Minotaur	GPAM	214
SoYb	Tdrd12	64
Vret	–	40
Shutdown	Fkbp6	38
Ago3	Hili	15
Spindle-E	Tdrd9	15
Tapas	Tdrd7	7
Aub	Hiwi	6
BoYb	Tdrd12	6
Papi	Tdrkh	5
Qin	Tdrd4	4
Piwi	Miwi2	3
Vasa	Ddx4	2

Table 2.2: piRNA factors co-immunoprecipitate with Flag-Myc-Armi

Ovaries from transgenic flies overexpressing germline-specific N-terminal 3×FLAG-6xMyc tagged Armi (FM-Armi) were crosslinked with DTME before cell lysis, followed by FLAG IP and mass spectrometry. Only known piRNA factors that passing the Fisher's exact test using weighted spectra and a threshold of Benjamini-Hochberg multiple test-corrected $p < 0.05$ are shown. Each *Drosophila* protein was searched in the Alliance of Genome Resources database (https://www.alliancegenome.org/) to obtain its human ortholog. Fold enrichment was calculated by dividing normalized iBAQ quantification score in the experimental IP by that in the control IP. A pseudo-count equals to the average of the lowest 10 iBAQ values in each sample was added to all proteins in that sample to eliminate zero. Shown are the mean fold enrichment from three biological replicates. piRNA factors that localize to mitochondria are shown in red. Those that localize to nuage are shown in blue. Piwi localizes to the nucleus, and Armi localizes to both nuage and mitochondria.

Table 2.3

Fly name	Mean FC	Fly name	Mean FC
Armi (bait)	12900	Marf	6
CG10880	105	Parp	6
Otu	77	Rpl1	6
CG9925	62	Shtd	6
Pzg	61	Usp16-45	6
CG2941	48	CG5205	5
CG7546	39	Glu	5
CG7686	38	CG1582	4
Gprk1	26	CG5604	4
CG3523	21	CG9247	4
CG2614	20	MBD-R2	4
CG10289	19	Mi-2	4
Lk6	19	Nop17l	4
eIF4G	18	CG10565	3
Rpl135	18	CG17514	3
CG11505	17	CG31368	3
CG6045	14	Clu	3
CG6204	13	Ctrip	3
Spd-2	13	DNApol-alpha180	3
CG16908	12	DNApol-epsilon	3
Cana	9	l(3)76BDr	3
CG9485	9	Pins	3
Nipped-A	9	Ade2	2
CG15618	8	Ade3	2
CG7261	8	Ago2	2
l(2)k09022	8	CG11148	2
CG8915	7	CG7878	2
Mask	7	CG8858	2
UBR1	7	eIF5B	2
Aft	6	Mle	2
Cap-D2	6	Poe	2
CG12499	6	Raptor	2
CG18596	6	Srp68	2
CG9674	6	Upf1	2

Table 2.3: Other proteins co-immunoprecipitate with Flag-Myc-Armi

Mean FC, mean fold change of iBAQ quantification scores in Armi IP relative to the negative control IP. See Table 2.2 for details.

Armi interacts with piRNA precursors

To test if Armi interacts with piRNA precursors, we immunoprecipitated germline-specific FM-Armi overexpressed in the wild-type background, and cloned RNAs bearing a 5′-monophosphate and longer than 200 nt for deep sequencing (degradome-seq (Han et al., 2015a)). A 5′-monophosphated end is the cleavage signature of both Argonaute proteins and the phasing endonuclease. The ovaries were crosslinked with PFA to stabilize protein-RNA interactions and to allow stringent washes (high salt and ionic detergent) for background reduction. Ovaries that lack the UAS-Armi transgene served as a negative control. Because the FM-Armi being immunoprecipitated is only expressed in the germline, we used the reads mapping uniquely to the somatic follicle cell-specific piRNA cluster *flam* to normalize Armi IP degradome libraries. Three lines of evidence suggest that Armi interacts with piRNA precursors. First, transposon-mapping reads are enriched in the Armi IP (Figure 2.8), in which genic mRNA degradation products were also enriched (discussed below). Second, the Armi IP enriches for antisense transposon reads (Armi IP: 61% of the transposon mappers are antisense; Control IP: 45%). Third, antisense transposon mappers from the Armi IP, but not the control IP, display 5′-uridine nucleotide bias, a hallmark for primary piRNA 5′-ends (Figure 2.9).

Figure 2.8

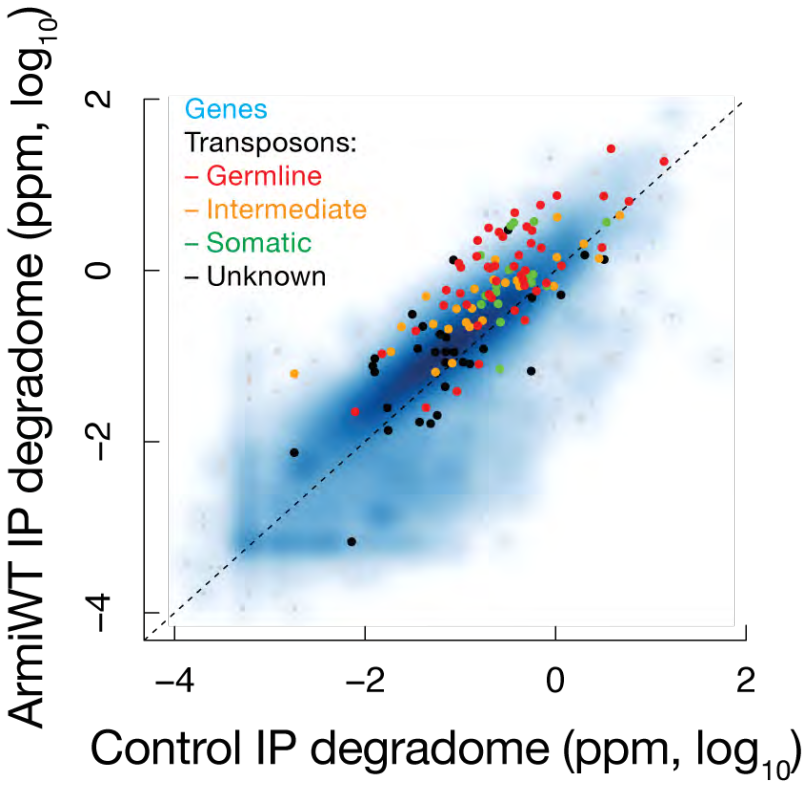

Genes
Transposons:
– Germline
– Intermediate
– Somatic
– Unknown
ArmiWT IP degradome (ppm, log10)
Control IP degradome (ppm, log10)
2
0
−2
−4
−4
−2
0
2

Figure 2.8: Armi interacts with piRNA precursors

Scatterplot showing the abundance of transposon- or gene-mapping degradome 5′-ends in ArmiWT IP versus Control IP in the wild-type background. Each dot represents a transposon family (red, germline-specific, 46 families; orange, intermediate, 25 families; green, soma-specific, 17 families; black, unknown, 33 families, a total of 121 transposon families (Wang et al., 2015)). Genes are shown in blue in a density plot.

Figure 2.9

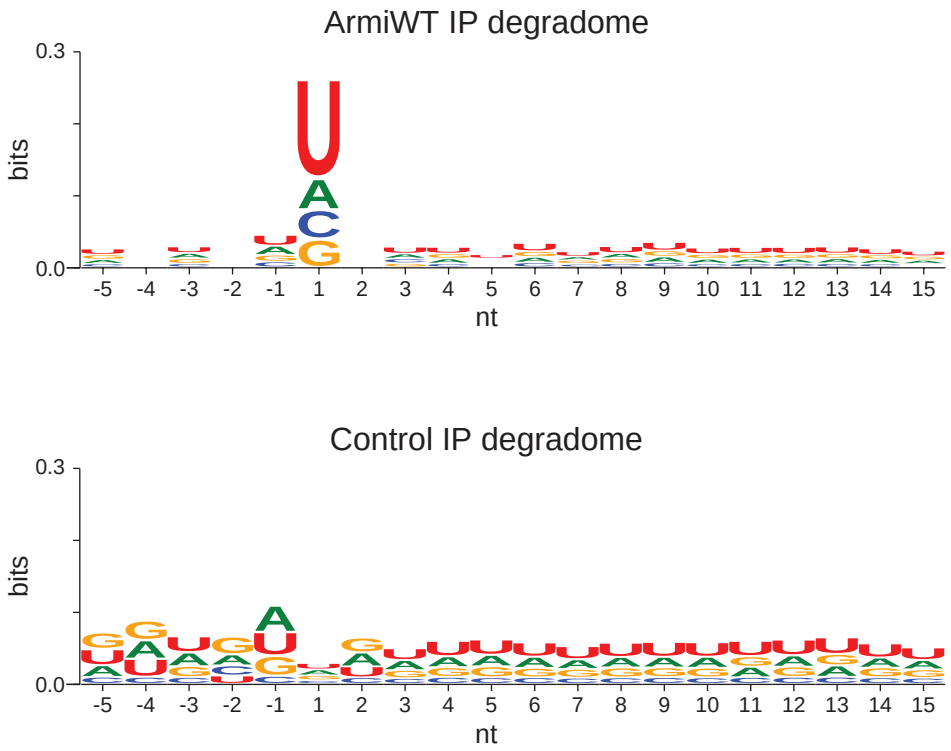

ArmiWT IP degradome
bits
0.3
0.0
-5 -4 -3 -2 -1 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
nt
Control IP degradome
bits
0.3
0.0
-5 -4 -3 -2 -1 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
nt

Figure 2.9: Armi IP degradome 5′-ends display uridine bias

Nucleotide bias of genomic sequences surrounding the degradome 5′-ends of Armi IP or Control IP (nt position 1). Analysis is restricted to antisense transposon mappers. Each 5′-end is used only once. Information content is shown in bits.

Armi interacts with phasing intermediates

According to the current model of piRNA biogenesis in *Drosophila* germline, the piRNA precursor is first cleaved by ping-pong partners (Aub or Ago3) in the nuage, then cleaved multiple times by Zuc on mitochondria with 5′-to-3′ directionality, where the first piRNA is loaded into Aub or Ago3 (Aub is preferred over Ago3), and the remaining piece is processively made into multiple phased piRNAs that load into Piwi or Aub (Piwi is preferred over Aub). If Armi transports precursor from nuage to mitochondria, we can predict that Armi interacts with piRNA precursors (long transposon-mapping degradome reads) of two types: 1) RNAs whose 5′-ends are made by a ping-pong partner (and will produce Aub- or Ago3-bound piRNAs, but NOT Piwi-bound piRNAs, directly from its 5′-end), referred to as “ping-pong intermediates”; and 2) RNAs whose 5′-ends are made by Zuc (and will produce Piwi- or Aub-bound piRNAs directly from its 5′-end), referred to as “phasing intermediates”.

To test if Armi binds to phasing intermediates, we aligned all transposon-mapping, Armi IP degradome 5′-ends onto the 0 position of an index (x-axis) and mapped Piwi-bound piRNAs (Han et al., 2015a) to it (y-axis). This reveals which position of the degradome more frequently coincides with the 5′-end of Piwi-bound piRNAs. The sharp peak at the 0 position indicates that Piwi-bound piRNAs are more likely to originate from the same 5′-end as the Armi-bound degradome reads (Figure 2.10A). The control IP degradome 5′-ends, sequenced

using FLAG IP eluate from ovaries that lack the UAS-Armi transgene, were also aligned into an index. As expected, the control IP degradome does not more frequently coincide with Piwi-bound piRNAs at the 5′-end (Figure 2.10B). Therefore, Armi binds to phasing intermediates.

Figure 2.10

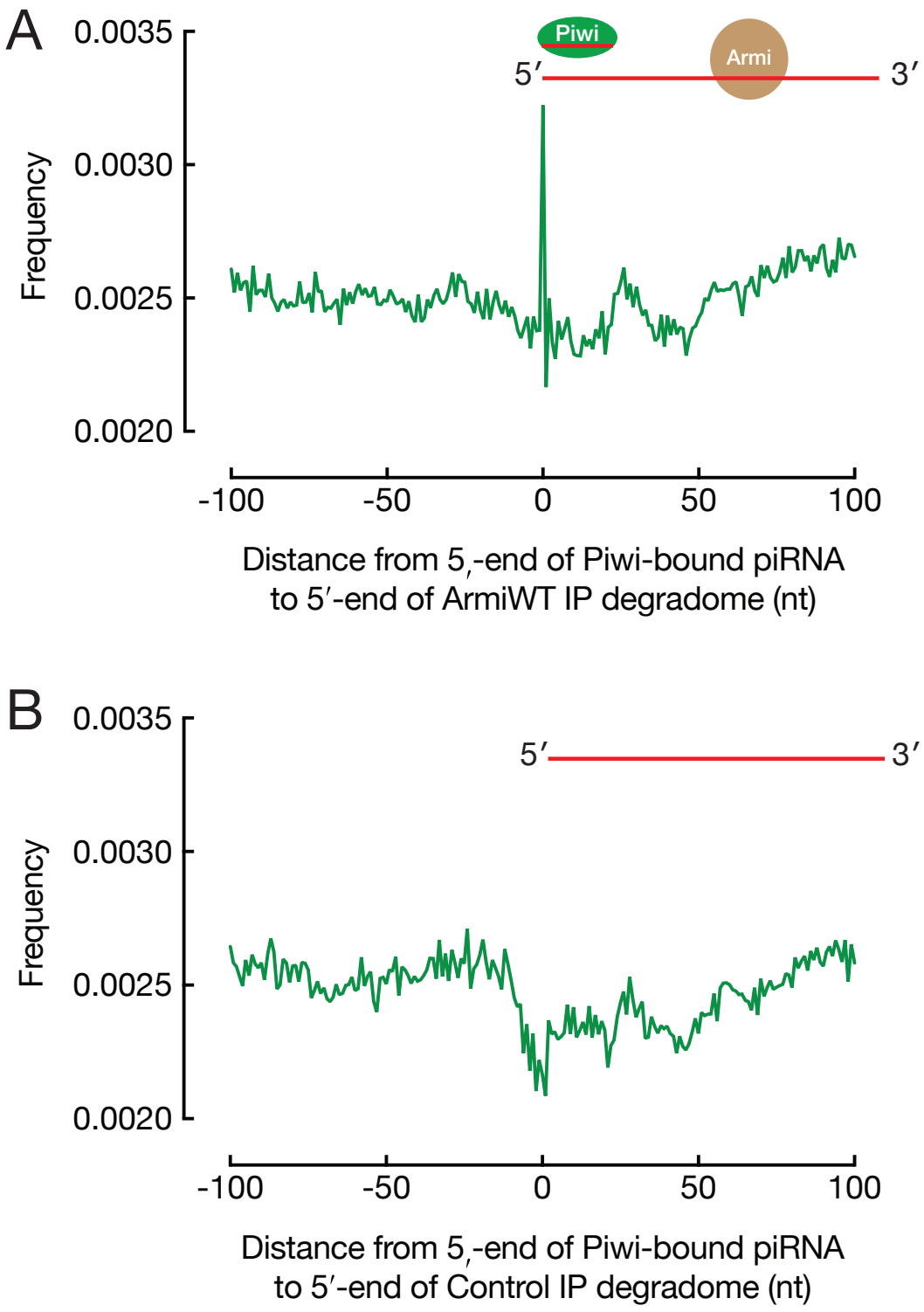

A
0.0035
0.0030
0.0025
0.0020
Frequency
Piwi
Armi
5′
3′
-100
-50
0
50
100
Distance from 5′-end of Piwi-bound piRNA
to 5′-end of ArmiWT IP degradome (nt)
B
0.0035
0.0030
0.0025
0.0020
Frequency
5′
3′
-100
-50
0
50
100
Distance from 5′-end of Piwi-bound piRNA
to 5′-end of Control IP degradome (nt)

Figure 2.10: Armi IP degradome reads frequently share 5′-ends with Piwi-bound piRNA

(A) Distance from the 5′-end of Piwi-bound piRNA to the 5′-end of Armi IP degradome reads on the same genomic strand. (B) Distance from the 5′-end of Piwi-bound piRNA to the 5′-end of control IP degradome reads on the same genomic strand. All datasets are from the wild-type genetic background.

Armi interacts with ping-pong intermediates

Similarly, to test if Armi binds to ping-pong intermediates, we mapped Aub-bound piRNAs (Han et al., 2015a) to the index constructed from either Armi IP or Control IP degradome, to ask which position of the degradome more frequently coincides with the 5′-end of Aub piRNAs. Despite of our best efforts to remove background-binding of RNA to the antibody or beads during immunoprecipitation, ping-pong intermediates that are already abundant in the background were still recovered in the control IP: Aub-bound piRNAs coincide frequently with the 5′-end of the control IP degradome (Figure 2.11B). The peak at position 0 in the Armi IP degradome index rises to a similar height, suggesting that the Armi IP does not enrich ping-pong intermediates above background (Figure 2.11A).

Both Aub- and Piwi-bound piRNAs predominantly begin with uridine (1U), so it is expected that ping-pong intermediates that produce Aub-bound piRNAs and phasing intermediates that produce Piwi-bound piRNAs both begin with uridine. It is therefore surprising that the control IP transposon-mapping degradome reads, which contain ping-pong intermediates that share 5′-end with Aub-bound piRNAs (Figure 2.11B), lack a 1U-bias (Figure 2.9). The discrepancy may be explained by that the percentage of Aub-piRNA-producing ping-pong intermediates in the control IP library is low (27% of the species and 30% of the reads), which is diverse enough to produce a peak that shares with Aub-bound piRNAs in index mapping analysis, but not abundant enough to display the 1U-

bias among all other reads that do not begin with uridine. If that is true, the 1U-bias detected in the Armi IP (Figure 2.9) is mainly contributed by phasing intermediates that give rise to Piwi-bound piRNAs (Figure 2.10A), and therefore suggests that the Armi IP library is predominantly occupied with phasing intermediates.

Figure 2.11

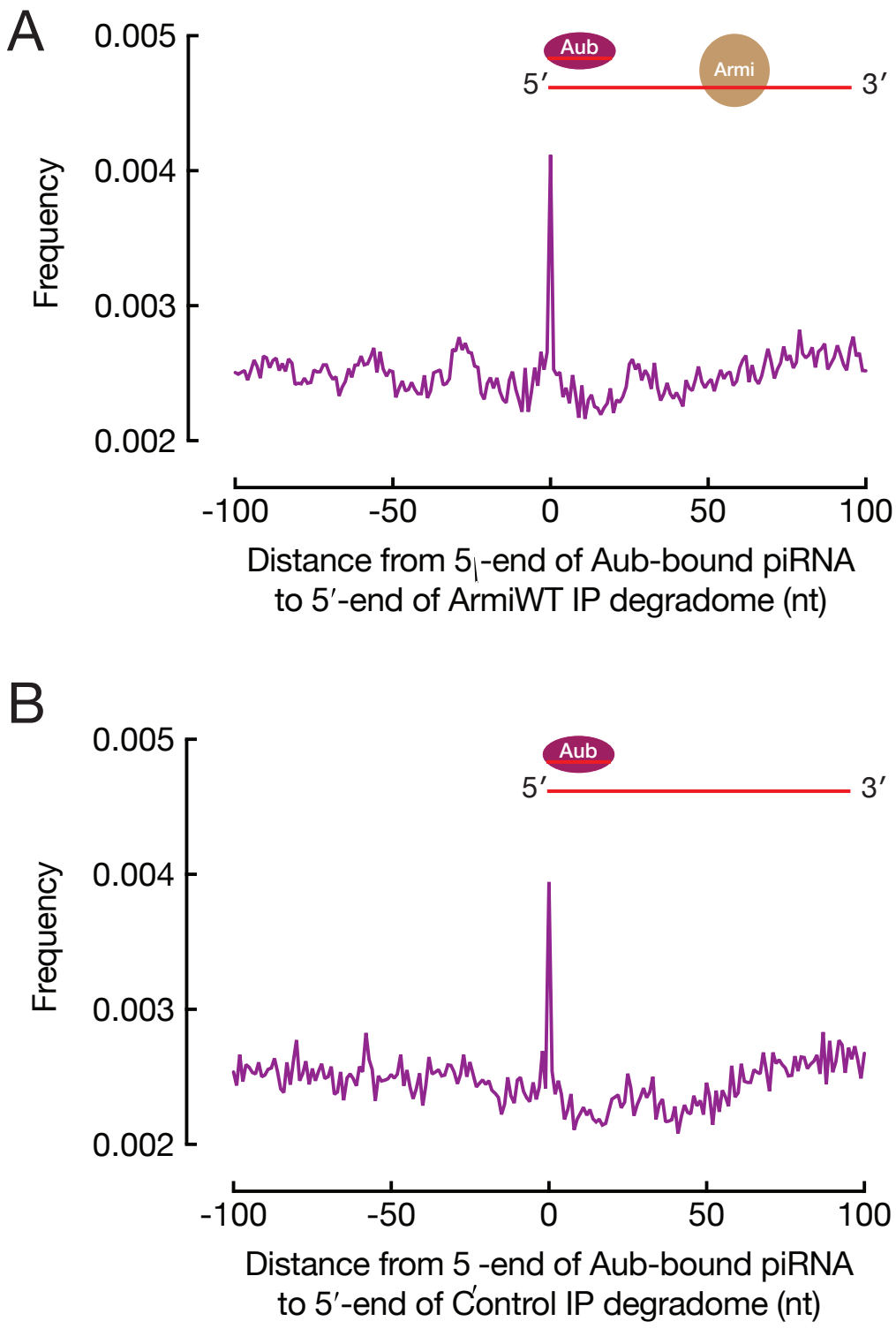

A
0.005
0.004
0.003
0.002
Frequency
Aub
Armi
5′
3′
-100
-50
0
50
100
Distance from 5′-end of Aub-bound piRNA
to 5′-end of ArmiWT IP degradome (nt)
B
0.005
0.004
0.003
0.002
Frequency
Aub
5′
3′
-100
-50
0
50
100
Distance from 5′-end of Aub-bound piRNA
to 5′-end of Control IP degradome (nt)

Figure 2.11: Armi IP degradome does not frequently share 5′-ends with Aub-bound piRNAs in the wild-type background

(A) Distance from the 5′-end of Aub-bound piRNA to the 5′-end of Armi IP degradome reads on the same genomic strand. (B) Distance from the 5′-end of Aub-bound piRNA to the 5′-end of control IP degradome reads on the same genomic strand. All datasets are from the wild-type genetic background.

Our data suggest that in the wild-type background, Armi is predominantly engaged with phasing intermediates, and rarely associates with ping-pong intermediates. This low signal, in combination with the abundant ping-pong intermediates in the background (i.e., high noise), could be the reason why ping-pong intermediates are not enriched in the IP. To increase the signal to noise ratio, we genetically removed phasing intermediates by performing Armi IP in fly ovaries that only express catalytically inactive Zuc (*zuc*^{H169Y}, referred to as *zucCD* hereinafter), which gets rid of the ability to make Piwi-bound, phased piRNAs (Han et al., 2015a).

We then asked if Armi binds to piRNA precursors in the *zucCD* background. Transposon-mapping reads occupied 29.5% of the library in the Armi IP degradome (of which 42.1% are antisense) and 22.9% in the control IP degradome (of which 29.1% are antisense). Furthermore, reads mapping to almost all transposon families are higher in the Armi IP than the control IP (Figure 2.12). Note that compared to the wild-type background, the percentage of transposon-mapping degradome reads goes up dramatically in the *zucCD* background, likely caused by both transposon activation (increase in sense transposon-mapping RNA) and piRNA pathway inactivation (increase in piRNA precursor intermediates, many of which are antisense to transposons).

Mapping Piwi-bound piRNAs from the wild-type background to Armi IP degradome from the *zucCD* background reveals a complete loss of the peak at 0

position, confirming the absence of phasing intermediates (Figure 2.13A). Importantly, the mapping of Aub-bound piRNAs instead reveals a higher peak at 0 position in Armi IP than Control IP, indicating that Armi does associate with ping-pong intermediates on *zucCD* background (Figure 2.13B). Consistently, the +27 nt peak in Piwi-bound piRNA mapping is more prominent in Armi IP than Control IP (Figure 2.13A), indicating that Armi binds to intermediates that could have produced Piwi-bound piRNAs about one piRNA length (~27 nt) downstream of its 5′-end, if Zuc is active (i.e., ping-pong intermediates). Therefore, we conclude that Armi does bind to ping-pong intermediates.

Taken together, our data indicate that Armi interacts with both ping-pong intermediates and phasing intermediates, in agreement with its role in shuttling ping-pong-cleaved precursors from nuage to mitochondria.

Figure 2.12

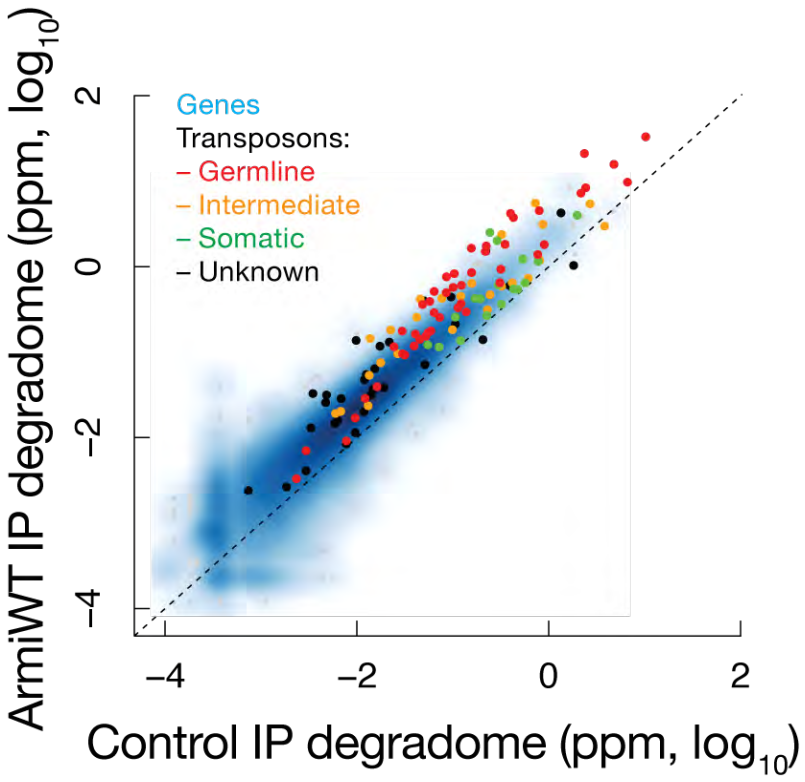

Genes
Transposons:
– Germline
– Intermediate
– Somatic
– Unknown
ArmiWT IP degradome (ppm, log10)
Control IP degradome (ppm, log10)
2
0
−2
−4
−4
−2
0
2

Figure 2.12: Armi interact with piRNA precursors on *zucCD* background

Scatterplot showing the abundance of transposon- or gene-mapping degradome 5′-ends in the Armi IP versus the control IP from the *zucCD* background. Transposon grouping is as in Figure 2.8.

Figure 2.13

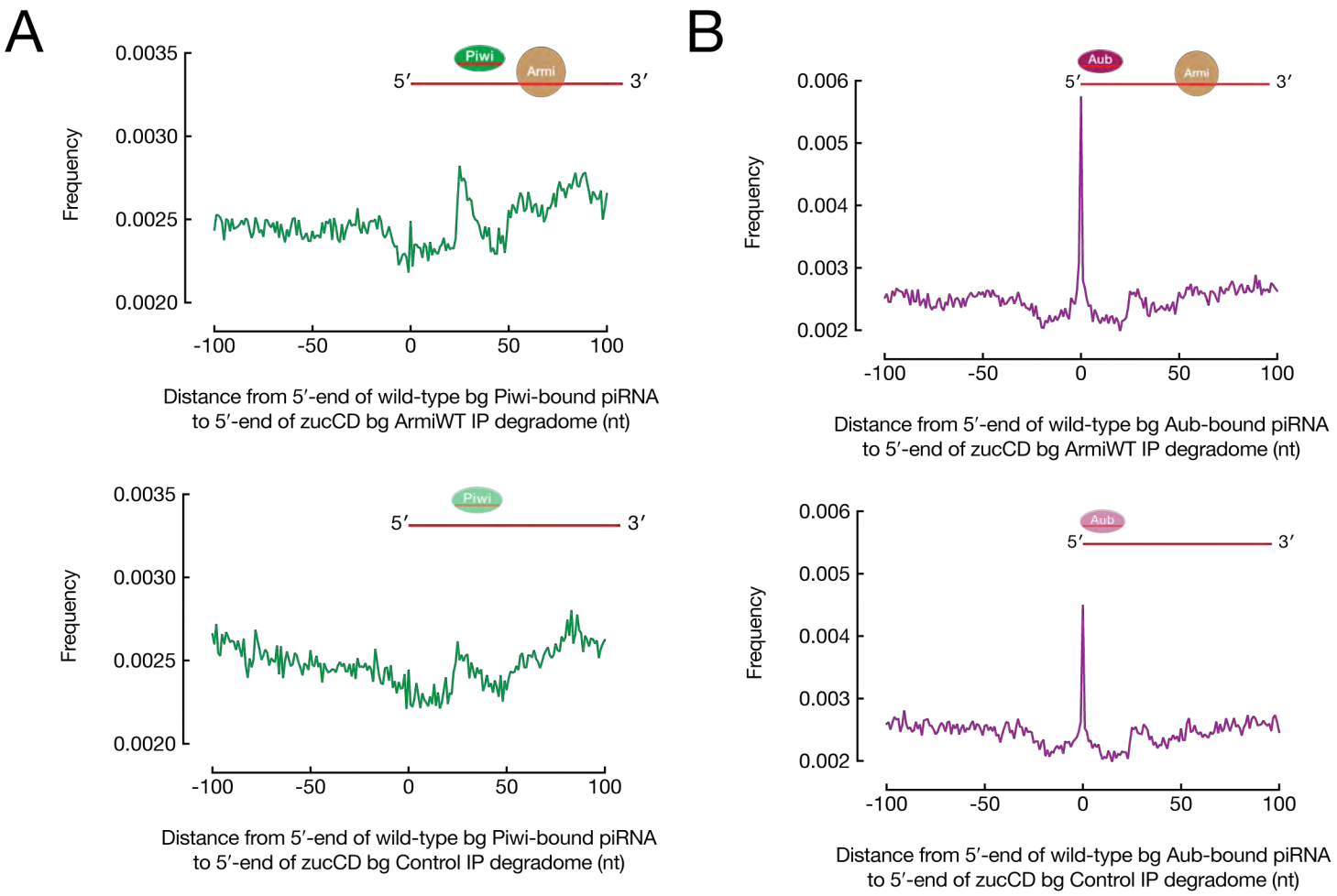

A
B
Piwi
Armi
Aub
5′
3′
Frequency
0.0035
0.0030
0.0025
0.0020
0.006
0.005
0.004
0.003
0.002
-100
-50
0
50
100
Distance from 5′-end of wild-type bg Piwi-bound piRNA to 5′-end of zucCD bg ArmiWT IP degradome (nt)
Distance from 5′-end of wild-type bg Aub-bound piRNA to 5′-end of zucCD bg ArmiWT IP degradome (nt)
Distance from 5′-end of wild-type bg Piwi-bound piRNA to 5′-end of zucCD bg Control IP degradome (nt)
Distance from 5′-end of wild-type bg Aub-bound piRNA to 5′-end of zucCD bg Control IP degradome (nt)

Figure 2.13: Armi IP degradome reads from zucCD background frequently share 5′-ends with Aub-bound piRNAs, but not Piwi-bound piRNAs

(A) Distance from the 5′-end of Piwi-bound piRNA from the wild-type background to the 5′-end of Armi IP or control IP degradome reads from the *zucCD* background. (B) Distance from the 5′-end of Aub-bound piRNA from the wild-type background to the 5′-end of Armi IP or control IP degradome reads from the *zucCD* background. Note that piRNA and piRNA precursors are cloned from different genetic backgrounds.

Armi localization is regulated by phasing activity

Armi localizes to nuage and mitochondria, binds to proteins and piRNA precursors in both compartments, but its mutant only affects the mitochondrial phase of piRNA production, not the nuage ping-pong cycle (Han et al., 2015a). Since ping-pong cycle products (ping-pong intermediates) are the starting material for phasing, the *armi* loss of function phenotype fits its proposed role as the coupler. It can be predicted that, in the absence of phasing (e.g., when Zuc is catalytically inactive), Armi may remain stuck with the unprocessed ping-pong intermediates on the mitochondrial surface.

To test this hypothesis, we comprehensively examined localization of Armi relative to nuage or mitochondria in wild-type or *zucCD* mutant ovaries. Three nuage markers, Aub, Ago3 and Vas, and two mitochondrial markers, Zuc and ATP5A, were each co-stained with Armi. Consistent with Figure 2.7, Armi colocalized with all the nuage and mitochondrial markers to a great extent in wild-type ovaries (Figure 2.14A, Figure 2.14B). In contrast, Armi was much less colocalized with nuage markers Aub, Ago3 or Vas in *zucCD* mutants, but still extensively colocalized with the mitochondrial markers (Figure 2.15A, Figure 2.15B, quantified in Figure 2.17). To test if a similar phenomenon holds true for other phasing mutants, we co-stained Armi with Aub or Zuc in *minotaur* mutants, which phenocopied *zucCD* with regard to Armi localization (Figure 2.16,

quantified in Figure 2.17). This suggests that the change in Armi localization is due to inactivation of the phasing pathway instead of specific mutations.

In agreement with our findings, it was previously shown that when the upstream ping-pong pathway is inactivated by mutating the catalytic residues of Ago3, Armi is trapped with the inactive Ago3 in nuage (Huang et al., 2014). Taken together, these results suggest that a dynamic distribution of Armi between nuage and mitochondria is maintained by active piRNA production in both nuage and mitochondria.

Figure 2.14

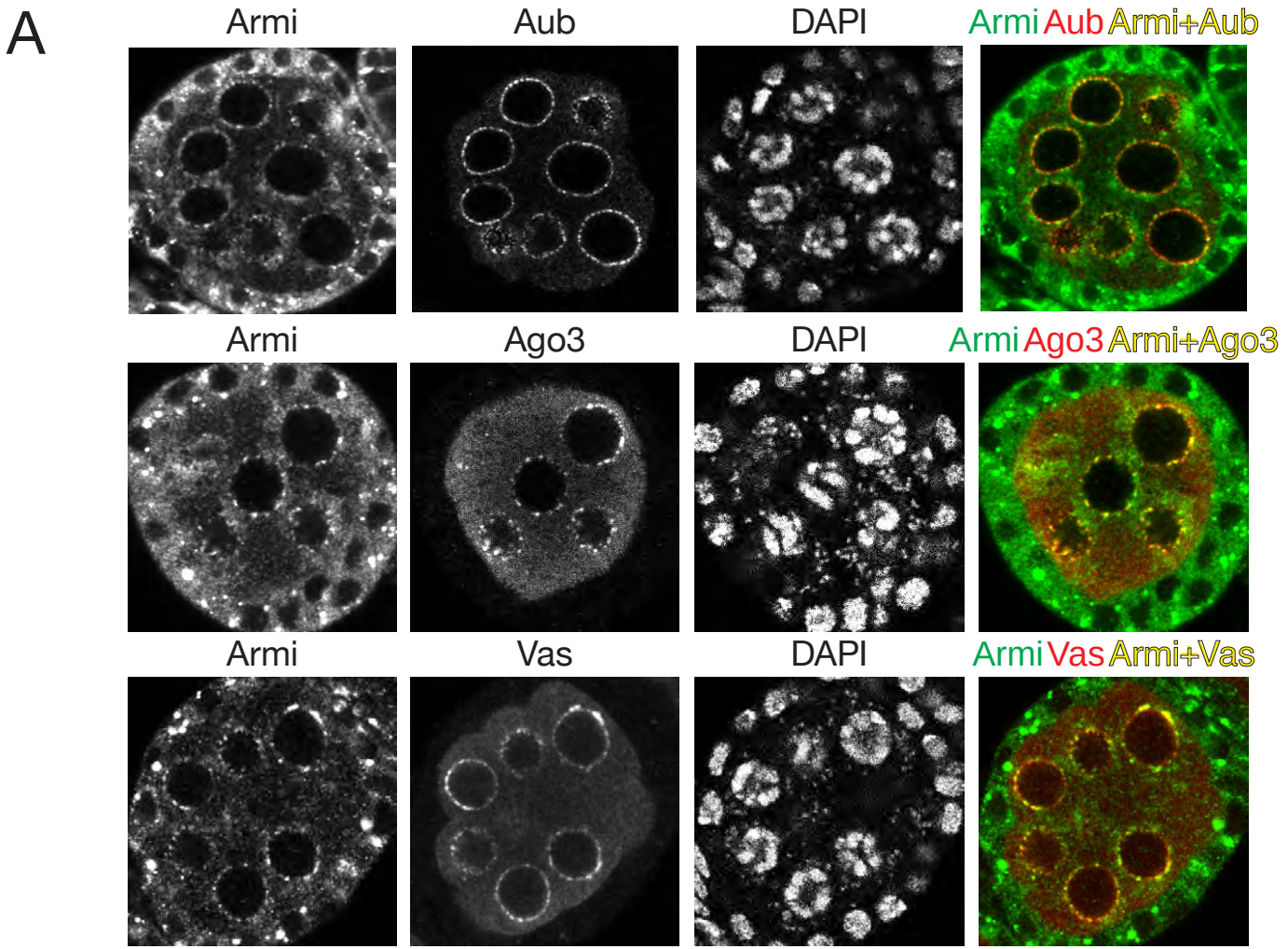

A
Armi
Aub
DAPI
Armi Aub Armi+Aub
Armi
Ago3
DAPI
Armi Ago3 Armi+Ago3
Armi
Vas
DAPI
Armi Vas Armi+Vas

Figure 2.14

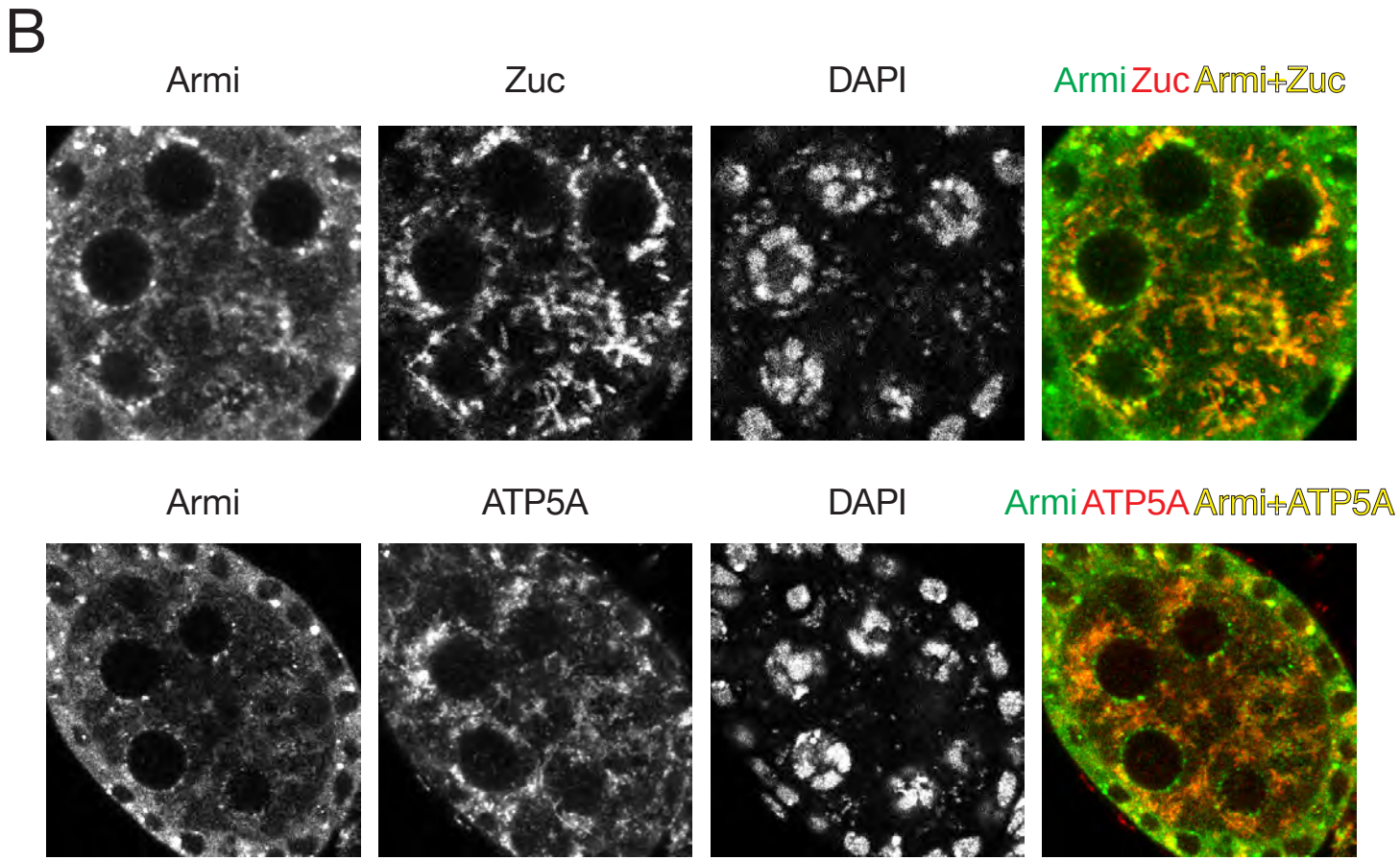

B
Armi
Zuc
DAPI
Armi Zuc Armi+Zuc
Armi
ATP5A
DAPI
Armi ATP5A Armi+ATP5A

Figure 2.14: Armi colocalizes with nuage and mitochondrial markers in wild-type ovaries

(A) Immunofluorescence detection of Armi with Aub, Ago3, Vas, and (B) Armi with Zuc-3×FLAG, ATP5A in stage 3 egg chambers. Each channel and the overlapping signals are color-coded.

Figure 2.15

A

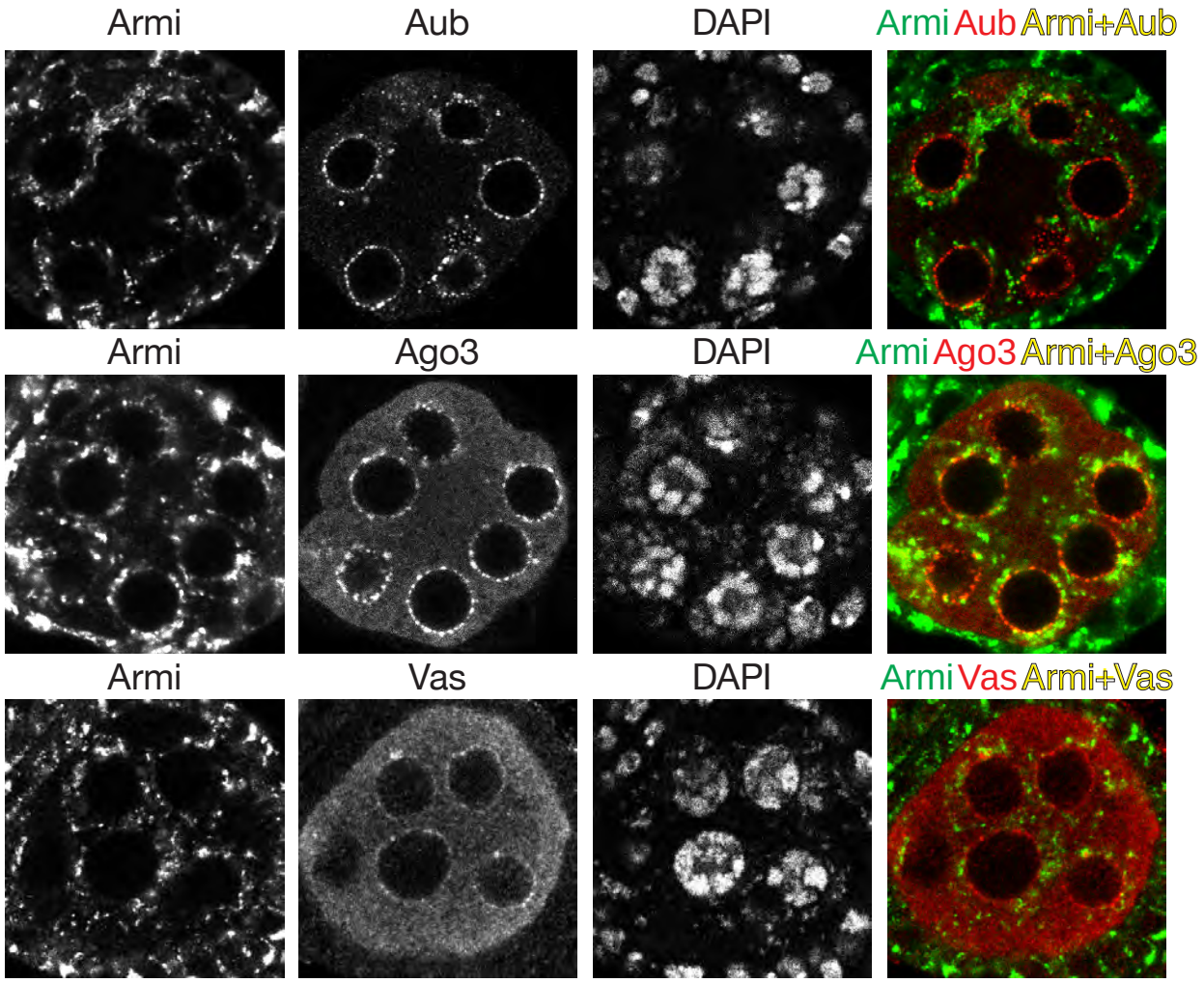

Armi
Aub
DAPI
Armi Aub Armi+Aub
Armi
Ago3
DAPI
Armi Ago3 Armi+Ago3
Armi
Vas
DAPI
Armi Vas Armi+Vas

Figure 2.15

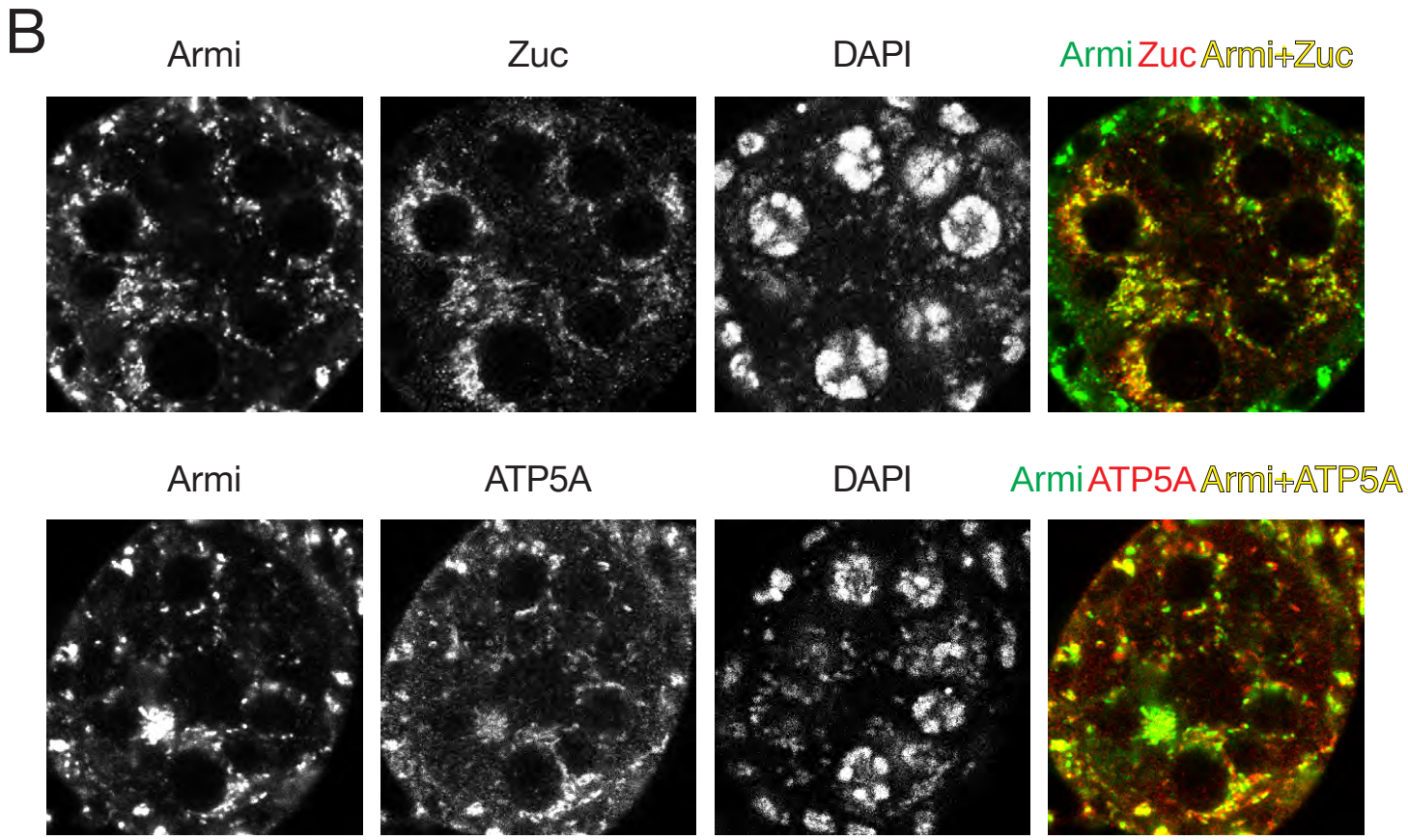

B
Armi
Zuc
DAPI
Armi Zuc Armi+Zuc
Armi
ATP5A
DAPI
Armi ATP5A Armi+ATP5A

Figure 2.15: Armi colocalization with nuage markers, but not mitochondrial markes, is reduced in *zucCD* ovaries

See Figure 2.14 for details.

Figure 2.16

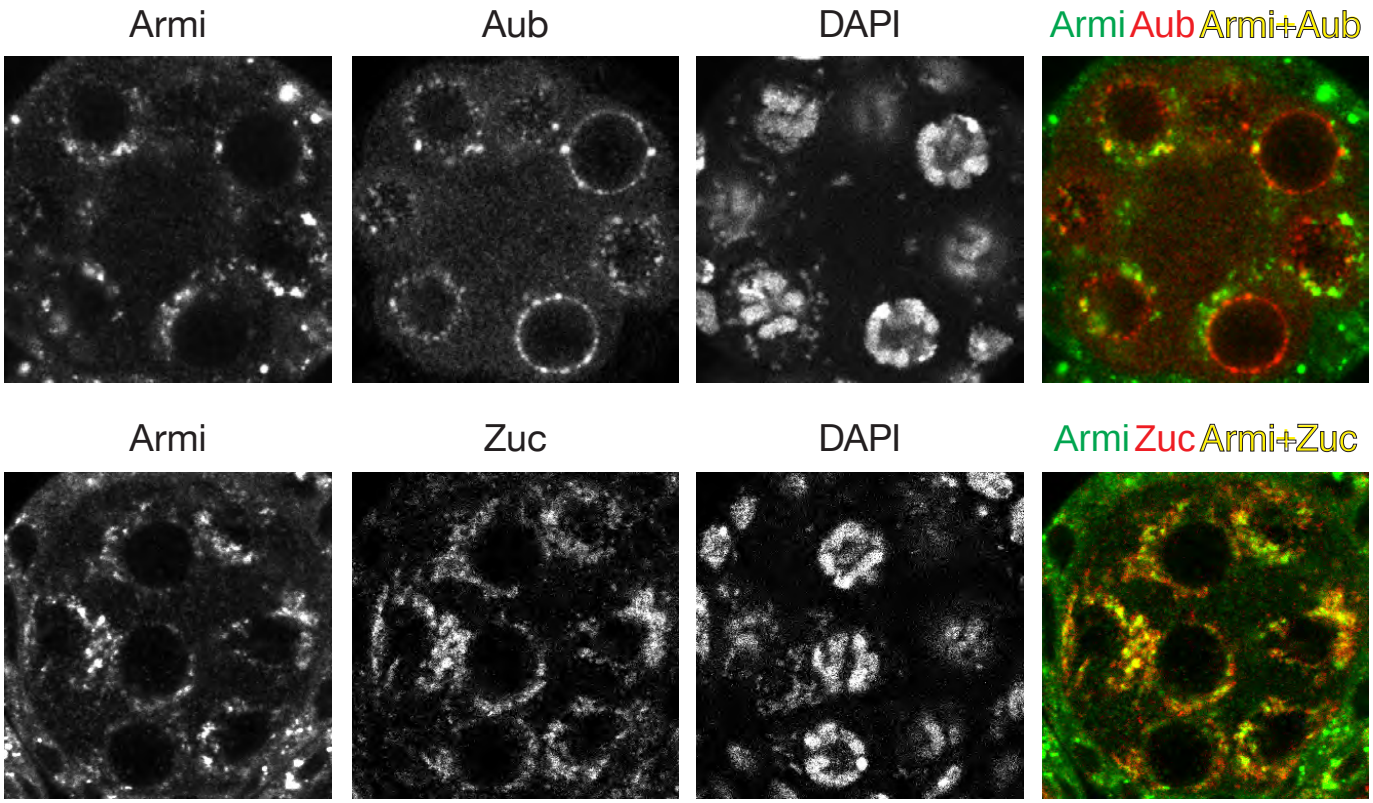

Armi
Aub
DAPI
Armi Aub Armi+Aub
Armi
Zuc
DAPI
Armi Zuc Armi+Zuc

Figure 2.16: Armi colocalization with Aub, but not Zuc, is reduced in *minotaur* ovaries

Immunofluorescence detection of Armi, Aub, Zuc-3×FLAG and nucleic acids (DAPI) in stage 3 egg chambers. Each channel and the overlapping signal are color-coded.

Figure 2.17

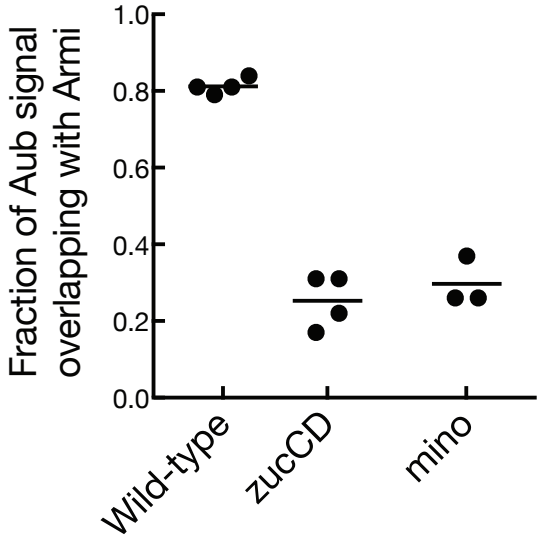

Fraction of Aub signal overlapping with Armi
1.0
0.8
0.6
0.4
0.2
0.0
Wild-type
zucCD
mino

Figure 2.17: Aub immunofluorescence signal that overlapping with Armi decreases in *zucCD* or *mino* mutants

Aub and Armi objects were computationally recognized using a custom script of the CellProfiler program (see *Experimental Procedures*). The amount of Aub immunofluorescence signal in areas that overlapped with Armi objects was divided by the sum of signal in all Aub objects in the same image. Confocal z-axis serial scans were taken at 1 µm intervals. 16 to 23 serial z scan images were analyzed for each stage 3 egg chamber and the average for the whole egg chamber is plotted. Each dot represents one egg chamber.

Armi ATPase activity is required for piRNA production

While the N-terminal half of Armi is not well conserved and does not contain predictable domains, the C-terminal half of Armi harbors a conserved helicase domain. By motif analysis, it belongs to the SF1A (Upf1-like) family. We modeled the 3-D structure of the Armi helicase core using the I-TASSER server (Roy et al., 2010), which predicts a structure that is highly similar to the published human Upf1 helicase core (Cheng et al., 2007) (root-mean-square deviation of atomic positions (RMSD) = 0.91 Å) (Figure 2.18). Upf1 is a core member of the nonsense mediated RNA decay (NMD) pathway, whose loss-of-function mutants have been characterized biochemically for ATPase or helicase activities (Weng et al., 1996). Changing the lysine 498 in the highly conserved motif I of the human Upf1 protein to alanine abolished its ATP binding and ATPase activities, but not RNA-binding in vitro (Cheng et al., 2007). The same Upf1 mutant binds RNA in cells, albeit losing the ability to discriminate between target and non-target mRNAs (Lee et al., 2015).

Figure 2.18

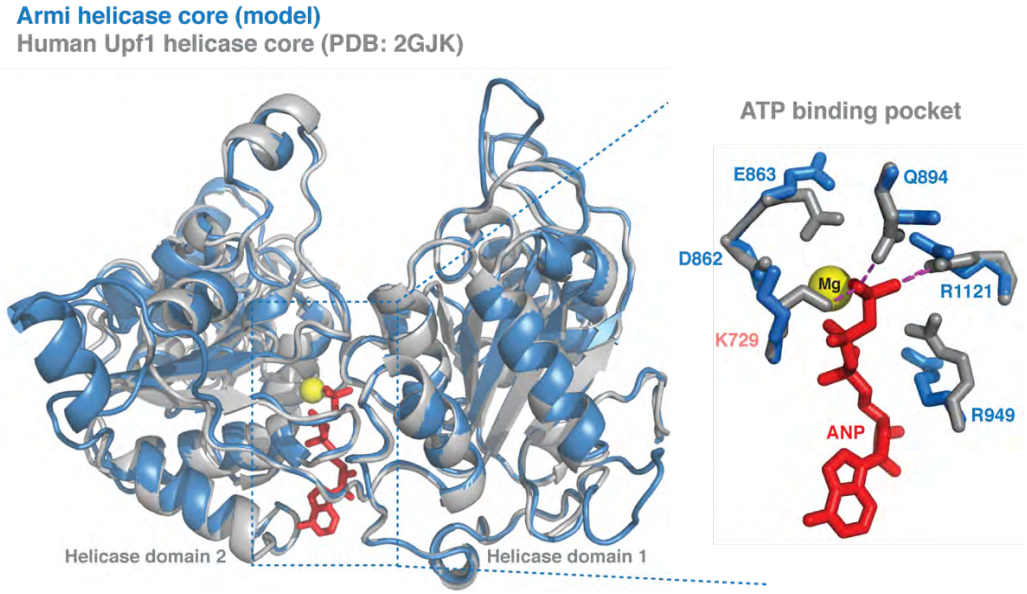

Armi helicase core (model)
Human Upf1 helicase core (PDB: 2GJK)
ATP binding pocket
E863
Q894
D862
Mg
R1121
K729
R949
ANP
Helicase domain 2
Helicase domain 1

Figure 2.18: Modeled structure of the Armi helicase core

The predicted helicase core, a 469 aa fragment at the C-terminal half of Armi protein (amino acid 692 to 1160) was submitted to the I-TASSER server. The modeled structure was superimposed on a published human Upf1 helicase core structure (PDB ID 2GJK) using PyMol v1.3 (Schrodinger, LLC). The ATP binding pocket of Armi was enlarged to show the residues surrounding the ANP and magnesium ion of the published Upf1 structure. Armi K729, D862 and E863 are the mutated residues in this study.

To ask whether Armi depends on the ATPase activity to function in piRNA biogenesis, we mutated the lysine at position 729 of the Armi protein to alanine, which corresponds to the lysine 498 of human Upf1. We expressed either the Armi wild-type cDNA or the K729A version specifically in the female germline that lacked endogenous Armi (germline null). ArmiK729A is expressed to similar levels as ArmiWT (Figure 2.19).

Figure 2.19

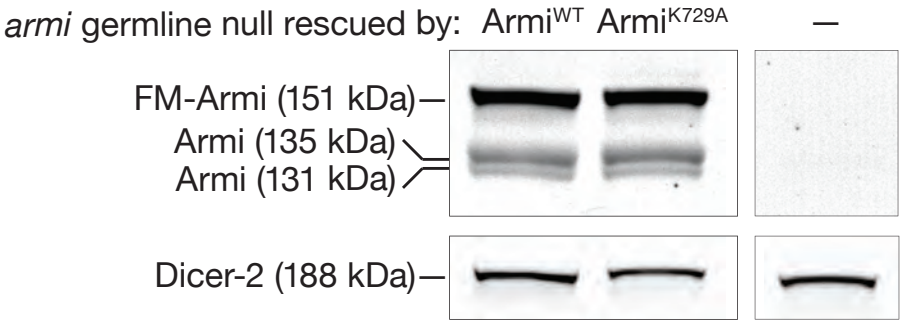

armi germline null rescued by: Armi^WT Armi^K729A —
FM-Armi (151 kDa)
Armi (135 kDa)
Armi (131 kDa)
Dicer-2 (188 kDa)

Figure 2.19: ArmiK729A and ArmiWT are similarly expressed on armi germline null mutant background

Western blot showing total protein lysate from ArmiWT rescue, ArmiK729A rescue or armi germline null ovaries (for genotypes see *Experimental Procedures*) probed with anti-Armi. Dicer-2 was used as a loading control.

Three lines of evidence show that ArmiK729A is unable to support piRNA production. First, while ArmiWT rescues steady-state piRNA abundance almost to wild-type levels, ArmiK729A does not support piRNA production above the level of the armi germline null (Figure 2.20A). Second, while ArmiWT partially rescues transposon silencing, ArmiK729A does not (Figure 2.20B). Third, ArmiWT, but not ArmiK729A, partially rescues female fertility as measured by the number of eggs laid and the percentage of eggs that hatch (Figure 2.21).

Figure 2.20

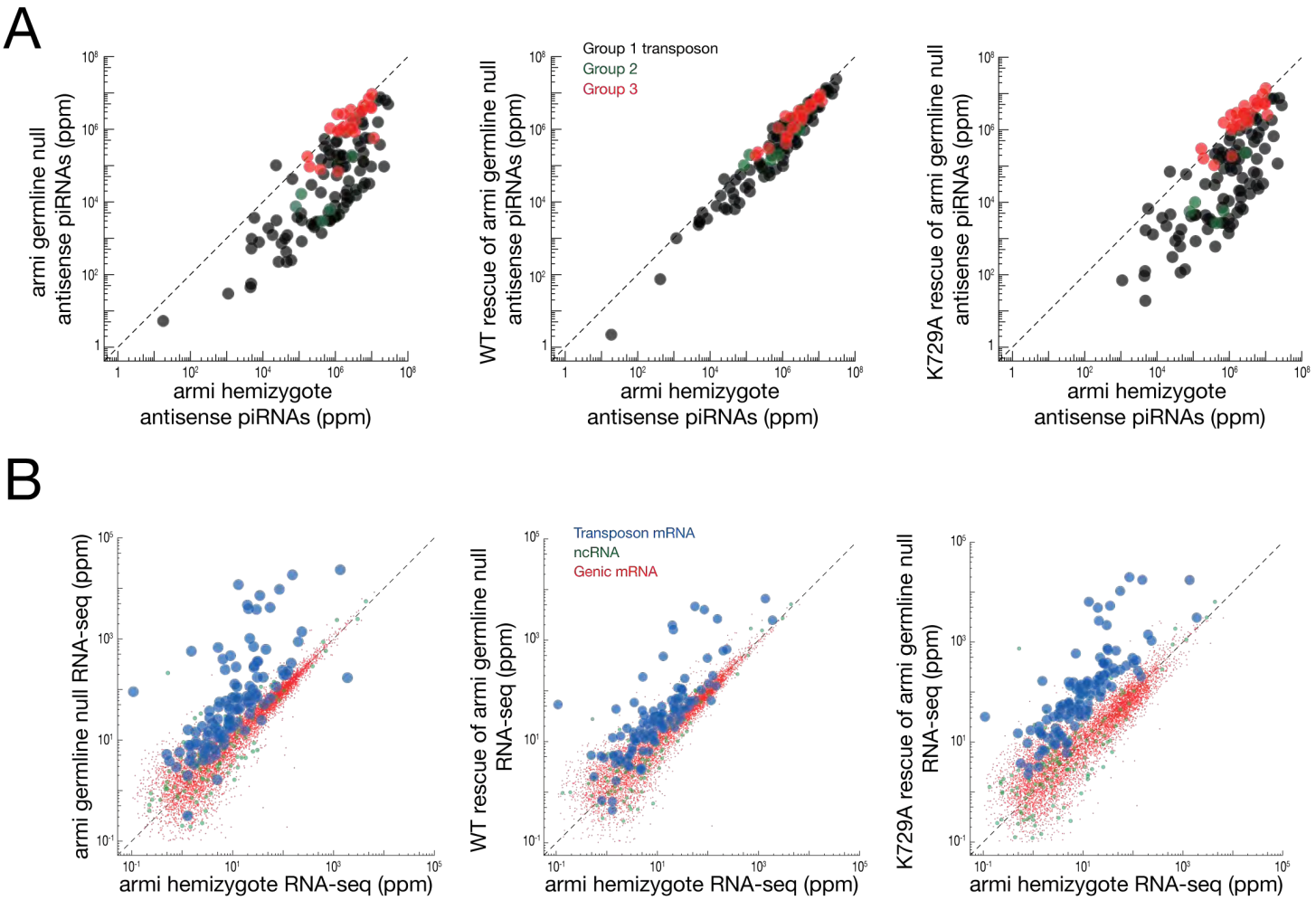

A
armi germline null antisense piRNAs (ppm)
armi hemizygote antisense piRNAs (ppm)
WT rescue of armi germline null antisense piRNAs (ppm)
armi hemizygote antisense piRNAs (ppm)
Group 1 transposon
Group 2
Group 3
K729A rescue of armi germline null antisense piRNAs (ppm)
armi hemizygote antisense piRNAs (ppm)
B
armi germline null RNA-seq (ppm)
armi hemizygote RNA-seq (ppm)
WT rescue of armi germline null RNA-seq (ppm)
armi hemizygote RNA-seq (ppm)
Transposon mRNA
ncRNA
Genic mRNA
K729A rescue of armi germline null RNA-seq (ppm)
armi hemizygote RNA-seq (ppm)

Figure 2.20: ArmiWT, but not ArmiK729A, partially rescues piRNA production and transposon silencing

(A) Scatterplot showing the level of piRNAs antisense to transposons in *armi* germline null, ArmiWT rescue, or ArmiK729A rescue ovaries comparing to *armi* hemizygous (*armi*$^{\Delta 1}$/+) control. *armi* hemizygous female flies have normal fertility and wild-type level of piRNAs. Each dot represents one transposon family, grouped according to (Li et al., 2009). (B) Scatterplot showing the level of long RNA in the genotypes described above. Each dot represents a transposon family, a non-coding RNA, or a genic mRNA. All genome mappers are shown and displayed as ppm (parts per million genome mappers).

Figure 2.21

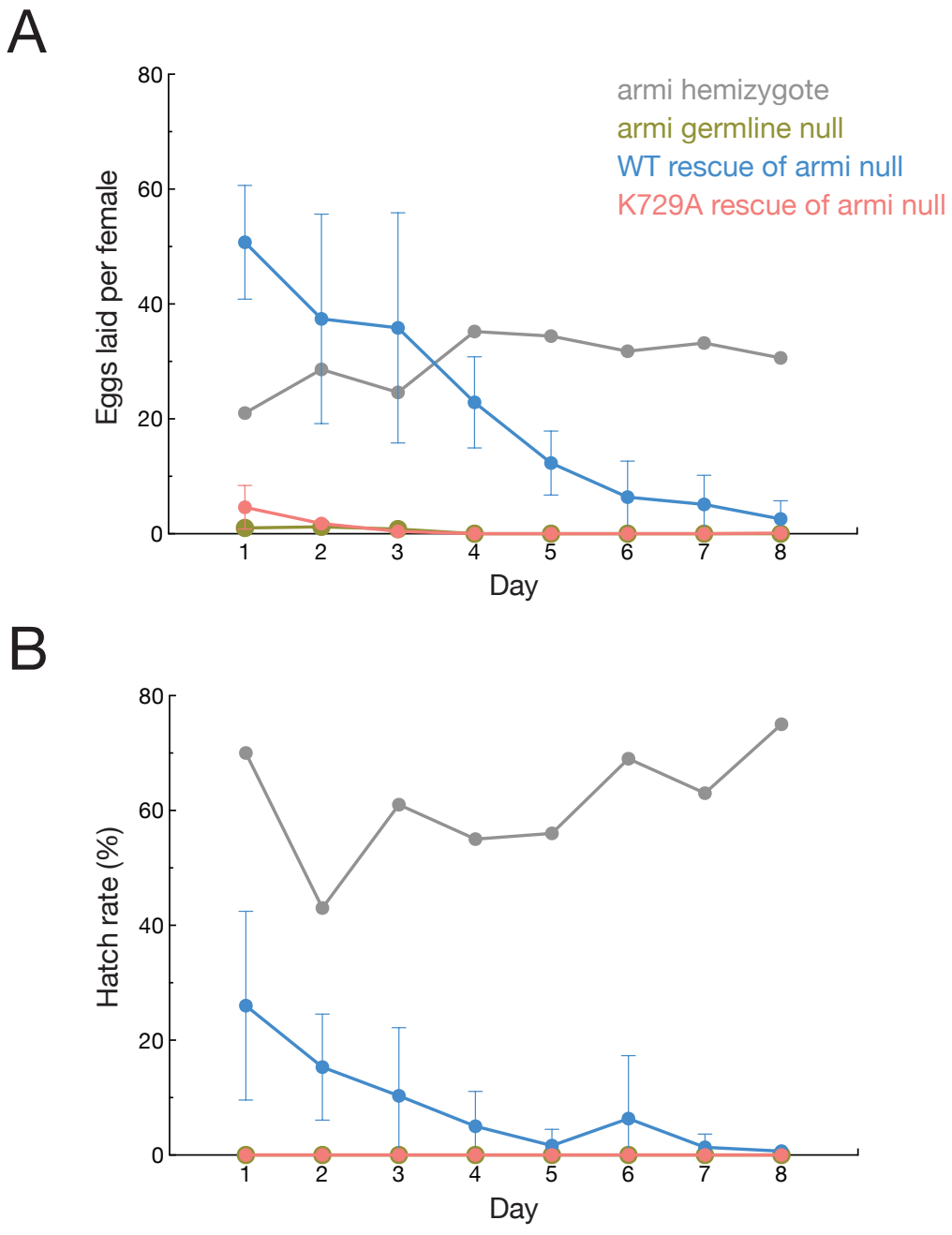

A
armi hemizygote
armi germline null
WT rescue of armi null
K729A rescue of armi null
Eggs laid per female
80
60
40
20
0
1
2
3
4
5
6
7
8
Day
B
Hatch rate (%)
80
60
40
20
0
1
2
3
4
5
6
7
8
Day

Figure 2.21: ArmiWT, but not ArmiK729A, partially rescues female fertility

(A) The number of eggs laid per female on each day of the fertility test is shown. (B) For eggs laid on each day of the fertility test, the percentage that hatch after 24 hours is shown.

Armi ATPase activity is required for localization

To understand why ArmiKA fails to rescue phased piRNA production, we examined its localization in nurse cells. Fly ovaries expressing Flag-Myc tagged ArmiWT, ArmiK729A, ArmiDE862AA or ArmiE863Q instead of the endogenous Armi were stained with anti-FLAG antibody (Figure 2.22). In contrast to the cloud-like aggregation of ArmiWT around the nucleus, all three ATPase mutants are dispersed in the cytoplasm. We hypothesized that the ATPase mutations affect either Armi-RNA interaction or Armi-protein interaction that normally stabilizes Armi in the nuage or on the mitochondrial surface.

Figure 2.22

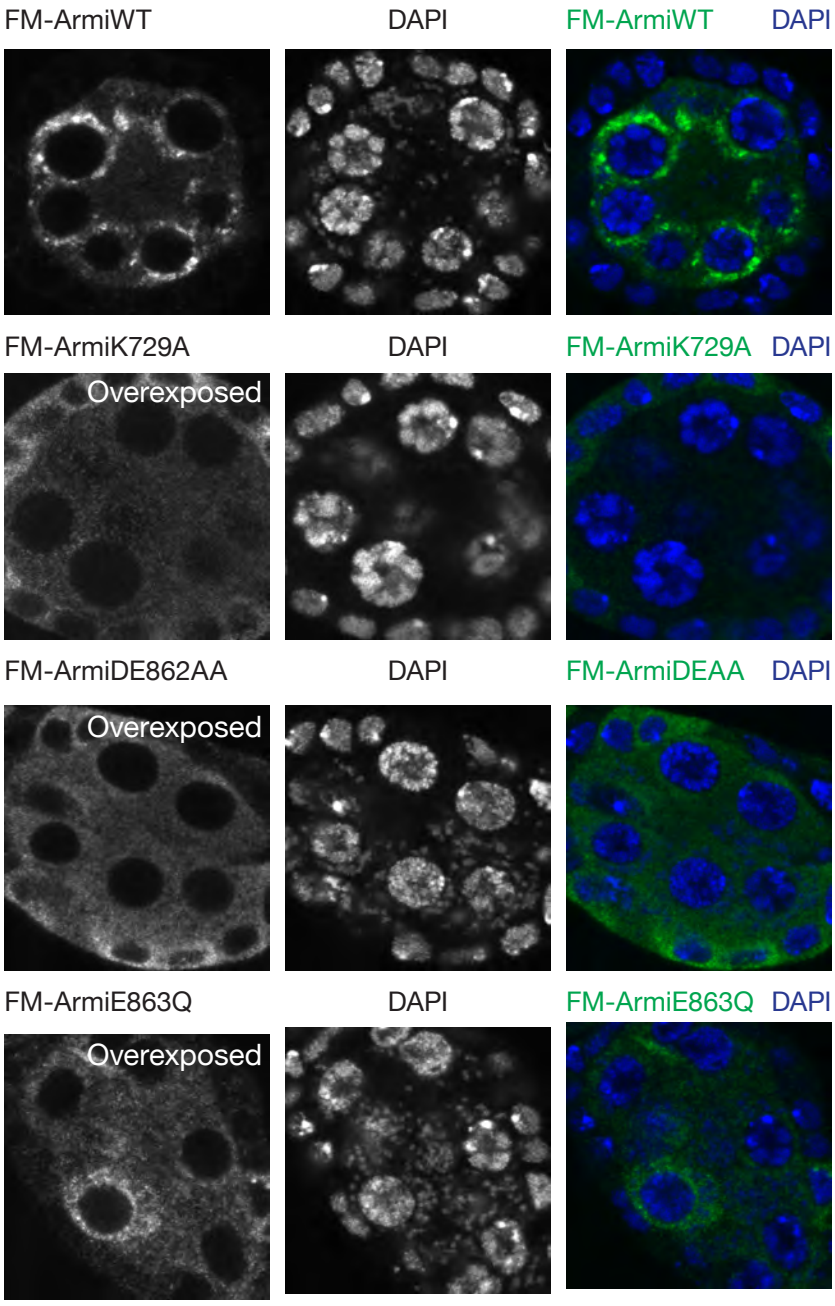

FM-ArmiWT
DAPI
FM-ArmiWT DAPI
FM-ArmiK729A
DAPI
FM-ArmiK729A DAPI
Overexposed
FM-ArmiDE862AA
DAPI
FM-ArmiDEAA DAPI
Overexposed
FM-ArmiE863Q
DAPI
FM-ArmiE863Q DAPI
Overexposed

Figure 2.22: Armi ATPase mutants are dispersed in the cytoplasm

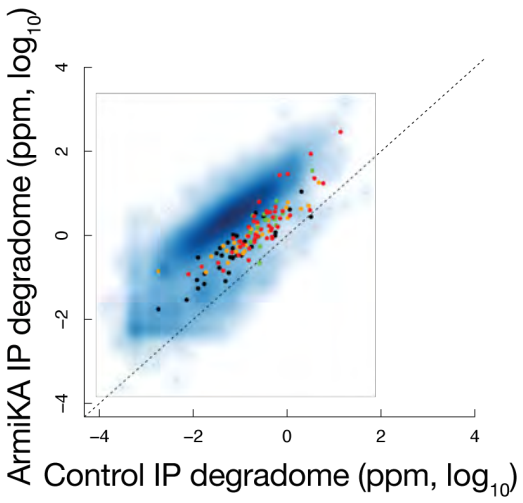
Immunofluorescence detection of Flag-Myc-Armi (anti-FLAG, green) and nucleic acids (DAPI, blue) in stage 3 egg chambers. The Flag-Myc-tagged Armi is the only Armi protein expressed in the germline. The single green channel images are overexposed to show Armi localization.

Armi ATPase activity is required for piRNA precursor selection

We first asked if the dispersed ArmiKA retains interaction with piRNA precursors. We immunoprecipitated germline-specific FM-ArmiK729A overexpressed in the wild-type genetic background (i.e., in the presence of endogenous wild-type Armi), extracted RNA from the IP eluate and constructed degradome libraries in the same way as in Figure 2.8. Degradome sequencing of ArmiKA immunoprecipitate revealed a similar degree of enrichment for transposon mappers as ArmiWT (Figure 2.23B), suggesting that ArmiKA retains piRNA precursor binding. Intriguingly, ArmiKA IP enriches for genic mRNA degradation products much more than transposon mappers (Figure 2.23A), suggesting that Armi ATPase mutant loses substrate selectivity. To test if the Armi RNA substrate discrimination depends on a full ATPase cycle or just on ATP binding, we overexpressed two other Armi transgenes in the wild-type background and studied their associated RNA: ArmiDE862AA (D862A, E863A), mutating the well-conserved motif II required for ATP hydrolysis but not ATP binding (Cheng et al., 2007); ArmiE863Q, shown in the DEAD-box protein Vasa to affect the release of ADP and Pi after ATP hydrolysis (Xiol et al., 2014). Degradome-sequencing from ArmiDEAA or ArmiEQ immunoprecipitates showed a similar enrichment for genic mRNA degradation products as ArmiKA (Figure 2.24), suggesting that not only ATP binding, but also ATP hydrolysis and ADP release steps are required to confer substrate selectivity.

Figure 2.23

A



B

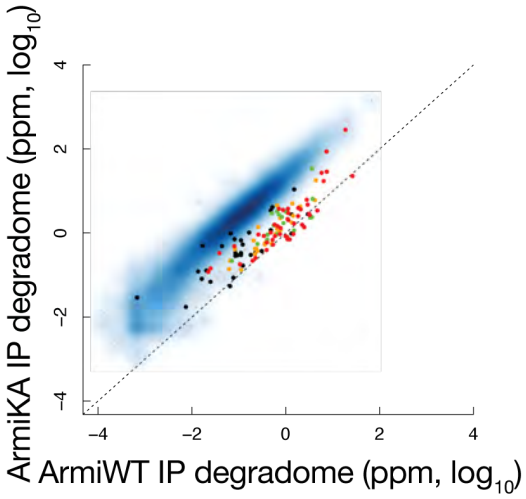
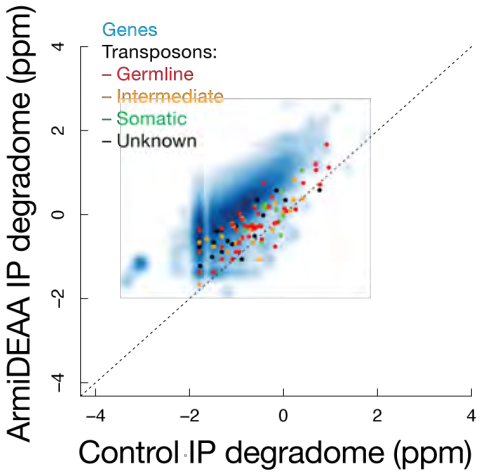


Figure 2.23: ArmiK729A gains promiscuous binding to genic mRNAs

Scatterplot showing the abundance of transposon- or gene-mapping degradome 5′-ends in ArmiKA IP versus control IP or ArmiWT IP in the wild-type background. Transposon grouping is as in Figure 2.8.

Figure 2.24

A



B

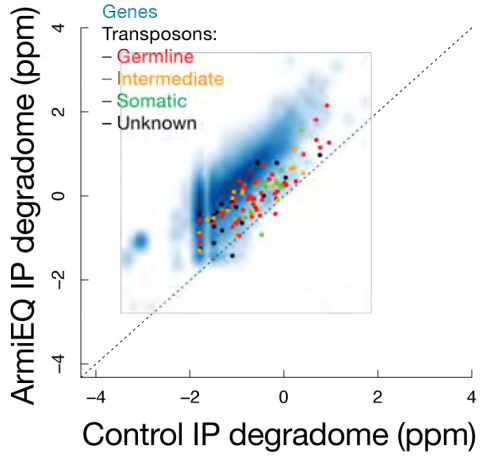


Figure 2.24: ArmiDE862AA and ArmiE863Q gain promiscuous binding to genic mRNAs

Scatterplot showing the abundance of transposon- or gene-mapping degradome 5′-ends in ArmiDE862AA IP (A) or ArmiE863Q IP (B) versus control IP in the wild-type background. Transposon grouping is as in Figure 2.8.

Armi ATPase activity is not required for interaction with piRNA factors

Next, we asked if the dispersed ArmiK729A mutant retains interaction with piRNA pathway factors by conducting mass spectrometry of the ArmiK729A immunoprecipitate. Surprisingly, despite of its dispersed localization, ArmiK729A remains associated with all the piRNA factors that associate with ArmiWT (Table 2.4), including the nuage factors Vret, Shutdown, Ago3, Spn-E, Tapas, Aub, BoYb, Qin, Vasa, and the mitochondrial factors Gasz, Minotaur, SoYb and Papi. Therefore, binding to piRNA pathway proteins does not determine the nuage or mitochondrial localization of Armi. Instead, Armi subcellular localization is likely determined its ATPase-dependent binding to piRNA precursor in the nuage or on mitochondria. Conversely, the dispersed localization of ArmiK729A is likely due to its promiscuous binding to genic mRNA throughout the cytoplasm.

Table 2.4

Name	ArmiWT IP mean fold enrichment	ArmiK729A IP mean fold enrichment
Armi	12900	9266
Gasz	238	339
Minotaur	214	259
SoYb	64	113
Vret	40	51
Shutdown	38	150
Ago3	15	3
Spindle-E	15	12
Tapas	7	18
Aub	6	6
BoYb	6	2
Papi	5	4
Qin	4	3
Piwi	3	4
Vasa	2	3
Tudor	1	3

Table 2.4: Flag-Myc-ArmiK729A remain associated with piRNA factors

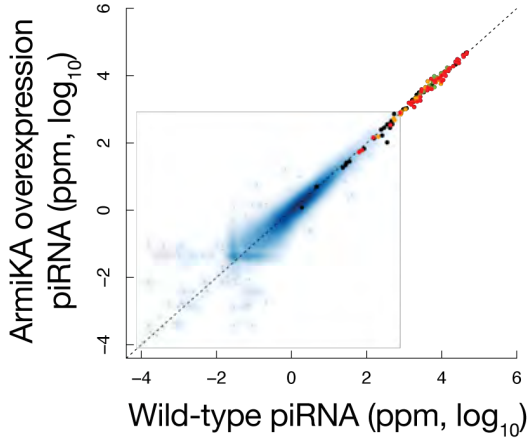
Ovaries from transgenic flies overexpressing germline-specific N-terminal 3×FLAG-6xMyc tagged ArmiK729A were crosslinked with DTME before cell lysis, followed by FLAG IP and mass spectrometry. Fold enrichment was calculated by dividing normalized iBAQ quantification score in the experimental IP by that in the control IP. A pseudo-count equals to the average of the lowest 10 iBAQ values in each sample was added to all proteins in that sample to eliminate zeros. Shown are the mean fold enrichment from three biological replicates. piRNA factors that localize to mitochondria are shown in red. Those that localize to nuage are shown in blue. ArmiWT IP mean fold enrichment is copied from Table 2.2 to compare to ArmiK729A IP.

Armi ATPase mutant-bound RNAs are not made into piRNAs

ArmiK729A mutant remains associated with the piRNA biogenesis machineries and retains the RNA-binding property. If all that Armi does is to bind the correct RNA precursor substrate and delivers it to the phasing machinery, the K729A mutant should deliver genic mRNAs to the piRNA production line and gene-derived piRNAs should increase. However, there is no change of the level of gene-derived piRNAs or transposon-derived piRNAs upon ArmiK729A overexpression (Figure 2.25A). In addition, ArmiK729A does not rescue piRNA production at all in the *armi* germline mutant background (Figure 2.20A, Figure 2.25B). These results suggest that the mere binding of ArmiK729A to a transcript and to other piRNA factors are not sufficient to trigger piRNA production. In light of the recent finding that artificial tethering of ArmiWT to a transcript triggers piRNA production (Rogers et al., 2017) (Pandey et al., 2017), our data suggest that the Armi ATPase is further required after the substrate binding step for piRNA biogenesis.

Figure 2.25

A



B

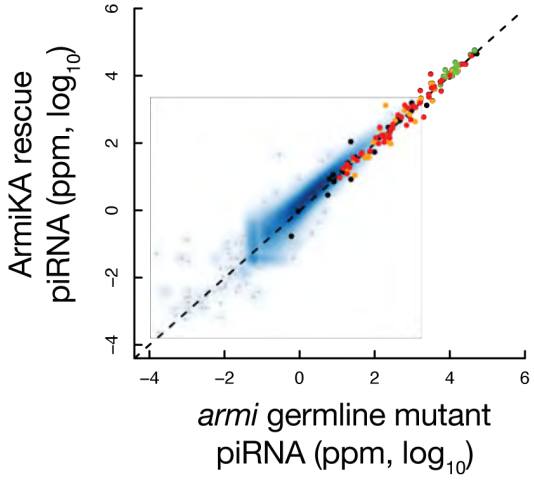


Figure 2.25. ArmiK729A-bound RNAs are not made into piRNAs

(A) Scatterplot showing the level of piRNAs mapping to transposons in ovaries overexpressing ArmiK729A comparing to the wild-type control. (B) Comparing the level of piRNAs mapping to transposons in *armi* germline mutant ovaries rescued with ArmiK729A to the mutant control. Transposon grouping is as in Figure 2.8.

Zuc does not interact with piRNA precursors

Given that Armi interacts with phasing intermediates in the wild-type background and interacts with ping-pong intermediates in the *zucCD* background, we asked if such intermediates co-immunoprecipitate with Zuc. Zuc is thought to be the phasing endonuclease that makes consecutive cuts on piRNA precursors, therefore we expected to capture transposon-mapping long RNAs in complex with Zuc. In our model, the piRNA precursors binding to Armi are delivered to the mitochondrial surface for Zuc to process, which predicts that the same piRNA precursors may be enriched in Zuc IP, with the proper crosslinking methods. We attempted the following crosslinking conditions for Zuc IP in the wild-type background: PFA (RNA-seq and degradome-seq), PFA+DTME (RNA-seq and degradome-seq), DTME+UV (RNA-seq). In all cases, when compared to the negative control IP (FLAG IP from wild-type ovary lysate without FLAG tag expression), there was no enrichment of any long RNAs mapping to transposons or piRNA clusters (data not shown), even though many piRNA factors working in the same pathway as Zuc are enriched in the IP under similar crosslinking conditions (e.g., Table 2.1).

Since phasing is predicted to be fast-acting (see *Discussions* for phasing kinetics), we reasoned that the absence of piRNA precursors in ZucWT IP may be due to transient Zuc-precursor contact. If that is the case, then the catalytically inactive ZucCD, in which only the catalytic residue is mutated, may be able to

trap the precursor in its complex. Furthermore, we already know that in this mutant Armi still binds to ping-pong intermediates (Figure 2.13) and colocalizes with ZucCD (Figure 2.15B), so likely ZucCD will contact the precursor. To control for the background, immunoprecipitation using anti-GFP antibody instead of the anti-FLAG antibody was performed in *zucCD-3xFLAG* ovary lysates as the negative control. Comparing FLAG IP to GFP IP shows that transposon-mapping reads are not enriched in ZucCD immunoprecipitates, no matter using PFA crosslinking (RNA-seq and degradome-seq) or DTME crosslinking (degradome-seq) (Figure 2.26 and data not shown). We conclude that either ZucCD does not interact with piRNA precursors, or the signal to noise ratio is too low for us to detect the interaction (see *Discussions* for the low expression of Zuc).

Figure 2.26

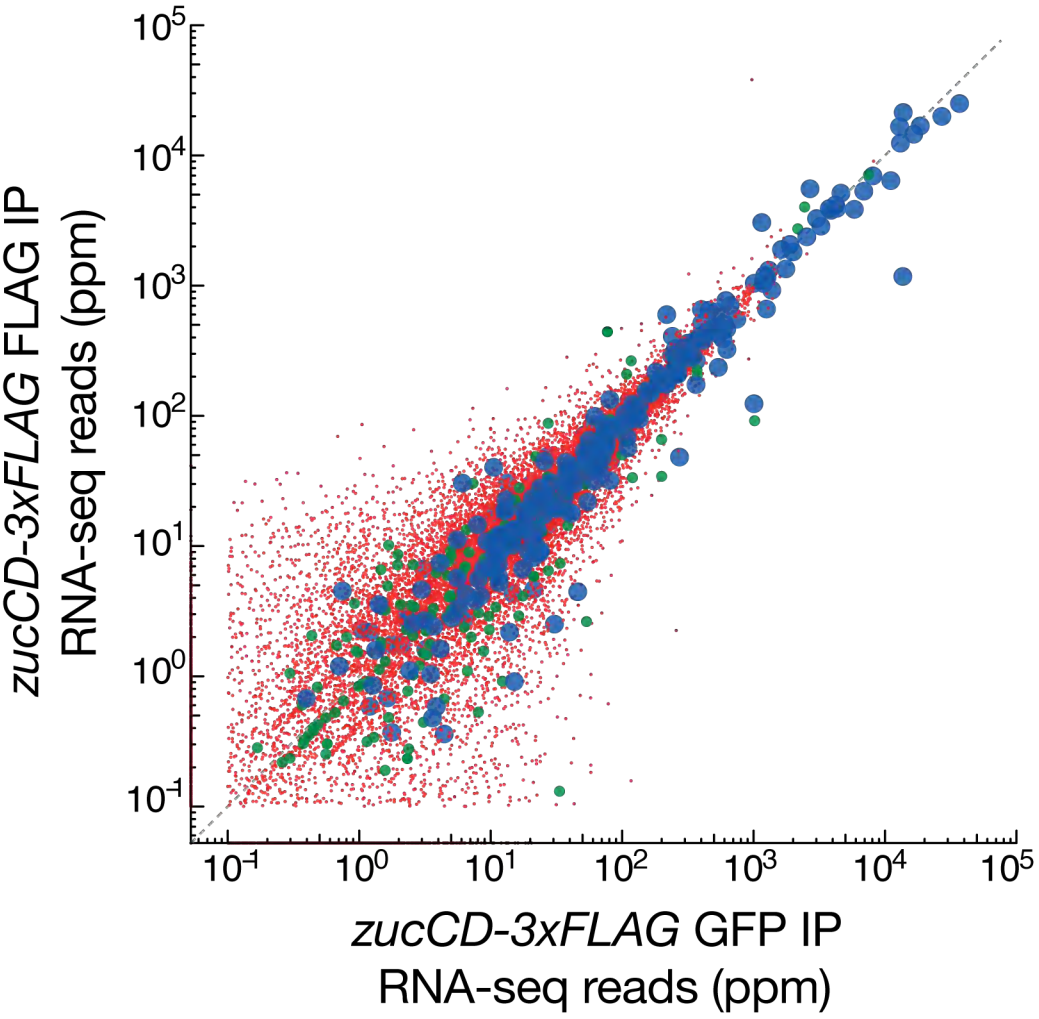

zucCD-3xFLAG FLAG IP
RNA-seq reads (ppm)
zucCD-3xFLAG GFP IP
RNA-seq reads (ppm)
10^-1
10^0
10^1
10^2
10^3
10^4
10^5

Figure 2.26: ZucCD does not interact with piRNA precursors

Scatterplot showing the level of long RNA in FLAG IP versus GFP IP from PFA-crosslinked *zucCD* ovaries. Each dot represents a transposon family (blue, a total of 238 families shown), a non-coding RNA (green), or a genic mRNA (red). All genome mappers are shown and displayed as ppm (parts per million genome mappers).

FISH fails to detect piRNA precursors on mitochondria

According to the model presented in this thesis, phased piRNA biogenesis happens on the mitochondrial outer membrane. We therefore asked if piRNA precursors can be detected on mitochondria. A published Stellaris fluorescent *in situ* hybridization (FISH) probe set (42AB_left2S, (Mohn et al., 2014)) was used to together with antibodies to simultaneously detect piRNA precursor RNA and piRNA factor proteins. The FISH probe set contains 46 non-overlapping 20 nt probes within a 2 kb unique region in cluster *42AB*—a piRNA cluster producing 30% of the germline piRNAs. This probe set specifically recognizes *42AB* transcripts, because no signal was detected in egg chambers of *42ABDf1.1* flies, in which the *42AB* cluster genomic fragment between Chr2R:6,271,759 and 2R:6,501,861 (~230 kbp) is deleted (gift from Ruth Lehmann). To be consistent with the previous immunofluorescence experiments, we focused on stage 3 egg chambers. However, we note that the *42AB* FISH signal is weaker at stage 3 than later stages. Consistent with the literature (Mohn et al., 2014), the *42AB* FISH signal is in general only detected inside the nucleus or at the nuclear periphery (the region occupied by nuage) (Figure 2.27), but not detected further out in the cytoplasm (where mitochondria are). This phenomenon can be interpreted by at least two possibilities, 1) the mitochondrial phasing machinery may process substrates much faster than the nuclear transcription or the nuage ping-pong machinery, or 2) because the fluorescence signal is proportional to the

number of Stellaris probes binding to the same RNA molecule, the *42AB* FISH probes may not be sensitive enough to detect phasing intermediates, which are conceivably shorter than piRNA precursors in the nucleus or at the nuage stage of piRNA biogenesis.

Figure 2.27

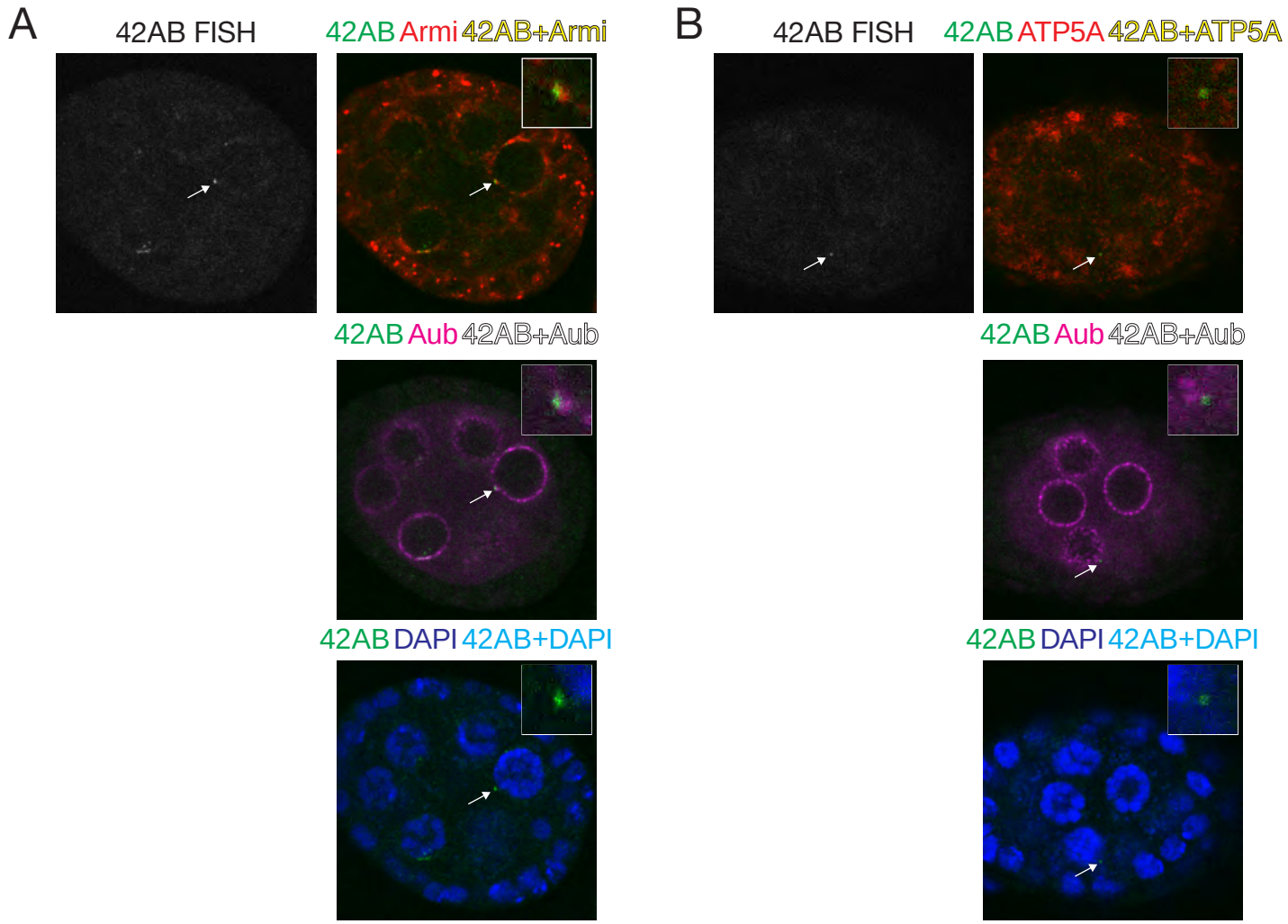

A
42AB FISH
42AB Armi 42AB+Armi
42AB Aub 42AB+Aub
42AB DAPI 42AB+DAPI
B
42AB FISH
42AB ATP5A 42AB+ATP5A
42AB Aub 42AB+Aub
42AB DAPI 42AB+DAPI

Figure 2.27: FISH does not detect *42AB* piRNA precursors on mitochondria in wild-type nurse cells

FISH/Immunofluorescence detection of *42AB* RNA, Armi, Aub, ATP5A and nucleic acids (DAPI) in wild-type stage 3 egg chambers. Each channel and the overlapping signal are color-coded. Arrows: perinuclear *42AB* puncta that are enlarged in the inset.

In *zucCD* mutants, the binding of Armi to piRNA precursors is not affected (Figure 2.12), and Armi is sequestered on mitochondria, away from the nuage (Figure 2.15). We hypothesized that piRNA precursor (ping-pong intermediates) are also stuck on mitochondria waiting to be processed. Because of the lack of phasing in *zucCD* mutants, Armi-bound ping-pong intermediates are not expected to get shorter and therefore are more likely to be detected by the Stellaris FISH probes. We therefore tested whether the same *42AB* probe set can detect piRNA precursors colocalizing with Armi or mitochondria in *zucCD* mutants.

While still being rare, *42AB* puncta outside of the nuclear periphery can be detected in *zucCD* mutants more frequently than the wild-type (Figure 2.28, arrows). Surprisingly, these puncta do not directly overlap with Armi or mitochondria, but are instead juxtaposed to them. Curiously, they almost always colocalize with Aub, in which case the Aub puncta travel outside of the nuclear periphery into the cytoplasm. Indeed, while Aub is almost exclusively perinuclear in wild-type cells (Figure 2.2, Figure 2.14, Figure 2.27), in *zucCD* mutants Aub puncta more frequently detach from the nuclear periphery, and are often surrounded by or juxtaposed to Armi puncta (Figure 2.28). The spatial organization of juxtaposed Armi-Aub foci are reminiscent of the liquid droplet assemblages recently described in the *C. elegans* germline, where the ZNFX-1 granule–P granule relationship changes from colocalization to juxtaposition

during germline development (Wan et al., 2017). Interestingly, ZNFX-1 is also a Upf1 family RNA-binding ATPase, and P granule is the *C. elegans* counterpart of the nuage.

The fact that *42AB* FISH signals are detected on cytoplasmic Aub puncta, but not the adjacent Armi puncta or mitochondria, suggest that the precursors being detected are still in the nuage stage of biogenesis and not yet bound to Armi. One interpretation for not detecting Armi-bound precursors by FISH in *zucCD* mutants is that in the absence of phasing, the ping-pong machinery may process the long precursor into multiple short pieces that are beyond the detection limit of the Stellaris FISH probes. This fits a previously proposed model in which ping-pong and phasing compete to process piRNA precursors (Hayashi et al., 2016).

To summarize, the FISH technique we used here is likely not sensitive enough to detect piRNA precursors less than 2 kb long. A better-suited FISH method is needed to study mitochondrial phasing intermediates in the future.

Figure 2.28

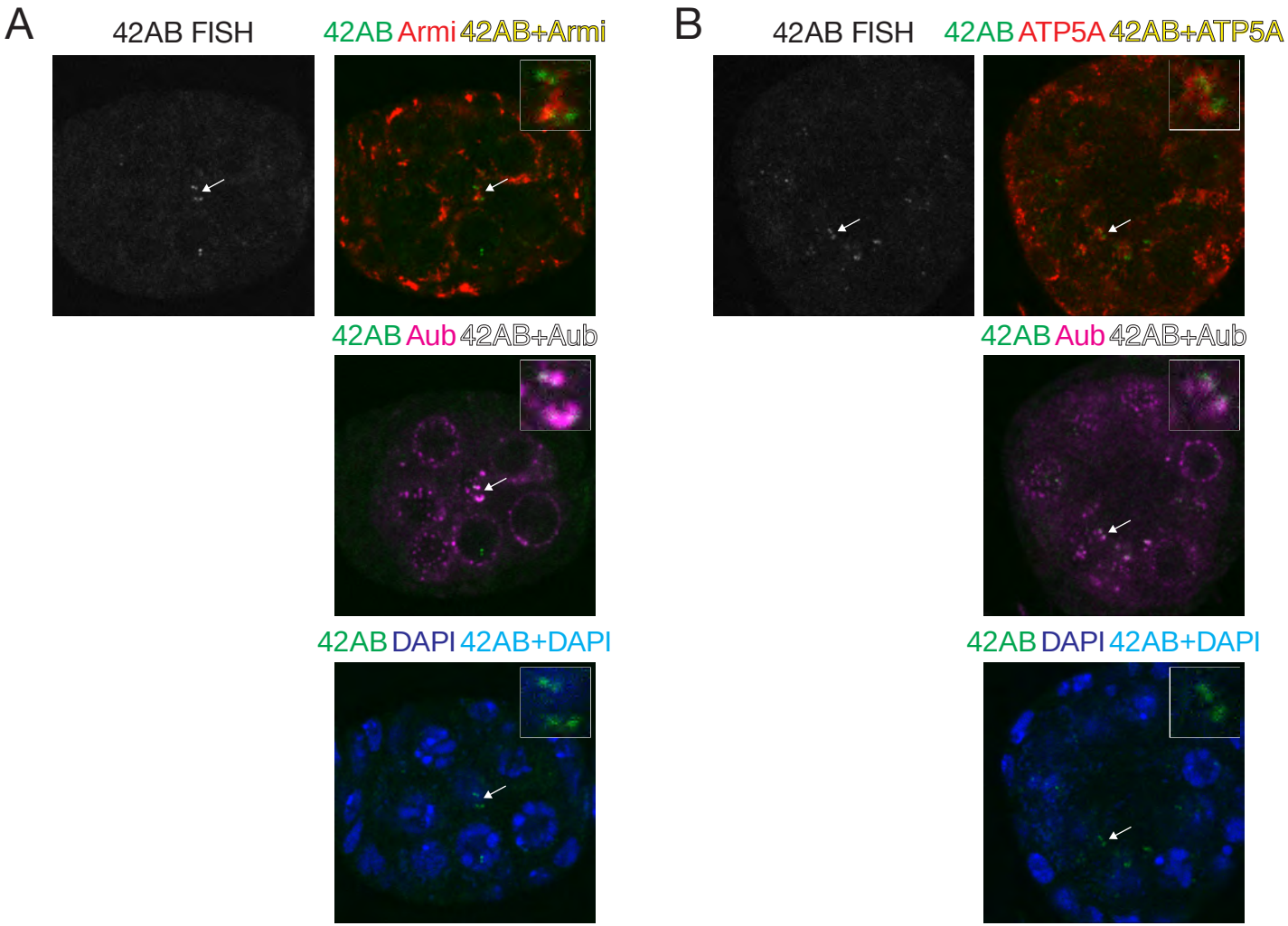

A
42AB FISH
42AB Armi 42AB+Armi
42AB Aub 42AB+Aub
42AB DAPI 42AB+DAPI
B
42AB FISH
42AB ATP5A 42AB+ATP5A
42AB Aub 42AB+Aub
42AB DAPI 42AB+DAPI

Figure 2.28: FISH detects cytoplasmic *42AB* precursors in complex with Aub in *zucCD* mutants

FISH/Immunofluorescence detection of *42AB* RNA, Armi, Aub, ATP5A and nucleic acids (DAPI) in *zucCD* stage 3 egg chambers. Each channel and the overlapping signal are color-coded. Arrows: cytoplasmic *42AB* puncta that are enlarged in the inset.

DISCUSSION

Dynamic Armi distribution between nuage and mitochondria

We have found that Armi links the nuage and mitochondrial phases of piRNA biogenesis by associating with piRNA precursors and protein factors in both compartments, and likely transfers piRNA precursors through passive diffusion and anchoring to the two compartments. In *zucCD* mutants where the downstream phasing is blocked, Armi is trapped on mitochondria; on the other hand, it was reported that in Ago3CD rescue mutants where the upstream ping-pong is blocked by the catalytically inactive Ago3, Armi is likely trapped in nuage (Huang et al., 2014). The ArmiK729A mutant that gained promiscuous binding to genic mRNA disperses in the cytoplasm. These results suggest that the dynamic distribution of Armi between nuage and mitochondria is regulated by 1) active piRNA production in both compartments, and likely 2) its association with bona fide piRNA precursors. We speculate that Armi first binds to piRNA precursors in the nuage; after ping-pong cleavage, the Armi-RNA complex is released, diffuses into the cytoplasm, and docks at the surface of mitochondria (likely through Armi-Gasz interaction (Handler et al., 2013)). The mitochondrial phasing machinery converts piRNA precursors into phased piRNAs and releases Armi to the cytoplasm. In RNA-free state, Armi may have higher affinity for nuage factors; after precursor RNA binding, Armi gains higher affinity for mitochondrial factors.

This may be explained by a conformational change on Armi between RNA-bound and RNA-free states. Armi ATPase mutants bind tightly to RNA, but it remains to be explored how to find a mutant Armi that is unable to bind RNA. Conducting the Armi IP–mass spectrometry experiments in those mutants can test if Armi-piRNA factor interactions are regulated by RNA binding.

In the fly germline nurse cells, nuage and mitochondria are physically separate, which facilitated our study of Armi localization dynamics between the two compartments. In other species whose germ cells have closely associated nuage and mitochondria (e.g., mice, frogs), we expect Armi to serve similar roles, but the much smaller spatial distance between nuage and mitochondria may not allow visualization of Armi dynamics.

Armi ATPase serves multiple roles in piRNA biogenesis

RNA-binding ATPases utilize the energy of ATP hydrolysis to perform an expanding repertoire of functions including ribonucleoprotein complex (RNP) remodeling and RNA duplex unwinding (Pyle, 2011). Upf1, a superfamily I RNA-binding ATPase, is essential for the nonsense-mediated decay (NMD) pathway—degradation of the premature stop codon-containing mRNAs (He and Jacobson, 2015). Upf1 hydrolyzes ATP and processively translocates on RNA in 5′-to-3′ manner in vitro (Fiorini et al., 2015). In cells, Upf1 ATPase activity is required to quickly dissociate from non-target mRNAs (Lee et al., 2015), disassemble the

premature translation termination complex (Serdar et al., 2016), and after NMD completion, disassemble the decay complex (Franks et al., 2010). Upf1 therefore plays multiple ATPase-dependent roles in the initiation, execution and termination of NMD.

Armi and Upf1 share a similar helicase core (Figure 2.18), and our findings suggest that Armi ATPase similarly serve multiple roles in piRNA biogenesis. First, the promiscuous binding of Armi ATPase mutants to genic mRNA (Figure 2.23, Figure 2.24) suggests a role of the ATPase in RNA substrate discrimination. This calls to mind the aberrant binding of Upf1 ATPase mutants to non-NMD-target mRNAs (Lee et al., 2015). It was proposed that Upf1 ATPase activity is rapidly activated on non-target mRNAs and promote dissociation, while on NMD targets the ATPase is temporarily inhibited to allow longer Upf1 dwelling time (Lee et al., 2015). This agrees with the observation that ATP or ADP decreases the association between the Upf1 helicase domain and ssRNA (Cheng et al., 2007). Similarly, it is possible that the ATPase activity of Armi enables it to dissociate from incorrect RNA substrates, thus confer substrate selectivity towards bona fide piRNA precursors.

Second, the futile binding of ArmiK729A to genic mRNA, which is not made into piRNAs (Figure 2.25), suggests that Armi ATPase is also required after the substrate discrimination step for phased piRNA production. Since ArmiK729A still interact with other piRNA factors (Table 2.4), Armi may use ATP

to remodel the RNA-phasing machinery complex, in order for the phasing complex to cleave the RNA. The Armi homolog Mov10 translocate on mRNA 3′-UTRs in ATPase-dependent manner (Gregersen et al., 2014). Consistently, Armi harbors ATP-dependent 5′-to-3′ RNA duplex unwinding activity in vitro (Pandey et al., 2017). Therefore, this phasing complex-remodeling activity may involve Armi translocation on the precursor RNA.

Based on these observations, we propose the following model: in the cytoplasm, Armi ATPase allows it to rapidly dissociate from genic mRNA; the RNA-free Armi concentrates in the nuage, likely through protein-protein interactions that also inhibits the Armi ATPase, to allow its stable binding to ping-pong intermediates. The RNA-bound Armi gains higher affinity for mitochondrial piRNA factors, which facilitated the localization of Armi-RNA complex to the mitochondrial surface. The new protein-protein interaction also activates the Armi ATPase and allows it to translocate on the RNA. As a result, the phasing endonuclease turns the precursor RNA into multiple, head-to-tail-linked piRNAs (Figure 2.29).

Figure 2.29

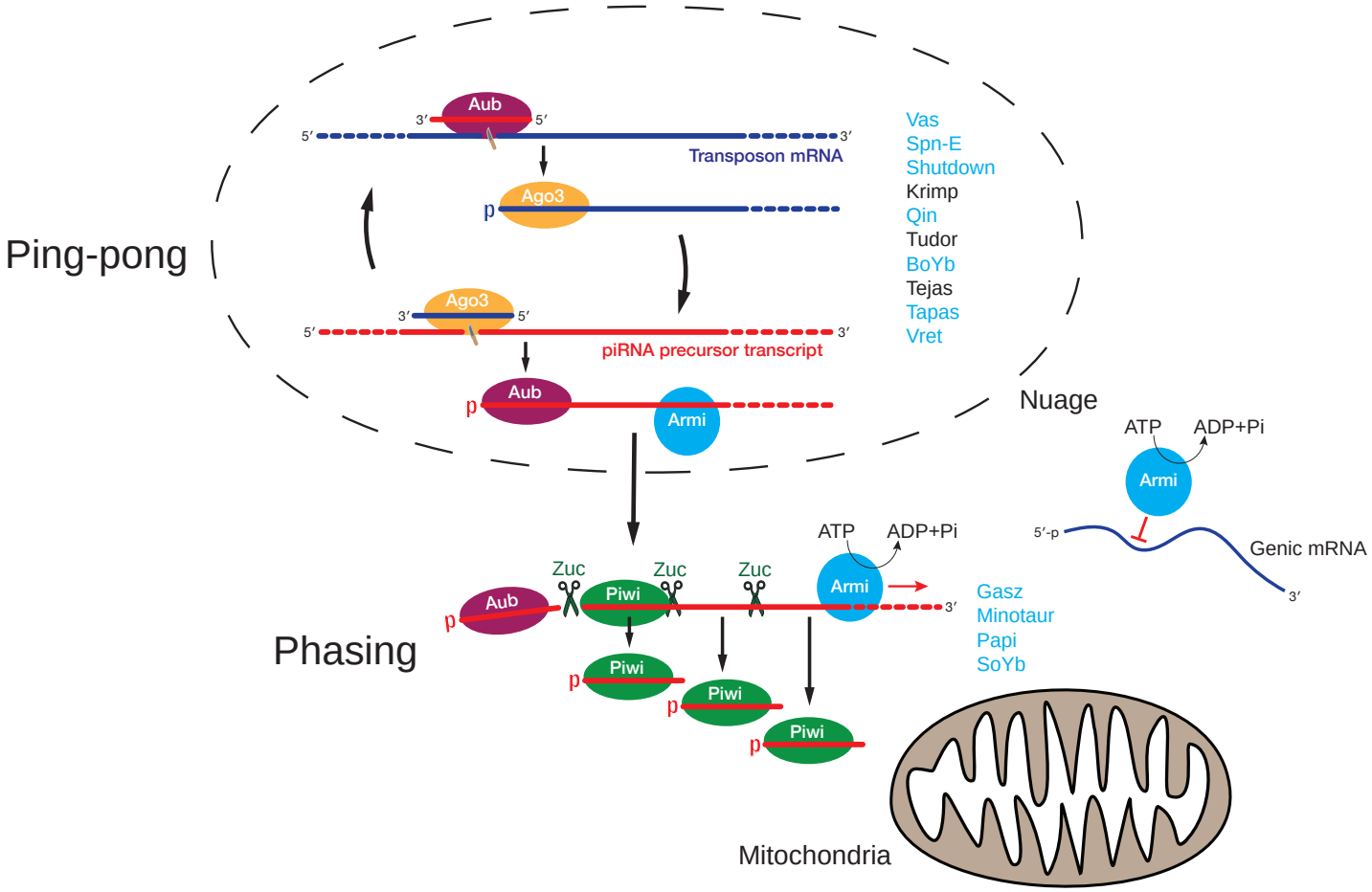

Ping-pong
Aub
Transposon mRNA
Ago3
piRNA precursor transcript
Armi
Vas
Spn-E
Shutdown
Krimp
Qin
Tudor
BoYb
Tejas
Tapas
Vret
Nuage
ATP
ADP+Pi
5′-p
Genic mRNA
Phasing
Zuc
Piwi
Gasz
Minotaur
Papi
SoYb
Mitochondria

Figure 2.29. A model for the role of Armi in piRNA biogenesis

The kinetics of ping-pong-phasing coupling

In the degradome index mapping analysis, phasing intermediates are not detected in the control IP (Figure 2.10B), but ping-pong intermediates are (Figure 2.11B). Assuming that the RNAs recovered from the control IP represent the RNA repertoire of the whole cell, these data indicate that ping-pong intermediates are much more abundant than phasing intermediates. The accumulation of ping-pong intermediates implies that the ping-pong cleavages happen faster than the downstream processing, i.e., phasing. Conversely, the depletion of phasing intermediates suggests that the phasing machinery consumes the intermediates at a fast rate. Importantly, these observations suggest that the step in between, i.e., the delivery of ping-pong intermediates to the phasing endonuclease, is the rate-limiting step, which is mediated by Armi. Two possibilities can explain this slow step: 1) diffusion of the Armi-ping-pong intermediate complex out of the nuage may be disfavored because of the phase difference between the nuage and the surrounding cytoplasm (Brangwynne et al., 2009); or 2) the low protein expression of Zuc (discussed more below) may reflect its low density on mitochondria and therefore harder for the Armi-RNA complex to anchor to.

Is Zuc an endonuclease or phospholipase?

Zuc is the top candidate for the phasing endonuclease for several reasons. First, recombinant Zuc shows endonuclease activity in vitro (Ipsaro et al., 2012; Nishimasu et al., 2012). Second, the crystal structure of Zuc resembles that of the nuclease Nuc and reveals a positively charged groove (Ipsaro et al., 2012; Nishimasu et al., 2012) and a CCCH zinc finger motif for potential ssRNA binding (Ipsaro et al., 2012). Third, no better candidate has been identified through three genome-wide RNAi screens (Handler et al., 2013; Baena-Lopez et al., 2013; Muerdter et al., 2013). However, other observations cast doubts on Zuc being the endonuclease. First, the in vitro RNA cleavage reaction is inhibited by 25 mM sodium chloride (Ipsaro et al., 2012; Nishimasu et al., 2012). Second, Zuc in vitro cleavage does not have a base selectivity, while it is well established that the phasing nuclease prefers to cut in front of uridine in vivo (Mohn et al., 2015; Gainetdinov et al., 2018). Third, the current model of phased piRNA production suggests that Piwi binds to the free 5′-monophosphorylated end of the precursor before the endonuclease cuts in front of the first uridine unprotected by the Piwi footprint (Gainetdinov et al., 2018). This predicts that Zuc and Piwi are at least transiently present in the same complex during the biogenesis cycle. However, Piwi does not co-IP with Zuc (Table 2.1), but instead co-IPs with Armi (Table 2.2, Table 2.4). Fourth, our attempts to capture RNA in ZucWT or ZucCD immunoprecipitates using different crosslinking methods were not fruitful (Figure

2.26 and data not shown). Therefore, either Zuc does not interact with RNA, or the interaction cannot be captured in our hands.

Alternatively, it has been suggested that Zuc/MitoPLD promotes piRNA biogenesis indirectly through the generation of the signaling lipid phosphatidic acid, which promotes membrane fusion (Huang et al., 2011; Watanabe et al., 2011). Cells overexpressing MitoPLD have hyper-fused and aggregated mitochondria, and cells depleted with MitoPLD have fragmented mitochondria (Choi et al., 2006). Our data also suggest a link between piRNA phasing and mitochondrial fusion: the core fusion factor mitofusin (Marf) (Kameoka et al., 2018), fusion factor mitoguardin (Zhang et al., 2016) and fusion inhibitor Usp30 (Yue et al., 2014) all co-immunoprecipitate with Zuc; Marf also co-immunoprecipitates with Armi (Table 2.3). In addition, CG3394, whose mammalian homolog, SLC27A1, has been implicated in phosphatidic acid biosynthesis (transport of long-chain fatty acids), also co-immunoprecipitates with Zuc. Interestingly, the other piRNA phasing factor Minotaur is also predicted to be a factor in phosphatidic acid biosynthesis, although the catalytic residue is dispensable for piRNA production (Vagin et al., 2013; Fu et al., 2013).

The controversy around Zuc being a nuclease or phospholipase arose because for HKD-motif containing proteins, the same catalytic site may be used to hydrolyze phosphodiester bonds in either phospholipids or nucleic acids (Selvy et al., 2011). A point mutation in the HKD motif of MitoPLD causes both piRNA

depletion and mitochondrial aggregation (Watanabe et al., 2011). This observation could be explained by: 1) the MitoPLD mutation affects piRNA biogenesis, and piRNA depletion causes mitochondrial aggregation, 2) the mutation affects mitochondrial dynamics, and aggregated mitochondria impair piRNA biogenesis, or 3) both pathways are affected separately by the same MitoPLD mutation.

Could it be that Zuc functions as endonuclease in piRNA production, and piRNAs are required for mitochondrial fusion? It is unlikely because cultured human cells that do not express piRNAs show changes in mitochondrial morphology upon MitoPLD/Zuc loss or overexpression (Huang et al., 2011; Choi et al., 2006). On the other hand, preliminary data show that other piRNA mutants, including *armi*, *mino* (data not shown) and *gasz* ((Handler et al., 2013)), also have changes in mitochondrial morphology, which may or may not be a result of transposon activation in these mutants.

On the other hand, could it be that Zuc functions as phospholipase in phosphatidic acid (PA) production, and PA-stimulated mitochondrial fusion is required for piRNA production? The loss of intermitochondrial cement (the mammalian equivalent of the nuage) in MitoPLD mutants (Watanabe et al., 2011; Huang et al., 2011), and the increase of intermitochondrial cement in mutants of Lipin 1, an enzyme that consumes PA (Huang et al., 2011), favor this model. However, Mitochondrial fusion does not seem to be essential for piRNA

biogenesis because RNAi-mediated knockdown of genes involved in mitochondrial fusion do not affect transposon silencing (Handler et al., 2013; Baena-Lopez et al., 2013; Muerdter et al., 2013). Nonetheless, it remains possible that the low amount of fusion protein still present in the RNAi knockdown experiments suffices for piRNA biogenesis, or that multiple fusion factors serve redundant roles and escape single gene knockdown screens. Alternatively, Zuc-mediated PA production may function in piRNA production independent of the role of PA in mitochondrial fusion.

The final possibility is that the same catalytic site of Zuc functions separately in piRNA production (as an endonuclease) and in mitochondrial fusion (as a phospholipase), in which case a separate mutation of Zuc that only affects one of the two pathways (e.g., disrupts a specific protein-protein interaction) needs to be identified. To conclude, the relationship between mitochondrial fusion and piRNA biogenesis is not yet resolved.

Stabilization and characterization of piRNA biogenesis factor interactions

Apart from the Zuc-Papi interaction in silk moth BmN4 cells (Nishida et al., 2018), no solid in vivo protein-protein interactions have been demonstrated for Zuc (Haase et al., 2010). Similarly, only Armi-Piwi and Armi-Ago3 interactions had previously been shown in fly ovaries (Olivieri et al., 2010; Huang et al., 2014). In our proteomics studies following immunoprecipitation, Zuc was found to interact

with many mitochondrial piRNA biogenesis factors (Armi, Gasz, Minotaur, Papi, SoYb, Table 2.1), and Armi was found to interact with even more piRNA factors, including Shu, Ago3, Spn-E, Tapas, Aub, BoYb, Qin, Vas, Gasz, Minotaur, SoYb, Papi, Vret and Piwi (Table 2.2).

Our success in identifying the in vivo interactome of Zuc and Armi can be attributed to two technical advancements. First, in vivo crosslinking using the sulfhydryl-to-sulfhydryl crosslinker DTME is critical. Without crosslinking or with other crosslinkers, the interactions among piRNA factors were not stabilized (Figure 2.6, mass spectrometry data not shown). As exposed cysteines are generally rare, it will be interesting to investigate what attributes to the crosslinking efficiency of DTME. For example, are there cysteine residues on the crosslinked proteins that are important to “lock” the complex? An initial attempt to find the reduced TME adduct on peptides mapping to enriched proteins from the mass spectrometry datasets was not fruitful.

Second, the advent of CRISPR gene editing technology (Jinek et al., 2012) allowed us to knock-in a 3×FLAG tag at the endogenous *zuc* coding sequence with relative ease (described in detail in Chapter III). The endogenously expressed Zuc-3×FLAG can be immunoprecipitated with a commercial monoclonal antibody whose affinity is in the pM range (Sigma anti-FLAG M2 antibody) (LaCava et al., 2015), and eluted under native conditions using the commercial 3×FLAG peptide. The high affinity of the anti-FLAG

antibody to 3×FLAG-tagged proteins allows it to efficiently deplete lowly expressed proteins from lysates, which is exactly the case of Zuc: the level of Zuc-3×FLAG in ~40 µg of total ovary protein is beyond the detection limit of the Western blot, and is only detectable after being concentrated 25-fold through IP. Because of the low target protein abundance, the low signal to noise ratio in the IP eluate posed a significant hurdle to the detection of proteins that co-immunoprecipitate with Zuc, which was overcome through rounds of optimization to decrease the noise (using the least amount of antibody and Protein G-conjugated magnetic beads, short incubation time, and varied amount of salt and ionic detergents in the IP/wash buffers). Finally, the ability to elute Zuc-3×FLAG under native conditions (in contrast to low pH elution or boiling) decreases the elution of nonspecifically-bound proteins (i.e., noise) and virtually eliminates the co-elution of the anti-FLAG antibody, therefore increasing the bandwidth of mass spectrometry detection toward bona fide co-immunoprecipitated proteins.

Since Zuc-3×FLAG is a small protein (predicted molecular weight: 31.5 kDa), the possibility that it may pass the nitrocellulose membrane with 0.45 µm pore size when being transferred from a polyacrylamide gel was tested by using two layers of membrane. The Zuc-3×FLAG signal in the second membrane was ~50% of the first, suggesting that using a membrane with smaller pore size (e.g., 0.2 µm) may be helpful to increase Zuc retention to the first membrane.

Potentially new piRNA factors revealed by IP-MS

Our mass spectrometric characterization of the proteins interacting with Zuc or Armi revealed many uncharacterized proteins (Table 2.1, Table 2.3). We expect some of them to be novel piRNA factors. Multiple whole-genome RNAi screens have been conducted to find piRNA factors (Handler et al., 2013; Baena-Lopez et al., 2013; Muerdter et al., 2013). However, two classes of genes can escape RNAi screens: genes with redundant functions, and genes essential for cell differentiation or viability. Proteomic characterization of the multi-protein complex by immunoprecipitating core members is an approach we used to validate and complement the RNAi screens. For example, Armi interacts with the Tudor domain-containing protein CG9925, which is predicted to be one of the two homologs of mammalian Tdrd1 (the other being CG9684, not enriched in Armi IP, Table 2.3). Zuc interacts with Tudor domain-containing proteins Spoon/Yu and Papi, a pair that predicted to serve similar functions (Handler et al., 2011). Single gene knockdown of CG9925, CG9684, Spoon or Papi did not derepress transposons, suggesting redundant roles between each pair (Handler et al., 2011). Indeed, the loss of Papi in the silk moth BmN4 cells severely affected piRNA levels (Izumi et al., 2016; Nishida et al., 2018), but not in *Drosophila* (Hayashi et al., 2016; Zhang et al., 2018). The fact that both Spoon and Papi interact with Zuc suggests that Spoon may compensate for the loss of Papi in *Drosophila*. A collaborative effort to characterize the function of these gene pairs

in piRNA production has started. In addition, we found that CG10880, a hit in one of the RNAi screens (Baena-Lopez et al., 2013) and remain uncharacterized so far, is one of the most enriched proteins in both Zuc IP and Armi IP (Table 2.1, Table 2.3). It may therefore serve important functions in phased piRNA production.

FUTURE PERSPECTIVE

How does Armi move between nuage and mitochondria?

While we have not directly shown the movement of Armi between nuage and mitochondria, our data strongly suggest so. Dynamic exchange of piRNA factors between nuage and the cytoplasm has been reported (Vasa (Xiol et al., 2014), Tudor, Aub, Tejas, Spn-E and Ago3, but not Krimper (Webster et al., 2015), and Spn-E (Andress et al., 2016)). We expect Armi to behave similarly in fluorescence recovery after photobleaching (FRAP) experiments. In addition, photoactivation of a subpopulation of Armi in the nuage or on mitochondria in live egg chambers and follow the activated Armi fluorescence may reveal how Armi moves between the two compartments.

How are the 5′-ends of ping-pong intermediates protected?

If Armi binds to the ping-pong intermediates in the nuage and traverses the cytoplasm to reach the mitochondria through passive diffusion, how are the 5′-monophosphorylated ends of the ping-pong intermediates protected from cytoplasmic 5′-to-3′ exonucleases such as Xrn1? Two possible models can be offered. First, Armi itself sits on the 5′-monophosphorylated ends made by Ago3 or Aub cleavages and protects them from Xrn1 digestion; after docking at the mitochondrial surface, Armi is activated by mitochondrial piRNA factors to translocate downstream, allowing Aub to bind to the unprotected 5′-end, before

Zuc makes the cut at the 3′-end of Aub footprint. However, this model cannot explain why it is Aub, but not Piwi, that binds to the first unprotected 5′-end, since after the first Zuc cleavage, all subsequent 5′-ends are predominantly bound by Piwi (Figure 2.29). In the second model, Aub or Ago3 binds to ping-pong intermediates in the nuage, and Armi binds immediately downstream of Aub or Ago3 on the same RNA molecule (by interacting with both Aub/Ago3 and the precursor), to facilitate the delivery of this complex to mitochondria. In the cytoplasm and at the mitochondrial surface, Aub or Ago3 protects the 5′-end of the precursor. After Zuc made the first cleavage on the 3′-end of Aub/Ago3 footprint and liberated them with the newly made piRNA, Armi translocates downstream, either resolving secondary structure or displacing other proteins (such as Papi (Nishida et al., 2018)), and allows Piwi to bind to the “naked” single-stranded region of the precursor. Zuc then cuts again and liberates Piwi with a newly made piRNA. Such cycle repeats until the whole precursor is converted into piRNAs, at which point Armi is released back to the cytoplasm.

While we favor the second model, many steps involved are pure conjectures at this point. For example, it has not been shown that Aub or Ago3 binds to the 5′-ends of ping-pong intermediates. Simply doing an IP on Aub or Ago3 and looking for an enrichment of ping-pong intermediates is not enough: Aub or Ago3 could use piRNA as a guide to bind ping-pong intermediates “in trans” as targets. Treating the IP fraction with a 5′-to-3′ exonuclease and ask if

the degradome 5′-ends are different from untreated controls can answer whether the 5′-end of immunoprecipitated ping-pong intermediates are bound by a protein, which is not necessarily Aub or Ago3. Repeating the experiment in *armi* null mutants can answer if Armi is the protein protecting the 5′-end. However, *aub* or *ago3* null mutants cannot be used to answer similar questions because the ping-pong cycle relies on both proteins. An experiment that may get us one step closer is the Armi-Aub or Armi-Piwi tandem IP, which may help to answer whether Armi-bound ping-pong intermediates or phasing intermediates also have Aub or Piwi bound.

How does Armi remodel the phasing complex?

During phased processing of a ping-pong intermediate (Figure 2.29), it is possible that after the first Zuc cleavage and the release of the Aub/Ago3-piRNA complex, Armi hydrolyzes ATP to translocate downstream, displacing proteins that were originally bound downstream of the Aub footprint (i.e., at the new 5′-end of the precursor), such as Papi (Nishida et al., 2018), to provide a “naked” single-stranded region immediately downstream of the new 5′-end that an empty Piwi protein can bind to. Piwi-binding stabilizes the new 5′-end and directs Zuc to cut at the first unprotected uridine downstream of the Piwi footprint (Gainetdinov et al., 2018). At the same time, Armi has to stop translocating to allow Papi to bind downstream of the Piwi footprint (Nishida et al., 2018). It is possible that the

binding of Piwi to Armi inhibits the Armi ATPase activity, strengthens Armi-RNA interaction and converts Armi from an RNA translocator to an RNA clamp. Following Zuc cleavage, only after the Piwi-piRNA complex departs can Armi return to the translocation mode and move downstream, starting the next phasing cycle. The inhibition of Armi ATPase activity by PIWI proteins may similarly happen in the nuage, where the binding of Ago3 or Aub to Armi increases its affinity for the ping-pong intermediates (i.e., as an RNA clamp), allowing their delivery from nuage to mitochondria through the diffusion of Armi.

Future work using purified Armi protein in single molecule studies, as has been done for Upf1 (Fiorini et al., 2015), may help to observe Armi translocation in vitro. Of note, full-length Armi has been successfully purified as a SUMO-fusion protein in insect cells (Pandey et al., 2017). On the other hand, using in vivo crosslinking methods such as photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) to study the footprint of ArmiWT or Armi ATPase mutants on piRNA precursors, as has been done for the mammalian Armi homolog Mov10 (Gregersen et al., 2014), may provide evidence for Armi translocation in vivo.

To test the inhibition of Armi ATPase activity by PIWI proteins, Piwi can be expressed in insect cell culture and purified as a Maltose Binding Protein (MBP)-fusion protein together with Armi (Pandey et al., 2017). It is interesting to note that that Piwi solubility increases when co-expressed with Armi (Pandey et al.,

2017). In the presence or absence of Piwi, the Armi ATPase and helicase activities can be measured by ATP hydrolysis and RNA unwinding assays, respectively.

Furthermore, the N-terminal region of Armi (amino acids 1-448) doesn't contain any known sequence motif, but is indispensable to induce piRNA production when Armi is tethered to a reporter mRNA (Pandey et al., 2017). The N-terminal region is therefore likely involved in protein-protein interactions that anchor Armi to the phasing complex. To test that, a series of N-terminal truncated versions of Armi can be expressed in fly ovaries, immunoprecipitated, and subjected to mass spectrometry. By using these truncated versions of Armi to rescue the *armi* germline mutant, the contribution of protein-protein interactions to phased piRNA production may be revealed.

Why does phasing happen on mitochondria?

One of the biggest mysteries in piRNA biogenesis is why the phasing complex localizes to the mitochondrial outer surface. Preliminary data indicate that Zuc has to anchor at the mitochondrial outer membrane to function in piRNA biogenesis: an N-terminal GFP-Zuc fusion transgene (Pane et al., 2007), in which the N-terminal mitochondrial localization sequence is masked by GFP, failed to rescue the fertility of *zucCD* mutant females. In contrast, a C-terminal Zuc-GFP transgene (Webster et al., 2015), driven by the same Nos-Gal4 driver,

rescued the fertility of *zucCD* mutant females (data not shown). Small RNA sequencing of the GFP-Zuc rescue ovaries can tell whether phasing is just impaired or completely inactive. Comparison of GFP-Zuc or Zuc-GFP localization in nurse cells is needed to correlate localization to function. Comparing GFP-Zuc IP-MS results to that of the Zuc-GFP may provide hints on the necessary protein-protein interactions that make phasing active. Furthermore, functional understanding of the mitochondrial proteins that associate with Zuc (Table 2.1) may provide insights on the role of mitochondrial outer membrane in phased piRNA biogenesis.

Some thoughts on the potential role of mitochondria in piRNA biogenesis can be offered. First, the mitochondrial outer membrane provides anchorage for both the phasing complex and the Armi-precursor RNP, increases the local concentration of protein/RNA components, thereby favoring intermolecular interactions. However, if only a two-dimensional membrane support is needed for this purpose, then why is the mitochondrial outer membrane preferred over the endoplasmic reticulum membrane, which is both larger in surface area and more closely associated with the nuage (Jaglarz et al., 2011)? Second, it is recently shown that mitochondria are physiologically maintained at 50˚C (Chrétien et al., 2018), a temperature more than 10˚C higher than the surrounding cytoplasm. It is possible that the higher temperature at the mitochondrial milieu aids the denaturation of secondary structures in ssRNA precursors, thereby facilitates

Armi translocation, Piwi binding and Zuc cleavage. As the higher temperature is maintained by oxidation of the respiratory substrates, it would be interesting to investigate the impact on piRNA production when the respiratory is inhibited by electron transport chain uncouplers such as 2,4-dinitrophenol. Third, some positive-strand RNA viruses, such as the flock house virus (FHV), replicate their RNA on the mitochondrial outer membrane (Miller et al., 2001). Given the resemblance of transposons to viruses, would it be possible that certain transposons also replicate their RNA or cDNA on the mitochondrial surface, and by localizing near the transposon replication site, the phasing machinery can directly process transposon RNA into piRNAs? Lastly, the unresolved relationship between mitochondrial fusion and piRNA production remain another possibility that can be explored.

To summarize, we have just scratched the surface of the complex mechanisms involved in the production of ping-pong or phased piRNAs, and the coordination between these two piRNA biogenesis machineries. Future endeavors on this topic hold the promise to further our understanding of the role of RNA-binding ATPases, PLD-family of nucleases, Tudor domain-containing proteins and Argonaute proteins in RNA metabolism, and the intriguing relationship between mitochondria, nuage and piRNA biogenesis.

EXPERIMENTAL PROCEDURES

Drosophila stocks

Construction of UASp-3×FLAG-6xMyc-Armi flies: a 3558 bp *armi* cDNA was amplified from Oregon R ovarian total cDNA, which corresponds to the *armi* mRNA isoform A (NM_001014556). The wild-type *armi* cDNA was cloned into pENTR-D-TOPO vector (Invitrogen). Site-directed mutagenesis introduced the K729A, DE862AA and E863Q mutations into the *armi* coding sequence in pENTR-D-TOPO-*armi*, respectively, and then subcloned into the modified Gateway vector pPFM-attB, which carries a *UASp-3×FLAG-6xMyc* N-terminal tag and an *attB* site for site-specific integration in the *PhiC31* integrase-mediated transgenesis system. The pPFM-attB-*armi* plasmid was injected into *attP40* flies carrying the *attP* landing site at cytological band 25C7 in chromosome 2L. Rainbow Transgenic Flies, Inc. (Camarillo, CA, USA) performed injections and screened for stocks with successful transgene integration. The established stocks are referred to as *UAS-FM-ArmiWT, UAS-FM-ArmiK729A, UAS-FM-ArmiDE862AA* or *UAS-FM-ArmiE863Q*, respectively.

Rescue of armi germline null mutants by UAS-FM-Armi: Armi function is necessary in the somatic follicle cells for proper ovary development, as *armi* null flies ($armi^{\Delta 1}$) develop rudimentary ovaries (Olivieri et al., 2010). $armi^{72.1}$ is an incomplete excision of the *P-element* inserted in *armi* 5′-UTR, resulting in the loss of Armi expression specifically in the germline (Olivieri et al., 2010). $armi^{\Delta 1}$

removes the coding sequence of *armi*, *CycJ* and *CG14971*. We therefore used a trans-heterozygous combination between $armi^{72.1}$ and $armi^{\Delta 1}$ as the *armi* germline null mutant background. To test if the *UAS-FM-Armi* transgene can rescue the *armi* germline null, $armi^{72.1}$ was recombined with a third-chromosome germline-specific Gal4 driver *P{GAL4::VP16-nos.UTR}CG6325[MVD1]* (Bloomington #4937) to yield $armi^{72.1}$*, Nos-Gal4-VP16*. A second-chromosome ubiquitously expressed Gal4 driver, *P{Act5C-GAL4}25FO1* (Bloomington #4937), was also crossed in to increase *UAS-FM-Armi* rescue efficiency. The final rescue flies have the following genotype: w^{1118}*; UAS-FM-ArmiWT/Act5C-Gal4;* $armi^{72.1}$*, Nos-Gal4-VP16/*$armi^{\Delta 1}$, or w^{1118}*; UAS-FM-ArmiK729A/Act5C-Gal4;* $armi^{72.1}$*, Nos-Gal4-VP16/*$armi^{\Delta 1}$, referrd to as “ArmiWT rescue” or “ArmiK729A rescue”, respectively.

Germline-specific overexpression of UAS-FM-Armi: *UAS-FM-Armi* was overexpressed on wild-type *armi* background using the third-chromosome germline-specific Gal4 driver *P{GAL4::VP16-nos.UTR}CG6325[MVD1]*. To overexpress *UAS-FM-Armi* on *zucCD* (zuc^{H169Y}) background, *UAS-FM-Armi* was recombined with zuc^{SG63} (Pane et al., 2007), which carries the H169Y point mutation in *zuc* coding sequence, then crossed in trans to zuc^{HM27}, a null allele (Pane et al., 2007). The same third-chromosome germline-specific Gal4 driver *P{GAL4::VP16-nos.UTR}CG6325[MVD1]* was used to express *UAS-FM-Armi* on *zucCD* background.

Other stocks: Endogenously tagged *zucWT-3×FLAG* or *zucCD-3×FLAG* strains are described in the next chapter. *minotaur*$^{z3\text{-}5967}$ (Vagin et al., 2013) was crossed in trans to *Df(3R)ED6280* (Bloomington #29667) to obtain *minotaur* mutants.

Female fertility assay

Female fertility was tested as described in (Li et al., 2009) with a few changes: five female virgins were mated to three Oregon R males in a small cage with a 60 mm diameter grape juice agar plate dabbed with yeast paste at 25°C. The virgin females are two days old at the onset of fertility test (Day 0). After 24 hours, the first plate was replaced with a fresh plate and the number of eggs on the first plate was counted. Plates were then changed and scored every subsequent day. The number of eggs that hatched were scored 24 h after the plate was changed out. Fertility was recorded for 8 days.

***Drosophila* ovary isolation and crosslinking**

Drosophila crosses were grown at 25°C. Unless otherwise noted, 0–3 days old female flies were fed on yeast paste for two days, before ovary dissection. Fly ovaries were quickly dissected in saline solution (5 mM HEPES, 128 mM NaCl, 2 mM KCl, 4 mM $MgCl_2$, 1.8 mM $CaCl_2$, 35.5 mM sucrose, pH 7.2) and transferred to 1.7 mL Eppendorf tubes on ice. After collecting 30–50 mg of ovaries in the

tube, saline solution was removed. One mL of crosslinking solution was added, the tube rotating at RT for 10 min (0.2% paraformaldehyde in 0.1 M sodium phosphate, pH 7.3), 15 min (2 mM DTME in saline solution), or 30 min (5 mM DST, EGS or DSP in saline solution), before the crosslinking solution removed and ovaries washed 3 × 5 min with 1× TBS at RT. The ovaries can be snap frozen in liquid nitrogen and stored at -80˚C until later use. Crosslinking was reversed in the following ways: PFA, heating at 95˚C for 30 min (for protein experiments) or 65˚C for 2 h (for RNA experiments); DTME, heating at 37˚C for 30 min in the presence of 10 mM DTT; DST, RT for 30 min in the presence of 15 mM sodium periodate; EGS, heating at 37˚C for 3 h in the presence of 1 M hydroxylamine HCl; DSP, heating at 95˚C for 5 min in the presence of 100 mM DTT.

Immunofluorescence

Intact ovaries were fixed in 4% methanol-free paraformaldehyde (Pierce #PI28908) in 0.1 M sodium phosphate, pH 7.3 (PB) for 10 min by rotating at RT. Ovaries were then washed three times, for 5, 10, and 15 min at RT in PB supplemented with 0.1% (w/v) Triton X-100 (PBT). After complete removal of the wash buffer, 100 µl of PB was added and ovaries separated into ovarioles by repetitive pipetting using a P200 pipette with a cut tip. The disrupted ovarioles were transferred to PCR tubes and incubated with PBT supplemented with 5%

(v/v) normal donkey serum (Sigma #D9663) and primary antibodies, rotating at 4˚C overnight. The following primary antibodies were used: rabbit anti-Armi C-terminal peptide (gift from William Theurkauf (Cook et al., 2004)) at 1:1000, mouse anti-Armi N-terminal peptide (clone 2F8A9, gift from Mikiko Siomi (Saito et al., 2010), purified with protein G and concentrated to 1.3 mg/mL) at 1:200, mouse anti-FLAG (clone M2, Sigma) at 1:500, mouse anti-ATP5A (Abcam 15H4C4) at 1:200, rabbit anti-Aub (MA514, 2.4 mg/mL, (Li et al., 2009)) at 1:2000, mouse anti-Ago3 (gift from Julius Brennecke, (Senti et al., 2015)) at 1:2000, rat anti-Vas (DSHB) at 1:50.

The next day, ovarioles were washed three times, for 5, 10, and 15 min at RT in PBT. Secondary antibodies diluted in PBT were then incubated with the ovarioles in the dark, rotating at RT for 2 hours. All secondary antibodies are from Thermo Fisher, produced in donkey or goat, against mouse, rabbit or rat IgG (H+L), highly cross-adsorbed against close species, and conjugated to Alexa Fluor® 488 or Alexa Fluor® 594 (for three color experiments with DAPI), or Alexa Fluor® 488, Alexa Fluor® 546 or Alexa Fluor® 633 (for four color experiments with DAPI).

Ovarioles were then washed twice in PBT for 10 min each at RT, incubated with 0.5 µg/mL of DAPI diluted in 2× SSC (0.3 M NaCl, 0.03 M sodium citrate) for 15 min at RT, washed again with PBT for 10 min. The wash buffer is completely removed and a drop of VECTASHIELD Mounting Medium (Vector

Laboratories #H-1000) was added. After a few gentle pipetting to mix the ovarioles with mounting medium, all contents were transferred to a glass slide using a P200 pipette with a cut tip and covered with a 0.13–0.17 mm thick cover slip (VWR #48393106). The cover slip was gently pressed with a Kim wipe to absorb extra mounting medium, before sealed with nail polish. Images were captured using a Leica TCS SP5 II laser scanning confocal microscope using the 63× objective at 1 µm-thick z-stacks.

Fluorescent *in situ* hybridization with immunofluorescence

Ovaries were fixed and stained with antibodies as described above. After secondary antibody washes, ovarioles were dehydrated in 70% ethanol overnight, rotating at 4˚C in the dark. They were rehydrated in Wash Buffer (2× SSC, 10% formamide) for 5 min twice. One hundred microliters of hybridization buffer (2× SSC, 10% formamide, 10% dextran sulfate [w/v]) containing 25 pmol of Stellaris probe were added and incubated at 37˚C with gentle rotation overnight. Ovarioles were briefly rinsed with Wash Buffer, then rotating in two changes of Wash Buffer at 37˚C, each for 15 min. The second wash contained 0.5 µg/mL of DAPI. After another two changes of wash in 2x SSC with 0.1% Triton X-100, each at RT for 5 min, the wash buffer is completely removed and ovarioles mounted.

CellProfiler image quantification

Confocal images were quantified for fluorescence signal using a custom-built CellProfiler script. CellProfiler is developed by Anne E. Carpenter and Thouis R. Jones in the laboratories of David M. Sabatini and Polina Golland. To measure the amount of Aub signal that overlaps with Armi, primary Armi and Aub objects in separate fluorescent channels were identified, using the adaptive Otsu thresholding method. “Threshold correction factor” and “lower bound on threshold” were empirically determined using representative test images. Once optimized, the same object identification settings were applied to all samples. Armi objects were then used as masking object to mask Aub objects. The amount of signal in masked Aub objects was divided by the amount of signal in total Aub objects for each image. The average of all the serial z scan images taken from the same egg chamber was reported.

Transmission electron microscopy

Fly ovaries were quickly dissected in saline solution (5 mM HEPES, 128 mM NaCl, 2 mM KCl, 4 mM MgCl2, 1.8 mM CaCl2, 35.5 mM Sucrose, pH 7.2) and transferred to 1.7 mL Eppendorf tubes on ice. Saline solution was removed and ovaries fixed in 2.5% glutaraldehyde in 0.1 M sodium cacodylate buffer, pH 7.2, overnight at 4˚C. Samples were processed and analyzed at the University of Massachusetts Medical School Electron Microscopy Core Facility according to

standard procedures. Briefly, intact ovaries were then rinsed three times in the same fixation buffer and post-fixed with 1% osmium tetroxide for 1 h at RT. Samples were then washed three times with water for 10 minutes each and dehydrated through a graded ethanol series (10, 30, 50, 70, 85, 95%), before three changes in 100% ethanol. Samples were then infiltrated first with two changes of 100% propylene oxide and then with a 50% / 50% propylene oxide / SPI-Pon 812 resin mixture. Over the following 2 days seven changes of fresh 100% SPI-Pon 812 resin were done before the samples were polymerized at 68˚C in flat molds. The samples were then reoriented for horizontal sections of the center of individual ovaries. The thin sections (approx. 70 nm) were placed on gold support grids, and contrasted with lead citrate and uranyl acetate. Sections were examined using the CM10 with 80Kv accelerating voltage, and images captured using a Gatan TEM CCD camera.

Immunoprecipitation

Zuc immunoprecipitation: freshly dissected *zucWT-3×FLAG* or *zucCD-3×FLAG* ovaries were crosslinked with either DTME or PFA, or both, and kept on ice. For UV crosslinking, DTME-crosslinked whole ovaries were separated into ovarioles by repetitive pipetting using a P200 pipette with a cut tip in saline solution. Around 170 mg of ovarioles were plated in single layer in ~1.2 mL of saline solution to a 60 mm diameter plastic plate and kept on ice. The plate was

exposed to 300 mJ/cm^2 of 254 nm UV light in a Stratalinker UV Crosslinker 2400 for three times on ice, with the ovarioles slightly mixed in between each UV exposure. The ovarioles were then removed from the plate and pelleted. For each volume of the ovary pellet, 4 volumes of ice-cold lysis buffer (50 mM Tris, 150 mM NaCl, 1 mM EDTA, 0.5% IGEPAL CA-630, 1% Empigen BB, 0.1% SDS, 0.5 mM DTT, and 1× home-made protease inhibitor cocktail) was added (i.e., 200 µl of lysis buffer to 50 mg of ovaries). 1× home-made protease inhibitor cocktail contained 1 mM AEBSF (4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride; EMD Millipore #101500), 0.3 µM Aprotinin (Bio Basic Inc #AD0153), 20 µM Bestatin (Sigma Aldrich #B8385), 10 µM E-64 ((1S,2S)-2-(((S)-1-((4-Guanidinobutyl)amino)-4-methyl-1-oxopentan-2-yl)carbamoyl)cyclopropanecarboxylic acid; VWR #97063), and 10 µM Leupeptin (Fisher Scientific #108975). The ovaries were homogenized with a motorized plastic pestle (Fisher Scientific #12141364) for 30 strokes on ice. The tube containing ovary lysates was then submerged in ice-cold water bath and sonicated by a Branson Digital Sonifier model 450 at 40% amplitude for total 2 min of sonication (8 cycles, one cycle includes 15 s of sonication followed by 1 min of rest). The lysate was centrifuged at 13,000 x *g* for 10 min at 4°C to remove insoluble parts. Mouse anti-FLAG antibody (clone M2, Sigma) or mouse anti-GFP antibody (clone GF28R, Invitrogen, as negative control) was added to the supernatant at 6 µg antibody per 1 mL of lysate. The tube was rotated for 2

hours at 4˚C, before contents were transferred to a new tube containing washed protein G Dynabeads (1/10 volume of the lysate, buffer removed). The tube was further rotated for 1 hour at 4˚C for the beads to capture antibodies. The beads were then separated using a magnetic stand, washed six times with lysate volume of wash buffer (WB)-1, -2, -3, -4, -1 and -5, each at RT for 1 min. WB-1 contained 50 mM Tris, 150 mM NaCl, 1 mM EDTA, 0.05% NP-40, and 0.1% SDS. WB-2, WB-3 and WB-4 contained same components as WB-1 except 300, 500 mM and 750 mM NaCl, respectively. WB-5 contained only 0.05% NP-40 in water. Finally, beads were eluted under native condition by incubating with elution buffer (50 mM Tris, 150 mM NaCl, 1 mM EDTA, 0.05% NP-40, 0.1% SDS, 1 mM DTT, and 0.5 µg/µl [175 µM] of 3×FLAG peptide [Sigma]) for 10 min at RT with occasional mixing to prevent beads sedimentation.

Armi immunoprecipitation: ovaries with germline-specific overexpression of Flag-Myc-tagged Armi was dissected and subjected to immunoprecipitation as described above, with the following modifications: after sonication and clearance of the lysate by centrifugation, 4 volumes of dilution buffer (50 mM Tris, 150 mM NaCl, 1 mM EDTA, 0.1% SDS, 0.5 mM DTT, and 1x home-made protease inhibitor cocktail) was added to 1 volume of cleared lysate. Mouse anti-FLAG antibody (clone M2, Sigma) was added at the amount of 6 µg antibody per 1 mL of diluted lysate. Protein G Dynabeads were used as 1/10 volume of the diluted lysate.

Western blotting

The ovaries were homogenized with a motorized plastic pestle (Fisher Scientific #12141364) in ice-cold lysis buffer (for each 30 mg ovaries, 120 µl of 100 mM potassium acetate, 30 mM HEPES-KOH [pH 7.4], 2 mM magnesium acetate, 1 mM DTT) containing 1× home-made protease inhibitor cocktail. Lysate was centrifuged at 13,000 x *g* for 10 min at 4°C and an equal volume of 2× loading dye (100 mM Tris-HCl [pH 6.8], 4% (w/v) SDS, 0.2% (w/v) bromophenol blue, 20% (v/v) glycerol, and 200 mM DTT) was added to the supernatant and heated to 95°C for 5 min.

The lysate was resolved through a 4–20% gradient PAGE (Bio-Rad Laboratories #5671085). Proteins were transferred to a 0.45 µm pore nitrocellulose membrane (Amersham #GE10600002) in a Mini Trans-blot tank at 15V overnight. The membrane was then blocked in Blocking Buffer (Rockland Immunochemicals #MB-070) at 4˚C for 5 h or overnight, before being incubated overnight at 4°C in with primary antibodies diluted in Blocking Buffer. The primary antibodies used were: mouse anti-Armi N-terminal peptide (clone 2F8A9, gift from Mikiko Siomi (Saito et al., 2010), concentrated to 1.3 mg/mL) at 1:2000, goat anti-Armi C-terminal peptide (Santa Cruz dD-17) at 1:500, mouse anti-FLAG (clone M2, Sigma) at 1:10,000, rabbit anti-Piwi (MA511, (Li et al., 2009)) at 1:20,000, rabbit anti-DcrII (Abcam ab4732) at 1:3000.

The membrane was washed 3x5 min with 1× TBST [50 mM Tris-HCl [pH 7.5], 150 mM NaCl, 0.1% Tween 20 (v/v)] at RT, incubated for 1 hour at RT with secondary antibodies diluted in TBST (conjugated to IRDye 680RD or 800CW, LICOR Biosciences, 1:20,000 dilution), and then washed 5 × 5 min with 1× TBST at RT in the dark. The signal was detected using the Odyssey Infrared Imaging System.

Mass spectrometry

FLAG IP was eluted as described above, and DTME was reverse-crosslinked. LC/MS/MS digestion and analysis was carried out by the University of Massachusetts Proteomics Core: the eluted immunoprecipitation reaction (for Zuc IP, eluate from ~350 mg of ovary tissue; for Armi IP, from ~40 mg ovary tissue) was denatured in 2× loading dye (100 mM Tris-HCl [pH 6.8], 4% (w/v) SDS, 0.2% (w/v) bromophenol blue, 20% (v/v) glycerol, and 200 mM DTT) and run for 20 min on an SDS-PAGE gel to separate proteins from lower molecular weight contaminants, and the entire protein region of the gel excised and subjected to in-gel trypsin digestion after reduction with DTT and alkylation with IAA. Peptides eluted from the gel were lyophilized and re-suspended in 25 µL of 5% acetonitrile/0.1% TFA. A 3 µL injection was loaded by a Waters NanoAcquity UPLC in 5% acetonitrile/0.1% formic acid at 4.0 µL/min for 4 min onto a 100 µm I.D. fused-silica pre-column packed with 2 cm of 5 µm (200Å) Magic C18AQ

(Bruker-Michrom). Peptides were eluted using a gradient at 300 nL/min from a 75 μm I.D. gravity-pulled analytical column packed with 25 cm of 3 μm (100Å) Magic C18AQ particles using a linear gradient from 5-35% of mobile phase B (acetonitrile + 0.1% formic acid) in mobile phase A (water + 0.1% formic acid) over 45 min. Ions were introduced by positive electrospray ionization via liquid junction into a Q Exactive hybrid mass spectrometer (Thermo). Mass spectra were acquired over *m/z* 300-1750 at 70,000 resolution (*m/z* 200) and data-dependent acquisition selected the top 10 most abundant precursor ions for tandem mass spectrometry by HCD fragmentation using an isolation width of 1.6 Da, max fill time of 110ms, AGC target of 1e6, collision energy of 27, and a resolution of 17,500 (*m/z* 200). For data analysis, raw data files were peak processed with Proteome Discoverer (version 2.1, Thermo) followed by identification using Mascot Server (version 2.5) against the *Drosophilia melanogaster* Uniprot FASTA file downloaded 5/2016. Search parameters included Trypsin/P specificity, up to 2 missed cleavages, fixed carbamidomethyl on cysteine, and the variable modifications of oxidized methionine, pyroglutamic acid for N-terminal glutamine peptides, and N-terminal acetylation of the protein. Assignments were made using a 10 ppm mass tolerance for the precursor and 0.05Da mass tolerance for the fragments. All non-filtered search results were then loaded into the Scaffold Viewer (Proteome Software, Inc.) with thresholding

to a peptide FDR of 1%, for subsequent peptide/protein validation and label free quantitation.

Small RNA-seq library preparation and analysis

Small RNA libraries were constructed as described (Han et al., 2015a). Briefly, total RNA (50 µg) was purified by 15% urea PAGE, selecting for 18–30 nt small RNAs using 18 nt and 30 nt size markers. Half of the purified sRNAs were oxidized with $NaIO_4$ to deplete miRNAs and enrich for siRNAs and piRNAs (Li et al., 2009). To reduce ligation bias, a 3′ adaptor with three random nucleotides at its 5′ end was used (5′-rApp NNN TGG AAT TCT CGG GTG CCA AGG /ddC/-3′). 3′ adaptor was ligated using truncated, K227Q mutant T4 RNA Ligase 2 (homemade) at 16°C overnight, sRNAs precipitated, and size selected using 3′ adaptor-ligated 18 nt and 30 nt size markers. To exclude 2S rRNA from sequencing libraries, 10 pmol 2S blocker oligo was annealed before the 5′ adaptor ligation step (Wickersheim and Blumenstiel, 2013). 5′ adaptor was ligated using T4 RNA ligase (Life Technologies #AM2141) at 25°C for 2 h, followed by reverse transcription using AMV reverse transcriptase (New England Biolabs #M0277L) and PCR using AccuPrime *Pfx* DNA Polymerase (Invitrogen #12344-024). PCR products purified using 2% agarose gel, and the gel slice extracted with QIAquick gel extraction kit (Qiagen). The length distribution and quality of the libraries were analyzed by Agilent 2100 Bioanalyzer. Libraries were

then quantified using KAPA library quantification kit, before sequenced on a NextSeq500 (Illumina) to obtain 75 nt single-end reads.

Barcodes were sorted by BaseSpace (Illumina), and the 3′ adaptors, including the three random nucleotides, were identified and removed using the first 15 nucleotides, allowing one mismatch. Inserts with length < 18 nt were discarded. sRNAs were analyzed with piPipes (v1.4; (Han et al., 2015b)). Briefly, reads were first aligned to rRNA or miRNA hairpin sequences using Bowtie2 (v2.2.0). Unaligned reads were mapped to the genome and 23–29 nt RNAs (fly piRNAs) were kept for analyses. The number of piRNAs were apportioned by the number of times they aligned to the genome. Division of transposon families into groups was according to Li et al., 2009.

RNA-seq library preparation and analysis

RNA-seq libraries were constructed as previously described (Zhang et al., 2012) with a few modifications (Fu et al., 2018). For ribosomal RNA depletion, RNA was hybridized in 10 µl with a pool of 186 rRNA antisense oligos (0.05 µM/each) with 10 mM Tris-HCl [pH 7.4] and 20 mM NaCl and heated to 95°C, then cooled at -0.1°C/sec to 22°C, and finally incubated at 22°C for 5 min. Ten units of Thermostable RNase H (Lucigen #H39500) were added and incubated at 45°C for 30 min in 20 µl containing 50 mM Tris-HCl [pH 7.4], 100 mM NaCl, and 20 mM $MgCl_2$. RNA was then treated with 4 units of Turbo DNase (Thermo Fisher

#AM2238) in 50 µl at 37°C for 20 min, before purified using RNA Clean & Concentrator-5 (Zymo Research R1016), which enriches for RNA longer than 200 nt. RNA-seq libraries were sequenced using a NextSeq500 (Illumina) to obtain 75 + 75 nt, paired-end reads.

Barcodes were sorted by BaseSpace (Illumina), and adaptors removed. RNA-seq analysis was performed with piPipes (v1.4; (Han et al., 2015b)). Briefly, RNAs were first aligned to rRNA sequences using Bowtie2 (v2.2.0). Unaligned reads were then mapped using STAR to the fly genome (v2.3.1). Counts were produced using the “strict” option on HTseq (v0.6.1). The number of reads were apportioned by the number of times they aligned to the genome.

Degradome-seq library preparation and analysis

Degradome-seq libraries were constructed as previously described (Han et al., 2015a) with a few modifications (Fu et al., 2018). PFA-crosslinked Armi IP eluate was reversed by mixing with equal volume of 2x PK buffer (200 mM Tris-Cl, pH 7.5, 25 mM EDTA, pH 8.0, 300 mM NaCl, 2 % SDS [w/v]) containing 0.4 mg/mL of proteinase K. The mixture was incubated at 50˚C for 1 h then 65˚C for 2 h, before extracted with equal volume of acid phenol:chloroform (5:1 by volume, pH 4.5; AMRESCO LLC, Solon, OH, USA), and centrifuged at 20,800 x g for 5 min at room temperature. The top aqueous layer was precipitated with one-tenth volume 3 M sodium acetate and three volumes 100% ethanol on ice for 1 h. The

precipitate was recovered by centrifugation (20,800 x g for 15 min at 4˚), washed with 70% (v/v) ethanol, air dried, and dissolved in water. rRNA depletion and RNA Clean & Concentrator-5 purification were done in the same way as RNA-seq. 5′-adapter with Unique Molecular Identifier (UMI) (an equimolar mix of two versions, 5′- GUUCAGAGUUCUACAGUCCGACGAUC(N3)CGA(N3)UAC(N3) -3′ and 5′- GUUCAGAGUUCUACAGUCCGACGAUC(N3)AUC(N3)AGU(N3) -3′ (Fu et al., 2018)) was ligated to 5′-monophosphorylated RNAs using T4 RNA ligase (Ambion) at 25°C for 2 h. The ligation reaction was purified using 1.5x volume of Ampure XP beads. Reverse transcription using SuperScript III (Life Technologies) employed a primer containing degenerate sequences at its 3′ end (5′-GCA CCC GAG AAT TCC ANN NNN NNN-3′). The reverse transcription reaction was digestd with 1 µl of RNase H (Ambion) at 37˚C for 20 min, and purified using 1.5x volume of Ampure XP beads. Purified cDNA was amplified by the first round of PCR, using a pair of primers that anneal to the 5′-adapter (5′-CTACACGTTCAGAGTTCTACAGTCCGA -3′) or to the 3′-adapter (5′-GCCTTGGCACCCGAGAATTCCA -3′). The PCR reaction was mixed with 0.7x volume of Ampure beads, and the supernatant was transferred to a new tube containing 0.5x volume of Ampure beads (total 1.2x volume) to purify 200–400 nt PCR products. The second round of PCR uses the same barcoded primer set as the small RNA library cloning protocol, and purified with 1.1x volume of Ampure beads. All PCRs were using Phusion polymerase (NEB). Degradome-seq

libraries were sequenced using a NextSeq500 (Illumina) to obtain 75 + 75 nt, paired-end reads.

Reads were aligned to a reference built from transcriptome, transposon family sequence and piRNA cluster sequences by eXpress, which is able to resolve multiple mapping of reads across gene families. Total count and unique count for each gene, transposon and cluster were then extracted from the eXpress output for downstream analysis. Total count for each feature were then normalized to total counts for all features from the sample and multiplied by 1 million. UMI sequences were extracted from read 1 via barcode pattern NNNCCCNNNCCCNNN (designed in house) and the UMI sequence for each read was attached to the read name of read 1. The sequence name of read 2 were then convert to the same name of its corresponding read 1. Paired-end reads with UMI extracted were then aligned to the genome. The aligned sequences were then sorted by genome coordinates and deduplicated by umi_tools (Fu et al., 2018) with method “directional”. Deduplicated reads were then sorted by name analyzed using piPipes (v1.4, (Han et al., 2015b)). Briefly, sequences were first aligned to ribosomal RNA using Bowtie2 (v2.2.0). Unaligned reads were then mapped using STAR (v2.3.1) to fly genome dm3 and alignments with soft clipping of ends were removed with SAMtools (v1.0.0). The numbers of reads overlapping genes and transposons were apportioned by the number of times each read aligned in the genome. Division of transposon

families into germline, somatic and intermediate groups was according to Wang et al., 2015.

ACKNOWLEDGEMENTS

We thank John Leszyk and Michelle Dubuke at UMass Proteomics Core for performing mass spectrometry experiments, Keith Reddig at UMass Electron Microscopy Core for performing TEM experiments, Cha San Koh for PyMol structural modeling, Christina Baer for initiating the CellProfiler script. We thank members of the Zamore laboratory and Haiwei Mou for helpful discussions. This work was supported by National Institutes of Health grant R37GM062862 to P.D.Z. The TEM work done at the UMass Electron Microscopy Core Facility was supported by Award S10RR027897 from the National Center for Research Resources.

CHAPTER III: TOWARD MORE EFFICIENT CRISPR EDITING AND SCREENING STRATEGIES

PREFACE

The work presented in this chapter was a collaborative effort: Cindy Tipping and I split the work of plasmid construction, fly cross, eye color counting, DNA extraction and PCR genotyping. I designed all the experiments and analyzed the data.

This Chapter has been published in:

Ge, D. T., Tipping, C., Brodsky, M. H., and Zamore, P. D. (2016). Rapid Screening for CRISPR-Directed Editing of the *Drosophila* Genome Using *white* Coconversion. G3 (Bethesda) *6*, 3197-3206.

SUMMARY

Adoption of a streamlined version of the bacterial CRISPR/Cas9 defense system has accelerated targeted genome engineering. The *Streptococcus pyogenes* Cas9 protein, directed by a simplified, CRISPR-like single guide RNA, catalyzes a double-stranded DNA break at a specific genomic site; subsequent repair by end joining can introduce mutagenic insertions or deletions, while repair by homologous recombination using an exogenous DNA template can incorporate new sequence at the target locus. However, the efficiency of Cas9-directed mutagenesis is low in *Drosophila melanogaster*. Here, we describe a strategy that reduces the time and effort required to identify flies with targeted genomic changes. The strategy uses editing of the *white* gene, evidenced by altered eye color, to predict successful editing of an unrelated gene-of-interest. The red eyes of wild-type flies are readily distinguished from white-eyed (end joining-mediated loss of White function) or brown-eyed (recombination-mediated conversion to the $white^{coffee}$ allele) mutant flies. When single injected G0 flies produce individual G1 broods, flies carrying edits at a gene-of-interest were readily found in broods in which all G1 offspring carried *white* mutations. Thus, visual assessment of eye color substitutes for wholesale PCR screening of large numbers of G1 offspring. We find that end joining-mediated mutations often show signatures of microhomology-mediated repair and that recombination-based mutations frequently involve donor plasmid integration at the target locus. Finally, we show

that gap repair induced by two guide RNAs more reliably converts the intervening target sequence, whereas the use of *Lig4*^{169} mutants to suppress end joining does not improve recombination efficacy.

INTRODUCTION

The ability to make targeted changes in the genome of virtually any organism is transforming biological research. Early genome editing strategies used zinc-finger nucleases (Bibikova et al., 2001; Kim et al., 1996; Smith et al., 1999) or transcription activator-like effector nucleases (Boch et al., 2009; Christian et al., 2010; Moscou and Bogdanove, 2009) that required the construction of unique proteins for each target site. In contrast, the discovery that a chimeric single-guide RNA (sgRNA) can direct the *Streptococcus pyogenes* type II clustered regular interspersed short palindromic repeat (CRISPR)-associated protein 9 (Cas9) to catalyze site-specific double-stranded DNA breaks (DSBs) has eliminated laborious protein construction (Jinek et al., 2012; Qi et al., 2013). To date, Cas9 is active in all tested organisms including bacteria, plants, fungi and animals (for reviews see (Sander and Joung, 2014; Govindan and Ramalingam, 2016; Sternberg and Doudna, 2015; Hsu et al., 2014).

DSBs induced by sgRNA-guided Cas9 stimulate host DNA repair pathways. In many cases the breaks are perfectly rejoined, recreating the original target site, which can be cut again. Occasionally, error-prone end joining inserts or deletes nucleotides at the target site thereby preventing re-cutting. Such insertions, deletions, and substitutions, collectively called indels, can disrupt a protein-coding sequence. When a DNA donor is supplied exogenously, the DSB can be repaired by homologous recombination (HR), allowing the incorporation of

novel sequences at the target site. Unlike sequences incorporated via transgenes, modifying an endogenous gene preserves the chromatin context, enhancers, promoters, introns, and post-transcriptional regulatory elements of the wild-type locus.

Cas9-mediated genome editing requires just three components: (1) Cas9, which can be provided as a purified protein, mRNA, or gene; (2) sgRNA, which can be provided as an RNA or transcribed in vivo from a DNA template; and (3) a DNA donor bearing the target sequence containing indels or novel sequences to be incorporated. In *Drosophila*, providing Cas9, sgRNA, and donor DNA transgenes efficiently triggers editing, but establishing the requisite fly stocks takes over a month (Port et al., 2015; Port et al., 2014; Chen et al., 2015; Izumi et al., 2013). Injecting sgRNA and donor DNA into Cas9-expressing embryos requires far less time but is also less efficient, making it necessary to screen large numbers of animals. Co-integrating a visible marker such as GFP into the target locus can speed the identification of recombinants (Port et al., 2015; Chen et al., 2015; Yu et al., 2014; Ren et al., 2014a; Port et al., 2014; Gratz et al., 2014; Ren et al., 2014b; Zhang et al., 2014b; Taylor et al., 2013). However, removing the GFP marker by site-specific recombination (e.g., Cre-*LoxP*) takes multiple generations, negating the time advantage of injection and leaving a "scar" sequence (e.g., *LoxP*) at the target site. Indels, of course, must be identified molecularly or through complementation analysis.

In *Caenorhabditis elegans*, co-conversion strategies targeting a marker gene together with the gene-of-interest speed the screening for indels and recombinants and avoid introducing an exogenous marker gene at the target locus (Kim et al., 2014; Ward, 2015; Arribere et al., 2014). The co-conversion strategy restricts molecular screening to marker-positive animals, substantially reducing the work required to find mutant or recombinant animals. In theory, a similar co-conversion system should speed genome editing in *Drosophila melanogaster*.

Here, we describe a strategy in which co-targeting the eye-color gene *white* (*w*) speeds identification of both mutants and recombinants at the gene-of-interest. In our strategy, indels generate loss-of-function *w* mutants whose eyes are white, instead of the wild-type red. In contrast, recombination with the exogenous w^{coffee} (w^{cf}) donor DNA produces flies with reddish-brown eyes. Mating the injected animals to w^{1118} null flies and examining the eye color of their offspring allows rapid identification of parents that produce only w^{-} or w^{cf} gametes. These flies have an enhanced frequency of indels or recombination at the co-targeted gene-of-interest.

While developing this co-conversion strategy for fly genome editing, we also discovered that Cas9-induced recombinants frequently harbor undesirable integration of the entire donor plasmid at the target locus. We find that inducing gap repair with a pair of sgRNAs increases the likelihood of conversion of the

intervening target region. Moreover, when DSBs are repaired by end joining, the junction site frequently contains microhomologies or templated insertions, suggesting that the Cas9-catalyzed DSBs are repaired by the microhomology-mediated end-joining pathway and not by the canonical Ligase 4 (Lig4)-dependent non-homologous end joining; injecting into Cas9-expressing *Lig4*^{169} mutants to block canonical end joining neither decreases the yield of indels nor increases the yield of recombinants. Our protocol should reduce the time and effort needed to modify specific loci in the *Drosophila* genome, especially when generating Cas9-induced recombinants.

RESULTS

white co-conversion facilitates screening for both indels and recombinants

Changes in eye color are among the most readily identified phenotypes in *Drosophila*. Wild-type eyes are bright red with an obvious pseudopupil. Mutations in *w* generate eye colors ranging from brown to yellow for hypomorphic alleles and white for null alleles. Among the alleles of *w* that are caused by point mutations, w^{coffee} (w^{cf}) (Zachar and Bingham, 1982) was chosen as the co-conversion marker because of its easy-to-screen, reddish brown eyes lacking a pseudopupil. We designed a *w* sgRNA that directs Cas9 to cut 5 bp upstream of the w^{cf} mutation (*w* sgRNA-1) and an HR donor comprising 2080 bp from the w^{cf} allele, which differs from wild-type *w* by both a GC-to-AA mutation that creates a G589E missense mutation in the White protein (Mackenzie et al., 1999) and silent mutations that confer resistance to the *w* sgRNAs (Figure 3.1A). HR-mediated repair of the Cas9-catalyzed DSB produces coffee-colored eyes, whereas imprecise end joining generates white eyes when an indel disrupts function of the *w* mRNA or protein. Importantly, ectopic insertion of the HR donor will not produce the coffee-eye phenotype, as the donor carries only 1,144 bp of the 2,064 bp *w* coding sequence.

To test this strategy, a plasmid containing the w^{cf} HR donor, a plasmid containing the donor for the gene-of-interest, and a plasmid engineered to express both the *w* sgRNA and an sgRNA targeting the gene-of-interest were co-

injected into *Lig4*$^{+}$ or *Lig4*$^{-}$ preblastoderm embryos that express *S. pyogenes* Cas9 (*vas*-Cas9) (Gratz et al., 2014). The adult flies that developed from the injected embryos were mated with w^{1118} flies; the eye-colors of the resulting G1 offspring revealed the *w* genotype of the germline stem cells of the G0 parent. The G1 progeny included coffee-, white-, and red-eyed flies (Figure 3.1B). Sequencing white and coffee G1 flies confirmed that white-eyed flies (*n* = 10/10) had indels at the target site in *w*, whereas flies with coffee-colored eyes contained the G1766A, C1767A w^{cf} mutation (*n* = 6/6). Thus, eye color provides an effective reporter for *w* sgRNA-directed mutagenesis in the fly germline.

Some G0 produced broods with uniformly red, white or coffee eyed-flies, while others produced broods comprising flies of all possible combinations of the three eye colors. Editing of *w* can occur early in any of the dozens of pole cells that form at the posterior pole of the syncytial blastoderm embryo or later in the descendants of these germ cell progenitors. Because individual G0 pole cells may incorporate different amounts of the injected plasmids, the frequency of DNA cleavage by sgRNA-guided Cas9 and the choice of repair pathways will differ among germ cells, generating variation in the ratio of red, white, and coffee-eyed G1 flies. The percentage of non-red G1 flies should reflect the allele frequency of mutant chromosomes in G0 germline stem cells, which in turn reflects the overall targeting efficiency.

Figure 3.1

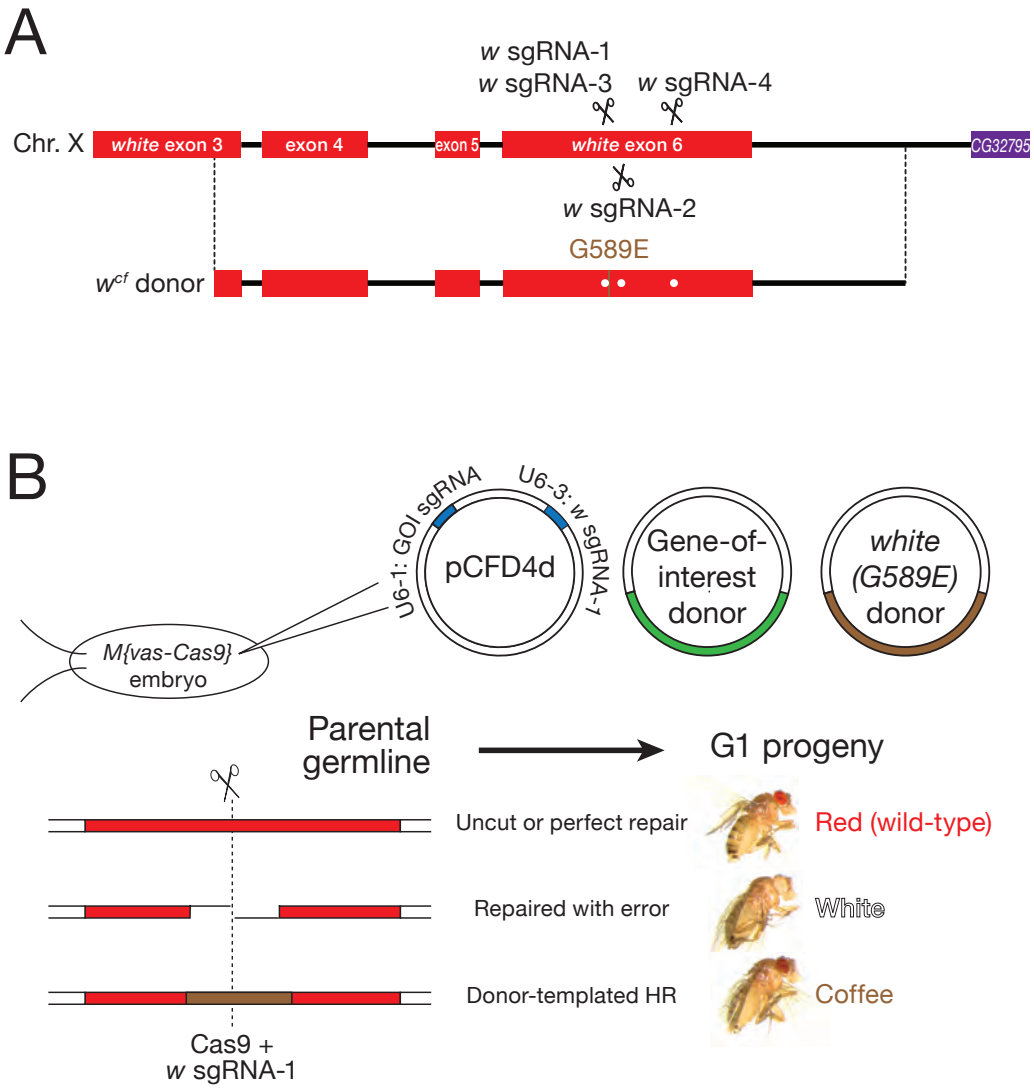

A
w sgRNA-1
w sgRNA-3
w sgRNA-4
Chr. X
white exon 3
exon 4
exon 5
white exon 6
CG32795
w sgRNA-2
G589E
w^cf donor
B
U6-1: GOI sgRNA
U6-3: w sgRNA-1
pCFD4d
Gene-of-interest donor
white (G589E) donor
M{vas-Cas9} embryo
Parental germline
G1 progeny
Uncut or perfect repair
Red (wild-type)
Repaired with error
White
Donor-templated HR
Coffee
Cas9 + w sgRNA-1

Figure 3.1: *white* co-conversion strategy

(A) The eye pigment gene *white* was co-targeted with the gene-of-interest. The w^{cf} HR donor carries a GC-to-AA mutation that creates a G589E missense mutation in the White protein. Flies homozygous or hemizygous for w^{cf} (i.e., w^{cf}/w^{1118} or w^{cf}/Y) have coffee, instead of the wild-type red, eyes. Scissors mark the target loci of the *white* sgRNAs. Dots on the donor plasmid mark silent mutations that confer resistance to the *white* sgRNAs. (B) Plasmids expressing *w* sgRNA-1 and an sgRNA targeting the gene-of-interest, a plasmid containing the donor for the gene of interest (GOI), and a plasmid containing the w^{cf} donor were co-injected into *Drosophila* syncytial blastoderm embryos that express transgenic Cas9 (*vas*-Cas9). The double-strand break created by *w* sgRNA-1-guided Cas9 may be repaired either perfectly, with nucleotide insertion or deletion (indels), or with sequence copied from the co-injected exogenous donor DNA. The eye color of the G1 progeny reflects the repair mechanism: red eyes indicate perfect repair or no cutting by Cas9; white indicates creation of an indel; and coffee reflects repair by HR.

To test this idea, we assigned each fertile G0 to one of six groups according to the eye color composition of its G1 brood: (1) all red; (2) white and red; (3) all white; (4) coffee and red or coffee, white, and red; (5) coffee and white; and (6) all coffee (Table 3.1). Six independent experiments co-targeted *w* and *armitage* (*armi*), a third chromosome gene; one experiment co-targeted *w* and *zucchini* (*zuc*), a second chromosome gene. Representative numbers of broods across the six eye color groups were screened by genotyping 9–10 G1 flies from each brood for sequence changes at the gene-of-interest (i.e., *armi* or *zuc;* Table 3.1). For simplicity, we combined the three groups containing no red-eyed progeny into a single category, “no red in broods,” and the three groups containing at least some red-eyed flies into a single category, “with red in broods.” The fraction of broods that yielded indels or recombinants was 21% ± 19% in the “with red” category, and 65% ± 34% (mean ± S.D.) in the “no red” category (Figure 3.2). Therefore, screening for mutations at a gene-of-interest can be restricted to the “no red” broods, which account for 6.3–21% of all broods (mean ± S.D. = 14% ± 6%, Table 3.1). For these seven experiments, *w* co-conversion would have successfully identified mutants in the gene-of-interest by screening just the 37 “no red” broods (14% of the total 272) using a simple genetic scheme (Figure 3.3 and *Experimental Procedures*).

Table 3.1

Shape	G0 *Lig4* genotype	sgRNA plasmid	Donor plasmids	Fertile G0 (*n*)	Number of G0 whose G1 offspring had eyes that were: All red	White & red	Coffee & red/white	All white	Coffee & white	All coffee
●	***Lig4*$^{169\ a}$**	armi-1 & w-1 (78 nM)	armi & w (132 nM ea.)	8.9% (25/281)	10 EJ: 1/10 HR: 0/10	10 EJ: 2/10 HR: 0/10	2 EJ: 0/2 HR: 0/2	1 EJ: 1/1 HR: 1/1	0	2 EJ: 1/2 HR: 2/2
■	***Lig4*^{169}**	armi-1 & w-1 armi-2 & armi-3 (26 nM ea.)	armi & w (33 nM ea.)	17% (43/260)	32 –	7 EJ: 2/7 HR: 2/7	1 EJ: 0/1 HR: 0/1	1 EJ: 0/1 HR: 1/1	1 EJ: 1/1 HR: 0/1	1 EJ: 1/1 HR: 1/1
▲	***Lig4*$^{+}$**	armi-3 & armi-4 w-1 & w-1 (26 nM ea.)	armi & w (132 nM ea.)	11% (25/230)	5 –	14 EJ: 0/5 HR: 0/5	2 EJ: 0/1 HR: 0/1	0	0	4 EJ: 1/3 HR: 0/3
◆	***Lig4*$^{+}$**	armi-3 & armi-4 w-1 & w-4 (26 nM ea.)	armi & w (132 nM ea.)	20% (44/220)	27 EJ: 0/13 HR: 0/13	5 EJ: 0/2 HR: 0/2	3 EJ: 0/3 HR: 1/3	3 EJ: 0/3 HR: 0/3	3 EJ: 1/3 HR: 1/3	3 EJ: 0/3 HR: 1/3
⬣	***Lig4*$^{+}$**	armi-5 & armi-6 w-1 & w-1 (26 nM ea.)	armi & w (132 nM ea.)	29% (64/220)	31 EJ: 0/9 HR: 0/9	20 EJ: 0/10 HR: 0/10	9 EJ: 3/9 HR: 0/9	0	1 EJ: 0/1 HR: 0/1	3 EJ: 1/2 HR: 0/2
▼	***Lig4*$^{+}$**	armi-5 & armi-6 w-1 & w-1 (26 nM ea.)	armi & w (132 nM ea.)	17% (38/220)	7 EJ: 0/5 HR: 0/5	12 EJ: 0/6 HR: 2/6	12 EJ: 2/6 HR: 0/6	1 EJ: 1/1 HR: 0/1	4 EJ: 3/3 HR: 0/3	2 EJ: 1/2 HR: 1/2
★	***Lig4*$^{+}$**	zuc-1 & zuc-2 w-1 & w-1 (26 nM ea.)	zuc & w (132 nM ea.)	15% (33/222)	8 EJ: 1/7 HR: 0/7	10 EJ: 4/6 HR: 1/6	8 EJ: 6/7 HR: 3/7	2 EJ: 2/2 HR: 2/2	3 EJ: 2/3 HR: 3/3	2 EJ: 1/2 HR: 1/2

Table 3.1: Co-targeting *white* and a gene-of-interest

sgRNA-expressing and HR donor plasmids were co-injected into *Lig4*^{169} or *Lig4*$^{+}$, *vas*-Cas9 G0 embryos. Shapes identify the corresponding experiment in Figure 3.2. *n*, total number of G0 embryos injected, irrespective of fertility or survival. "Coffee & red/white" includes G0 with coffee and red-eyed, or with coffee-, white-, and red-eyed G1 broods. EJ, broods yielding indels; HR, broods yielding homologous recombinants (plasmid integration and gene conversion); conversion tracts were analyzed only for gene conversion events (Figure 3.6 and Figure 3.7). [a]Co-injected with 1.2 µM of NLS-Cas9 protein (PNA-Bio, Inc., Thousand Oaks, CA, USA), which had no observable effect.

Figure 3.2

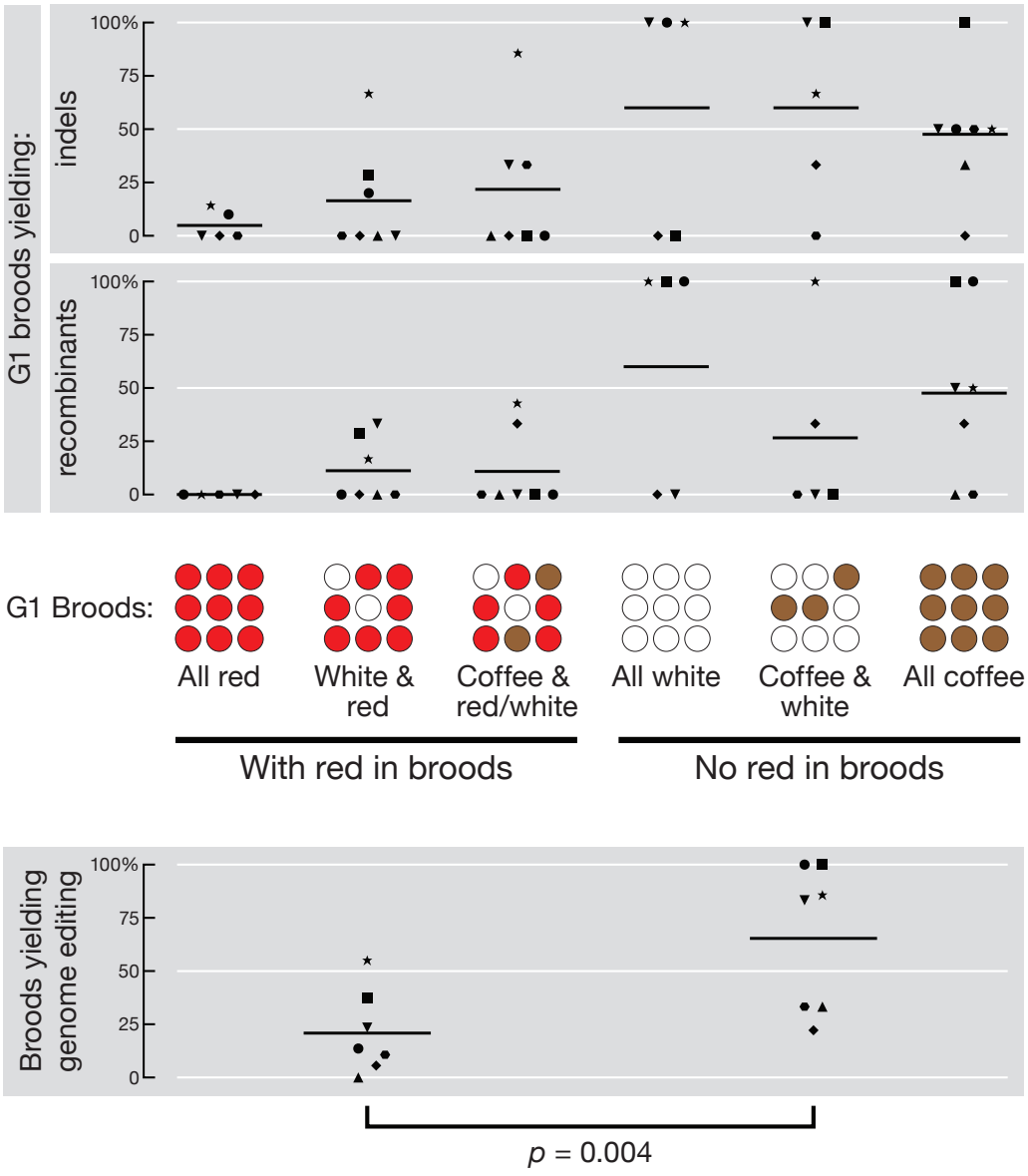

G1 broods yielding:
indels
recombinants
100%
75
50
25
0
G1 Broods:
All red
White & red
Coffee & red/white
All white
Coffee & white
All coffee
With red in broods
No red in broods
Broods yielding genome editing
p = 0.004

Figure 3.2: Co-occurrence of *w* and gene-of-interest genomic editing events

Adults from injected G0 embryos that produce G1 broods are either divided into six groups according to their eye color composition: (1) all red; (2) white and red; (3) coffee and red or coffee, white, and red; (4) all white; (5) coffee and white; and (6) all coffee (upper panel), or divided into two categories, “with red in broods” (groups 1–3) and “no red in broods” (groups 4–6; lower panel). For each experiment, the number of broods yielding indels, recombinants (upper panel) or both editing events (lower panel) at the gene-of-interest, as identified by PCR screening of individual G1 progeny, is reported as percentage of total broods sampled. Shapes of data points represent individual experiments described in Table 3.1. Line presents the mean across all seven experiments. *p*: two-tailed, paired *t*-test.

Figure 3.3

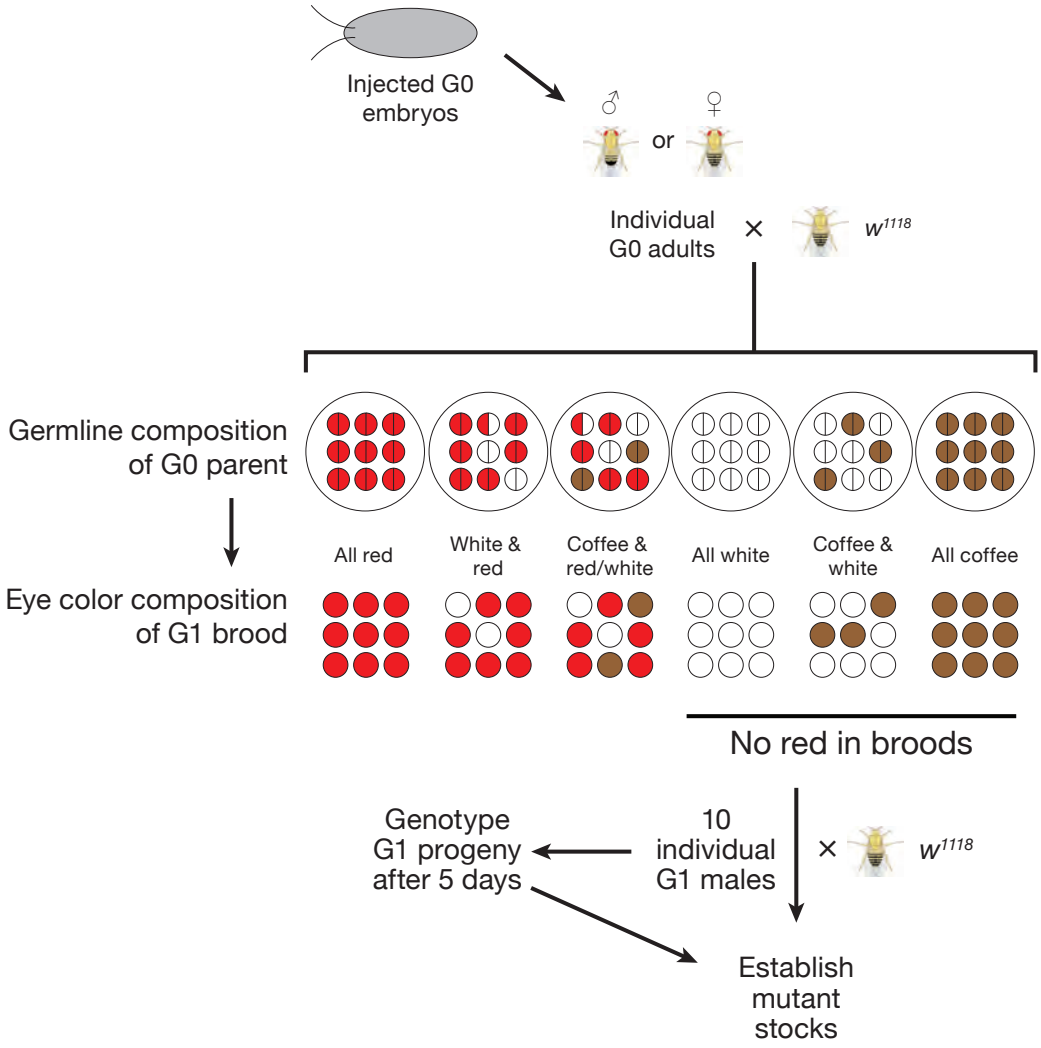

Injected G0 embryos
♂
or
♀
Individual G0 adults
×
w1118
Germline composition of G0 parent
Eye color composition of G1 brood
All red
White & red
Coffee & red/white
All white
Coffee & white
All coffee
No red in broods
Genotype G1 progeny after 5 days
10 individual G1 males
×
w1118
Establish mutant stocks

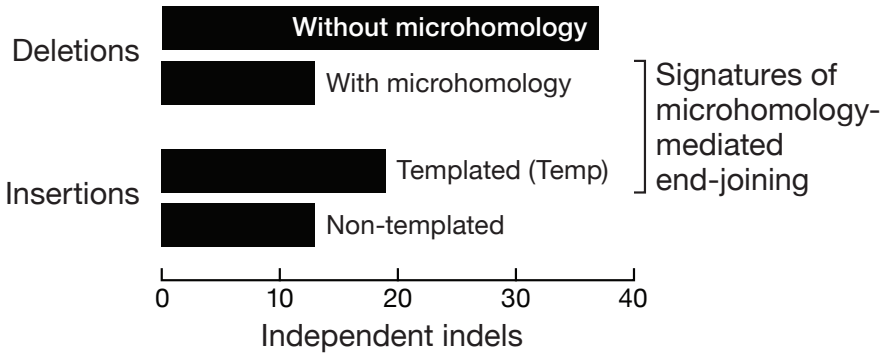
Figure 3.3: Genetic scheme used to screen and establish CRISPR-edited stocks

Microhomology-mediated end joining is frequent

We identified 82 independent indels at seven sgRNA target sites (Figure 3.4B and Table 3.2 to Table 3.7), and grouped them by ligation junction signatures. Two types of deletions were observed: 13 events showed a pair of ≥ 2 nt long, identical sequences (microhomology) being reduced to a single sequence via sticky-end ligation; the other 37 events reflected either blunt junctions or only 1 nt of microhomology (Figure 3.4A). Two types of insertions, often after a deletion, were observed: for 19 events, a sequence ≥ 3 nt long near the cleavage site appeared to have served as a template for the inserted nucleotides; in the other 13 events, the insertions lacked an obvious template, or were shorter than 3 nucleotides (Figure 3.4A). Both junctional microhomologies (16% of all events) and templated insertions (23%) are likely products of the microhomology-dependent end joining pathway, a form of alternative end joining that does not require the canonical non-homologous end joining proteins Ku70/80 or Ligase 4 (Sfeir and Symington, 2015; Yu and McVey, 2010; Chan et al., 2010). Consistently, injecting *Lig4*^{169} null mutant embryos (McVey et al., 2004b) produced microhomologies and templated insertions at *white* or *armi* (Figure 3.4B and Table 3.4 and Table 3.5).

Figure 3.4

A



B

N	PAM-distal	Junction	PAM-proximal	Type
	CATACCATTCCTGCTCTT |||||||||||||||||| GTATGGTAAGGACGAGAA	- -	TGGCGGCTTCT ||||||||||| ACCGCCGAAGA	
1	CATACCATTCCTGCTC--	(T)	-GGCGGCTTCT	No MH
1	CATACCATTCCT------	(GC)	----GGCTTCT	MH
1	CATACCA----------	(TTC)	----------T	MH
1	CATACCATTCCTGCTCTT	CTTC	TGGCGGCTTCT	Temp Ins
1	CATACCATTCCTGC----	AGGAA	TGGCGGCTTCT	Temp Ins
1	CATACCATTCCTGCT---	TCAGGAAGCC	--GCGGCTTCT	Temp Ins
1	CATACCATTCCTGCTC--	CATACCATTCC	-GGCGGCTTCT	Temp Ins
1	CATACCATTCCTGCT---	TC	-GGCGGCTTCT	Ins
1	CATACCATTCCTGCTCTT	CC	TGGCGGCTTCT	Ins

Figure 3.4: Indel junctional signatures suggest the involvement of microhomology-mediated end joining

(A) Eighty-two independent indels at seven DSBs (Figure 3.4B, Table 3.2 to Table 3.7) were classified as deletion without microhomology when there was ≤ 1 nt of microhomology; as deletion with microhomology when there were ≥ 2 nt of microhomology; as templated insertion when there where ≥ 3 nt of inserted nucleotides with identifiable template; or as non-templated insertion when nucleotide insertions were present without an identifiable template. (B) Indels at the *white* sgRNA-1 target site. The 20 nt sgRNA target sequence is in grey. The PAM sequence is in red. The DSB junction is 3 bp away from the PAM. Dash: deleted nucleotide. Underline: templated insertions at the junction. Nucleotides in parentheses identify microhomologies that can be mapped to either the PAM-distal or PAM-proximal side of the DSB. WT, wild type; No MH, deletion without microhomology; MH, deletion with microhomology; Temp Ins, templated insertion; Ins, non-templated insertion. *N*, number of independent events. G0 embryos were *vas-Cas9, lig4*^{169}*.*

Table 3.2

N	PAM-distal	Junction	PAM-proximal	Type
	GTGATTATCCGGATCATCAG	**–**	**CGACGGGGAGAT**	**WT**
1	GTGATTATCCGGATCATCAG	–	----GGGGAGAT	No MH
6	GTGATTATCCGGATCATC--	–	CGACGGGGAGAT	No MH
1	GTGATTATCCGGATCAT---	–	-----GGGAGAT	No MH
1	GTGATTATCCGGATC-----	–	CGACGGGGAGAT	No MH
1	GTGATTATCCGGATC-----	–	---CGGGGAGAT	No MH
1	GTGATTATCCGGA-------	–	---CGGGGAGAT	No MH
1	GTGATTATCCGGATCATC--	(A)	---CGGGGAGAT	No MH
1	GTGATTATCCGGATC-----	(A)	---CGGGGAGAT	No MH
1	GTGATTATCCGGATCAT---	(C)	-GACGGGGAGAT	No MH
2	GTGATTATCCGGATCAT---	(C)	----GGGGAGAT	No MH
2	GTGATTATCCGGATCATCA-	(G)	-----GGGAGAT	No MH
4	GTGATTATCCG---------	(GA)	---CGGGGAGAT	MH
1	GTGATTATCCGGATCATC--	G	----GGGGAGAT	Ins
1	GTGATTATCCGGAT------	TT	----GGGGAGAT	Ins
1	GTGATTATCCGGATC-----	CG	-GACGGGGAGAT	Ins
1	GTGATTATCCGGATCATCAG	GGA	-GACGGGGAGAT	Temp Ins
1	GTGATTATCCGGATC-----	CCCG	-GACGGGGAGAT	Temp Ins
1	GTGATTATCCGGATCATCAG	GACG	-GACGGGGAGAT	Temp Ins
1	GTGATTATCCGGATCATCAG	GGAGAT	------GGAGAT	Temp Ins
1	GTGATTATCCGGATCAT---	TATTATC	CGACGGGGAGAT	Temp Ins
1	GTGATTATC----------	GGGATTAT	CGACGGGGAGAT	Temp Ins

Table 3.2: Indels at the zuc sgRNA-1 target site

See Figure 3.4 for details. G0 embryos were *vas-Cas9, lig4*$^{+}$.

Table 3.3

N	PAM-proximal	Junction	PAM-distal	Type
	TTCCTCAG	**–**	**TATGCCAAATAATTATAG**	**WT**
2	TTCCTCAG	–	–ATGCCAAATAATTATAG	No MH
1	TTCCTC––	(A)	––TGCCAAATAATTATAG	No MH
1	TTCCTCA–	(G)	––––CCAAATAATTATAG	No MH
1	TTCCTCAG	T	TATGCCAAATAATTATAG	Ins
1	TTCCTCAG	T	––TGCCAAATAATTATAG	Ins
1	TTCCTCAG	TTCCTC	––TGCCAAATAATTATAG	Temp Ins
1	TTCCTCA–	AATAAATAATAATTCCTCAA	–ATGCCAAATAATTATAG	Temp Ins

Table 3.3. Indels at the zuc sgRNA-2 target site

See Figure 3.4 for details. G0 embryos were *vas-Cas9, lig4*$^{+}$.

Table 3.4

N	PAM-distal	Junction	PAM-proximal	Type
	ATCAGAAAGATGAGGTCCCCGTTT	**–**	**GCTTGGAAACATTC**	**WT**
1	ATCAGAAAGATGAGGTCC------	–	---TGGAAACATTC	No MH
1	ATCAGAAAGATGAGGT--------	–	--TTGGAAACATTC	No MH
1	----(43 bp deletion)----	–	--TTGGAAACATTC	No MH
1	ATCAGAAAGATGAGGTCCC-----	(C)	--TTGGAAACATTC	No MH
5	ATCAGAAAGATGAGGT--------	(C)	--TTGGAAACATTC	No MH
1	ATCAGAAAGATGAG----------	(G)	-CTTGGAAACATTC	No MH
1	A-----------------------	(T)	---TGGAAACATTC	No MH
1	ATCAGAAAGATGA----------	(GG)	------AAACATTC	MH
1	ATCAGAAAGATGAGGT--------	CGG	----GGAAACATTC	Ins
1	ATCAGAAAGATGAGGT--------	GGG	----GGAAACATTC	Ins
1	ATCAGAAAGATGA----------	TGTCGTGAATGTGG	---TGGAAACATTC	Temp Ins

Table 3.4: Indels at the armi sgRNA-1 target site

See Figure 3.4 for details. G0 embryos were *vas-Cas9, lig4*^{169}.

Table 3.5

sgRNA-2

N	PAM-proximal	Junction	PAM-distal	Type
	TTCCAGGA	**–**	**ACGTGACATTATACTAA**	**WT**
1	TT------	AT	-----ACATTATACTAA	Ins

sgRNA-3

N	PAM-distal	Junction	PAM-proximal	Type
	ACTTAACGTTCGTTATT	**–**	**TCCAGGAA**	**WT**
1	ACTTAACGTTCGTTA--	ACGTTAA	--CAGGAA	Temp Ins

Table 3.5: Indels at armi sgRNA-2 or sgRNA-3 target site

See Figure 3.4 for details. G0 embryos were *vas-Cas9, lig4*^{169}.

Table 3.6

N	PAM-distal	Junction	PAM-proximal	Type
	ATATGTGCTAATTTTAGC	**–**	**GAATGGTA**	**WT**
1	ATATGTGCT---------	(AAT)	----GGTA	MH
1	ATATGTGCTAAT------	GGTA	--ATGGTA	Temp Ins

Table 3.6: Indels at armi sgRNA-4 target site

See Figure 3.4 for details. G0 embryos were *vas-Cas9, lig4*$^{+}$.

Table 3.7

N	PAM-distal	Junction	PAM-proximal	PAM-distal	Junction	PAM-proximal	Type
	GAAACTTTTTACAATTGGGGTTTC	**–**	**CAGCGGGCCATTTCACTCACGTGCTGTTCGATGA**		**–**	**AGCTGGTCAATGCA**	**WT**
1	GAAACTTTTTACAATT--------	(GGG)	-------CCATTTCACTCATGTGCTGTTCAACCA		–	AGCTGGTCAATGCA	MH;HR
1	GAAACTTTTTACAATTGGGGTTTC	–	CAGCGGGCCATTTCACTCACGTGCTGTTCGA---		(TG)	-----GTCAATGCA	MH
3	GAAACTTTTTACAATTGGGGTTTC	–	CAGCGGGCCATTTCACTCACGTGCTGTTCG----		(ATG)	----------CA	MH
1	GAAACTTTTTACAATTGGGGTTTC	ATTCCATGGGGAATGAAACGTGG	----GGGCCATTTCACTCACGTGCTGTTCGATGA		–	AGCTGGTCAATGCA	Ins
1	GAAACTTTTTACAATTGGGGTTTC	–	–		–	AGCTGGTCAATGCA	No MH
2	GAAACTTTTTACAATTGGGGTTT-	–	(C)		–	---TGGTCAATGCA	No MH
1	GAAACTTTTTACAATTGGGGTT--	–	GG		–	-GCTGGTCAATGCA	Ins
1	GAAACTTTTTACAATTGGGGTT--	–	GTAAA		–	---------ATGCA	TempIns
1	GAAACTTTTTACAATTGGGGT---	–	CATTGACC		–	AGCTGGTCAATGCA	TempIns
1	GAAACTTTTTACAATTGGGG----	–	GATCCTTTTTACAATT		–	-GCTGGTCAATGCA	TempIns
1	GAAACTTTTTACAATTGG------	–	TCAATG(..68 nt..)GTGATC		–	--CTGGTCAATG(..68nt..)GTGATCCTC	TempIns

Table 3.7: Indels at armi sgRNA-5 or sgRNA-6 target site.

Nucleotides with boarders: substitutions as a result of HR using exogenous donor as template. HR: homologous recombination. See Figure 3.4 for details. G0 embryos were *vas-Cas9, lig4*$^{+}$.

A circular plasmid donor frequently integrates at the target locus

HR in the gene-of-interest was identified by PCR screening using a primer that binds within both the donor and the genomic locus and a primer that binds exclusively to the genomic sequence. This primer pair can amplify the original or the edited genomic locus, but not donor DNA present extra-chromosomally or integrated at an ectopic location. As previously reported (Yu et al., 2014), some of the recombinants identified by this strategy corresponded to genomic integration at the gene-of-interest of the entire donor, including the plasmid backbone. In addition to converting the genomic locus to the donor sequence, these recombination events also duplicate the genomic sequence present in the donor (Figure 3.5A). To distinguish between gene conversion and plasmid integration, we repeated the PCR using primers binding only to the genome and not to sequence present in the HR donor. This strategy readily identified plasmid integration events by their lack of a PCR product or the amplification of a larger-than-expected product. Of the 16 independent HR events identified at *armi*, seven reflected gene conversion while nine integrated the plasmid, a 56% false-positive rate; of the 12 independent HR events identified at *zuc*, ten underwent gene conversion while two integrated the plasmid, a 17% false-positive rate (Figure 3.5B).

Figure 3.5

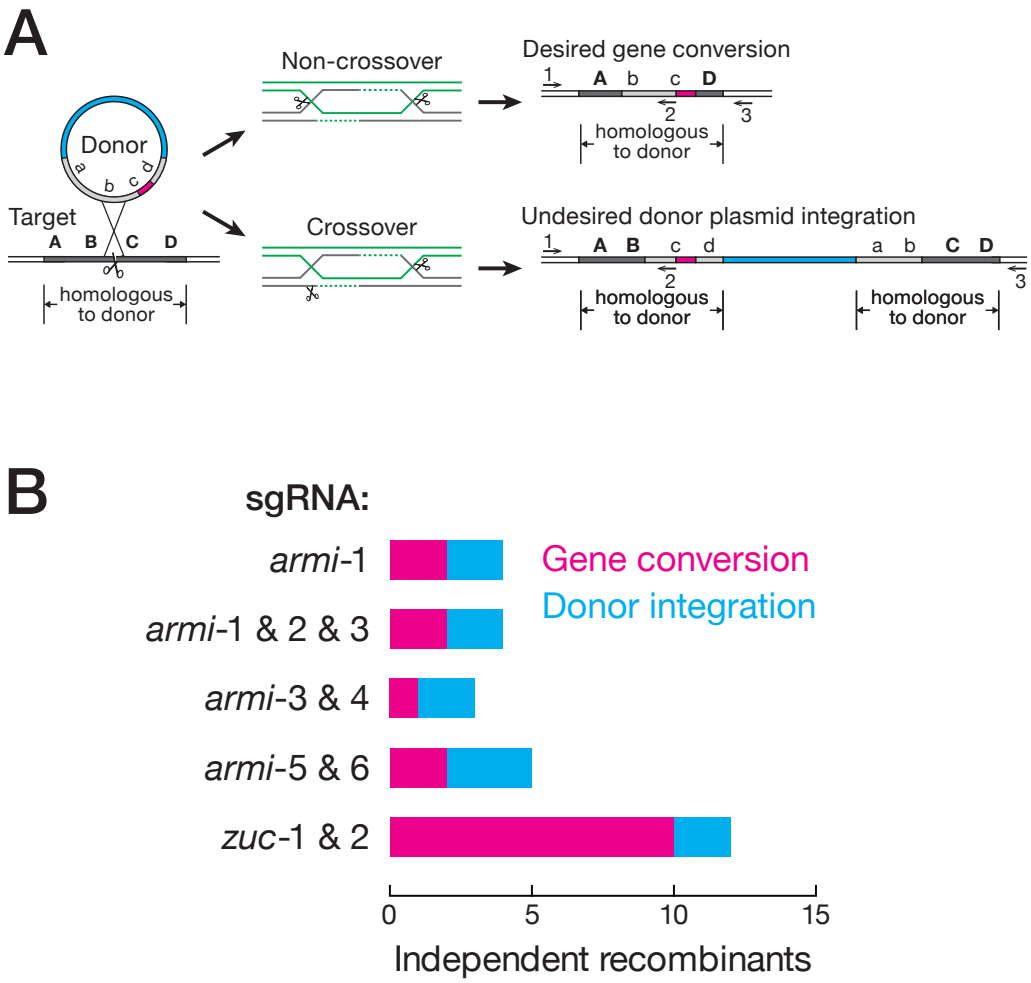

A
Donor
a
b
c
d
Target
A B C D
homologous to donor
Non-crossover
Crossover
Desired gene conversion
1
A b c D
2
3
homologous to donor
Undesired donor plasmid integration
1
A B c d a b C D
2
3
homologous to donor
homologous to donor
B
sgRNA:
armi-1
armi-1 & 2 & 3
armi-3 & 4
armi-5 & 6
zuc-1 & 2
Gene conversion
Donor integration
0
5
10
15
Independent recombinants

Figure 3.5: HR using a circular plasmid donor produces either gene conversion or plasmid integration

(A) Two possible outcomes for HR depending on the resolution of double Holliday junctions. PCR primers 1 and 2 can exclude donors present extra-chromosomally or ectopically integrated, but cannot differentiate between gene conversion and plasmid integration at the target locus. PCR primers 1 and 3 both bind to the genome and not the donor, allowing unambiguous detection of gene conversion events. (B) Number of gene conversion versus plasmid integration events obtained using different sgRNA combinations. See also Table 3.1.

Gap repair reliably converts the intervening sequence

When gene conversion occurs, the genomic sequence replaced by donor sequence is termed the “conversion tract.” If the conversion tract is short, mutations can only be introduced near the DSB. On the other hand, long conversion tracts allow a single HR event to introduce multiple mutations that are distant from the sgRNA-complementary site. Given that each gene-targeting experiment in *Drosophila* takes two to three months to accomplish, the ability to introduce two or more edits via a long conversion tract is advantageous. We therefore determined the length of conversion tracts in our experiments.

To introduce a peptide tag at the carboxy terminus of the Armi protein, we assembled a donor plasmid harboring 2,280 bp of sequence from the endogenous *armi* locus and introducing a Strep-tag II peptide tag before the stop codon (Figure 3.6). The donor harbored nineteen sites different in sequence from the injected strain, allowing measurement of the length of the conversion tract. We first designed *armi* sgRNA-1 to target a sequence near the end of *armi* exon 8. The HR donor contained 1,809 bp upstream and 484 bp downstream of the predicted DSB, and templated two gene conversion events (Table 3.1). One tract was unidirectional: only the sequence downstream of the DSB (≥ 77 bp) was converted; the other tract had between 1,396–1,804 bp upstream and ≥ 77 bp downstream of the DSB converted to the sequence of the donor DNA (Figure 3.6). We then used two adjacent sgRNAs, sgRNA-5 and -6, targeting sequences

in *armi* exon 7. Because the two sgRNA have predicted cleavage sites separated by just 34 bp, we considered them to be a single target site. The same HR donor now contains 790 bp upstream and 1,503 bp downstream of the target site, and templated four gene conversion events (Table 3.1). The first tract converted between 377–785 bp upstream and between 68–263 bp downstream of the target site; the second between 154–377 bp upstream and between 989–1,068 bp downstream; and the third between 377–785 bp upstream and between 989–1,068 bp downstream (with 37 bp deleted in the middle of the downstream conversion tract). The fourth tract only converted 15 bp upstream of the predicted DSB generated by armi sgRNA-6, and carried a 12 bp-deletion at the predicted DSB generated by sgRNA-5, suggesting independent repair events induced by the two guides (Table 3.7). Therefore, conversion tracts initiated from the sgRNA-5/6 target site were unpredictable in directionality and length, just like the *armi* sgRNA-1 site (Figure 3.6).

In order to more reliably predict the coverage of conversion tracts, we reasoned that by deleting the entire target region, HR could be directed to replace the missing gap using the supplied donor DNA. To achieve this, we targeted armi exon 8 with a pair of guides, sgRNA-3 and -4, whose predicted cleavage sites were separated by 454 bp. The donor includes 1,530 bp upstream of the first target site and 286 bp downstream of the second, and templated one gene conversion event (Table 3.1). As expected when both guides direct Cas9 to

cleave the genome, the 454 bp interval between the two DSBs was fully replaced with the sequence contained in the HR template plasmid (Figure 3.6).

We repeated the same strategy with three sgRNAs whose target sites were separated by 280 bp (sgRNA-1, sgRNA-2 and sgRNA-3; sgRNA-2 and -3 had predicted cleavage sites separated only by 7 bp therefore can be considered as a single target site). The donor included 1,530 bp upstream of the first target site and 484 bp downstream of the second, and templated three gene conversion events (Table 3.1). The first tract reliably replaced the 280 bp gap with that of the donor; the second tract converted between 1,117–1,525 bp upstream of the first target site in addition to a full replacement of the 280 bp gap. The third tract lacked gap repair: the first target site harbored a 2 bp insertion after an 11 bp deletion (Table 3.5); the second site harbored a ≥ 77 bp conversion tract downstream of the DSB. The 280 bp gap was not converted, suggesting separate repair events at the two target sites.

We observed a similar gap repair phenomenon when introducing sequence encoding a carboxy terminal 3×FLAG peptide tag into the *zucchini* genomic locus (Table 3.1 and Figure 3.7). The two guides, *zuc* sgRNA-1 and -2, targeted sites 395 bp apart. The *zucchini* HR template included 970 bp upstream of the first target site and 760 bp downstream of the second and templated 18 gene conversion events. Of the two gap repair events, one reliably converted the predicted gap, and the other converted ≥720 bp upstream of the first target site in

addition to fully replacing the 395 bp gap. The remaining 16 gene conversion events lacked gap repair: only markers near the zuc sgRNA-1 target site were converted. At the zuc sgRNA-2 target site, six contained indels, and ten had wild-type sequence, suggesting separate repair events at the two target sites.

Figure 3.6

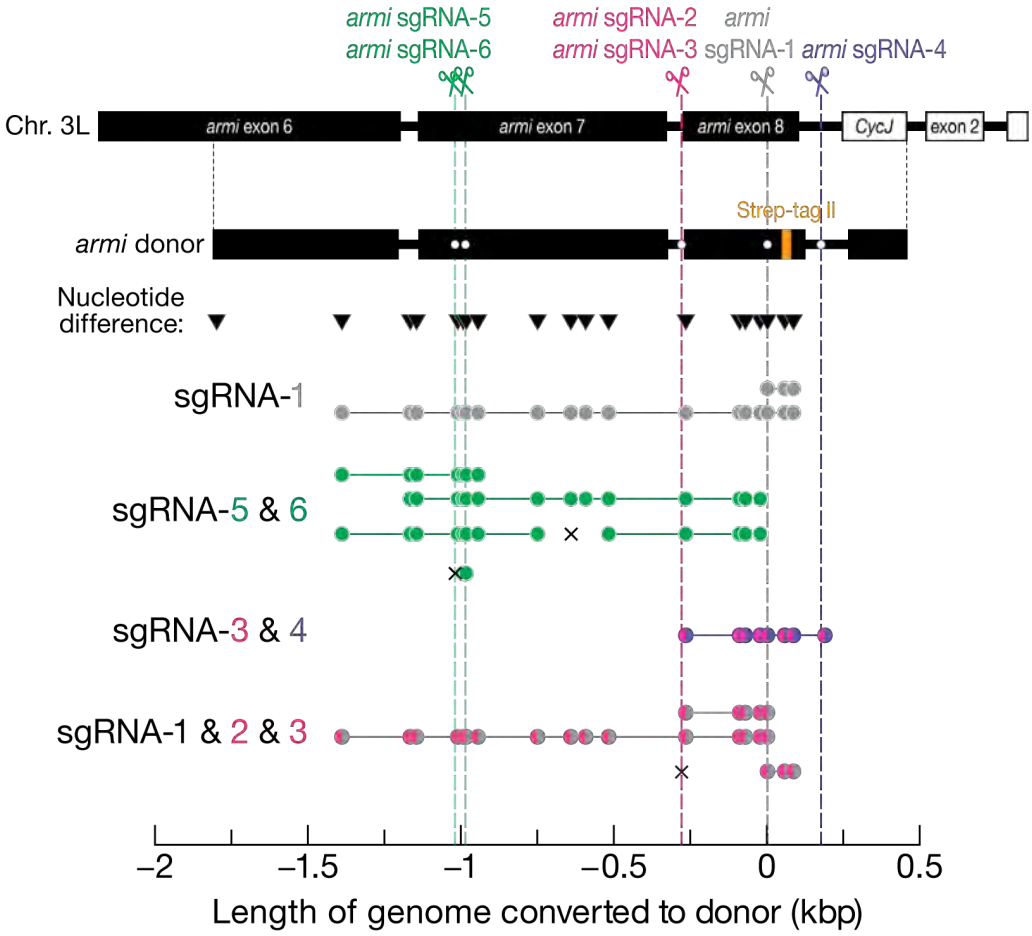

armi sgRNA-5
armi sgRNA-6
armi sgRNA-2
armi sgRNA-3
armi
sgRNA-1
armi sgRNA-4
Chr. 3L
armi exon 6
armi exon 7
armi exon 8
CycJ
exon 2
armi donor
Strep-tag II
Nucleotide difference:
sgRNA-1
sgRNA-5 & 6
sgRNA-3 & 4
sgRNA-1 & 2 & 3
−2
−1.5
−1
−0.5
0
0.5
Length of genome converted to donor (kbp)

Figure 3.6: Conversion tracts in *armi* recombinants

The homologous donor carried part of the *armi* gene, a Streptag II peptide tag at the end of coding sequence, and 18 sites (inverted triangles) differing in sequence from the endogenous locus that allowed mapping of conversion tracts. Dots on the donor plasmid mark silent mutations that confer resistance to the *armi* sgRNAs. Closed circle: site converted to the donor sequence. Each line presents one recombinant, and the color of closed circles corresponds to the color of the DSB(s) from which HR was initiated (dotted vertical lines). An × indicates an indel.

Figure 3.7

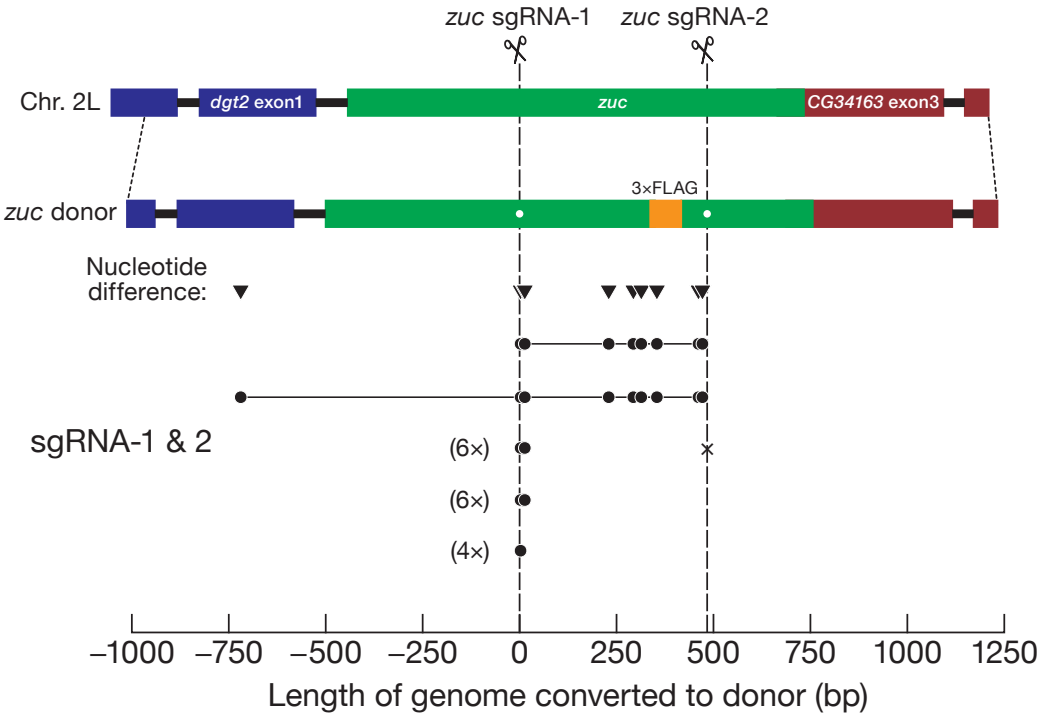

zuc sgRNA-1
zuc sgRNA-2
Chr. 2L
dgt2 exon1
zuc
CG34163 exon3
3×FLAG
zuc donor
Nucleotide difference:
sgRNA-1 & 2
(6×)
(6×)
(4×)
–1000
–750
–500
–250
0
250
500
750
1000
1250
Length of genome converted to donor (bp)

Figure 3.7: Conversion tracts in *zuc* recombinants.

See Figure 3.6 for details.

Ligase 4 mutation does not inhibit end joining or improve HR

In flies, mutation of *Ligase 4* (*Lig4*^{169}), a key enzyme in the canonical non-homologous end-joining pathway, has been proposed to promote HR by suppressing end joining. Zinc-finger nuclease-catalyzed DSBs yield a greater proportion of recombinants in *Lig4*^{169} null mutant embryos than in wild-type, but at the cost of decreased fitness of the injected animals (Ran et al., 2013; Bozas et al., 2009; Beumer et al., 2008). Inhibition of Ligase 4 using RNA interference or small molecule protein inhibitors similarly increased HR efficiency in mosquitos (Basu et al., 2015), mice (Maruyama et al., 2015), and cultured *Drosophila*, human, or mouse cells (Böttcher et al., 2014; Chu et al., 2015).

To test whether *Lig4*^{169} null mutants increased the yield of recombinants, we co-injected sgRNA-expressing and HR donor plasmids targeting *w* into *vas-Cas9, Lig4*^{169} or *vas-Cas9 Lig4*$^{+}$ embryos. We used the fraction of coffee-producing broods and percentage of coffee-eyed G1 in such broods to score for HR efficiency (Table 3.8 and Figure 3.8). Three independent comparisons were conducted, each with a unique sgRNA targeting *white*. *w* sgRNA-1 and sgRNA-3 were provided on the pCFD4d vector together with *armi* sgRNA-1. *w* sgRNA-2 was provided using the pDCC6 vector, which also encodes the Cas9 mRNA (*Experimental Procedures,* and Figure 3.1A). We detected no statistically significant difference between *Lig4*$^{+}$ and *Lig4*^{169} embryos in producing recombinant, coffee-eyed G1. Similarly, we observed no significant difference

between *Lig4*$^{+}$ and *Lig4*^{169} embryos in producing indels (white-eyed G1, Figure 3.8). Mothers homozygous for *vas*-Cas9 and either *Lig4*^{169} or *Lig4*$^{+}$ produce the expected 1:1 Mendelian ratio of red/coffee-eyed or red/white-eyed siblings, excluding the formal possibility that the Cas9-expressing, *Lig4*^{169} background affects the recovery of *w* mutant flies. We conclude that the use of *Lig4*^{169} embryos does not reduce the recovery of Cas9-induced indels or increase the rate of HR.

Table 3.8

			Percent of fertile G0 whose G1 offspring had eyes that were:					
***w* sgRNA**	**G0 *Lig4* genotype**	**Fertile G0 (*n*)**	**All red**	**White & red**	**Coffee & red/white**	**All white**	**Coffee & white**	**All coffee**
pCFD4d-1 (26 nM)	*Lig4*$^{+}$	23% (255)	48%	12%	8.6%	10%	17%	3.4%
	Lig4^{169}	14% (310)	76%	10%	7.1%	2.4%	0%	4.8%
pCFD4d-3 (26 nM)	*Lig4*$^{+}$	28% (240)	42%	26%	6.1%	4.5%	7.6%	14%
	Lig4^{169}	6.3% (240)	73%	0%	6.7%	6.7%	0%	13%
pDCC6-2 (26 nM)	*Lig4*$^{+}$	23% (230)	15%	15%	52%	1.9%	7.4%	9.3%
	Lig4^{169}	31% (235)	19%	18%	58%	1.4%	4.1%	0%

Table 3.8: Targeting *w* in *Lig4*$^{+}$ or *Lig4*^{169}, *vas-Cas9* G0 embryos

sgRNA templates were co-injected with 33 nM pUC-w HR donor plasmid DNA. *n*, the number of G0 embryos injected, irrespective of fertility or survival. The coffee & red/white group includes G0 with coffee and red-eyed, or with coffee-, white-, and red-eyed G1 broods. pCFD4d also carries *armi* sgRNA-1, and pDCC6 also carries a Cas9 gene expression unit.

Figure 3.8

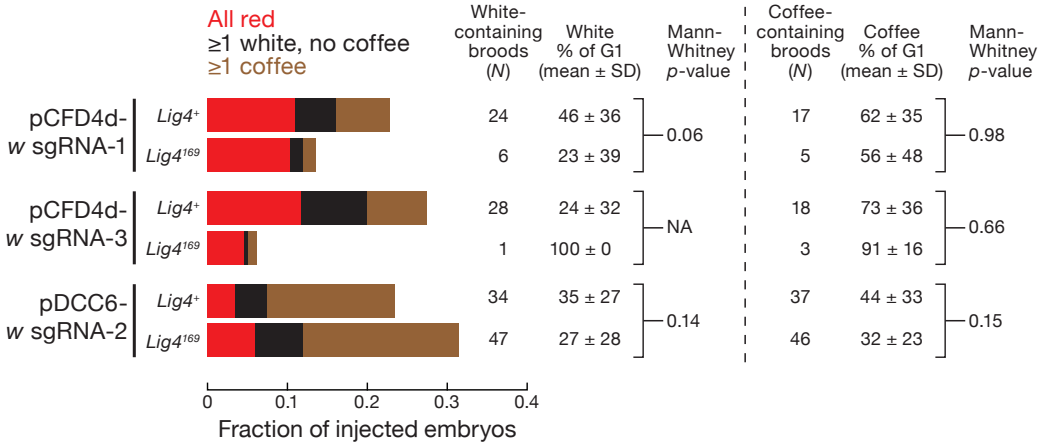

All red
≥1 white, no coffee
≥1 coffee
White-containing broods (N)
White % of G1 (mean ± SD)
Mann-Whitney p-value
Coffee-containing broods (N)
Coffee % of G1 (mean ± SD)
Mann-Whitney p-value
pCFD4d-w sgRNA-1
Lig4+
Lig4169
24
46 ± 36
6
23 ± 39
0.06
17
62 ± 35
5
56 ± 48
0.98
pCFD4d-w sgRNA-3
Lig4+
Lig4169
28
24 ± 32
1
100 ± 0
NA
18
73 ± 36
3
91 ± 16
0.66
pDCC6-w sgRNA-2
Lig4+
Lig4169
34
35 ± 27
47
27 ± 28
0.14
37
44 ± 33
46
32 ± 23
0.15
0
0.1
0.2
0.3
0.4
Fraction of injected embryos

Figure 3.8: *Lig4*^{169} mutant does not inhibit end joining or improve HR

Adults from injected G0 embryos that produce G1 broods were divided into three groups according to their eye color composition: (1) all red; (2) having at least one white, but no coffee, G1; and (3) having at least one coffee G1. For each *w* sgRNA, the percentage of coffee G1 in individual group 3 broods was compared between *Lig4*$^{+}$ or *Lig4*^{169} embryos. Similarly, the percentage of white G1 in broods with at least one white G1 (with or without coffee G1) was compared. All datasets failed the Shapiro-Wilk normality test, and therefore the two-tailed Mann-Whitney Rank Sum test was used to calculate *p* value. NA: *N* was too small to compute a *p*-value.

DISCUSSION

Our data demonstrate that the co-conversion strategy previously used in *C. elegans* (Ward, 2015; Arribere et al., 2014; Kim et al., 2014) can be successfully applied to *Drosophila*, reducing the burden of screening for mutations at the gene-of-interest. The co-conversion strategy worked equally well for the generation of indels or recombinants: both types of mutations were enriched in the broods that had no red-eyed progeny (Figure 3.2 and Table 3.1). The absence of red-eyed G1 flies in a brood indicates that all germline alleles in the G0 animal underwent targeted genome modification at *w*, reflecting efficient delivery of the guide plasmid to all the pole cells after injection. Our data suggest that when this happens, regardless of the choice of repair pathway, the co-targeted gene-of-interest is more likely to be modified. It is worth noting that Cas9-catalyzed DSBs at *w* and the gene-of-interest were correlated, but we did not observe a correlation between the repair pathways used at *w* and at the gene-of-interest: broods with HR at *w* did not necessarily produce recombinants at the gene-of-interest.

We frequently recovered more than one type of mutation at the gene-of-interest from a single G1 brood, evidence that independent repair events occurred among the dozens of germline stem cells of the G0 founder parent. In other words, the G0 germline is frequently mosaic. As an extreme example, five different indels and three different HR events at *zuc* were identified in the ten G1

flies we genotyped from a brood consisting of 15% white-eyed and 85% coffee-eyed offspring.

At the seven sgRNA target sites we tested, 39% of the 82 independent indels had junctional microhomologies or templated insertions (Figure 3.4 and Table 3.2 to Table 3.7), signatures of the *Lig4*-independent, microhomology-dependent end-joining pathway (Yu and McVey, 2010; Sfeir and Symington, 2015). We recovered many indels containing such signatures from *Lig4*$^{+}$ embryos, suggesting that the microhomology-mediated end-joining pathway normally operates even in the presence of Ligase 4. In fact, *Lig4*^{169} mutant embryos produced no fewer indels than *Lig4*$^{+}$ embryos (Figure 3.8), suggesting that a Ligase 4-independent end-joining pathway predominates at generating indels. In *C. elegans*, polymerase theta, but not Lig4, is used to repair Cas9-induced DSBs (van Schendel et al., 2015). As in worms, the *Drosophila* polymerase theta (*mus308*) is important for *Lig4*-independent end joining (Chan et al., 2010). Future experiments to test whether inactivation of *mus308,* alone or together with *Lig4*, reduces indel mutations in flies are clearly needed.

Eliminating donor integration, in which the plasmid integrates into the target locus instead of promoting the desired gene conversion, remains a challenge for Cas9-targeted HR: in our experiments, such integration accounted for 17% to 67% (median, 50%) of all HR events (Figure 3.5). Plasmid integration has been reported to account for 70% to 100% of Cas9-targeted recombinants

and was proposed to reflect the outcome of the resolution of double Holliday junctions formed between the donor and the genome (Figure 3.5A) (Yu et al., 2014). “Ends-in” targeting, in which the circular plasmid donor is linearized in vivo using the *I-Sce*I endonuclease to generate DSB in the center of the homologous arm, produced plasmid integration 66% of the time (Rong and Golic, 2000). For Cas9-induced HR, the DSB is in the genomic locus instead of the extra-chromosomal donor, but is otherwise analogous to “ends-in” targeting. For both, distinguishing gene conversion from plasmid integration is essential.

In theory, a linear donor whose sequence is restricted to the target genomic locus should eliminate the problem of integration. Plasmid donors containing a pair of *w* sgRNA-1 target sites—one before the upstream homology arm and one after the downstream arm, both in the same orientation—are predicted to be cleaved twice by *w* sgRNA-1-guided Cas9, liberating the HR donor from the plasmid DNA. Unfortunately, this donor design was inefficient in producing recombinants in our experiments (data not shown).

Variability in conversion tract length was observed in regions flanking a single DSB or flanking the gap deleted by two concomitant DSBs. Measured from the breaks, some tracts were ~1,000 bp, while others were less than 50 bp (Figure 3.6); some were even < 7 bp (Figure 3.7). The conversion of the region flanking the DSB(s) is therefore unpredictable. In contrast, when a pair of sgRNAs was used to direct two DSBs, the intervening sequence was reliably

replaced with that of the donor (Figure 3.6 and Figure 3.7). Pairs of sgRNAs have been used to change or insert 1–3 kbp of novel sequence into a gene in *Drosophila*, presumably through the same gap repair mechanism (Port et al., 2015; Yu et al., 2014; Ren et al., 2014a; Gratz et al., 2014; Ren et al., 2014b; Zhang et al., 2014b). Using the sister chromatid as a repair template, gap repair readily restores a 9 kbp gap following *P* element excision (McVey et al., 2004a). Alternatively, the conversion of intervening sequence between two DSBs may result from two convergent HR events initiated from each DSB separately. In this scenario, the two DSBs do not have to be created concomitantly. It is worth noting that gap repair does not always happen when two sgRNAs were co-injected, as we frequently observed gene conversion at one target site and either an indel or wild-type sequence at the other (Figure 3.6 and Figure 3.7). One possibility is that one of the two sgRNAs was more active than the other, reducing the chance of generating two DSBs at the same time—a prerequisite of gap repair. Thus, it may be prudent to carry out two experiments each using a unique pair of sgRNAs to ensure successful gap repair, which also offers the opportunity to generate two independent recombinants with non-overlapping potential off-target mutations.

Previous studies with zinc-finger nucleases suggested that *Lig4*^{169} mutant embryos promote HR (Ran et al., 2013; Bozas et al., 2009; Beumer et al., 2008). Surprisingly, the use of *Lig4*^{169} embryos did not increase HR efficiency in our

experiments (Figure 3.8), perhaps because Cas9, unlike zinc-finger nucleases, leaves blunt ends (Jinek et al., 2012; Kim et al., 1996).

In conclusion, co-targeting the *w* gene in *Drosophila* when using Cas9 to alter the fly genome substantially reduces the time and effort required for the molecular identification of mutations in the gene-of-interest. Other organisms with available endogenous or transgenic marker genes should be able to adopt a similar co-conversion strategy.

ACKNOWLEDGEMENTS

We thank Wen xue and members of the Zamore laboratory for critical comments on the manuscript, and the UMass CRISPR community for helpful discussions. This work was supported by NIH R37 grant GM62862 to P.D.Z.

EXPERIMENTAL PROCEDURES

Fly stocks

vas-Cas9 (*y^1, M{vas-Cas9}ZH-2A*) was generated by recombining *y^1, M{vas-Cas9}ZH-2A, w^{1118}* (Bloomington #51323; Gratz et al., 2014) with Oregon-R. *vas-Cas9, $Lig4^{169}$* (*y^1, M{vas-Cas9}ZH-2A, $Lig4^{169}$*) was generated by recombining *y^1, M{vas-Cas9}ZH-2A* with *w^{1118}, $Lig4^{169}$* (Bloomington #28877; McVey et al., 2004b). Rainbow Transgenic Flies, Inc. (Camarillo, CA, USA) performed injections. Strains are available upon request.

sgRNA-expressing plasmid construction

sgRNA design: Target loci of the injection strains were sequenced before 20 nt sgRNAs designed using crispr.mit.edu (Beumer et al., 2013). Guides were preferred if nucleotides 19 and 20 were purines (Farboud and Meyer, 2015); positions 15–20, the protospacer-adjacent motif-proximal nucleotides, were >33% GC (Ren et al., 2014b); and the sequence placed the guide close to the site of modification.

sgRNA cloning: pCFD4, which expresses one sgRNA from a U6:3 promoter and another sgRNA from a *U6:1* promoter (Addgene #49411; Port et al., 2014), was modified to remove *vermillion* and *attB* (pCFD4d). Sequence- and ligation-independent cloning (Jeong et al., 2012) was used to clone two guides into *Bbs*I-digested pCFD4d following a PCR incorporating one guide after the

U6:1 promoter, and the other after the *U6:3* promoter (Port et al., 2014). The 20 nt *w* sgRNA-2 template was inserted into the *Bbs*I sites of pDCC6, which expresses sgRNA from a *U6:2* promoter and Cas9 mRNA from the *hsp70Bb* promoter (Gokcezade et al., 2014). Plasmids were purified (Plasmid Midi Kit; QIAGEN, Hilden, Germany) and dissolved in water.

Donor template construction

pUC-w: A 2,080 bp fragment, spanning genomic nucleotides x:2,792,206–2,790,141 (*Drosophila melanogaster* genome release r6.07) was amplified by PCR from w^{cf} genomic DNA, sequenced to confirmed the w^{cf} point mutation and identify natural polymorphisms, and inserted into pUC57 between the *Sac*I and *Sph*I sites to produce pUC-*w*. Site-directed mutagenesis was used to mutate the sites targeted by *w* sgRNAs-1, -2, -3, and -4.

pUC-armi: a 2,280 bp DNA (synthesized at GenScript, Inc., Piscataway, NJ, USA) spanning genomic nucleotides 3L:3,464,383–3,466,434 was inserted into pUC57 between the *Sac*I and *Sph*I sites. The sequence included silent mutations, a naturally occurring nine-nucleotide deletion polymorphism in *armi* exon 8 that disrupts the *armi* sgRNA-1 target site, a naturally occurring 12-nucleotide deletion polymorphism in the *armi* 3′ UTR, and a 36 nt C-terminal Strep-tag II peptide tag.

pCR-zuc: a 2,120 bp PCR fragment spanning genomic nucleotides 2L:11,990,382– 11,988,263 was inserted into pCR-Blunt II-TOPO to make pCR-*zuc*^{WT}. A 991 bp fragment containing a 3×FLAG peptide tag before the stop codon of *zuc* and silent mutations disrupting four potential sgRNAs binding sites were synthesized as a gBlock (Integrated DNA Technologies, Coralville, IA, USA), digested with *Nde*I and *Pac*I, and inserted into pCR-*zuc*^{WT} between the *Nde*I and *Pac*I sites to produce pCR-*zuc*.

Screening for mutations at *white*

For *armi* targeting, individual injected G0 adults were mated with two *w*^{1118}*; + ; Dr/TM3, Sb* males or virgin females. For *zuc* targeting, *w*^{1118}*; Sp/CyO;* + was used in place of *w*^{1118}*; + ; Dr/TM3, Sb*. Three-to-five day-old G1 progeny (25˚) were assessed by light microscopy (MZ6 Stereomicroscope, Leica Microsystems GmbH, Wetzlar, Germany).

Screening for mutations at the gene-of-interest

Due to the large number of all-red, white-and-red, and coffee-and-red broods, and their lower chance of harboring gene-of-interest conversion (see text), not all G1 broods were PCR screened. Instead, 44 all-red (37% of total), 46 white-and-red (59%), 8 all-white (100%), 29 coffee-and-red (78%), 11 coffee-and-white (92%), and 15 all-coffee broods (88%) were picked for genotyping. Anesthetized

G1 male flies were deposited on a CO_2 pad, and the 9–10 flies closest to the front edge of the pad were individually mated to corresponding balancer virgin females to generate stocks. After five days, the G1 males were removed from the crosses, and 1–3 flies from the same brood were homogenized (Gloor et al., 1993) in 30 µl per fly “squishing buffer” (10 mM Tris-Cl, pH 8.0, 1 mM EDTA, 25 mM NaCl, 200 µg/ml freshly diluted Proteinase K solution [AM2546; Thermo Fisher Scientific]) with a plastic pestle (Kimble-Chase Kontes, Vineland, NJ, USA) in 1.7 ml microcentrifuge tubes, incubated at 37˚ for 30 min, and then the Proteinase K inactivated at 95˚ for 5 min. PCR was used to amplify 505–1,225 bp amplicons spanning the target loci from 1 µl homogenate (15 µl final reaction volume; MeanGreen 2x Taq Master Mix, Empirical Bioscience, Inc., Grand Rapids, MI, USA). We note that using this experimental setup, PCR efficiency drops for amplicons longer than 1 kbp. Because different sgRNAs targeted different regions of *armi* or *zuc*, different PCR primers were designed for each target locus. Whenever possible, one of the two primers bound only to the genome and not the donor, to avoid amplifying extra-chromosomal or ectopically inserted donor DNA. When screening for recombinants with novel sequences knocked-in at the target locus, PCR with one primer bound to the novel sequence (e.g., 3×FLAG) and another primer bound only to the genome and not the donor can quickly identify the positive recombinants. When screening for indels or

recombinants with point mutations at the target loci, we used the following strategies to identify PCR products that contained such mutations.

Restriction enzyme digestion: Because G1 flies inherit one chromosome from the injected G0 embryo and the other from the balancer fly, at least half of the PCR products were amplified from the wild-type gene. We digested the PCR reaction with a restriction enzyme that cleaves adjacent to the predicted DSB in the wild-type amplicon: PCR products resistant to the restriction digestion should harbor mutations at the recognition site. The uncut PCR product was then gel isolated (QIAquick Gel Extraction Kit, QIAGEN) and sequenced to identify the underlying mutation. This approach ensures that the wild-type PCR products does not confound the sequencing trace and allows the detection of one mutant allele among ≥ 6 alleles, allowing multiple G1 flies to be pooled in the same PCR. In addition to indels, HR can also be detected by this method, as long as the HR donors are engineered to contain silent mutations that disrupt the restriction enzyme site. A drawback is that the deletion or HR must affect the restriction enzyme recognition sequence; those that do not will remain undetected. The following restriction digestions were used:

armi sgRNA-1 DSB: an AvaII site 6 bp away; 5 µl of PCR digested with *Ava*II (0.2 U/µl final concentration [f.c.]) in 0.5x CutSmart Buffer (New England Biolabs, Inc., Ipswich, MA, USA) in 10 µl final volume (f.v.) at 37˚ for 2 h;

armi sgRNA-2/3 DSBs: a BstNI site 1 bp (sgRNA-2) or 1 bp (sgRNA-3) away; 5 µl of PCR with *Bst*NI (0.5 U/µl f.c.) in 1x NEBuffer 3.1 (NEB) in 10.5 µl f.v. at 60˚ for 1 h;

armi sgRNA-4 DSB: no restriction enzyme site nearby; digested with T7E1 as described below;

armi sgRNA-5/6 DSBs: a PmlI site 17 bp (sgRNA-5) or 11 bp (sgRNA-6) away; 10 µl PCR with *Eco*72I (0.5 U/µl f.c., Thermo Fisher) in 12.5 µl f.v. at room temperature for 1 h;

zuc sgRNA-1 DSB: a BccI site 9 bp away; 5 µl of PCR with *Bcc*I (0.5 U/µl f.c.) in 0.5x CutSmart Buffer in 10 µl f.v. at 37˚ for 1 h;

zuc sgRNA-2 DSB: a *Hpy*CH4III site 7 bp away; 5 µl of PCR with *Hpy*CH4III (0.25 U/µl f.c.) in 0.5x CutSmart Buffer in 10 µl f.v. at 37˚ for 2 h.

T7 endonuclease I (T7E1) digestion: to complement the restriction enzyme digestion, the same PCR products were denatured, re-annealed to form heteroduplex, and digested with the mismatch-specific, sequence-independent T7E1. In G1 single-fly PCR, either 0% (both alleles are wild-type) or 50% (one allele is mutant) of re-annealed products will be substrates for T7E1. The drawback of this approach is: (1) some small sequence changes may escape T7E1 detection (Vouillot et al., 2015); (2) lower sensitivity and higher background

prevents the pooling of G1 flies in the same PCR; and (3) as the wild-type PCR products cannot be specifically destroyed, the sequencing trace has to be manually inspected to detect a mutation. To digest with T7E1, 5 µl PCR product was denatured at 95˚ for 5 min, re-annealed by reducing the temperature 0.1˚/sec to 25˚ to allow heteroduplex to form, and then digested with T7E1 (0.125 U/µl f.c.) in 1x NEBuffer 2 (NEB) in 10 µl f.v. at 37˚ for 15 min, as previously described (Zhang et al., 2014a).

Differentiating gene conversion from plasmid integration

The homozygous G3 descendants of the G1 flies carrying HR were further analyzed by PCR to distinguish between gene conversion and plasmid integration. To ensure efficient amplification of PCR amplicons >2 kbp, genomic DNA from G3 homozygotes was isolated by homogenizing ten flies in 200 µl 2x PK buffer (200 mM Tris-Cl, pH 7.5, 25 mM EDTA, 300 mM NaCl, 2% [w/v] SDS), incubated with 200 µg/ml (f.c.) proteinase K at 65˚ for 30 min, extracted with 200 µl buffer-equilibrated phenol:chloroform:isoamyl alcohol (25:24:1 by volume, pH 8.0; AMRESCO LLC, Solon, OH, USA), and centrifuged at 20,800 x *g* for 5 min at room temperature. The top aqueous layer was precipitated with one-tenth volume 3 M sodium acetate and three volumes 100% ethanol on ice for 1 h. The precipitate was recovered by centrifugation (20,800 x *g* for 15 min at 4˚), washed with 70% (v/v) ethanol, air dried, and dissolved in water. To detect gene

conversion events, PCR was performed using forward and reverse primers binding exclusively to the genome and the PCR product sequenced to differentiate between gene conversion and plasmid integration. For armi, armi-exon6 forward and CycJ-exon2 reverse primers generated a 2,539 bp amplicon; for zuc, dgt2-exon2 forward and CG34163-upstream reverse primers generated a 2,450 bp amplicon (Phusion DNA Polymerase, NEB; 200 ng genomic DNA, 50 µl reaction volume).

Statistical analysis

Two-tailed tests were performed using Prism 6 (GraphPad Software, Inc., La Jolla, CA, USA).

Plasmids and fly strains have been deposited to Addgene and Bloomington Stock Center, respectively.

sgRNA guide and PCR primer sequences can be found at: http://www.g3journal.org/highwire/filestream/473041/field_highwire_adjunct_files/1/SupplementalTables.xlsx

APPENDIX A: DEVELOPMENT OF SHORT DEGRADOME-SEQ TO CAPTURE PIRNA INTERMEDIATES

Short degradome-seq captures piRNA precursors

Degradome-seq was developed to capture long RNAs (>200 nt) that bear a 5′-monophosphate (Han et al., 2015a), the cleavage signature of both PIWI proteins and the phasing endonuclease. However, in the process of producing piRNAs, piRNA precursors are expected to get shorter, by at least two mechanisms: 1) downstream ping-pong cleavages on a long precursor, and 2) phased endonucleolytic cleavages spreading from 5′ to 3′. In both cases, new 5′ ends should always begin with a 5′-monophosphate. In order to test if such shorter intermediates can be detected, we developed a modified small RNA cloning protocol to clone 32–85 nt, 5′-monophosphorylated RNAs (see below for a detailed protocol). From wild-type ovarian total RNA, this protocol yielded 31% genome mapping reads, out of which 3.1% map to transposons (54% sense, 46% antisense) and 87.9% map to genes (99.3% sense and 0.7% antisense). The ratio of transposons to genes is similar to that in the long degradome library.

To test if the antisense transposon reads correspond to piRNA precursors, we mapped total small RNA cloned from wild-type ovaries to an index constructed using these short degradome (s-Deg) 5′-ends (Figure 4.1A). As a positive control, we mapped the same small RNA dataset to an index constructed using long degradome (l-Deg) 5′-ends (Figure 4.1B). Both s-Deg and l-Deg were cloned from the same total RNA sample. The height of the peak at 0, denoting the frequency of piRNA sharing 5′-ends with the degradome index, is

alternatively an indication of the diversity of degradome species whose 5′-ends share with piRNA. In this sense, s-Deg captures some, but not as many different degradome 5′-ends as the l-Deg library. Nonetheless, s-Deg does capture piRNA precursors in the 32–85 nt size range.

Figure 4.1

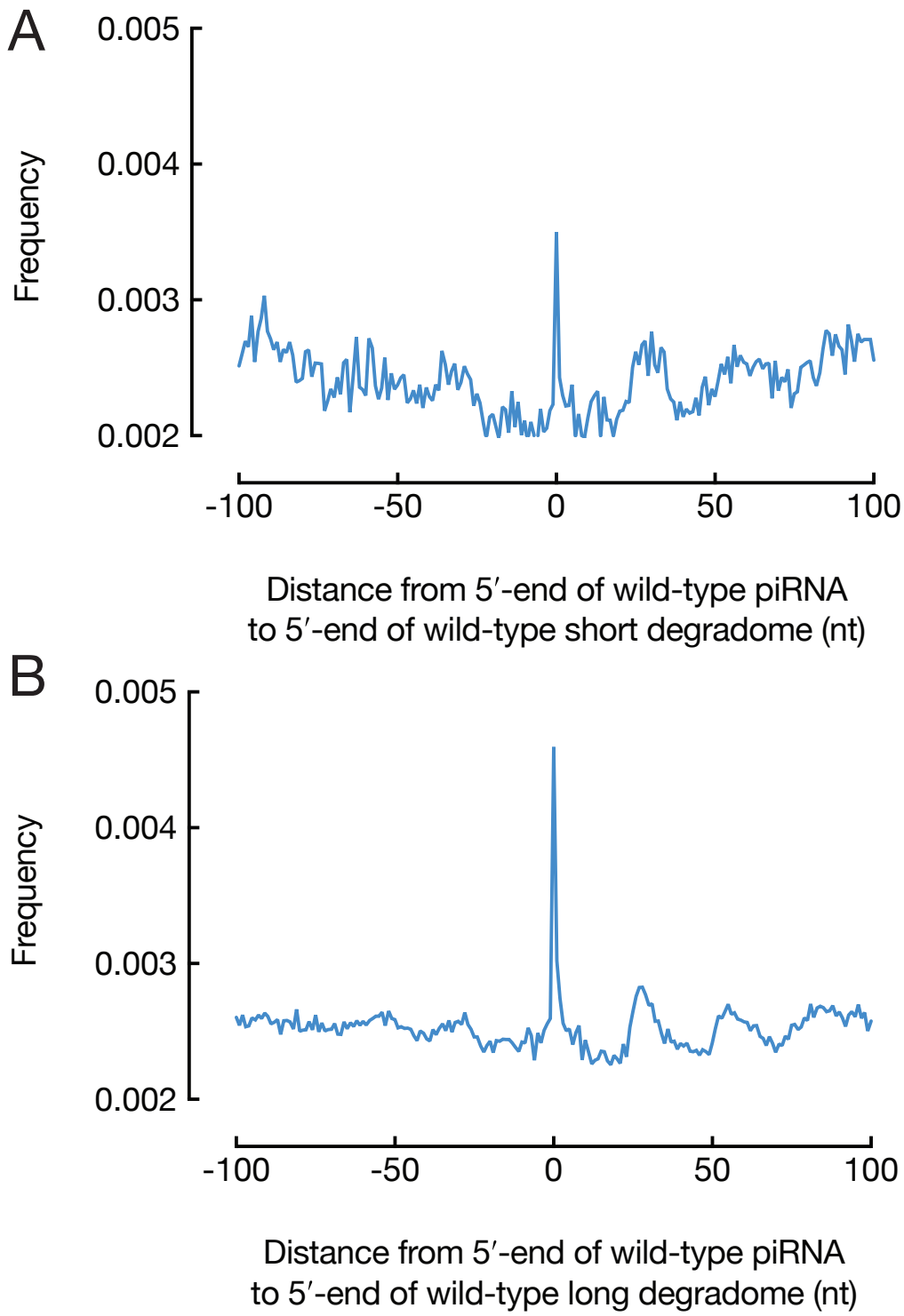

A
0.005
0.004
0.003
0.002
Frequency
-100
-50
0
50
100
Distance from 5′-end of wild-type piRNA
to 5′-end of wild-type short degradome (nt)
B
0.005
0.004
0.003
0.002
Frequency
-100
-50
0
50
100
Distance from 5′-end of wild-type piRNA
to 5′-end of wild-type long degradome (nt)

Figure 4.1: Short degradome cloning captures piRNA precursors in the 32–85 nt size range

(A) Distance from the 5′-end of wild-type piRNA to the 5′-end of wild-type short degradome (32–85 nt) on the same genomic strand. (B) Distance from the 5′-end of wild-type piRNA to the 5′-end of wild-type long degradome (>200 nt) on the same genomic strand.

Short piRNA precursors may be more prevalent in *zucCD*

It is suggested that in the absence of phasing, the competing ping-pong machinery processes the precursor more frequently and leaves shorter fragments (Hayashi et al., 2016). However, there is no direct evidence showing the increase of short precursor fragments in phasing mutants. To test this hypothesis, we cloned s-Deg from *zucCD* mutant ovaries, where only ping-pong is active.

The *zucCD* s-Deg library yielded 25% genome mapping reads, out of which 15% map to transposons (90% sense, 10% antisense) and 63% map to genes (99.1% sense and 0.9% antisense). Again, the ratio of transposons to genes is similar to that in the long degradome library.

As described above, a more sensitive way to detect piRNA precursor 5′-end diversity in degradome libraries is to map small RNAs to the degradome index. To test if short piRNA precursors are more abundant in *zucCD* mutants, we mapped the same *zucCD* small RNA dataset to either s-Deg or l-Deg from *zucCD* mutants. The similar heights of the peak at 0 suggests that s-Deg and l-Deg libraries contain similarly diverse piRNA precursor 5′-ends (Figure 4.2). In wild-type, s-Deg contains less diverse piRNA precursor 5′-ends when compared to l-Deg (Figure 4.1). Therefore, it is likely that piRNA precursor length distribution shifts to the shorter side in *zucCD* ovaries.

Figure 4.2

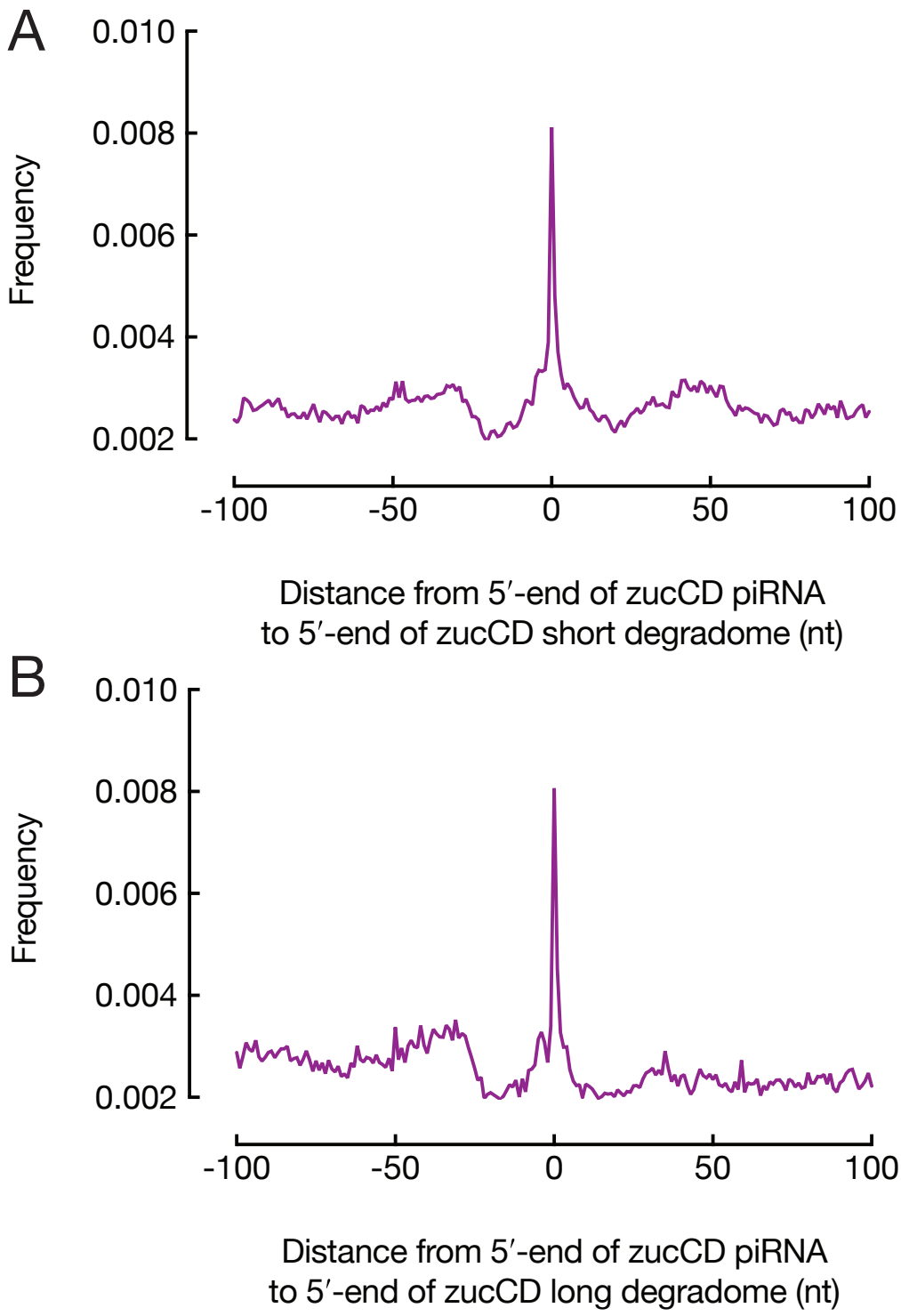

A
0.010
0.008
0.006
0.004
0.002
Frequency
-100
-50
0
50
100
Distance from 5′-end of zucCD piRNA
to 5′-end of zucCD short degradome (nt)
B
0.010
0.008
0.006
0.004
0.002
Frequency
-100
-50
0
50
100
Distance from 5′-end of zucCD piRNA
to 5′-end of zucCD long degradome (nt)

Figure 4.2: Short piRNA precursors are equally well captured as long precursors in *zucCD* mutants

(A) Distance from the 5′-end of *zucCD* piRNA to the 5′-end of *zucCD* short degradome (32–85 nt) on the same genomic strand. (B) Distance from the 5′-end of *zucCD* piRNA to the 5′-end of *zucCD* long degradome (>200 nt) on the same genomic strand.

Short piRNA precursors are generated by ping-pong cleavages in *zucCD*

An advantage of s-Deg over the l-Deg library is that the 3′-ends of the intermediates are preserved. In the l-Deg protocol, a random primer fused to a 3′-adaptor was used to create artificial 3′-ends of the inserts; in the s-Deg protocol, a 3′-adaptor is directly ligated to the RNA. With the 3′-end information, we can now ask whether the s-Deg fragments themselves are made by endonucleolytic cleavages. In the silk moth BmN4 cell line, Papi-bound short intermediates (model length ~65 nt) were found to be head-to-tail connected (i.e., made by endonucleolytic cleavages) and such connection was not affected by Zuc depletion (Nishida et al., 2018), consistent with the model that ping-pong cleavages generate short (~65 nt) fragments.

To test if fly s-Deg fragments are made by endonucleolytic cleavage, we measured the distance between the 3′-end of the preceding s-Deg and the 5′-end of the following s-Deg on the same genomic strand (3-5 phasing analysis, (Han et al., 2015a)). An obvious peak at d = 1 means that head-to-tail connected s-Deg fragments are more frequently observed than any other distances. In *zucCD* mutants, this peak stays strong (Figure 4.3), implying that these short (32–85 nt) fragments are generated by ping-pong cleavages, consistent with observations made in BmN4 cells (Nishida et al., 2018).

In conclusion, we have developed a deep sequencing protocol to clone 32–85 nt piRNA intermediates, filling in the gap of the current sequencing

protocols (small RNA-seq, 18–30 nt; RNA-seq/degradome-seq, >200 nt). The advantage of preserving the 3′-end info of piRNA intermediates may be used to answer future questions arise in the study of piRNA biogenesis.

Figure 4.3

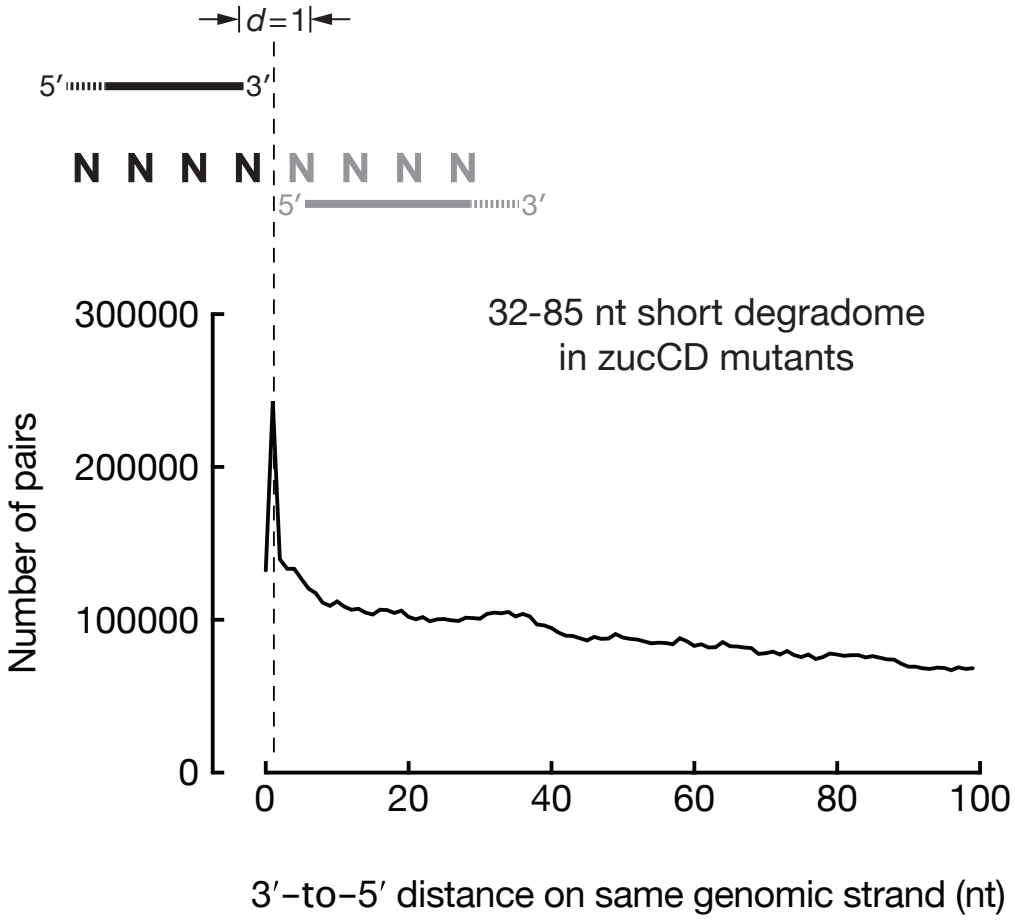

d = 1
5′
3′
N N N N N N N N
5′
3′
32-85 nt short degradome
in zucCD mutants
Number of pairs
300000
200000
100000
0
0
20
40
60
80
100
3′–to–5′ distance on same genomic strand (nt)

Figure 4.3: Short piRNA precursors in *zucCD* mutants are likely made by ping-pong cleavages

3′-to-5′ distance between *zucCD* short degradome inserts on the same genomic strand.

Short degradome-seq library preparation and analysis

Short degradome-seq libraries were constructed using a modified small RNA cloning protocol (Ge & Zamore, unpublished). Starting material can be 2 µg of total fly RNA, or RNA extracted from immunoprecipitation in the ng range. rRNAs were depleted using oligos that tile the *Drosophila* 18S or 28S rRNAs (Fu et al., 2018). DNase-treated RNA was purified using RNA Clean & Concentrator-5, keeping the fraction with <200 nt RNAs. The RNA was then treated with T4 polynucleotide kinase (PNK) under acidic conditions in buffer containing 50 mM sodium acetate, pH 6, 12.5 mM $MgCl_2$, 1 mM EDTA, 0.1% beta-mercaptoethanol, 0.01% Triton X-100, 5 U of T4 PNK and 20 U of RNasin Plus in 20 µl reaction volume, at 37˚C for 2 hours. The reaction was extracted with an equal volume of acid phenol:chloroform (5:1 by volume, pH 4.5; AMRESCO), and centrifuged at 20,800 × g for 5 min at room temperature. The top aqueous layer was precipitated with one-tenth volume 3 M sodium acetate and three volumes 100% ethanol on ice for 1 h. The precipitate was recovered by centrifugation (20,800 × g for 15 min at 4˚C), washed with 70% (v/v) ethanol, air dried, and dissolved in water. To prevent 2S rRNA from ligating to the 3′ adaptor, 10 pmol of 2S blocker II oligo (TACAACCCTCAACCATATGTAGTCCAAGCA blocked at both 5′ and 3′ ends by a C3 Spacer) was added before 3′ adaptor ligation. 3′ adaptor (5′-rApp NNN TGG AAT TCT CGG GTG CCA AGG /ddC/-3′) was ligated using truncated, K227Q mutant T4 RNA Ligase 2 (homemade) at 16°C overnight. The ligation

mixture was ethanol precipitated, and separated on a 10% PAGE. The band between 57–110 nt was cut out, using both the 3′ adaptor-ligated 30 nt size marker and the RNA Century Markers (Ambion). The band was transferred to a pre-wet Pur-A-Lyser Midi 3500 dialysis tube together with 300 µl of water, submerged in a horizontal agarose gel tank with 0.5× TBE, and electroeluted at 150V for 30 min. The current was reversed and run for two more minutes, before the eluate was removed for ethanol precipitation. To exclude 2S rRNA from sequencing libraries, 10 pmol 2S blocker oligo was added before 5′ adaptor ligation (Wickersheim and Blumenstiel, 2013). 5′ adaptor was ligated using T4 RNA ligase (Life Technologies, #AM2141) at 25°C for 2 h, followed by ethanol precipitation and reverse transcription using Superscript III reverse transcriptase. The reverse transcription reaction was amplified by a first round of PCR, using a pair of primers that anneal to the 5′-adapter (5′-CTACACGTTCAGAGTTCTACAGTCCGA -3′) or to the 3′-adapter (5′-GCCTTGGCACCCGAGAATTCCA -3′) and NEBNext (NEB). The PCR reaction was separated on a 10% PAGE. The band between 83–126 nt was cut out, using RNA Century Markers (Ambion). The band was transferred to a pre-wet Pur-A-Lyser Midi 3500 dialysis tube and electroeluted as described above. The second round of PCR uses the same barcoded primer set as the small RNA library cloning protocol, and purified using 2% agarose gel. The gel slice was extracted with QIAquick gel extraction kit (Invitogen). The length distribution and quality of

the libraries were analyzed by Agilent 2100 Bioanalyzer. Libraries were then quantified using KAPA library quantification kit, before being sequenced on a NextSeq500 (Illumina) to obtain 150 nt single-end reads.

Barcodes were sorted by BaseSpace (Illumina), and adaptors were removed in the same way as the small RNA-seq, explained in details in Chapter II. Short degradome-seq analysis was performed in the same way as the long degradome-seq, explained in details in Chapter II.

BIBLIOGRAPHY

Andress, A., Bei, Y., Fonslow, B. R., Giri, R., Wu, Y., Yates, J. R., and Carthew, R. W. (2016). Spindle-E cycling between nuage and cytoplasm is controlled by Qin and PIWI proteins. J Cell Biol *213*, 201-211.

Aravin, A. A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K. F., Bestor, T., and Hannon, G. J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol Cell *31*, 785-799.

Aravin, A. A., van der Heijden, G. W., Castañeda, J., Vagin, V. V., Hannon, G. J., and Bortvin, A. (2009). Cytoplasmic compartmentalization of the fetal piRNA pathway in mice. PLoS Genet *5*, e1000764.

Arribere, J. A., Bell, R. T., Fu, B. X., Artiles, K. L., Hartman, P. S., and Fire, A. Z. (2014). Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in Caenorhabditis elegans. Genetics *198*, 837-846.

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

Bastock, R., and St Johnston, D. (2008). Drosophila oogenesis. Curr Biol *18*, R1082-7.

Basu, S., Aryan, A., Overcash, J. M., Samuel, G. H., Anderson, M. A., Dahlem, T. J., Myles, K. M., and Adelman, Z. N. (2015). Silencing of end-joining repair for efficient site-specific gene insertion after TALEN/CRISPR mutagenesis in Aedes aegypti. Proc Natl Acad Sci U S A *112*, 4038-4043.

Beumer, K. J., Trautman, J. K., Bozas, A., Liu, J. L., Rutter, J., Gall, J. G., and Carroll, D. (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., Trautman, J. K., Mukherjee, K., and Carroll, D. (2013). Donor DNA Utilization During Gene Targeting with Zinc-Finger Nucleases. G3 (Bethesda) *3*, 657-664.

Bibikova, M., Carroll, D., Segal, D. J., Trautman, J. K., Smith, J., Kim, Y. G., and Chandrasegaran, S. (2001). Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. Mol Cell Biol *21*, 289-297.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. Science *326*, 1509-1512.

Böttcher, R., Hollmann, M., Merk, K., Nitschko, V., Obermaier, C., Philippou-Massier, J., Wieland, I., Gaul, U., and Förstemann, K. (2014). Efficient chromosomal gene modification with CRISPR/cas9 and PCR-based homologous recombination donors in cultured Drosophila cells. Nucleic Acids Res *42*, e89.

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

Brangwynne, C. P., Eckmann, C. R., Courson, D. S., Rybarska, A., Hoege, C., Gharakhani, J., Jülicher, F., and Hyman, A. A. (2009). Germline P granules are liquid droplets that localize by controlled dissolution/condensation. Science *324*, 1729-1732.

Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G. J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell *128*, 1089-1103.

Brennecke, J., Malone, C. D., Aravin, A. A., Sachidanandam, R., Stark, A., and Hannon, G. J. (2008). An epigenetic role for maternally inherited piRNAs in transposon silencing. Science *322*, 1387-1392.

Carmell, M. A., Girard, A., van de Kant, H. J., Bourc'his, D., Bestor, T. H., de Rooij, D. G., and Hannon, G. J. (2007). MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. Dev Cell *12*, 503-514.

Castel, S. E., and Martienssen, R. A. (2013). RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. Nat Rev Genet *14*, 100-112.

Chan, S. H., Yu, A. M., and McVey, M. (2010). Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in Drosophila. PLoS Genet *6*, e1001005.

Chen, H. M., Huang, Y., Pfeiffer, B. D., Yao, X., and Lee, T. (2015). An enhanced gene targeting toolkit for Drosophila: Golic+. Genetics *199*, 683-694.

Cheng, Z., Muhlrad, D., Lim, M. K., Parker, R., and Song, H. (2007). Structural and functional insights into the human Upf1 helicase core. EMBO J *26*, 253-264.

Choi, S. Y., Huang, P., Jenkins, G. M., Chan, D. C., Schiller, J., and Frohman, M. A. (2006). A common lipid links Mfn-mediated mitochondrial fusion and SNARE-regulated exocytosis. Nat Cell Biol *8*, 1255-1262.

Chrétien, D., Bénit, P., Ha, H. H., Keipert, S., El-Khoury, R., Chang, Y. T., Jastroch, M., Jacobs, H. T., Rustin, P., and Rak, M. (2018). Mitochondria are physiologically maintained at close to 50 °C. PLoS Biol *16*, e2003992.

Christian, M., Cermak, T., Doyle, E. L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A. J., and Voytas, D. F. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. Genetics *186*, 757-761.

Chu, V. T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., and Kühn, R. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat Biotechnol *33*, 543-548.

Cook, H. A., Koppetsch, B. S., Wu, J., and Theurkauf, W. E. (2004). The Drosophila SDE3 homolog armitage is required for oskar mRNA silencing and embryonic axis specification. Cell *116*, 817-829.

Dapples, C. C., and King, R. C. (1970). The development of the nucleolus of the ovarian nurse cell of Drosophila melanogaster. Z Zellforsch Mikrosk Anat *103*, 34-47.

Ecco, G., Imbeault, M., and Trono, D. (2017). KRAB zinc finger proteins. Development *144*, 2719-2729.

Eddy, E. M. (1974). Fine structural observations on the form and distribution of nuage in germ cells of the rat. Anat Rec *178*, 731-757.

Eddy, E. M. (1975). Germ plasm and the differentiation of the germ cell line. Int Rev Cytol *43*, 229-280.

Farboud, B., and Meyer, B. J. (2015). Dramatic enhancement of genome editing by CRISPR/Cas9 through improved guide RNA design. Genetics *199*, 959-971.

Feschotte, C., and Pritham, E. J. (2007). DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet *41*, 331-368.

Fiorini, F., Bagchi, D., Le Hir, H., and Croquette, V. (2015). Human Upf1 is a highly processive RNA helicase and translocase with RNP remodelling activities. Nat Commun *6*, 7581.

Franks, T. M., Singh, G., and Lykke-Andersen, J. (2010). Upf1 ATPase-dependent mRNP disassembly is required for completion of nonsense-mediated mRNA decay. Cell *143*, 938-950.

Friedli, M., and Trono, D. (2015). The developmental control of transposable elements and the evolution of higher species. Annu Rev Cell Dev Biol *31*, 429-451.

Frost, R. J., Hamra, F. K., Richardson, J. A., Qi, X., Bassel-Duby, R., and Olson, E. N. (2010). MOV10L1 is necessary for protection of spermatocytes against retrotransposons by Piwi-interacting RNAs. Proc Natl Acad Sci U S A *107*, 11847-11852.

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. Nat Biotechnol *31*, 822-826.

Fu, Y., Wu, P. H., Beane, T., Zamore, P. D., and Weng, Z. (2018). Elimination of PCR duplicates in RNA-seq and small RNA-seq using unique molecular identifiers. BMC Genomics *19*, 531.

Gainetdinov, I., Colpan, C., Cecchini, K., and Zamore, P. (2018). A Single Mechanism of Biogenesis, Initiated and Directed by PIWI Proteins, Explains piRNA Production in Most Animals. bioRxiv doi: https://doi.org/10.1101/261545

Girard, A., and Hannon, G. J. (2008). Conserved themes in small-RNA-mediated transposon control. Trends Cell Biol *18*, 136-148.

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

Gokcezade, J., Sienski, G., and Duchek, P. (2014). Efficient CRISPR/Cas9 plasmids for rapid and versatile genome editing in Drosophila. G3 (Bethesda) *4*, 2279-2282.

Govindan, G., and Ramalingam, S. (2016). Programmable Site-Specific Nucleases for Targeted Genome Engineering in Higher Eukaryotes. J Cell Physiol *231*, 2380-2392.

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.

Gregersen, L. H., Schueler, M., Munschauer, M., Mastrobuoni, G., Chen, W., Kempa, S., Dieterich, C., and Landthaler, M. (2014). MOV10 Is a 5′ to 3′ RNA helicase contributing to UPF1 mRNA target degradation by translocation along 3′ UTRs. Mol Cell *54*, 573-585.

Gunawardane, L. S., Saito, K., Nishida, K. M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M. C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in Drosophila. Science *315*, 1587-1590.

Haase, A. D., Fenoglio, S., Muerdter, F., Guzzardo, P. M., Czech, B., Pappin, D. J., Chen, C., Gordon, A., and Hannon, G. J. (2010). Probing the initiation and effector phases of the somatic piRNA pathway in Drosophila. Genes Dev *24*, 2499-2504.

Han, B. W., Wang, W., Li, C., Weng, Z., and Zamore, P. D. (2015a). Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. Science *348*, 817-821.

Han, B. W., Wang, W., Zamore, P. D., and Weng, Z. (2015b). piPipes: a set of pipelines for piRNA and transposon analysis via small RNA-seq, RNA-seq, degradome- and CAGE-seq, ChIP-seq and genomic DNA sequencing. Bioinformatics *31*, 593-595.

Han, B. W., and Zamore, P. D. (2014). piRNAs. Curr Biol *24*, R730-3.

Handler, D., Meixner, K., Pizka, M., Lauss, K., Schmied, C., Gruber, F. S., and Brennecke, J. (2013). The genetic makeup of the Drosophila piRNA pathway. Mol Cell *50*, 762-777.

Handler, D., Olivieri, D., Novatchkova, M., Gruber, F. S., Meixner, K., Mechtler, K., Stark, A., Sachidanandam, R., and Brennecke, J. (2011). A systematic analysis of Drosophila TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. EMBO J *30*, 3977-3993.

Hay, B., Jan, L. Y., and Jan, Y. N. (1988). A protein component of Drosophila polar granules is encoded by vasa and has extensive sequence similarity to ATP-dependent helicases. Cell *55*, 577-587.

Hayashi, R., Schnabl, J., Handler, D., Mohn, F., Ameres, S. L., and Brennecke, J. (2016). Genetic and mechanistic diversity of piRNA 3′-end formation. Nature *539*, 588-592.

He, F., and Jacobson, A. (2015). Nonsense-Mediated mRNA Decay: Degradation of Defective Transcripts Is Only Part of the Story. Annu Rev Genet *49*, 339-366.

Honda, S., Kirino, Y., Maragkakis, M., Alexiou, P., Ohtaki, A., Murali, R., Mourelatos, Z., and Kirino, Y. (2013). Mitochondrial protein BmPAPI modulates the length of mature piRNAs. RNA *19*, 1405-1418.

Hsu, P. D., Lander, E. S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262-1278.

Huang, C. R., Burns, K. H., and Boeke, J. D. (2012). Active transposition in genomes. Annu Rev Genet *46*, 651-675.

Huang, H., Gao, Q., Peng, X., Choi, S. Y., Sarma, K., Ren, H., Morris, A. J., and Frohman, M. A. (2011). piRNA-associated germline nuage formation and spermatogenesis require MitoPLD profusogenic mitochondrial-surface lipid signaling. Dev Cell *20*, 376-387.

Huang, H., Li, Y., Szulwach, K. E., Zhang, G., Jin, P., and Chen, D. (2014). AGO3 Slicer activity regulates mitochondria-nuage localization of Armitage and piRNA amplification. J Cell Biol *206*, 217-230.

Ipsaro, J. J., Haase, A. D., Knott, S. R., Joshua-Tor, L., and Hannon, G. J. (2012). The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. Nature *491*, 279-283.

Izumi, N., Kawaoka, S., Yasuhara, S., Suzuki, Y., Sugano, S., Katsuma, S., and Tomari, Y. (2013). Hsp90 facilitates accurate loading of precursor piRNAs into PIWI proteins. RNA *19*, 896-901.

Izumi, N., Shoji, K., Sakaguchi, Y., Honda, S., Kirino, Y., Suzuki, T., Katsuma, S., and Tomari, Y. (2016). Identification and Functional Analysis of the Pre-piRNA 3′ Trimmer in Silkworms. Cell *164*, 962-973.

Jaglarz, M. K., Kloc, M., Jankowska, W., Szymanska, B., and Bilinski, S. M. (2011). Nuage morphogenesis becomes more complex: two translocation pathways and two forms of nuage coexist in Drosophila germline syncytia. Cell Tissue Res *344*, 169-181.

Jeong, J. Y., Yim, H. S., Ryu, J. Y., Lee, H. S., Lee, J. H., Seen, D. S., and Kang, S. G. (2012). One-step sequence- and ligation-independent cloning as a rapid and versatile cloning method for functional genomics studies. Appl Environ Microbiol *78*, 5440-5443.

Jin, Z., Flynt, A. S., and Lai, E. C. (2013). Drosophila piwi mutants exhibit germline stem cell tumors that are sustained by elevated Dpp signaling. Curr Biol *23*, 1442-1448.

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.

Kameoka, S., Adachi, Y., Okamoto, K., Iijima, M., and Sesaki, H. (2018). Phosphatidic Acid and Cardiolipin Coordinate Mitochondrial Dynamics. Trends Cell Biol *28*, 67-76.

Kaminker, J. S., Bergman, C. M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D. A., Lewis, S. E., Rubin, G. M., Ashburner, M., and Celniker, S. E. (2002). The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective. Genome Biol *3*, RESEARCH0084.

Khurana, J. S., Wang, J., Xu, J., Koppetsch, B. S., Thomson, T. C., Nowosielska, A., Li, C., Zamore, P. D., Weng, Z., and Theurkauf, W. E. (2011). Adaptation to P element transposon invasion in Drosophila melanogaster. Cell *147*, 1551-1563.

Kim, H., Ishidate, T., Ghanta, K. S., Seth, M., Conte, D., Shirayama, M., and Mello, C. C. (2014). A co-CRISPR strategy for efficient genome editing in Caenorhabditis elegans. Genetics *197*, 1069-1080.

Kim, Y. G., Cha, J., and Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci U S A *93*, 1156-1160.

Klattenhoff, C., Bratu, D. P., McGinnis-Schultz, N., Koppetsch, B. S., Cook, H. A., and Theurkauf, W. E. (2007). Drosophila rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. Dev Cell *12*, 45-55.

Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Khurana, J. S., Zhang, F., Schultz, N., Koppetsch, B. S., Nowosielska, A., Seitz, H., Zamore, P. D., Weng, Z., and Theurkauf, W. E. (2009). The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. Cell *138*, 1137-1149.

Kloc, M., Jedrzejowska, I., Tworzydlo, W., and Bilinski, S. M. (2014). Balbiani body, nuage and sponge bodies--term plasm pathway players. Arthropod Struct Dev *43*, 341-348.

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Takamatsu, K., Chuma, S., Kojima-Kita, K., Shiromoto, Y., Asada, N., Toyoda, A., Fujiyama, A., Totoki, Y., Shibata, T., Kimura, T., Nakatsuji, N., Noce, T., Sasaki, H., and Nakano, T. (2010). MVH in piRNA processing and gene silencing of retrotransposons. Genes Dev *24*, 887-892.

LaCava, J., Molloy, K. R., Taylor, M. S., Domanski, M., Chait, B. T., and Rout, M. P. (2015). Affinity proteomics to study endogenous protein complexes: pointers, pitfalls, preferences and perspectives. Biotechniques *58*, 103-119.

Le Thomas, A., Rogers, A. K., Webster, A., Marinov, G. K., Liao, S. E., Perkins, E. M., Hur, J. K., Aravin, A. A., and Tóth, K. F. (2013). Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. Genes Dev *27*, 390-399.

Lee, S. R., Pratt, G. A., Martinez, F. J., Yeo, G. W., and Lykke-Andersen, J. (2015). Target Discrimination in Nonsense-Mediated mRNA Decay Requires Upf1 ATPase Activity. Mol Cell *59*, 413-425.

Levin, H. L., and Moran, J. V. (2011). Dynamic interactions between transposable elements and their hosts. Nat Rev Genet *12*, 615-627.

Lewis, S. H., Quarles, K. A., Yang, Y., Tanguy, M., Frézal, L., Smith, S. A., Sharma, P. P., Cordaux, R., Gilbert, C., Giraud, I., Collins, D. H., Zamore, P. D., Miska, E. A., Sarkies, P., and Jiggins, F. M. (2018). Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. Nat Ecol Evol *2*, 174-181.

Li, C., Vagin, V. V., Lee, S., Xu, J., Ma, S., Xi, H., Seitz, H., Horwich, M. D., Syrzycka, M., Honda, B. M., Kittler, E. L., Zapp, M. L., Klattenhoff, C., Schulz, N., Theurkauf, W. E., Weng, Z., and Zamore, P. D. (2009). Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. Cell *137*, 509-521.

Liang, L., Diehl-Jones, W., and Lasko, P. (1994). Localization of vasa protein to the Drosophila pole plasm is independent of its RNA-binding and helicase activities. Development *120*, 1201-1211.

Lim, A. K., and Kai, T. (2007). Unique germ-line organelle, nuage, functions to repress selfish genetic elements in Drosophila melanogaster. Proc Natl Acad Sci U S A *104*, 6714-6719.

Liu, L., Qi, H., Wang, J., and Lin, H. (2011). PAPI, a novel TUDOR-domain protein, complexes with AGO3, ME31B and TRAL in the nuage to silence transposition. Development *138*, 1863-1873.

Ma, L., Buchold, G. M., Greenbaum, M. P., Roy, A., Burns, K. H., Zhu, H., Han, D. Y., Harris, R. A., Coarfa, C., Gunaratne, P. H., Yan, W., and Matzuk, M. M. (2009). GASZ is essential for male meiosis and suppression of retrotransposon expression in the male germline. PLoS Genet *5*, e1000635.

Mackenzie, S. M., Brooker, M. R., Gill, T. R., Cox, G. B., Howells, A. J., and Ewart, G. D. (1999). Mutations in the white gene of Drosophila melanogaster affecting ABC transporters that determine eye colouration. Biochim Biophys Acta *1419*, 173-185.

Mahowald, A. P. (1970). INTERCELLULAR MIGRATION OF CENTRIOLES IN THE GERMARIUM OF DROSOPHILA MELANOGASTER: An Electron Microscopic Study. The Journal of Cell Biology *45*, 306-320.

Mahowald, A. P. (1962). Fine structure of pole cells and polar granules inDrosophila melanogaster. Journal of Experimental Zoology *151*, 201-215.

Mahowald, A. P. (1971). Polar granules of Drosophila. 3. The continuity of polar granules during the life cycle of Drosophila. J Exp Zool *176*, 329-343.

Malone, C. D., Brennecke, J., Dus, M., Stark, A., McCombie, W. R., Sachidanandam, R., and Hannon, G. J. (2009). Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. Cell *137*, 522-535.

Maruyama, T., Dougan, S. K., Truttmann, M. C., Bilate, A. M., Ingram, J. R., and Ploegh, H. L. (2015). Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. Nat Biotechnol *33*, 538-542.

McCue, A. D., and Slotkin, R. K. (2012). Transposable element small RNAs as regulators of gene expression. Trends Genet *28*, 616-623.

McVey, M., Adams, M., Staeva-Vieira, E., and Sekelsky, J. J. (2004a). Evidence for multiple cycles of strand invasion during repair of double-strand gaps in Drosophila. Genetics *167*, 699-705.

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

Miller, D. J., Schwartz, M. D., and Ahlquist, P. (2001). Flock house virus RNA replicates on outer mitochondrial membranes in Drosophila cells. J Virol *75*, 11664-11676.

Mohn, F., Handler, D., and Brennecke, J. (2015). Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. Science *348*, 812-817.

Mohn, F., Sienski, G., Handler, D., and Brennecke, J. (2014). The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in Drosophila. Cell *157*, 1364-1379.

Moscou, M. J., and Bogdanove, A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. Science *326*, 1501.

Muerdter, F., Guzzardo, P. M., Gillis, J., Luo, Y., Yu, Y., Chen, C., Fekete, R., and Hannon, G. J. (2013). A genome-wide RNAi screen draws a genetic framework for transposon control and primary piRNA biogenesis in Drosophila. Mol Cell *50*, 736-748.

Nishida, K. M., Okada, T. N., Kawamura, T., Mituyama, T., Kawamura, Y., Inagaki, S., Huang, H., Chen, D., Kodama, T., Siomi, H., and Siomi, M. C. (2009). Functional involvement of Tudor and dPRMT5 in the piRNA processing pathway in Drosophila germlines. EMBO J *28*, 3820-3831.

Nishida, K. M., Sakakibara, K., Iwasaki, Y. W., Yamada, H., Murakami, R., Murota, Y., Kawamura, T., Kodama, T., Siomi, H., and Siomi, M. C. (2018). Hierarchical roles of mitochondrial Papi and Zucchini in Bombyx germline piRNA biogenesis. Nature *555*, 260-264.

Nishimasu, H., Ishizu, H., Saito, K., Fukuhara, S., Kamatani, M. K., Bonnefond, L., Matsumoto, N., Nishizawa, T., Nakanaga, K., Aoki, J., Ishitani, R., Siomi, H., Siomi, M. C., and Nureki, O. (2012). Structure and function of Zucchini endoribonuclease in piRNA biogenesis. Nature *491*, 284-287.

Olivieri, D., Sykora, M. M., Sachidanandam, R., Mechtler, K., and Brennecke, J. (2010). An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila. EMBO J *29*, 3301-3317.

Olivieri, D., Senti, K. A., Subramanian, S., Sachidanandam, R., and Brennecke, J. (2012). The cochaperone shutdown defines a group of biogenesis factors essential for all piRNA populations in Drosophila. Mol Cell *47*, 954-969.

Pandey, R. R., Homolka, D., Chen, K. M., Sachidanandam, R., Fauvarque, M. O., and Pillai, R. S. (2017). Recruitment of Armitage and Yb to a transcript triggers its phased processing into primary piRNAs in Drosophila ovaries. PLoS Genet *13*, e1006956.

Pane, A., Wehr, K., and Schüpbach, T. (2007). zucchini and squash encode two putative nucleases required for rasiRNA production in the Drosophila germline. Dev Cell *12*, 851-862.

Patil, V. S., Anand, A., Chakrabarti, A., and Kai, T. (2014). The Tudor domain protein Tapas, a homolog of the vertebrate Tdrd7, functions in the piRNA pathway to regulate retrotransposons in germline of Drosophila melanogaster. BMC Biol *12*, 61.

Patil, V. S., and Kai, T. (2010). Repression of retroelements in Drosophila germline via piRNA pathway by the Tudor domain protein Tejas. Curr Biol *20*, 724-730.

Port, F., Chen, H. M., Lee, T., and Bullock, S. L. (2014). Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. Proc Natl Acad Sci U S A *111*, E2967-76.

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

Preall, J. B., Czech, B., Guzzardo, P. M., Muerdter, F., and Hannon, G. J. (2012). shutdown is a component of the Drosophila piRNA biogenesis machinery. RNA *18*, 1446-1457.

Pyle, A. M. (2011). RNA helicases and remodeling proteins. Curr Opin Chem Biol *15*, 636-642.

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.

Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., Scott, D. A., Inoue, A., Matoba, S., Zhang, Y., and Zhang, F. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell *154*, 1380-1389.

Ren, X., Yang, Z., Mao, D., Chang, Z., Qiao, H. H., Wang, X., Sun, J., Hu, Q., Cui, Y., Liu, L. P., Ji, J. Y., Xu, J., and Ni, J. Q. (2014a). Performance of the Cas9 nickase system in Drosophila melanogaster. G3 (Bethesda) *4*, 1955-1962.

Ren, X., Yang, Z., Xu, J., Sun, J., Mao, D., Hu, Y., Yang, S. J., Qiao, H. H., Wang, X., Hu, Q., Deng, P., Liu, L. P., Ji, J. Y., Li, J. B., and Ni, J. Q. (2014b). Enhanced specificity and efficiency of the CRISPR/Cas9 system with optimized sgRNA parameters in Drosophila. Cell Rep *9*, 1151-1162.

Rogers, A. K., Situ, K., Perkins, E. M., and Toth, K. F. (2017). Zucchini-dependent piRNA processing is triggered by recruitment to the cytoplasmic processing machinery. Genes Dev *31*, 1858-1869.

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

Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc *5*, 725-738.

Rozhkov, N. V., Hammell, M., and Hannon, G. J. (2013). Multiple roles for Piwi in silencing Drosophila transposons. Genes Dev *27*, 400-412.

Saito, K., Ishizu, H., Komai, M., Kotani, H., Kawamura, Y., Nishida, K. M., Siomi, H., and Siomi, M. C. (2010). Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in Drosophila. Genes Dev *24*, 2493-2498.

Sander, J. D., and Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol *32*, 347-355.

Selvy, P. E., Lavieri, R. R., Lindsley, C. W., and Brown, H. A. (2011). Phospholipase D: enzymology, functionality, and chemical modulation. Chem Rev *111*, 6064-6119.

Senti, K. A., Jurczak, D., Sachidanandam, R., and Brennecke, J. (2015). piRNA-guided slicing of transposon transcripts enforces their transcriptional silencing via specifying the nuclear piRNA repertoire. Genes Dev *29*, 1747-1762.

Serdar, L. D., Whiteside, D. L., and Baker, K. E. (2016). ATP hydrolysis by UPF1 is required for efficient translation termination at premature stop codons. Nat Commun *7*, 14021.

Sfeir, A., and Symington, L. S. (2015). Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway. Trends Biochem Sci *40*, 701-714.

Sienski, G., Batki, J., Senti, K. A., Dönertas, D., Tirian, L., Meixner, K., and Brennecke, J. (2015). Silencio/CG9754 connects the Piwi-piRNA complex to the cellular heterochromatin machinery. Genes Dev *29*, 2258-2271.

Sigova, A., Rhind, N., and Zamore, P. D. (2004). A single Argonaute protein mediates both transcriptional and posttranscriptional silencing in Schizosaccharomyces pombe. Genes Dev *18*, 2359-2367.

Smith, J., Berg, J. M., and Chandrasegaran, S. (1999). A detailed study of the substrate specificity of a chimeric restriction enzyme. Nucleic Acids Res *27*, 674-681.

Snee, M. J., and Macdonald, P. M. (2004). Live imaging of nuage and polar granules: evidence against a precursor-product relationship and a novel role for Oskar in stabilization of polar granule components. J Cell Sci *117*, 2109-2120.

Sternberg, S. H., and Doudna, J. A. (2015). Expanding the Biologist's Toolkit with CRISPR-Cas9. Mol Cell *58*, 568-574.

Taylor, M. S., LaCava, J., Mita, P., Molloy, K. R., Huang, C. R., Li, D., Adney, E. M., Jiang, H., Burns, K. H., Chait, B. T., Rout, M. P., Boeke, J. D., and Dai, L. (2013). Affinity proteomics reveals human host factors implicated in discrete stages of LINE-1 retrotransposition. Cell *155*, 1034-1048.

Vagin, V. V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P. D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. Science *313*, 320-324.

Vagin, V. V., Yu, Y., Jankowska, A., Luo, Y., Wasik, K. A., Malone, C. D., Harrison, E., Rosebrock, A., Wakimoto, B. T., Fagegaltier, D., Muerdter, F., and Hannon, G. J. (2013). Minotaur is critical for primary piRNA biogenesis. RNA *19*, 1064-1077.

van Schendel, R., Roerink, S. F., Portegijs, V., van den Heuvel, S., and Tijsterman, M. (2015). Polymerase Θ is a key driver of genome evolution and of CRISPR/Cas9-mediated mutagenesis. Nat Commun *6*, 7394.

Voigt, F., Reuter, M., Kasaruho, A., Schulz, E. C., Pillai, R. S., and Barabas, O. (2012). Crystal structure of the primary piRNA biogenesis factor Zucchini reveals similarity to the bacterial PLD endonuclease Nuc. RNA *18*, 2128-2134.

Voronina, E., Seydoux, G., Sassone-Corsi, P., and Nagamori, I. (2011). RNA granules in germ cells. Cold Spring Harb Perspect Biol *3*, pii: a002774.

Vouillot, L., Thélie, A., and Pollet, N. (2015). Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. G3 (Bethesda) *5*, 407-415.

Vourekas, A., Zheng, K., Fu, Q., Maragkakis, M., Alexiou, P., Ma, J., Pillai, R. S., Mourelatos, Z., and Wang, P. J. (2015). The RNA helicase MOV10L1 binds piRNA precursors to initiate piRNA processing. Genes Dev *29*, 617-629.

Wan, G., Fields, B., Spracklin, G., Phillips, C., and Kennedy, S. (2017). Transgenerational Epigenetic Inheritance Factors Localize to Spatially and Temporally Ordered Liquid Droplet Assemblages. bioRxiv doi: https://doi.org/10.1101/220111

Wang, W., Han, B. W., Tipping, C., Ge, D. T., Zhang, Z., Weng, Z., and Zamore, P. D. (2015). Slicing and Binding by Ago3 or Aub Trigger Piwi-Bound piRNA Production by Distinct Mechanisms. Mol Cell *59*, 819-830.

Ward, J. D. (2015). Rapid and precise engineering of the Caenorhabditis elegans genome with lethal mutation co-conversion and inactivation of NHEJ repair. Genetics *199*, 363-377.

Watanabe, T., Chuma, S., Yamamoto, Y., Kuramochi-Miyagawa, S., Totoki, Y., Toyoda, A., Hoki, Y., Fujiyama, A., Shibata, T., Sado, T., Noce, T., Nakano, T., Nakatsuji, N., Lin, H., and Sasaki, H. (2011). MITOPLD is a mitochondrial protein essential for nuage formation and piRNA biogenesis in the mouse germline. Dev Cell *20*, 364-375.

Webster, A., Li, S., Hur, J. K., Wachsmuth, M., Bois, J. S., Perkins, E. M., Patel, D. J., and Aravin, A. A. (2015). Aub and Ago3 Are Recruited to Nuage through Two Mechanisms to Form a Ping-Pong Complex Assembled by Krimper. Mol Cell *59*, 564-575.

Weng, Y., Czaplinski, K., and Peltz, S. W. (1996). Genetic and biochemical characterization of mutations in the ATPase and helicase regions of the Upf1 protein. Mol Cell Biol *16*, 5477-5490.

White-Cooper, H. (2012). Tissue, cell type and stage-specific ectopic gene expression and RNAi induction in the Drosophila testis. Spermatogenesis *2*, 11-22.

Wickersheim, M. L., and Blumenstiel, J. P. (2013). Terminator oligo blocking efficiently eliminates rRNA from Drosophila small RNA sequencing libraries. Biotechniques *55*, 269-272.

Wilsch-Bräuninger, M., Schwarz, H., and Nüsslein-Volhard, C. (1997). A sponge-like structure involved in the association and transport of maternal products during Drosophila oogenesis. J Cell Biol *139*, 817-829.

Xiol, J., Spinelli, P., Laussmann, M. A., Homolka, D., Yang, Z., Cora, E., Couté, Y., Conn, S., Kadlec, J., Sachidanandam, R., Kaksonen, M., Cusack, S., Ephrussi, A., and Pillai, R. S. (2014). RNA clamping by Vasa assembles a piRNA amplifier complex on transposon transcripts. Cell *157*, 1698-1711.

Yu, A. M., and McVey, M. (2010). Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions. Nucleic Acids Res *38*, 5706-5717.

Yu, Y., Gu, J., Jin, Y., Luo, Y., Preall, J. B., Ma, J., Czech, B., and Hannon, G. J. (2015). Panoramix enforces piRNA-dependent cotranscriptional silencing. Science *350*, 339-342.

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. Biol Open *3*, 271-280.

Yue, W., Chen, Z., Liu, H., Yan, C., Chen, M., Feng, D., Yan, C., Wu, H., Du, L., Wang, Y., Liu, J., Huang, X., Xia, L., Liu, L., Wang, X., Jin, H., Wang, J., Song, Z., Hao, X., and Chen, Q. (2014). A small natural molecule promotes mitochondrial fusion through inhibition of the deubiquitinase USP30. Cell Res *24*, 482-496.

Zachar, Z., and Bingham, P. M. (1982). Regulation of white locus expression: the structure of mutant alleles at the white locus of Drosophila melanogaster. Cell *30*, 529-541.

Zamparini, A. L., Davis, M. Y., Malone, C. D., Vieira, E., Zavadil, J., Sachidanandam, R., Hannon, G. J., and Lehmann, R. (2011). Vreteno, a gonad-specific protein, is essential for germline development and primary piRNA biogenesis in Drosophila. Development *138*, 4039-4050.

Zamudio, N., Barau, J., Teissandier, A., Walter, M., Borsos, M., Servant, N., and Bourc'his, D. (2015). DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination. Genes Dev *29*, 1256-1270.

Zhang, F., Wang, J., Xu, J., Zhang, Z., Koppetsch, B. S., Schultz, N., Vreven, T., Meignin, C., Davis, I., Zamore, P. D., Weng, Z., and Theurkauf, W. E. (2012). UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. Cell *151*, 871-884.

Zhang, H., Liu, K., Izumi, N., Huang, H., Ding, D., Ni, Z., Sidhu, S. S., Chen, C., Tomari, Y., and Min, J. (2017). Structural basis for arginine methylation-independent recognition of PIWIL1 by TDRD2. Proc Natl Acad Sci U S A *114*, 12483-12488.

Zhang, X., Ferreira, I. R., and Schnorrer, F. (2014a). A simple TALEN-based protocol for efficient genome-editing in Drosophila. Methods *69*, 32-37.

Zhang, X., Koolhaas, W. H., and Schnorrer, F. (2014b). A versatile two-step CRISPR- and RMCE-based strategy for efficient genome engineering in Drosophila. G3 (Bethesda) *4*, 2409-2418.

Zhang, Y., Liu, W., Li, R., Gu, J., Wu, P., Peng, C., Ma, J., Wu, L., Yu, Y., and Huang, Y. (2018). Structural insights into the sequence-specific recognition of Piwi by Drosophila Papi. Proc Natl Acad Sci U S A *115*, 3374-3379.

Zhang, Y., Liu, X., Bai, J., Tian, X., Zhao, X., Liu, W., Duan, X., Shang, W., Fan, H. Y., and Tong, C. (2016). Mitoguardin Regulates Mitochondrial Fusion through MitoPLD and Is Required for Neuronal Homeostasis. Mol Cell *61*, 111-124.

Zhang, Z., Theurkauf, W. E., Weng, Z., and Zamore, P. D. (2012). Strand-specific libraries for high throughput RNA sequencing (RNA-Seq) prepared without poly(A) selection. Silence *3*, 9.

Zhang, Z., Xu, J., Koppetsch, B. S., Wang, J., Tipping, C., Ma, S., Weng, Z., Theurkauf, W. E., and Zamore, P. D. (2011). Heterotypic piRNA Ping-Pong requires qin, a protein with both E3 ligase and Tudor domains. Mol Cell *44*, 572-584.

Zheng, K., Xiol, J., Reuter, M., Eckardt, S., Leu, N. A., McLaughlin, K. J., Stark, A., Sachidanandam, R., Pillai, R. S., and Wang, P. J. (2010). Mouse MOV10L1 associates with Piwi proteins and is an essential component of the Piwi-interacting RNA (piRNA) pathway. Proc Natl Acad Sci U S A *107*, 11841-11846.