

# Molecular Function and Regulation of Aub Arginine Methylation in the piRNA Pathway

Thesis by  
Xiawei Huang

In Partial Fulfillment of the Requirements for the  
degree of  
Doctor of Philosophy in Biology

The Caltech logo is displayed in a bold, orange, sans-serif font. The word "Caltech" is centered within a light orange rectangular background.

CALIFORNIA INSTITUTE OF TECHNOLOGY  
Pasadena, California

2021  
Defended May 5<sup>th</sup>, 2021

© 2021

Xiawei Huang  
ORCID: 0000-0001-9084-0510

## ACKNOWLEDGEMENTS

Being a Ph.D. student at Caltech and one of the members of the AAA lab is a memorable period in my life. I can't get so far and complete the project successfully without the support and guidance of many people—families, advisors, teachers, and friends.

I want to express my deep and sincere gratitude to my research supervisor Alexei Aravin for allowing me to be a member of the AAA team. He had provided invaluable guidance throughout the research, including spending time to help me familiar with the piRNA field when I just joined the lab—showing incredible patience to discuss and design the project and experiments. He has taught me how to think logically, present the work, and, most importantly, ask a scientific question and find a way to solve it. Such scientific logic will be a priceless treasure throughout my career, no matter where I am. And thank him for tolerating lots of rookie mistakes I have made during the Ph.D. study and correcting them. I would also thank Katalin Fejes Tóth for the project discussion and important experiment designs to help me with the data visualization. Thank you, Kata.

I also want to express my appreciation to my committee. Thank you for your support and advice. The chair of my committee, David Chan, it was a great honor to being the head TA in your course, encouraged by the attitude to the scientific question. Shu-ou Shan for the critical question addressed during the annual meeting. Rebecca Voorhees, thank you for joining my committee board after Deshaies left Caltech, and thank you for the kindness and let Angel help me with my project. The feedback from each of you is important for me to finish my project, thank you.

Thank you, all members I have met in the AAA team and KFT lab: Dubravka Pezic, Evelyn Stuwe, Adrien Le Thomas, Alicia Rogers, Ariel Chen, Alex Webster, Sergei Manakov, Junho Hur, Hamada Masakazu, Yicheng Luo, Chen-yin Ou, Jae Cho, Maria Ninova, Riley Galton, April Jauhal, Sharan Prakash, Evita Varela, Zsofia Torok, Tang Qing, Peiwei Chen, Elena Fefelova, Elena Udartseva, Esiyunina, Daria, Baira Godneeva. You have made life in the lab lots of fun, and thank you for tolerating my messed bench.

To Alexandre Webster, my mentor when I rotated in the Aravin lab. Thank you for leading me to initiate my lab career in Caltech. Thank you for your previous work, which is the fundamental of my Ph.D. project.

To Alicia Rogers, thank you for teaching me the preparation of small RNA libraries and data processing. As a senior student in the lab, your advice is extremely important and helpful for my project. I also want to express my gratitude for taking care of my cat when I was off the town.

To Yicheng Luo, my labmate and teammate on the basketball court. Special thanks to you for the project discussion and the construction of fly strain.

To Marica Ninova, who is always providing generous help with bioinformatics data analysis and teaching me some background knowledge. Thank you!

To my best friends in China. Wenqiang Jian, Jiesi, Chen, Jianchong, Wei and Zepeng, Huang. Even though separated by the Pacific Ocean, you guys always gave me the strength and courage when I was down at those difficult times during my Ph.D. You are like my families and thank you for your love and support.

Finally, I am extraordinarily grateful to my family. My parents Zhiqiang Huang and Jian Liu, for their love, prayers, caring, and sacrifice. Especially, I want to thank my auntie Lily Huang and Uncle Charlie Trimble. I still remember the time we met in Xiamen, China, nine years ago. That meeting completely changed my life path. Your generous help supports me in seeking an opportunity to study in America and offering me the opportunity to have interned in a biotech company. Uncle Charlie, your wisdom and passion for science will always encourage me in the future of my life. And auntie Lily, you are the most special woman I have met in my life. Your life philosophy and experience will be the priceless knowledge guiding me and teaching me to become a useful and wise man no matter where I am and what I do.

## ABSTRACT

Transposon elements (TEs, Transposons) are DNA sequences that can change their position within the genome. TEs, so-called 'jump genes', sometimes create mutation which will disrupt genes or damage the genome integrity by causing double-stranded DNA breaks and germ cell death. It is important for living animals to maintain the integrity of genetic information during reproduction. In Metazoa germline, cells use the piwi-interacting RNA (piRNA) pathway, which is an RNA – interference (RNAi) based defense strategy to protect the genome from the attacking of the “selfish” transposons. The core unit of the piRNA pathway is the RNA-induced silencing complex (RISC), a conserved family of Argonaute protein that interacts with small (19–33 nt) RNA guides in eukaryotic species. In *Drosophila melanogaster*, three PIWI-clade Argonaute proteins are present in the germline – Aubergine (Aub), Argonaute 3 (Ago3) and Piwi. PIWI proteins together with their substrate piRNAs forming the RISC to suppress the TE activity. Arginine (Arg) methylation is an important post-translational modification among Argonaute proteins. Defects of Arginine methylation cause the de-repression of deleterious TE.

The work present in this thesis examines the molecular function and regulation mechanism of Aub Arginine methylation in *Drosophila* germline cells. Chapter I presents a general introduction to the TEs, key components of the piRNA pathway, potential piRNA processing site, “nuage”, and the correlation between arginine methylation and its interaction partner, Tudor domain-containing proteins. Chapter II presents the piRNA biogenesis in *Drosophila* germline and somatic cells, dividing the piRNA pathway into cytoplasmic and nuclear branches. We describe the mechanism of piRNA 5' end and 3' end formations. Chapter III explores the specific molecular function of Aub arginine methylation in the

piRNA ping-pong cycle. Further, we decipher the regulation mechanism of Aub Arginine methylation, addressing its biological meaning for the piRNA biogenesis. In chapter 4, we developed a heterologous two-hybrid system to identify factors that directly interact with Piwi, which can further be applied to elucidate the interaction network of the piRNA pathway. In chapter 5, we discuss the potential role of phase separation in the assembly of ping-pong processing granule and the biological meaning in the piRNA biogenesis. We also propose the future plan and the protocol to examine the hypothesis in the future.

## PUBLISHED CONTENT AND CONTRIBUTIONS

Huang, Xiawei, Katalin Fejes Tóth, and Alexei A. Aravin. "piRNA Biogenesis in *Drosophila melanogaster*." *Trends in Genetics* 33.11 (2017): 882-894.

**doi:** <https://doi.org/10.1016/j.tig.2017.09.002>

**XH** wrote the manuscript

Huang, Xiawei, et al. "Binding of guide piRNA triggers methylation of the unstructured N-terminal region of Aub leading to assembly of the piRNA amplification complex."

**bioRxiv: doi:** <https://doi.org/10.1101/2020.07.14.203323> (Submitted)

**XH** generated flies, constructs, design and executed experiments, analyzed data, wrote the manuscript

## TABLE OF CONTENTS

Acknowledgements.....	iii
Abstract .....	vi
Published Content and Contributions.....	viii
Table of Contents.....	ix
List of Illustrations and/or Tables.....	xi
<b>Chapter 1: INTRODUCTION.....</b>	<b>1</b>
Overview of transposable elements .....	1
piRNA and piRNA clusters.....	3
The role of PIWI-piRNA complexes in <i>Drosophila melanogaster</i> .....	5
The Nuage: a potential ping-pong processing site .....	8
Tudor-domain-containing proteins and PIWI sDMA modification .....	10
The role of phase separation in the formation of membraneless organelles	11
References .....	14
Figures and Figure Legends .....	21
<b>Chapter 2: PIRNA BIOGENESIS IN DROSOPHILA MELANOGASTER..</b>	<b>25</b>
Abstract .....	26
Introduction.....	26
Nuclear steps of piRNA biogenesis .....	29
piRNA biogenesis in the cytoplasm.....	35
Licensing of piRNA precursors for processing .....	41
References .....	46
Figures and Figure legends .....	56
<b>Chapter 3: BINDING OF GUIDE PIRNA TRIGGERS METHYLATION OF THE UNSTRUCTURED N-TERMINAL REGION OF AUB LEADING TO ASSEMBLY OF THE PIRNA AMPLIFICATION COMPLEX.....</b>	<b>62</b>
Abstract .....	63
Introduction.....	64
Result.....	67
Discussion.....	84
Materials and Methods .....	93
References .....	105
Figures and Figure legends .....	112
Extended Data Figures.....	128
<b>Chapter 4: IDENTIFICATION OF PIWI DIRECT INTERACTORS USING HETEROLOGOUS CELL CULTURE ASSAY .....</b>	<b>142</b>
Abstract .....	143
Introduction.....	144
Result.....	146
Discussion.....	151
Materials and Methods .....	155
References .....	158
Figures and Figure legends .....	165

Supplemental Material.....	173
Chapter 5: CONCLUDING REMARKS AND FUTURE WORK .....	177
Methods and Materials .....	183
References .....	185

## LIST OF ILLUSTRATIONS AND/OR TABLES

	<i>Page</i>
Chapter 1 INTRODUCTION	
Figure 1. The Domain structure of PIWI-clade Argonaute proteins.....	21
Figure 2. The Nuage in <i>Drosophila</i> ovaries.....	23
Chapter 2 PIRNA BIOGENESIS IN DROSOPHILA MELANOGASTER	
Figure 1. Transcription of PIWI-Interacting RNA (piRNA) clusters .....	56
Figure 2. Processing of the 5'- and the 3'-end of PIWI-Interacting RNA .....	58
Figure 3. Selection of PIWI-Interacting RNA precursors for processing.....	60
Chapter 3 BINDING OF GUIDE PIRNA TRIGGERS METHYLATION OF THE UNSTRUCTURED N-TERMINAL REGION OF AUB LEADING TO ASSEMBLY OF THE PIRNA AMPLIFICATION COMPLEX	
Figure 1. Krimper monomer simultaneously binds N-terminal regions of Aub and Ago3 .....	112
Figure 2. Structure of Krimper eTud2-Aub and eTud1-Ago3 complex.....	114
Figure 3. Arginine methylation of Aub is required for fertility and TE repression .....	116
Figure 4. Arginine methylation of Aub is dispensable for piRNA binding but required for piRNA expression and ping-pong processing .....	118
Figure 5. RNA binding promotes Aub arginine methylation .....	121
Figure 6. RNA binding triggers conformational change in Aub, exposing its N terminus to the methyltransferase CsuI/Vls complex.....	124
Figure 7. Model for sDMA regulation and its function in ping-pong cycle ..	126
Extended Data Figure 1. Interaction between Krimp, Aub and Ago3 .....	128
Extended Data Figure 2. Detailed analysis of Krimper eTud1 apo structure	130
Extended Data Figure 3. Structural comparison of Krimper eTud1-Ago3-2 complex with other extended Tudor domains .....	132
Extended Data Figure 4. Localization, mobility and slicer activity of wtAub and mdAub proteins .....	134
Extended Data Figure 5. Analysis of piRNAs in flies expressing wtAub and mdAub proteins .....	136

Extended Data Figure 6. Oligo binding causes Aub conformational change	138
Extended Data Table 1. Structural Data collection and refinement statistics	140
Extended Data Table 2. Oligo and primer sequences .....	141
<b>Chapter 4 IDENTIFICATION OF PIWI DIRECT INTERACTORS USING HETEROLOGOUS CELL CULTURE ASSAY</b>	
Figure 1. A heterologous cell culture assay to distinguish between direct and indirect protein-protein interactions with <i>Drosophila melanogaster</i> Piwi ....	165
Figure 2. Piwi directly interacts with nuage components Vasa, Shutdown, Papi, Squash, Qin, Spindle-E, and Aubergine.....	167
Figure 3. Piwi directly interacts with the exon junction complex (EJC) component Mago .....	169
Figure 4. Piwi directly interacts with the Rhino-Deadlock-Cutoff (RDC) complex component Cutoff and the novel interactor Arp6 .....	171
Supplementary Figure S1. Piwi does not directly interact with many nuage localized factors involved in piRNA biogenesis .....	173
Supplementary Figure S2. Piwi does not directly interact with several factors involved in transcript export from the nucleus .....	175
<b>Chapter 5 CONCLUDING REMARKS AND FUTURE WORK</b>	
Preliminary Data .....	182

## *Chapter 1*

### INTRODUCTION

#### **Overview of transposable elements**

Transposable elements (TEs) are selfish DNA elements that exploit the genome and replicative machinery of host cells to survive and proliferate (O'Donnell & Boeke, 2007). These elements occupy nearly one-third to half of the fruit fly *Drosophila* and human genomes, respectively. The activity of TEs within the germline genome will have a deleterious effect to the host (Doolittle & Sapienza, 1980; Orgel & Crick, 1980). Three distinct mechanisms cause such harmful effect. First, TEs insertions can directly disrupt gene expression by inserting into the exon or promoter region. Second, ectopic recombination between insertions at different sites may result in deleterious genomic rearrangements. Third, the product of the TEs, such as transposase, will generate DNA double-stranded break, which further damages the genome integrity (Hedges & Deininger, 2007; Moon et al., 2018; Slotkin & Martienssen, 2007). Although TEs are selfish elements, some are important in genome function and evolution (Bucher, Reinders, & Mirouze, 2012; Feschotte & Pritham, 2007). Several beneficial TEs insertions have been reported, such as increasing resistance to insecticides (Aminetzach, Macpherson, & Petrov, 2005; Casacuberta & Gonzalez, 2013).

Due to the deep evolutionary origins and diversifications, TE families consist of various forms and shapes. TEs are divided into two major classes based on transposition mechanisms, and each class can be subdivided into subclasses based on the mechanisms of chromosomal integration (Abrusan, Grundmann, DeMester, & Makalowski, 2009;

Kapitonov & Jurka, 2008). Class I TEs, which are also called retrotransposons, mobilize through three steps. First, they are transcribed from DNA to RNA, and then the transcripts are reverse transcribed to DNA sequences. The copied DNA sequences are finally inserted back to the genome at a new locus (Dombroski et al., 1994). Such copy and paste mechanism is very similar to retrovirus (Drost & Sanchez, 2019). Retrotransposons can be subclassified into three groups. Long terminal repeats (LTRs) encode reverse transcriptase (Brown, Bowerman, Varmus, & Bishop, 1987). Long interspersed nuclear elements (LINEs) which don't contain the LTR but encode reverse transcriptase. Short interspersed nuclear elements (SINEs) neither include the LTR nor encode the transcriptase (Luan, Korman, Jakubczak, & Eickbush, 1993). Class II TEs also known as DNA transposons, replicate via a cut and paste mechanism (Lambowitz & Zimmerly, 2011). The transposition process doesn't require RNA intermediates; instead, Two transposases recognize and bind to TIR sequences, join together and promote DNA double-strand cleavage. The DNA-transposase complex then inserts its DNA cargo at specific DNA motifs elsewhere in the genome (Munoz-Lopez & Garcia-Perez, 2010). TE sequences are grouped into families or subfamilies in the detailed classification, which is defined as closely related groups of sequences that can be traced as descendants of a single ancestral sequence (Britten & Kohne, 1968). This ancestral unit is regarded as a consensus sequence that presents the entire family (Jurka & Smith, 1988; Smit, 1999). Thus, any sequence in the genome can be affiliated to a TE class.

With the ability to proliferate within the genome, TEs frequently invade novel populations and species (Kidwell, 1983; Kofler, Hill, Nolte, Betancourt, & Schlotterer, 2015; Montchamp-Moreau, 1990; Rozhkov et al., 2013a). Uncontrolled TEs invasions will potentially lead to the extinction of the host population (Le Rouzic & Deceliere, 2005).

Therefore, it's essential for the organism to control the spread of TEs. Historically, it was thought that the proliferation of TEs is suppressed within the population by natural selection acting against the TEs invasions (Charlesworth & Charlesworth, 1983; Charlesworth & Langley, 1989; Russell & Woodruff, 1999). This model was described as 'transposition-selection balance'. TEs copy numbers within the population are at an equilibrium state between TEs transpositions events that generate new deleterious insertions and negative selections that remove the insertions (Petrov, Fiston-Lavier, Lipatov, Lenkov, & Gonzalez, 2011). This model was prevalent until the discovery of the piRNA pathway; a small RNA-based defense system dramatically changed our understanding of TE dynamics (Brennecke et al., 2007; Gunawardane et al., 2007).

### **piRNA and piRNA clusters**

People found that the spread of TEs was not solely counteracted by the natural selection but actively combated by the host (Blumenstiel, 2011; Lee & Langley, 2010). The host defense system relies on the 23-29nt small RNAs, so-called Piwi-interacting RNAs (piRNAs) (Brennecke et al., 2007; Gunawardane et al., 2007). Deep sequencing results show that a large fraction of piRNA sequences can be mapped to a few discrete loci locating at the heterochromatin region close to the euchromatin boundary. Such genomic loci are called piRNA clusters (Brennecke et al., 2007; Malone et al., 2009). piRNA clusters contain a large number and various types of TEs, most of which are inactive copies or truncated fragments (Brennecke et al., 2007). Thus, the sequences of piRNAs derived from these clusters are homologous to TEs in the clusters and related TEs located elsewhere in the genome and can therefore act as guides to repress TEs. Thus, piRNA clusters are genetic elements that regulate the activity of TEs. Several studies have shown novel TE insertion will trigger the

production of de novo piRNAs (Brennecke et al., 2008; Khurana et al., 2011; Rozhkov et al., 2013b). Furthermore, full-length TE sequences could be inserted into pre-existing piRNA clusters (Lu & Clark, 2010). Such observation gave rise to the 'trap model'. TEs keep proliferating until at least one TE sequence 'jumps' into the piRNA cluster (trap), triggering the production of piRNAs that silence themselves and homologous TEs as well. piRNA clusters employ small RNA-based immunity to silence homologous TEs selectively. It is believed that previously inserted TE sequences in the clusters will leave a memory of invasion. If the same TE attacks the host, the piRNA clusters will generate corresponding piRNAs based on the memory to limit the TE activity (Malone & Hannon, 2009; Matranga & Zamore, 2007; Saito & Siomi, 2010). piRNAs and piRNA clusters have been found in many species such as flies, worms, and mouse (Aravin, Hannon, & Brennecke, 2007; Girard, Sachidanandam, Hannon, & Carmell, 2006; Ruby et al., 2006; Watanabe et al., 2006), suggesting the small RNA-based defend system holds for most invertebrates and mammals.

TEs can be activated in either germline cells or somatic cells surrounding the germline, termed germline TEs and somatic TEs. These two types of TEs are controlled by two sets of piRNA machinery that depend on distinct sets of piRNA clusters (C. Li et al., 2009; Malone et al., 2009). In *Drosophila melanogaster*, the flamenco (flam) produces major piRNAs in the somatic support cells in the ovary (Malone et al., 2009). Flam was originally discovered to suppress the gypsy, idfix, and ZAM TE families (Desset, Meignin, Dastugue, & Vaury, 2003; Pelisson et al., 1994). piRNA precursors from flam, which span around 150kb, are generated through unidirectional transcription oriented in the anti-sense direction to the TEs. Such anti-sense orientation was not found in the germline (Malone et al., 2009). The

predominant fraction of germline piRNA clusters is dual-stranded clusters that produce piRNA from both strands of the same genomic region. 42AB, for example, is a representative dual-stranded cluster in *Drosophila* germline spans around 240kb locating near the pericentromeric heterochromatin boundary. The orientation of truncated TEs within 42AB is random rather than anti-sense bias, and piRNAs are produced from both sense and anti-sense strands.

### **The role of PIWI-piRNA complexes in *Drosophila melanogaster***

RNA-induced silencing complex (RISC) is a multiprotein complex, specifically a ribonucleoprotein, which functions in gene silencing via various pathways at the transcriptional and translational levels (Pratt & MacRae, 2009). RISCs find targets via base-pairing and promote the inactivation of homologous sequences (Carthew & Sontheimer, 2009; Ghildiyal & Zamore, 2009; Siomi & Siomi, 2009). piRNAs are produced and loaded onto PIWI-clade Argonaute proteins to form piRNA-induced silencing complexes (piRISCs). In *Drosophila melanogaster*, there are three PIWI-clade Argonaute proteins – Piwi, Aubergine (Aub), and Argonaute3 (Ago3). The PIWI proteins have a structure composed of four domains similar to other Argonaute proteins (Fig. 1). MID domain anchors 5' end of piRNA, PAZ domain is required for binding of 3' end of piRNAs, PIWI domain contains RNase H-like fold, providing endonuclease activity, allowing PIWI protein to cleave target sequences (Ma, Ye, & Patel, 2004; Ma et al., 2005; Song, Smith, Hannon, & Joshua-Tor, 2004; Wang et al., 2008; Yuan et al., 2005). The structure of the disordered N terminal region remains unknown. Genetic studies of three PIWI revealed that mutations in these genes affect germline development (Cox et al., 1998; Harris & Macdonald, 2001; C. Li et al., 2009; Lin & Spradling, 1997). TEs are derepressed in mutant ovaries defective in

these genes, suggesting a model in which TE overexpression and mobilization triggers DNA damage signaling-dependent defects in an early step in the germ cell patterning cascade (Khurana & Theurkauf, 2010). Despite their similarity, the profiles of small RNA binding partners of three PIWI clade proteins are quite different. Deep sequencing and computational analysis showed strong orientation biases within their bound piRNAs. Piwi and Aub bound piRNAs are predominantly anti-sense orientation to TEs, while Ago3 bound piRNAs demonstrate sense bias (Brennecke et al., 2007; Gunawardane et al., 2007). Furthermore, Piwi and Aub bound piRNAs show a pronounced tendency to begin with Uridine (1U), whereas Ago3-associated piRNAs featured an adenine at position 10 (10A). Bioinformatics analysis revealed significant 10-nt overlaps between Aub- and Ago3-associated sequences. These findings led to the proposal of the ping-pong model and, later on, similar amplification loops were identified in silkworm, fish, mouse, and many other organisms (De Fazio et al., 2011; Kaaij, Hoogstrate, Berezikov, & Ketting, 2013; Katsuma, Kawamoto, & Kiuchi, 2014). Due to the differences between their bound piRNAs, three PIWI-piRISCs work non-redundantly. piRISCs mediated TE suppression can be divided into two groups. Piwi-piRISCs suppress TE proliferation through transcriptional silencing while Aub- and Ago3-piRISCs apply post-transcriptional silencing (Brennecke et al., 2007; Gunawardane et al., 2007; Le Thomas et al., 2013; Sienski, Donertas, & Brennecke, 2012).

Piwi-piRISCs are genetic elements that mediate and maintain epigenetic chromatin modifications of target TE loci in somatic support cells and germline cells. Piwi-piRISCs assembled in the cytoplasm (Saito et al., 2009) are imported into the nucleus. Elimination of Piwi from the nucleus will cause a change in histone marks on TEs loci (Klenov et al., 2011; Poyhonen et al., 2012), suggesting a potential role of Piwi in chromatin regulation. Later,

people found that Piwi exclusively represses loci targeted by piRNAs. Piwi-mediated silencing correlates with the installment of repressive chromatin marks tri-methylated histone 3 lysine 9 (H3K9me3) at targeted loci (Huang et al., 2013; Le Thomas et al., 2013). TE silencing requires the continual activity of Piwi-piRISCs (Huang et al., 2013; Le Thomas et al., 2013).

Unlike Piwi protein, Aub and Ago3 are exclusively expressed in the germline cells. Piwi is a nuclear protein, while Aub and Ago3 are cytoplasmic (Brennecke et al., 2007; Gunawardane et al., 2007; Saito et al., 2006). Aub-piRISCs containing anti-sense piRNAs in the cytoplasm will target and cleave both sense precursor transcripts from dual-stranded piRNA clusters and transcripts from active TEs. Upon recognition of target sequences, Aub will use the slicer activity exhibited by the PIWI domain (Ma et al., 2004; Ma et al., 2005; Song et al., 2004). The cleavage product is the generation of sense piRNAs, which will be loaded onto Ago3. This process triggers a feed-forward amplification loop of piRNA production, the so-called 'ping-pong cycle'. The functional ping-pong cycle requires the catalytic activity of both Ago3-piRISCs and Aub-piRISCs. The cycle leads to repeated rounds of piRNA production by consuming both cluster transcripts and TE transcripts, thereby silencing TEs at post-transcriptional levels in the cytoplasm. The recent finding suggested that cleavage in the ping-pong cycle by Ago3-bound piRNAs produces not only a corresponding ping-pong partner (Aub-bound piRNA) trigger piRNA (bound to Ago3) produces not only a corresponding ping-pong partner piRNA but also piRNA precursors from the remaining 3' portion of the cleaved transcripts. The cleaved 3' transcript with 1U will be loaded onto Piwi, which will trigger repeated rounds of cleavage with a ~27-nucleotide interval. The interval may be determined by a combination of Piwi's footprint,

Zuc cleavage of Piwi-bound trail sequences, and the preference of Piwi to bind 1U (Han, Wang, Li, Weng, & Zamore, 2015; Mohn, Handler, & Brennecke, 2015).

### **The Nuage: a potential ping-pong processing site**

Germline cells of metazoa are characterized by the presence of an electron-dense unique membraneless organelle called the nuage (which means 'cloud' in French). The nuage forms an amorphous and fibrous structure that is localized to the cytoplasmic face of the nuclear envelope when visualized by electron microscopy. Nuage is featured by the absence of surrounding membranes, an abundance of RNAs and proteins, and a close association with mitochondria clusters or immediately adjacent to the nuclear envelope of germ cells. Similar structures are also known by various names in different animals, including intermitochondrial cement (IMC, also named pi-body), perinuclear P granules in *Caenorhabditis elegans*, and chromatoid body (CB) in mammalian germ cells, according to their localization, morphology, and biochemical properties (Aravin et al., 2009; Brangwynne et al., 2009; Eddy, 1974, 1975). The conservation of nuage across species in the animal kingdom suggests a pivotal role in the germline. In *Drosophila*, the nuage is first detectable when the primordial germ cells (PGCs) form and persist throughout oogenesis in the adult germline cells (Saffman & Lasko, 1999). RNase and protease treatments demonstrated that the nuage is enriched with RNAs and proteins (Mahowald, 1968, 1971).

In *Drosophila*, the nuage can be observed only in the nurse cells but not in the oocytes. Oocyte differentiation starts with cyst formation, where a germline stem cell divides asymmetrically to give rise to two daughter cells – a stem cell and a differentiating cystoblast. The cystoblast undergoes four rounds of mitosis and generates a 16-cell cyst in which all of the resulting germline cells are interconnected. Only one of those 16 cells eventually

becomes an oocyte, whereas the others differentiate into nurse cells. The nurse cells provide abundant RNAs, which are deposited into the oocyte through cytoplasmic bridges. During this stage, TEs may be actively transcribed in the nurse cells, and their transcripts could, thus, be delivered to the oocyte. Thus, it's vital to clean the deleterious transcripts before they leave the nurse cell. Consistent with the high transcriptional activity of transposons in nurse cells, the nuage can be seen only in the nurse cells but not in the oocyte (Fig. 2). Aub-piRISCs and Ago3-piRISCs both localize to the nuage, suggesting nuage is the potential site where post-transcriptional silencing occurs. The piRISCs appear to target such transposon transcripts and piRNA precursors as they are being shuttled from the nucleus to the cytoplasm of the nurse cells.

In the *Drosophila* germline, the nuage is large enough to be visible by light microscopy (Fig.2). Many of the proteins that localize to the nuage have been reported involved in the production of germline piRNAs, including Vasa (Vas), a conserved DEAD box RNA helicase (Liang, Diehljones, & Lasko, 1994); Aub and Ago3, two PIWI-clade proteins (Gunawardane et al., 2007; Harris & Macdonald, 2001; C. Li et al., 2009); Krimper (Krimp), a Tudor domain-containing protein (Lim & Kai, 2007); Maelstrom (Mael), an HMG box-containing protein (Findley, Tamanaha, Clegg, & Ruohola-Baker, 2003); Spindle-E (Spn-E), a DExH box putative RNA helicase (Gillespie & Berg, 1995); Tejas (Tej), a Tudor domain protein (Patil & Kai, 2010), and Vreteno (Vret), another tudor domain protein (Handler et al., 2011). Mutant of any of these factors in nuage will derepress majority TEs in ovary due to the insufficient production of piRNAs. Deep sequencing analysis of piRNA profiles from the mutant ovaries revealed the collapse of the ping-pong cycle (C. Li et al., 2009; Malone et al., 2009).

Several studies showed that nuage components involved in the ping-pong amplification of piRNAs exhibit a hierarchical genetic interaction for their localizations (Findley et al., 2003; Lim & Kai, 2007; Saito & Siomi, 2010). Vas appears to be the most upstream, followed by Tej/Spn-E, Krimp, Ago3, and Mael. In the absence of Vas, components above are de-localized from the perinuclear nuage, and Vas remains in nuage in all of the other mutant ovaries. In addition, Aub and Ago3, which are the main components of the ping-pong amplification loop, are recruited to nuage in two distinct mechanisms. Aub requires a piRNA guide for nuage recruitment, indicating that its localization depends on the recognition of RNA targets. Ago3 is recruited to nuage independently of a piRNA cargo and relies on interaction with Krimper. Aub is required for the proper localization of the remaining downstream components, Krimp and Mael (C. Li et al., 2009; Webster et al., 2015). These hierarchical relationships among the piRNA components indicate their sequential order of function in the biogenesis of piRNAs and nuage assembly.

### **Tudor-domain-containing proteins and PIWI sDMA Modification**

Although molecular functions of nuage components haven't been fully elucidated, people proposed that they may play important roles in the ping-pong cycle based on their genetic interaction for their nuage localization, predicted domain as well as computational analysis of ping-pong piRNA profiles upon mutant of components. Among those known nuage components, the most predominant components are Tudor-domain-related proteins (TDRDs). TDRD proteins belong to the TUDOR protein family, which received the name from the *Drosophila* tudor gene. Tudor domain has been found in proteins from a broad range of eukaryotes, including fission yeast, fungi, plants, and animals. Based on the sequence and structural similarity, Tudor domain, together with chromatin-binding

(Chromo), malignant brain tumor (MBT), PWWP (conserved Proline and Tryptophan), and plant Agenet domains, comprises the Tudor domain 'Royal Family' (Maurer-Stroh et al., 2003). The structural data from many of the Tudor domain proteins demonstrated that the domain has a ~60 amino acid core structure composed of four antiparallel  $\beta$ -strands that form a barrel-like structure with an aromatic binding pocket at the surface to accommodate methylated lysine/arginine ligands (Botuyan & Mer, 2016; Cote & Richard, 2005). Nuage components such as Spn-E, Krimp, Tej, Papi, and Vret harbor single or multiple copies of Tudor domains. The versatile protein domain architecture confers a diverse protein-protein/RNA interacting network and endows various functions in the piRNA pathway (Amikura, Hanyu, Kashikawa, & Kobayashi, 2001; Boswell & Mahowald, 1985; Thomson & Lasko, 2004). Indeed, recent studies have revealed that PIWI subfamily proteins contain RA/RG arginine-rich motif within their N terminal regions, which are modified to symmetric di-methyl arginines (sDMAs) by protein methyltransferase 5 (dPRMT5, also known as Capsuleen/Dart5) (Kirino et al., 2010; Nishida et al., 2009; Vagin et al., 2009). Interactions between some of the Tudor domain proteins and Piwi-family proteins are mediated by the sDMAs of the Piwi family of proteins (Kirino et al., 2010; Nishida et al., 2009; Webster et al., 2015). Depletion of sDMA methyltransferase Capsuleen will cause the collapse of the piRNA pathway and the depression of germline TEs, suggesting a potential role of arginine methylation in the piRNA biogenesis. However, the molecular and regulation mechanism remains unknown.

### **The role of phase separation in the formation of membraneless organelles**

Cellular compartments and organelles provide the site for biological processing. Most well-known organelles are separated by a membrane boundary. Recently, people found that

there are also many so-called membraneless organelles and suggested that these organelles, which are supramolecular assemblies of proteins and RNA molecules, form via protein phase separation. Nuage, as one of the well-known membraneless cytoplasmic compartments, has been studied for years. However, the forces driving their formation remained unclear. Several early studies of other membranless structures (Andrei et al., 2005; Handwerger, Cordero, & Gall, 2005; Patel et al., 2015). P granules (RNA and protein-containing bodies in embryos of *Caenorhabditis elegans*) have liquid-like properties and form by phase separation (Brangwynne et al., 2009). The physical nature of the formation occurs when a supersaturated solution of components spontaneously separates into two phases, a dense phase and a dilute phase, that then stably coexist. The proposed liquid-like nature of P granules was evident from their round appearance (the result of minimizing surface tension), deformability (fusion and fission events), and dynamic exchange of components (Boeynaems et al., 2018).

Phase separation is a well-known phenomenon in polymer chemistry. However, its application to biomacromolecules is a much more recent development. Hemoglobin, for example, had previously been reported to undergo phase separation at high concentrations in vitro (Dumetz, Chockla, Kaler, & Lenhoff, 2008). Still, the significance of these observations remained unclear. Liquid droplet formation lowers the free energy of nucleation and is often a desired phenomenon in crystallization experiments (ten Wolde & Frenkel, 1997). However, the hypothesis that phase separation may mediate the formation of membraneless organelles to regulate biological functions and activities has emerged only recently. As strong support for this idea, data showed that proteins and RNA-containing bodies could be reconstituted from purified components; they further provided evidence that these reconstituted liquid

bodies can promote the nucleation of actin polymers (P. Li et al., 2012). Membraneless organelles are known more generally as biomolecular condensates, and the constituent biomolecules obey the same physical principles as other polymers.

One force that drives the liquid-liquid phase separation is multivalent interaction. Many proteins and RNA molecules in the cell can exhibit multivalent interactions with each other. On the molecular level, such interactions can be mediated by electrostatic attraction between charged residues, dipoles, or aromatic groups, which are also present on the nucleic acid and proteins (Erdel & Rippe, 2018). Multivalent interactions make simultaneous contacts among several, promoting the formation of supramolecular clusters. Suppose the interactions can avoid the irreversible aggregation and stronger comparing to the interactions with solvent molecules. In that case, these clusters will undergo liquid-liquid phase separation (LLPS): a dense liquid phase is formed and coexists with a more dilute liquid phase (Berry, Brangwynne, & Haataja, 2018; Michieletto et al., 2018) (15–17,65). As a result, certain molecules become enriched or depleted in one of the two phases, thereby facilitating or inhibiting particular biochemical reactions (Hyman, Weber, & Julicher, 2014). Nuage is enriched with proteins and RNAs, the components of nuage such as Krimper, Vasa providing an interaction network for the multivalent interactions with other piRNA factors. These observations suggest phase separation may govern its assemblies.

## References

- Abrusan, G., Grundmann, N., DeMester, L., & Makalowski, W. (2009). TEclass--a tool for automated classification of unknown eukaryotic transposable elements. *Bioinformatics*, *25*(10), 1329-1330. doi:10.1093/bioinformatics/btp084
- Amikura, R., Hanyu, K., Kashikawa, M., & Kobayashi, S. (2001). Tudor protein is essential for the localization of mitochondrial RNAs in polar granules of *Drosophila* embryos. *Mechanisms of Development*, *107*(1-2), 97-104. doi:Doi 10.1016/S0925-4773(01)00455-5
- Aminetzach, Y. T., Macpherson, J. M., & Petrov, D. A. (2005). Pesticide resistance via transposition-mediated adaptive gene truncation in *Drosophila*. *Science*, *309*(5735), 764-767. doi:10.1126/science.1112699
- Andrei, M. A., Ingelfinger, D., Heintzmann, R., Achsel, T., Rivera-Pomar, R., & Luhrmann, R. (2005). A role for eIF4E and eIF4E-transporter in targeting mRNPs to mammalian processing bodies. *Rna-a Publication of the Rna Society*, *11*(5), 717-727. doi:10.1261/rna.2340405
- Aravin, A. A., Hannon, G. J., & Brennecke, J. (2007). The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science*, *318*(5851), 761-764. doi:10.1126/science.1146484
- Aravin, A. A., van der Heijden, G. W., Castaneda, J., Vagin, V. V., Hannon, G. J., & Bortvin, A. (2009). Cytoplasmic compartmentalization of the fetal piRNA pathway in mice. *PLoS Genet*, *5*(12), e1000764. doi:10.1371/journal.pgen.1000764
- Berry, J., Brangwynne, C. P., & Haataja, M. (2018). Physical principles of intracellular organization via active and passive phase transitions. *Rep Prog Phys*, *81*(4), 046601. doi:10.1088/1361-6633/aaa61e
- Blumenstiel, J. P. (2011). Evolutionary dynamics of transposable elements in a small RNA world. *Trends in Genetics*, *27*(1), 23-31. doi:10.1016/j.tig.2010.10.003
- Boeynaems, S., Alberti, S., Fawzi, N. L., Mittag, T., Polymenidou, M., Rousseau, F., . . . Fuxreiter, M. (2018). Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol*, *28*(6), 420-435. doi:10.1016/j.tcb.2018.02.004
- Boswell, R. E., & Mahowald, A. P. (1985). Tudor, a Gene Required for Assembly of the Germ Plasm in *Drosophila-Melanogaster*. *Cell*, *43*(1), 97-104. doi:Doi 10.1016/0092-8674(85)90015-7
- Botuyan, M. V., & Mer, G. (2016). Tudor Domains as Methyl-Lysine and Methyl-Arginine Readers. *Chromatin Signaling and Diseases*, 149-165. doi:10.1016/B978-0-12-802389-1.00008-3
- Brangwynne, C. P., Eckmann, C. R., Courson, D. S., Rybarska, A., Hoege, C., Gharakhani, J., . . . Hyman, A. A. (2009). Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science*, *324*(5935), 1729-1732. doi:10.1126/science.1172046
- Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., & Hannon, G. J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell*, *128*(6), 1089-1103. doi:10.1016/j.cell.2007.01.043

- Brennecke, J., Malone, C. D., Aravin, A. A., Sachidanandam, R., Stark, A., & Hannon, G. J. (2008). An Epigenetic Role for Maternally Inherited piRNAs in Transposon Silencing. *Science*, *322*(5906), 1387-1392. doi:10.1126/science.1165171
- Britten, R. J., & Kohne, D. E. (1968). Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science*, *161*(3841), 529-540. doi:10.1126/science.161.3841.529
- Brown, P. O., Bowerman, B., Varmus, H. E., & Bishop, J. M. (1987). Correct integration of retroviral DNA in vitro. *Cell*, *49*(3), 347-356. doi:10.1016/0092-8674(87)90287-x
- Bucher, E., Reinders, J., & Mirouze, M. (2012). Epigenetic control of transposon transcription and mobility in Arabidopsis. *Curr Opin Plant Biol*, *15*(5), 503-510. doi:10.1016/j.pbi.2012.08.006
- Carthew, R. W., & Sontheimer, E. J. (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell*, *136*(4), 642-655. doi:10.1016/j.cell.2009.01.035
- Casacuberta, E., & Gonzalez, J. (2013). The impact of transposable elements in environmental adaptation. *Mol Ecol*, *22*(6), 1503-1517. doi:10.1111/mec.12170
- Charlesworth, B., & Charlesworth, D. (1983). The Population-Dynamics of Transposable Elements. *Genetics Research*, *42*(1), 1-27. doi:10.1017/S0016672300021455
- Charlesworth, B., & Langley, C. H. (1989). The Population-Genetics of Drosophila Transposable Elements. *Annual Review of Genetics*, *23*, 251-287. doi:10.1146/annurev.ge.23.120189.001343
- Cote, J., & Richard, S. (2005). Tudor domains bind symmetrical dimethylated arginines. *Journal of Biological Chemistry*, *280*(31), 28476-28483. doi:10.1074/jbc.M414328200
- Cox, D. N., Chao, A., Baker, J., Chang, L., Qiao, D., & Lin, H. (1998). A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes Dev*, *12*(23), 3715-3727. doi:10.1101/gad.12.23.3715
- De Fazio, S., Bartonicek, N., Di Giacomo, M., Abreu-Goodger, C., Sankar, A., Funaya, C., . . . O'Carroll, D. (2011). The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature*, *480*(7376), 259-U148. doi:10.1038/nature10547
- Desset, S., Meignin, C., Dastugue, B., & Vaury, C. (2003). COM, a heterochromatic locus governing the control of independent endogenous retroviruses from *Drosophila melanogaster*. *Genetics*, *164*(2), 501-509. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/12807771>
- Dombroski, B. A., Feng, Q., Mathias, S. L., Sassaman, D. M., Scott, A. F., Kazazian, H. H., Jr., & Boeke, J. D. (1994). An in vivo assay for the reverse transcriptase of human retrotransposon L1 in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*, *14*(7), 4485-4492. doi:10.1128/mcb.14.7.4485
- Doolittle, W. F., & Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. *Nature*, *284*(5757), 601-603. doi:10.1038/284601a0
- Drost, H. G., & Sanchez, D. H. (2019). Becoming a Selfish Clan: Recombination Associated to Reverse-Transcription in LTR Retrotransposons. *Genome Biol Evol*, *11*(12), 3382-3392. doi:10.1093/gbe/evz255

- Dumetz, A. C., Chockla, A. M., Kaler, E. W., & Lenhoff, A. M. (2008). Protein phase behavior in aqueous solutions: crystallization, liquid-liquid phase separation, gels, and aggregates. *Biophys J*, *94*(2), 570-583. doi:10.1529/biophysj.107.116152
- Eddy, E. M. (1974). Fine structural observations on the form and distribution of nuage in germ cells of the rat. *Anat Rec*, *178*(4), 731-757. doi:10.1002/ar.1091780406
- Eddy, E. M. (1975). Germ plasm and the differentiation of the germ cell line. *Int Rev Cytol*, *43*, 229-280. doi:10.1016/s0074-7696(08)60070-4
- Erdel, F., & Rippe, K. (2018). Formation of Chromatin Subcompartments by Phase Separation. *Biophysical Journal*, *114*(10), 2262-2270. doi:10.1016/j.bpj.2018.03.011
- Feschotte, C., & Pritham, E. J. (2007). DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet*, *41*, 331-368. doi:10.1146/annurev.genet.40.110405.090448
- Findley, S. D., Tamanaha, M., Clegg, N. J., & Ruohola-Baker, H. (2003). Maelstrom, a Drosophila spindle-class gene, encodes a protein that colocalizes with Vasa and RDE1/AGO1 homolog, Aubergine, in nuage. *Development*, *130*(5), 859-871. doi:10.1242/dev.00310
- Ghildiyal, M., & Zamore, P. D. (2009). Small silencing RNAs: an expanding universe. *Nat Rev Genet*, *10*(2), 94-108. doi:10.1038/nrg2504
- Gillespie, D. E., & Berg, C. A. (1995). Homeless Is Required for Rna Localization in Drosophila Oogenesis and Encodes a New Member of the De-H Family of Rna-Dependent ATPases. *Genes & Development*, *9*(20), 2495-2508. doi:DOI 10.1101/gad.9.20.2495
- Girard, A., Sachidanandam, R., Hannon, G. J., & Carmell, M. A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature*, *442*(7099), 199-202. doi:10.1038/nature04917
- Gunawardane, L. S., Saito, K., Nishida, K. M., Miyoshi, K., Kawamura, Y., Nagami, T., . . . Siomi, M. C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. *Science*, *315*(5818), 1587-1590. doi:10.1126/science.1140494
- Han, B. W., Wang, W., Li, C., Weng, Z., & Zamore, P. D. (2015). Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. *Science*, *348*(6236), 817-821. doi:10.1126/science.aaa1264
- Handler, D., Olivieri, D., Novatchkova, M., Gruber, F. S., Meixner, K., Mechtler, K., . . . 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 Journal*, *30*(19), 3977-3993. doi:10.1038/emboj.2011.308
- Handwerker, K. E., Cordero, J. A., & Gall, J. G. (2005). Cajal bodies, nucleoli, and speckles in the Xenopus oocyte nucleus have a low-density, sponge-like structure. *Molecular Biology of the Cell*, *16*(1), 202-211. doi:10.1091/mbc.E04-08-0742
- Harris, A. N., & Macdonald, P. M. (2001). Aubergine encodes a Drosophila polar granule component required for pole cell formation and related to eIF2C. *Development*, *128*(14), 2823-2832. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/11526087>

- Hedges, D. J., & Deininger, P. L. (2007). Inviting instability: Transposable elements, double-strand breaks, and the maintenance of genome integrity. *Mutat Res*, *616*(1-2), 46-59. doi:10.1016/j.mrfmmm.2006.11.021
- Huang, X. A., Yin, H., Sweeney, S., Raha, D., Snyder, M., & Lin, H. (2013). A major epigenetic programming mechanism guided by piRNAs. *Developmental Cell*, *24*(5), 502-516. doi:10.1016/j.devcel.2013.01.023
- Hyman, A. A., Weber, C. A., & Julicher, F. (2014). Liquid-liquid phase separation in biology. *Annu Rev Cell Dev Biol*, *30*, 39-58. doi:10.1146/annurev-cellbio-100913-013325
- Jurka, J., & Smith, T. (1988). A fundamental division in the Alu family of repeated sequences. *Proc Natl Acad Sci U S A*, *85*(13), 4775-4778. doi:10.1073/pnas.85.13.4775
- Kaaij, L. J., Hoogstrate, S. W., Berezikov, E., & Ketting, R. F. (2013). piRNA dynamics in divergent zebrafish strains reveal long-lasting maternal influence on zygotic piRNA profiles. *Rna-a Publication of the Rna Society*, *19*(3), 345-356. doi:10.1261/rna.036400.112
- Kapitonov, V. V., & Jurka, J. (2008). A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat Rev Genet*, *9*(5), 411-412; author reply 414. doi:10.1038/nrg2165-c1
- Katsuma, S., Kawamoto, M., & Kiuchi, T. (2014). Guardian small RNAs and sex determination. *RNA Biol*, *11*(10), 1238-1242. doi:10.1080/15476286.2014.996060
- Khurana, J. S., & Theurkauf, W. (2010). piRNAs, transposon silencing, and Drosophila germline development. *Journal of Cell Biology*, *191*(5), 905-913. doi:10.1083/jcb.201006034
- Khurana, J. S., Wang, J., Xu, J., Koppetsch, B. S., Thomson, T. C., Nowosielska, A., . . . Theurkauf, W. E. (2011). Adaptation to P Element Transposon Invasion in Drosophila melanogaster. *Cell*, *147*(7), 1551-1563. doi:10.1016/j.cell.2011.11.042
- Kidwell, M. G. (1983). Evolution of hybrid dysgenesis determinants in Drosophila melanogaster. *Proc Natl Acad Sci U S A*, *80*(6), 1655-1659. doi:10.1073/pnas.80.6.1655
- Kirino, Y., Vourekas, A., Sayed, N., de Lima Alves, F., Thomson, T., Lasko, P., . . . Mourelatos, Z. (2010). Arginine methylation of Aubergine mediates Tudor binding and germ plasm localization. *Rna-a Publication of the Rna Society*, *16*(1), 70-78. doi:10.1261/rna.1869710
- Klenov, M. S., Sokolova, O. A., Yakushev, E. Y., Stolyarenko, A. D., Mikhaleva, E. A., Lavrov, S. A., & Gvozdev, V. A. (2011). Separation of stem cell maintenance and transposon silencing functions of Piwi protein. *Proc Natl Acad Sci U S A*, *108*(46), 18760-18765. doi:10.1073/pnas.1106676108
- Kofler, R., Hill, T., Nolte, V., Betancourt, A. J., & Schlotterer, C. (2015). The recent invasion of natural Drosophila simulans populations by the P-element. *Proc Natl Acad Sci U S A*, *112*(21), 6659-6663. doi:10.1073/pnas.1500758112
- Lambowitz, A. M., & Zimmerly, S. (2011). Group II Introns: Mobile Ribozymes that Invade DNA. *Cold Spring Harbor Perspectives in Biology*, *3*(8). doi:ARTN a003616  
10.1101/cshperspect.a003616

- Le Rouzic, A., & Deceliere, G. (2005). Models of the population genetics of transposable elements. *Genet Res*, 85(3), 171-181. doi:10.1017/S0016672305007585
- Le Thomas, A., Rogers, A. K., Webster, A., Marinov, G. K., Liao, S. E., Perkins, E. M., . . . Toth, K. F. (2013). Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev*, 27(4), 390-399. doi:10.1101/gad.209841.112
- Lee, Y. C. G., & Langley, C. H. (2010). Transposable elements in natural populations of *Drosophila melanogaster*. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 365(1544), 1219-1228. doi:10.1098/rstb.2009.0318
- Li, C., Vagin, V. V., Lee, S., Xu, J., Ma, S., Xi, H., . . . Zamore, P. D. (2009). Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell*, 137(3), 509-521. doi:10.1016/j.cell.2009.04.027
- Li, P., Banjade, S., Cheng, H. C., Kim, S., Chen, B., Guo, L., . . . Rosen, M. K. (2012). Phase transitions in the assembly of multivalent signalling proteins. *Nature*, 483(7389), 336-340. doi:10.1038/nature10879
- Liang, L., Diehljones, W., & Lasko, P. (1994). Localization of Vasa Protein to the *Drosophila* Pole Plasm Is Independent of Its Rna-Binding and Helicase Activities. *Development*, 120(5), 1201-1211. Retrieved from <Go to ISI>://WOS:A1994NM70200015
- Lim, A. K., & 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(16), 6714-6719. doi:10.1073/pnas.0701920104
- Lin, H., & Spradling, A. C. (1997). A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the *Drosophila* ovary. *Development*, 124(12), 2463-2476. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/9199372>
- Lu, J., & Clark, A. G. (2010). Population dynamics of PIWI-interacting RNAs (piRNAs) and their targets in *Drosophila*. *Genome Research*, 20(2), 212-227. doi:10.1101/gr.095406.109
- Luan, D. D., Korman, M. H., Jakubczak, J. L., & Eickbush, T. H. (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell*, 72(4), 595-605. doi:10.1016/0092-8674(93)90078-5
- Ma, J. B., Ye, K., & Patel, D. J. (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature*, 429(6989), 318-322. doi:10.1038/nature02519
- Ma, J. B., Yuan, Y. R., Meister, G., Pei, Y., Tuschl, T., & Patel, D. J. (2005). Structural basis for 5'-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. *Nature*, 434(7033), 666-670. doi:10.1038/nature03514
- Mahowald, A. P. (1968). Polar granules of *Drosophila*. II. Ultrastructural changes during early embryogenesis. *J Exp Zool*, 167(2), 237-261. doi:10.1002/jez.1401670211
- Mahowald, A. P. (1971). Polar granules of *Drosophila*. 3. The continuity of polar granules during the life cycle of *Drosophila*. *J Exp Zool*, 176(3), 329-343. doi:10.1002/jez.1401760308
- Malone, C. D., Brennecke, J., Dus, M., Stark, A., McCombie, W. R., Sachidanandam, R., & Hannon, G. J. (2009). Specialized piRNA pathways act in germline and somatic

- tissues of the *Drosophila* ovary. *Cell*, *137*(3), 522-535. doi:10.1016/j.cell.2009.03.040
- Malone, C. D., & Hannon, G. J. (2009). Small RNAs as Guardians of the Genome. *Cell*, *136*(4), 656-668. doi:10.1016/j.cell.2009.01.045
- Matranga, C., & Zamore, P. D. (2007). Small silencing RNAs. *Current Biology*, *17*(18), R789-R793. doi:DOI 10.1016/j.cub.2007.07.014
- Maurer-Stroh, S., Dickens, N. J., Hughes-Davies, L., Kouzarides, T., Eisenhaber, F., & Ponting, C. P. (2003). The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends in Biochemical Sciences*, *28*(2), 69-74. doi:10.1016/S0968-0004(03)00004-5
- Michieletto, D., Chiang, M., Coli, D., Papantonis, A., Orlandini, E., Cook, P. R., & Marenduzzo, D. (2018). Shaping epigenetic memory via genomic bookmarking. *Nucleic Acids Research*, *46*(1), 83-93. doi:10.1093/nar/gkx1200
- Mohn, F., Handler, D., & Brennecke, J. (2015). Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. *Science*, *348*(6236), 812-817. doi:10.1126/science.aaa1039
- Montchamp-Moreau, C. (1990). Dynamics of P-M Hybrid Dysgenesis in P-Transformed Lines of *Drosophila Simulans*. *Evolution*, *44*(1), 194-203. doi:10.1111/j.1558-5646.1990.tb04289.x
- Moon, S., Cassani, M., Lin, Y. A., Wang, L., Dou, K., & Zhang, Z. Z. (2018). A Robust Transposon-Endogenizing Response from Germline Stem Cells. *Developmental Cell*, *47*(5), 660-671 e663. doi:10.1016/j.devcel.2018.10.011
- Munoz-Lopez, M., & Garcia-Perez, J. L. (2010). DNA transposons: nature and applications in genomics. *Curr Genomics*, *11*(2), 115-128. doi:10.2174/138920210790886871
- Nishida, K. M., Okada, T. N., Kawamura, T., Mituyama, T., Kawamura, Y., Inagaki, S., . . . Siomi, M. C. (2009). Functional involvement of Tudor and dPRMT5 in the piRNA processing pathway in *Drosophila* germlines. *EMBO J*, *28*(24), 3820-3831. doi:10.1038/emboj.2009.365
- O'Donnell, K. A., & Boeke, J. D. (2007). Mighty Piwis defend the germline against genome intruders. *Cell*, *129*(1), 37-44. doi:10.1016/j.cell.2007.03.028
- Orgel, L. E., & Crick, F. H. (1980). Selfish DNA: the ultimate parasite. *Nature*, *284*(5757), 604-607. doi:10.1038/284604a0
- Patel, A., Lee, H. O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M. Y., . . . Alberti, S. (2015). A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell*, *162*(5), 1066-1077. doi:10.1016/j.cell.2015.07.047
- Patil, V. S., & Kai, T. (2010). Repression of retroelements in *Drosophila* germline via piRNA pathway by the Tudor domain protein Tejas. *Current Biology*, *20*(8), 724-730. doi:10.1016/j.cub.2010.02.046
- Pelisson, A., Song, S. U., Prud'homme, N., Smith, P. A., Bucheton, A., & Corces, V. G. (1994). Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the *Drosophila flamenco* gene. *EMBO J*, *13*(18), 4401-4411. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/7925283>

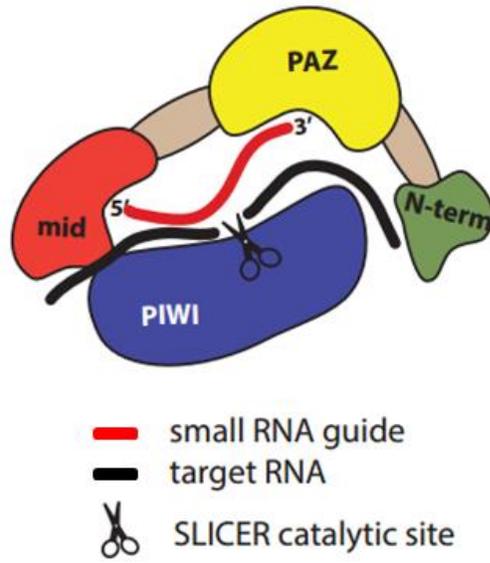
- Petrov, D. A., Fiston-Lavier, A. S., Lipatov, M., Lenkov, K., & Gonzalez, J. (2011). Population genomics of transposable elements in *Drosophila melanogaster*. *Mol Biol Evol*, 28(5), 1633-1644. doi:10.1093/molbev/msq337
- Poyhonen, M., de Vanssay, A., Delmarre, V., Hermant, C., Todeschini, A. L., Teysset, L., & Ronsseray, S. (2012). Homology-dependent silencing by an exogenous sequence in the *Drosophila* germline. *G3 (Bethesda)*, 2(3), 331-338. doi:10.1534/g3.111.001925
- Pratt, A. J., & MacRae, I. J. (2009). The RNA-induced silencing complex: a versatile gene-silencing machine. *Journal of Biological Chemistry*, 284(27), 17897-17901. doi:10.1074/jbc.R900012200
- Rozhkov, N. V., Schostak, N. G., Zelentsova, E. S., Yushenova, I. A., Zatsepina, O. G., & Evgen'ev, M. B. (2013a). Evolution and dynamics of small RNA response to a retroelement invasion in *Drosophila*. *Mol Biol Evol*, 30(2), 397-408. doi:10.1093/molbev/mss241
- Rozhkov, N. V., Schostak, N. G., Zelentsova, E. S., Yushenova, I. A., Zatsepina, O. G., & Evgen'ev, M. B. (2013b). Evolution and Dynamics of Small RNA Response to a Retroelement Invasion in *Drosophila*. *Molecular Biology and Evolution*, 30(2), 397-408. doi:10.1093/molbev/mss241
- Ruby, J. G., Jan, C., Player, C., Axtell, M. J., Lee, W., Nusbaum, C., . . . Bartel, D. P. (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell*, 127(6), 1193-1207. doi:10.1016/j.cell.2006.10.040
- Russell, A. L., & Woodruff, R. C. (1999). The genetics and evolution of the mariner transposable element in *Drosophila simulans*: worldwide distribution and experimental population dynamics. *Genetica*, 105(2), 149-164. doi:10.1023/A:1003760123429
- Saffman, E. E., & Lasko, P. (1999). Germline development in vertebrates and invertebrates. *Cell Mol Life Sci*, 55(8-9), 1141-1163. doi:10.1007/s000180050363
- Saito, K., Inagaki, S., Mituyama, T., Kawamura, Y., Ono, Y., Sakota, E., . . . Siomi, M. C. (2009). A regulatory circuit for piwi by the large Maf gene traffic jam in *Drosophila*. *Nature*, 461(7268), 1296-1299. doi:10.1038/nature08501
- Saito, K., Nishida, K. M., Mori, T., Kawamura, Y., Miyoshi, K., Nagami, T., . . . Siomi, M. C. (2006). Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev*, 20(16), 2214-2222. doi:10.1101/gad.1454806
- Saito, K., & Siomi, M. C. (2010). Small RNA-Mediated Quiescence of Transposable Elements in Animals. *Developmental Cell*, 19(5), 687-697. doi:10.1016/j.devcel.2010.10.011

## Figures and Figure Legends

### **Figure 1. The Domain structure of PIWI-clade Argonaute proteins**

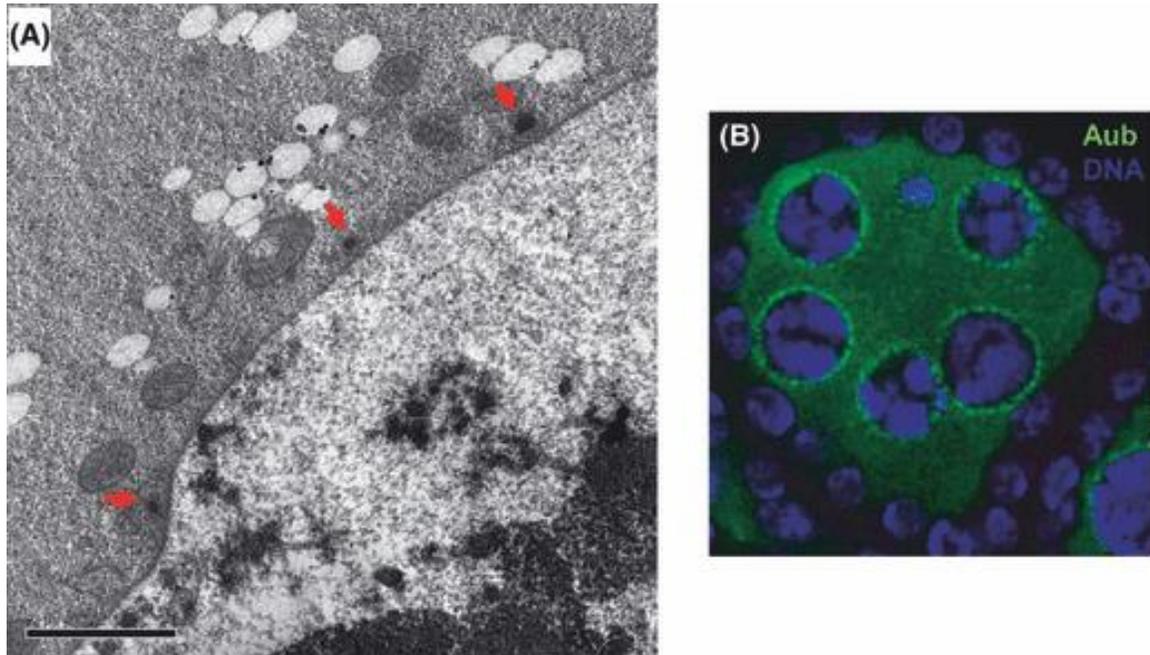
The domains of PIWI-clade Argonaute proteins in *Drosophila* share similar structures and orders to other Argonaute proteins. A disordered N-terminal region whose structural remains unknown, PAZ domain holds the 3' end of the piRNA PAZ, MID domain anchors piRNA at its 5' end , PIWI is the largest domain, and its catalytic site responsible for 'slicer' activity is positioned to cleave the backbone of annealed target RNA exactly 10 nt relative to the 5' end of the piRNA.

Figure 1



**Figure 2 The Nuage in *Drosophila* ovaries**

(A) An electron micrograph showing the perinuclear electron-dense nuage structures (arrows) on the cytoplasmic face of the nuclear envelope in *Drosophila* ovaries. (B) Immunostaining using an antibody against Aubergine to visualize the perinuclear nuage (green) in the *Drosophila* egg chamber. The DNA (blue) is labeled by 4'6'-diamidino-2-phenylindole dihydrochloride (DAPI).

**Figure 2**

Development, Growth & Differentiation, Volume: 54, Issue: 1, Pages: 66-77, First published: 16 January 2012, DOI: (10.1111/j.1440-169X.2011.01316.x)

*Chapter 2*

## PIRNA BIOGENESIS IN DROSOPHILA MELANOGASTER

Xiawei Huang<sup>1</sup>, Katalin Fejes Toth<sup>1&</sup> and Alexei A. Aravin<sup>1&</sup>

<sup>1</sup>California Institute of Technology,  
Division of Biology and Biological Engineering, 147-75  
1200 E. California Blvd.,  
Pasadena, CA 91125, USA

<sup>&</sup>To whom correspondence should be addressed:

Katalin Fejes Toth [kft@caltech.edu](mailto:kft@caltech.edu)

Alexei Aravin [aaa@caltech.edu](mailto:aaa@caltech.edu)

This chapter was published in: *Trends in Genetics* 33.11 (2017): 882-894

## **Abstract**

The PIWI-interacting RNA (piRNA) pathway is a conserved defense system that protects the genome integrity of animal germline from deleterious transposable elements. Targets of silencing are recognized by small non-coding piRNAs that are processed from long precursor molecules. Though piRNA and other classes of small non-coding RNAs, such as miRNA and siRNAs, interact with members of the same family of Argonaute proteins and their function in targets repression is similar, the biogenesis of piRNA differs from those of the other two small RNA. Recently, many aspects of piRNA biogenesis have been revealed in *Drosophila melanogaster*. In this review, we elaborate on piRNA biogenesis in *Drosophila* somatic and germline cells. We focus on the mechanisms by which piRNA precursor transcription is regulated and highlight recent work that advanced our understanding of piRNA precursor processing to mature piRNAs. We finish with discussing current models to the still unresolved question of how piRNA precursors are selected and channeled into the processing machinery.

## **Introduction**

The conserved family of Argonaute proteins interacts with small (19-33 nt) nucleic acid (single-stranded DNA or RNA) guides in eukaryotic and prokaryotic species. The guide enables the Ago complex to recognize its DNA or RNA targets with high level of specificity using complementary interactions. After recognition of the target it is cleaved by the intrinsic endonuclease activity of the Argonaute protein (J. B. Ma et al., 2005; Meister, 2013; Rivas et al., 2005). However, not all Agos are active nucleases and incomplete complementarity between the guide and the target can impair cleavage even by nuclease-competent

Argonautes. Suppression of the targets can be achieved without cleavage through the recruitment of additional effector proteins by Ago (Kim, Han, & Siomi, 2009).

While in some prokaryotic species Argonautes use small single-stranded DNA guides (J. B. Ma et al., 2005; Song, Smith, Hannon, & Joshua-Tor, 2004), eukaryotic Agos partner exclusively with small RNAs (Hur, Olovnikov, & Aravin, 2014; Swarts et al., 2014). There are three major classes of Ago-associated small RNAs in Metazoa: small interfering (si)RNA, microRNA and piwi-interacting (pi)RNA. siRNA were first discovered during studies of transgene-induced gene silencing in plants (Hamilton & Baulcombe, 1999), followed by RNAi experiments in *C. elegans* where introduction of long double-stranded (ds)RNA induced potent silencing of homologous endogenous genes (Fire et al., 1998). Soon it was discovered that the phenomenon of RNAi is conserved in diverse eukaryotic species and is mediated by siRNA, which are processed from long dsRNA by the endonuclease Dicer and subsequently loaded into Ago proteins (Bernstein, Caudy, Hammond, & Hannon, 2001). Dicer also cleaves endogenous miRNA precursors in the cytoplasm (Lund & Dahlberg, 2006), following processing by the nuclear nuclease Drosha (Lee et al., 2003).

In contrast to the ubiquitously expressed miRNA, piRNA and their protein partners, members of the Piwi clade of the Argonaute protein family, are predominantly expressed in gonads of Metazoa (A. Aravin et al., 2006; Carmell, Xuan, Zhang, & Hannon, 2002; Cox et al., 1998; Cox, Chao, & Lin, 2000; Girard, Sachidanandam, Hannon, & Carmell, 2006; Grivna, Beyret, Wang, & Lin, 2006; Lau et al., 2006; Peters & Meister, 2007; Watanabe et al., 2006). Accordingly, Piwi mutants in flies and mice – two model organisms that were extensively used to understand this pathway – have normal somatic development, but show gametogenesis defects which result in sterility (Carmell et al., 2007; Cox et al., 1998; Girard

et al., 2006; Kuramochi-Miyagawa et al., 2008; Thomson & Lin, 2009). In both organisms, piRNA and Piwi proteins are required in germ cells for suppression of transposable elements, selfish genomic elements that are able to move in the genome (A. A. Aravin et al., 2001; Brennecke et al., 2007; Gunawardane et al., 2007; Kalmykova, Klenov, & Gvozdev, 2005; C. Li et al., 2009; Nishida et al., 2007; Vagin et al., 2004; Vagin et al., 2006). The gametogenesis defects observed in Piwi mutants are likely to be the direct result of transposon activation and the associated double-stranded DNA breaks and concomitant activation of DNA damage checkpoint (Y. Chen, Pane, & Schupbach, 2007; Klattenhoff et al., 2007; Klattenhoff & Theurkauf, 2008).

While only a few hundred distinct miRNAs are encoded in genomes of different Metazoa, piRNAs have an amazing sequence diversity: deep sequencing of small RNA identified millions of unique piRNA reads that do not have much in common except for a bias in the first nucleotide at 5' end (A. Aravin et al., 2006; A. A. Aravin et al., 2008; Brennecke et al., 2007; Lau et al., 2006). Attempts to classify piRNA using the approach that was used for miRNA – giving every piRNA sequence its unique ID – were soon abandoned. The diversity of piRNA reflects the difference in their biogenesis from that of both miRNA and siRNA. All three types of small RNA are processed from longer precursors, however, precursors for miRNAs and siRNAs have double-stranded structures, which are recognized by Dicer, while piRNA precursors are single-stranded. piRNA biogenesis is independent of Dicer and involves a set of proteins that are unique to this pathway (Czech & Hannon, 2016b; Gunawardane et al., 2007; Iwasaki, Siomi, & Siomi, 2015; Kalmykova et al., 2005; Vagin et al., 2004; Vagin et al., 2006). piRNA biogenesis can be divided into two stages. First, long RNA precursors (pre-piRNAs) are transcribed in the nucleus and exported into cytoplasm.

In the cytoplasm, pre-piRNAs are further processed to generate mature piRNAs that get loaded into Piwi proteins.

### **Nuclear steps of piRNA biogenesis**

To understand biogenesis of piRNAs, it is necessary to explore their genomic origin. Since the majority of piRNA are derived from sequences of transposable elements that are present in many copies throughout the genome, this task is not as simple as it seems. The majority of piRNA can't be mapped to a unique position in the fly genome, making conclusions about their origin ambiguous (Brennecke et al., 2007). However, when only piRNA that can be uniquely mapped are considered, it is apparent that the majority of piRNA originate from a number of extended genomic loci, with sizes up to 200 kb in length. Such genomic regions were named piRNA clusters (A. Aravin et al., 2006; Brennecke et al., 2007; Girard et al., 2006; Lau et al., 2006). Each piRNA cluster produces thousands of piRNA sequences that are not arranged in any particular pattern and can even overlap with each other. Major piRNA clusters do not overlap with protein-coding genes but are comprised of a diverse set of TE fragments. Individual transposons – inserted in gene-dense euchromatic areas – also generate piRNA (Brennecke et al., 2007; Saito et al., 2006), but the precise quantification of the relative fractions of piRNA derived from individual transposons versus piRNA clusters is impossible due to the significant fraction of piRNA that map to both. It is worth mentioning that in addition to individual transposon insertions and piRNA clusters, some protein-coding genes also produce piRNA, predominantly from their 3'UTR region (Hirano et al., 2014; Robine et al., 2009; Saito et al., 2009).

Independently of the genomic origin – piRNA clusters, individual transposons or genes – mature (23-30nt) piRNA are processed from longer precursors termed pre-piRNA. Genic piRNA correspond to exonic sequences and are in sense orientation relative to gene's mRNA (Robine et al., 2009; Saito et al., 2009), suggesting that they are likely processed from spliced and processed mRNA. Currently it is not clear if mRNAs destined for piRNA processing vs. translation have specific marks that define their fate. Enrichment of piRNA at the 3'UTRs suggests that the same mRNA can be used for both translation and piRNA processing and that active translation interferes with piRNA biogenesis.

piRNA clusters are transcribed as long non-coding RNAs by RNA polymerase II (Goriaux, Desset, Renaud, Vaury, & Brassat, 2014; X. Z. G. Li et al., 2013; Mohn, Sienski, Handler, & Brennecke, 2014). Despite this fact, chromatin of piRNA clusters is enriched in the histone 3 lysine 9 tri-methylation (H3K9me3) mark, which is usually found on silent, heterochromatic regions and is thought to be a repressive mark that suppresses transcription (Klenov et al., 2014; Le Thomas et al., 2013; Mohn et al., 2014; Rangan et al., 2011; Rozhkov, Hammell, & Hannon, 2013; Sienski, Donertas, & Brennecke, 2012; Z. Zhang, Wang, et al., 2014). Surprisingly, presence of the H3K9me3 mark does not interfere with pre-piRNA transcription but is in fact required for piRNA expression (Le Thomas, Stuwe, et al., 2014; Mohn et al., 2014; Rangan et al., 2011). Depletion of one of the enzymes that installs the H3K9me3 mark, SetDB1/Egg, leads to decrease in pre-piRNA expression (Rangan et al., 2011). The level of H3K9me3 signal also positively correlates with piRNA generation from clusters that are differentially expressed between two *D. virilis* strains (Le Thomas, Marinov, & Aravin, 2014).

Two types of clusters were described in flies: uni-strand, for which the vast majority of piRNA map to one genomic strand, and dual-strand, for which piRNA map to both genomic strands. The differences between the two cluster types extend far beyond the strand of transcription (Figure 1A). The transcription of uni-strand clusters seems to be similar to canonical mRNA transcription as these clusters have unique promoters and produce 5' capped and polyadenylated RNA that are sometime spliced (Goriaux et al., 2014; Mohn et al., 2014; Zanni et al., 2013). Dual-strand clusters produce non-polyadenylated RNA and the majority of dual-strand clusters does not have clear signatures of Pol II promoters, such as peaks of Pol II and H3K4me2 (Y. C. A. Chen et al., 2016; Le Thomas et al., 2013; Mohn et al., 2014). The absence of clear promoters led to the proposal that transcription of dual-strand clusters may be driven by the promoters of flanking protein-coding genes (Mohn et al., 2014). However, deletion of a putative promoter next to the major 42AB piRNA cluster did not affect piRNA production, arguing against this possibility (Y. C. A. Chen et al., 2016). It is possible that both transcription initiation and termination occur at multiple positions inside dual-strand clusters. As transcription of dual-strand clusters seems to be distinct from canonical Pol II transcription, study of dual-strand cluster transcription is not only important for a better understanding of piRNA biogenesis but will shed light on general mechanisms of Pol II transcriptional control.

The unique nature of dual-strand clusters became obvious when several proteins required for biogenesis of piRNA from these regions, but not from uni-strand clusters and genic piRNAs were identified (Y. Chen et al., 2007; Czech, Preall, McGinn, & Hannon, 2013; Klattenhoff et al., 2009; Pane, Jiang, Zhao, Singh, & Schupbach, 2011). Rhino was the first protein reported to be required for piRNA biogenesis exclusively from dual-strand

clusters (Klattenhoff et al., 2009). Rhino is a paralog of the well-characterized heterochromatin protein HP1, but in contrast to HP1 it is only expressed in the germline. Similar to HP1, Rhino also binds the H3K9me3 mark through its chromodomain (Le Thomas, Stuwe, et al., 2014; Mohn et al., 2014; Patel et al., 2015). Rhino is enriched on chromatin of dual-strand clusters and forms the RDC complex with Deadlock (Del) and Cutoff (Cuff), two other proteins that are required for biogenesis of piRNA from dual- but not uni-strand clusters (Y. C. A. Chen et al., 2016; Klattenhoff et al., 2009; Le Thomas, Stuwe, et al., 2014; Mohn et al., 2014; Z. Zhang, Wang, et al., 2014). The mechanism for specific enrichment of Rhino on dual-strand piRNA clusters and not other genomic regions with high level of H3K9me3 is currently unknown. Such specificity could be achieved if loading of RDC on its genomic target sites was guided by the nuclear Piwi protein. Piwi is responsible for installment of the H3K9me3 mark on genomic targets of the piRNA pathway, including piRNA clusters, while heterochromatic marks at other genomic regions that are not targeted by piRNAs are established in a Piwi-independent fashion (Le Thomas et al., 2013; Rozhkov et al., 2013; Sienski et al., 2012). It is attractive to speculate that recognition of piRNA targets by the Piwi/piRNA complex not only lead to deposition of the H3K9me3 mark but also recruits component(s) of the RDC complex to the locus; however, this hypothesis awaits experimental verification.

Two – not necessarily mutually exclusive - models were proposed for the role of the RDC complex in piRNA biogenesis (Figure 1B). According to one model, components of RDC suppress splicing of pre-piRNA, probably by competing with the nuclear cap binding complex CBC for binding mRNAs (Z. Zhang, Wang, et al., 2014). However, the direct effect of RDC on CBC binding was not tested and it is not clear how the absence of splicing might

be important for piRNA processing, as intronless mRNAs are not more prone to be processed into piRNA. According to the other model, RDC is necessary for transcription of pre-piRNAs. Transcription of pre-piRNA decreases notably upon known-down or mutation of RDC components (Y. C. A. Chen et al., 2016; Mohn et al., 2014). RDC and associated proteins could enhance transcription of pre-piRNA through several mechanisms, including suppression of termination and promotion of initiation. Cutoff suppresses termination at polyA sites inserted into dual-strand clusters and when recruited to heterologous reporter mRNAs (Y. C. A. Chen et al., 2016). Cuff suppresses termination by preventing cleavage by the CPSF complex at the polyA site and, in case cleavage does occur, Cuff prevents degradation of the cleaved 5'monophosphorylated RNA by the nuclear exonuclease Rat1/Xrn2. Since Cuff is homologous to the Rai1/Dxo protein, which has a binding pocket for the 5' end of RNA, Cuff might protect RNA degradation by preventing RNA recognition by Rat1. It was proposed that the anti-termination activity of Cuff allows dual-strand clusters to function as traps for transposon insertions: if transcription of pre-piRNA would be canonical, insertion of new TEs that carry their own polyA/termination signal would cause premature termination and collapse of piRNA biogenesis (Y. C. A. Chen et al., 2016).

As clusters are in genomic regions characterized by H3K9me3 enrichment and lack a canonical promoter an important question is how transcription of clusters initiates. Genetic screens have identified Moonshiner, a protein that interacts with core transcription machinery and therefore might play a role in transcription initiation in the absence of proper promoter elements (Andersen et al., in press). Moonshiner physically interacts with the RCD complex, which likely explains its recruitment to clusters and the consequential initiation of

transcription at multiple sites within H3K9me3-rich clusters, where the repressive mark would otherwise antagonize transcription initiation.

The transcription initiation complex, TREX, composed of multiple proteins including UAP56 and THO subunits, is also required for piRNA biogenesis from dual-strand, but not uni-strand clusters (Hur et al., 2016; F. Zhang et al., 2012; Z. Zhang, Wang, et al., 2014). In contrast to RDC, which is not conserved outside of Diptera and seems to be exclusively present at genomic regions that generate piRNAs, TREX is a conserved complex from yeast to humans and is co-transcriptionally loaded on many – if not all – nascent pre-mRNA transcripts transcribed by Pol II (Katahira, 2012; Masuda et al., 2005; Reed, 2003; Strasser et al., 2002). However, TREX seems to be enriched on pre-piRNA compared to other Pol II nascent transcripts. In yeasts and mammals TREX mediates nuclear export of RNA (Katahira, 2012; Masuda et al., 2005; Reed, 2003; Strasser et al., 2002), therefore it is plausible that TREX also promotes export of pre-piRNA, although this has not yet been tested. Additionally, in TREX mutants, transcription of pre-piRNA is reduced (Hur et al., 2016). This effect might be mediated by the proposed ability of TREX to prevent formation of R-loops, hybrids between nascent RNA and DNA that inhibit transcription. Alternatively, proteins associated with nascent RNAs such as TREX might enhance transcription by preventing backtracking of Pol II, a role played by co-transcriptionally translating ribosomes in prokaryotes (Nudler, 2012).

Loading of TREX – at least in mammals – is dependent on splicing, which raises the question how it can be enriched on cluster transcripts that are not spliced (Katahira, 2012; Masuda et al., 2005). The genetic and physical interaction between RDC and TREX and their co-localization in nuclear foci suggests that chromatin-bound RDC might recruit TREX on

nascent pre-piRNA transcripts (Figure 1B) (Hur et al., 2016). Tethering of Cuff to a reporter enhances loading of TREX on the nascent reporter RNA. Thus, the study of piRNA biogenesis revealed an alternative, splicing-independent, mechanism of TREX loading on RNA that depends on the H3K9me3 chromatin mark and the associated RDC complex. These results suggest that chromatin marks can reach beyond regulation of transcription and affect – through guiding the loading of a particular set of proteins onto RNA – the post-transcriptional fate of RNA. It is remarkable how many unexpected insights into the mechanisms of transcription and early RNA processing were gained through studies of the piRNA pathway. Beyond revealing the role of chromatin in loading of RNA-binding proteins on nascent RNAs, these studies uncovered novel mechanisms to control termination and induce transcription initiation in a hostile chromatin environment.

### **piRNA biogenesis in the cytoplasm**

Once long piRNA precursors – transcripts from piRNA clusters, transposons and mRNA of genes – are exported across the nuclear envelope, further processing in the cytoplasm leads to generation of mature 23-29nt piRNA. The enzymatic machinery that processes pre-piRNA to generate the 5' and 3' ends of mature piRNA is different from enzymes that process miRNA and siRNA (Dicer and Drosha) (Le Thomas, Toth, & Aravin, 2014; Vagin et al., 2006). It is likely that 5' end of piRNA is formed first and the product might even be loaded into Piwi proteins prior to 3'end processing.

Early work on piRNA biogenesis revealed that 5'-end formation of piRNAs can occur through two distinct mechanisms, the so-called primary biogenesis and ping-pong processing (A. A. Aravin, Hannon, & Brennecke, 2007; Brennecke et al., 2007; Gunawardane et al.,

2007) (Figure 2A). The ping-pong pathway requires the existence of a mature piRNA that is complementary to the pre-piRNA. When such piRNA is loaded in either Aub or Ago3 – the two cytoplasmic Piwi proteins – these proteins induce endonucleolytic (slicer) cleavage of complementary transcripts. Unlike the siRNA pathway where a similar ‘slicing’ event simply leads to target degradation, here the newly formed fragment – after further 3’ end processing – becomes a new piRNA. Products of Aub-guided cleavage are usually loaded into Ago3, while products of Ago3 cleavage are loaded in to Aub, hence the ping-pong name (Brennecke et al., 2007; Gunawardane et al., 2007). In addition to the heterotypic ping-pong between Aub and Ago3, homotypic Aub/Aub ping-pong also operates although with less efficiency, and thus only becoming apparent when Ago3 is absent (Huang et al., 2014; C. Li et al., 2009; Z. Zhang et al., 2011).

The ping-pong mechanism seems to be a conserved feature of the piRNA pathway present in many organisms from Hydra to human (A. A. Aravin, Sachidanandam, Girard, Fejes-Toth, & Hannon, 2007; Grimson et al., 2008; Houwing et al., 2007; R. S. M. Lim, Anand, Nishimiya-Fujisawa, Kobayashi, & Kai, 2014). This level of conservation likely reflects the crucial function of ping-pong in the pathway: it enables the amplification of piRNA that target actively expressed transposons. Upon recognition of the mRNA of active transposons by antisense piRNA processed from cluster pre-piRNA, ping-pong will work as a cycle to generate more piRNA from the cluster transcript to target the transposon. It was proposed that this mechanism, which can conceptually be compared to expansion of cells that produce antigen to pathogens during an immune response, allows for fine-tuning piRNA populations to fight active transposons (Brennecke et al., 2007). In agreement with this,

transposon activation leads to increase in ping-pong and generation of piRNA against the active transposon (A. A. Aravin et al., 2008).

While piRNA loaded into Aub and Ago3 demonstrate clear signatures of slicer-dependent processing, Piwi-loaded piRNA lack these signatures (Figure 2A). This observation and the need of initial ‘starter’ piRNA for the ping-pong cycle lead to the postulation of another mechanism, which generates ‘primary’ piRNA in a slicer-independent fashion (Brennecke et al., 2007; C. Li et al., 2009). Though the nuclease that generate piRNA in this pathway was not known at the time, later it was identified as Zucchini (Zuc), an endoribonuclease that is anchored to the outer mitochondrial surface. The role of Zuc in piRNA processing is supported by genetic experiments – piRNA loaded into Piwi are eliminated in Zuc mutant flies – as well as structural and biochemical studies that demonstrate the endoribonuclease activity of Zuc in vitro (Ipsaro, Haase, Knott, Joshua-Tor, & Hannon, 2012; Nishimasu et al., 2012; Pane, Wehr, & Schupbach, 2007; Voigt et al., 2012). Despite these findings some questions about Zuc processing remain unanswered. In vitro Zuc does not show any preference for RNA cleavage at any specific nucleotide residue (Ipsaro et al., 2012; Nishimasu et al., 2012), while the 5’ ends of slicer-independent piRNA show a strong bias for uridine. It is possible that the observed preference to cleave in front of uridine is determined by co-factor of Zuc. Alternatively, Zuc-mediated cleavage might be truly unspecific and the bias might be created by the selectivity of Piwi for 5’U RNA.

The mechanisms responsible for 3’ end formation are even more diverse than those that lead to generation of the 5’ end of piRNA (Figure 2A). The 3’ end can be generated through cleavage by Zuc (Han, Wang, Li, Weng, & Zamore, 2015; Mohn, Handler, & Brennecke, 2015). Slicer-independent piRNA exhibit a ‘phasing’ signature, i.e. when mapped to the

sequence of the precursor one piRNA is immediately followed by another piRNA suggesting that a single Zuc cleavage can simultaneously generate the 3' end of an upstream and the 5' end of a downstream piRNA (Czech & Hannon, 2016a; Han et al., 2015; Mohn et al., 2015; W. Wang et al., 2015). 3' end formation can also be induced by slicer (ping-pong) cleavage. In this case, the mature piRNA is generated by two closely spaced slicer cleavages (Hayashi et al., 2016). Finally, 3' end formation might require processing by exonuclease(s). In this case, a 3'-to-5' exonuclease trims a longer precursor formed by cleavage by an endonuclease (slicer or Zuc) to make mature piRNA of the correct size. Such 'Trimmer' activity was first observed in vitro using lysate from silkworm cell line (Kawaoka, Izumi, Katsuma, & Tomari, 2011) and later associated with the exonuclease Nibbler (Nbr) (Feltzin et al., 2015; Hayashi et al., 2016; H. Wang et al., 2016). Independently of the mechanism by which the 3' end of piRNA is generated, the last step is the 2'OMe-modification of the last nucleotide by Hen1, which is thought to increase the stability of piRNA (Horwich et al., 2007; Saito et al., 2007; H. Wang et al., 2016).

Slicer-dependent (ping-pong) biogenesis was initially dubbed secondary pathway, while the slicer-independent processing was respectively named primary biogenesis as it was proposed to generate piRNA that feed into the ping-pong cycle (A. A. Aravin, Hannon, et al., 2007; Brennecke et al., 2007). Initiation of ping-pong requires pre-existing piRNA and it was logical to propose that they are generated by another mechanism. Indeed, piRNA biogenesis can operate independently of the ping-pong cycle as follicular cells of the fly ovary, which have somatic origin, do not express Aub and Ago3 and therefore do not have ping-pong, yet generate Piwi-loaded piRNA in a Zuc-dependent fashion (Malone et al., 2009; Zamore, 2010). However, recent research suggests that the interaction between the two

biogenesis pathways is more complicated and in fact – at least in some cases – is inverted. First, it was found that piRNA biogenesis in germ cells requires maternally inherited piRNA (Brennecke et al., 2008; Le Thomas, Marinov, et al., 2014; Le Thomas, Toth, et al., 2014). In addition to initiating installment of the H3K9me3 mark on piRNA clusters (described above), they can start the ping-pong cycle, eliminating the need for ‘starter’ piRNA formed by a different mechanism. Second it was reported that elimination of the ping-pong cycle in germ cells in Aub/Ago3 double mutants leads to loss of Piwi-loaded piRNA, which were thought to be generated by ping-pong-independent primary biogenesis (Han et al., 2015; Homolka et al., 2015; Mohn et al., 2015; W. Wang et al., 2015). Furthermore, slicer cleavage by Aub or Ago3 not only generates one ping-pong piRNA from the cleaved precursor, but leads to Zuc-dependent processing of the downstream fragment, which generates Piwi-bound piRNA. Thus, the ping-pong pathway, which was thought to be secondary, in fact initiates the Zuc-dependent ‘primary’ pathway. Simply swapping the names of the two pathways is, however, not fixing the problem, as – at least in follicular cells where ping-pong does not work – the Zuc-dependent pathway can be truly independent of ping-pong. Therefore, our deeper understanding of piRNA biogenesis calls for abandoning the terms of ‘primary’ and ‘secondary’ biogenesis. Instead we propose to base the nomenclature on the nature of the enzymatic machinery that generates the 5’ end of piRNA: slicer-dependent (or ping-pong) and Zuc-dependent processing (Figure 2B).

In addition to the above described proteins – Aub, Ago3, Zuc and Hen1 – piRNA biogenesis requires many other proteins of which the functions are less understood (Czech et al., 2013; Handler et al., 2013; Muerdter et al., 2013). As only the Zuc-dependent pathway operates in follicular cells, less proteins are involved in piRNA biogenesis in these cells.

Many factors required for piRNA biogenesis in follicular cells such as the putative RNA helicase Armitage (Armi), the Tudor-domain protein Vreteno (Vret) and Yb co-localize in perinuclear foci that were termed Yb bodies (Handler et al., 2011; Johnson, Wayne, & Nagoshi, 1995; Murota et al., 2014; Qi et al., 2011; Saito et al., 2010; Zamparini et al., 2011). Though at steady-state Piwi is present exclusively in the nucleus of follicular cells, deletion of its nuclear localization signal leads to Piwi localization to Yb bodies (Klenov et al., 2014). Therefore, it is plausible that Yb bodies represent a site of piRNA processing and loading into Piwi before the Piwi/piRNA complex relocates to the nucleus. In agreement with this, transcripts from the somatic piRNA cluster flamenco were reported to localize close to Yb body (Murota et al., 2014). The molecular functions of protein components of Yb bodies remain largely unknown. Additionally, Zuc, as well as other proteins with unknown functions required for piRNA biogenesis, such as GASZ and Minotaur (mino), are anchored on the mitochondrial surface (L. Ma et al., 2009; Shiromoto et al., 2013; Vagin et al., 2013), suggesting that mitochondria also play a key role in piRNA precursor processing. The interplay between Yb bodies and mitochondria merits further investigation.

In the germ cells, proteins involved in piRNA biogenesis also localize to a distinct cytoplasmic compartment that surrounds the nuclei of nurse cells called nuage (A. K. Lim & Kai, 2007; Pek, Patil, & Kai, 2012). Krimp, one of the most stable components of nuage that is able to form granules in the absence of other nuage proteins, recruits piRNA-loaded Aub and unloaded Ago3 to form a complex (Webster et al., 2015). The spatial proximity of Aub and Ago3 in such complex was proposed to facilitate delivery of Aub-cleaved product into Ago3 during ping-pong (Sato et al., 2015; Webster et al., 2015). Qin - another Tudor domain protein - has a role in maintaining heterotypic ping-pong by preventing the loading of Aub

cleavage products into Piwi or Aub (W. Wang et al., 2015; Z. Zhang, Koppetsch, et al., 2014; Z. Zhang et al., 2011). Qin forms a complex with Vasa (Nishida et al., 2015; Xiol et al., 2014), a member of the DEAD box helicase family, with ATPase, RNA binding and RNA unwinding activity (Dehghani & Lasko, 2015; Jeske et al., 2015; Sengoku, Nureki, Nakamura, Satoru, & Yokoyama, 2006). Two roles were proposed for Vasa in piRNA processing. First, the RNA-unwinding activity of Vasa helps to release cleaved products from the piRNA-protein complex to facilitate the ping-pong cycle (Nishida et al., 2015). Second, Vasa was proposed to participate in assembly of the ping-pong complex (Xiol et al., 2014). The molecular function of numerous other nuage factors remains unknown, despite the fact that they genetically and physically interact with piwi proteins and other nuage components. In the future, the challenge will be to elucidate their functions and the molecular interplay between these factors.

### **Licensing of piRNA precursors for processing**

One of the most important remaining unresolved questions is how piRNA precursors are discriminated from other cellular RNA and directed for processing into piRNA. In other examples of RNA processing, such as splicing or processing of CRISPR RNA, specific sequence and/or structure motifs in precursor RNA are recognized by the processing machinery. Similarly, Drosha and Dicer, the two key enzymes in miRNA processing, recognize the secondary structure of pre-miRNA (S. S. Li & Patel, 2016; Park et al., 2011; Zeng, Yi, & Cullen, 2005). To date, no common sequence or structural motifs that are shared by all piRNA precursors were identified. Inserting an extended artificial sequence into natural piRNA precursors results in its processing into piRNA arguing against the

requirement of local sequence or structural motifs for processing (Le Thomas et al., 2013; Muerdter et al., 2012).

In follicular cells some piRNA precursors – such as the mRNA of traffic jam (tj) and transcripts from the uni-strand flamenco piRNA cluster – contain sequences that target them for processing (Homolka et al., 2015; Ishizu et al., 2015). Inserting a fragment of the tj 3'UTR or flamenco into an unrelated RNA transcript is sufficient to trigger production of piRNA from this transcript (Homolka et al., 2015; Ishizu et al., 2015). The tj- and flamenco- derived sequences that trigger piRNA generation associate with the RNA-binding protein Yb, although a specific motif that is recognized by Yb has not been determined (Ishizu et al., 2015). Whether recruitment of Yb to RNA is sufficient to trigger its processing into piRNA has not yet been directly tested. Nor is it known whether other somatic piRNA precursors also harbor sequence motifs that are bound by Yb. Nonetheless, the above results suggest that Yb may recognize specific sequence motifs in transcripts and recruit them to the processing machinery (Figure 3).

The mechanism of piRNA precursor selection seems to be different in germ cells and somatic follicular cells as no sequences that would trigger processing were identified in piRNA precursors expressed in germline. Two models were proposed to explain how piRNA precursors can be selected in the absence of any sequence motifs (Figure 3). The first – which can be called 'persistent nuclear mark' - model suggests that a specific protein (or proteins) tightly associates with piRNA precursors in the nucleus and remains associated in the cytoplasm where it activates the processing machinery. The model relies on the fact that piRNA processing in germ cells depend on the RDC complex, which is enriched on chromatin of piRNA-generating loci (Y. C. A. Chen et al., 2016; Klattenhoff et al., 2009; Le

Thomas, Stuwe, et al., 2014; Mohn et al., 2014), and the TREX complex, which co-transcriptionally binds pre-piRNA in an RDC-dependent fashion (Hur et al., 2016; F. Zhang et al., 2012). It was proposed that components of either the RDC or the TREX complex might constitute the mark that triggers cytoplasmic piRNA processing (F. Zhang et al., 2012; Z. Zhang, Wang, et al., 2014). As localization of the RDC complex seems to be highly specific to genomic regions that generate piRNA, this model explains how piRNA precursors can be discriminated and targeted to processing in the absence of sequence motifs. However, both the nature of the mark and the mechanism by which it engages the processing machinery remain unclear. Evidence that components of RDC or TREX remain associated with piRNA precursors after their export to the cytoplasm is also lacking. Finally, tethering of Rhino, a component of the RDC complex, to a single-stranded transgene does not trigger piRNA biogenesis (Z. Zhang, Wang, et al., 2014), arguing against the idea that binding of RDC by itself is sufficient to specify piRNA precursors.

An alternative model, which we will call ‘selection by pre-existing piRNA’, suggests that precursors are specified in the cytoplasm by complementary piRNA associated with the cytoplasmic piwi proteins. The model relies on the observation that a transcript that is recognized by complementary piRNA residing in Aub or Ago3 is first cleaved by their slicer activities to generate a single responder piRNA followed by Zuc-dependent processing of the remainder of the transcript to multiple piRNAs (Han et al., 2015; Mohn et al., 2015). Insertion of a single piRNA target sequence into a heterologous transcript leads to efficient processing of the transcript into piRNA (Hayashi et al., 2016; Mohn et al., 2015; Siomi & Siomi, 2015). This model raises the obvious question of how the very first piRNAs – which subsequently recognize piRNA precursors - are made. The answer seems to be inheritance

of piRNA from the previous generation. Maternally expressed piRNA are present in the early *Drosophila* embryo (Brennecke et al., 2008). Furthermore, trans-generational inheritance of piRNA is necessary for piRNA biogenesis in germ cells of the new generation (Le Thomas et al., 2013; Le Thomas, Stuwe, et al., 2014; Le Thomas, Toth, et al., 2014; Sienski et al., 2012), suggesting that maternally provided piRNA initiate piRNA biogenesis.

The above models suggest that the mechanism of precursor selection in the soma and the germline is radically different: in the soma selection relies on recognition of sequence motifs in precursor RNA, while in the germline selection is sequence-independent. However, the core piRNA processing machinery composed of Zuc and a number of other proteins operates in both cell types, suggesting that a common principle for precursor selection should exist. We propose that sequestration of RNA into a distinct cellular compartment might be such a central principle that is shared by both cell types. The central postulate of this proposal is that any RNA that is localized to the processing compartment will be processed to piRNAs in a sequence-independent fashion. Recruitment of RNA to this compartment might be achieved by different mechanisms including recognition of sequence motifs either by RNA-binding proteins or by complementary piRNA associated with the cytoplasmic piwi proteins. Both Yb and the cytoplasmic piwi proteins Aub and Ago3 localize in a distinct compartment – the Yb body in follicular cells or nuage granules in the germline – and therefore might be able to recruit RNA to these structures. The sequestration hypothesis puts the fact that Zuc-dependent piRNA biogenesis in germ cells requires Aub and Ago3 in new light. While originally this result was interpreted as a requirement for slicer cleavage to trigger Zuc-dependent processing (Han et al., 2015; Mohn et al., 2015), it is possible that piRNA-loaded Aub and Ago3 are necessary not (only) to slice the precursor, but to recruit the substrate into

the compartment. This model is supported by the finding that expression of catalytically-impaired Ago3 and Aub can at least partially rescue Zuc-dependent processing of Piwi-associated piRNA (W. Wang et al., 2015). We propose that this alternative, slicer-independent, mechanism to initiate Zuc-processing is due to recruitment of precursor RNAs to the processing machinery by the slicer-impaired Aub and Ago3.

It should be noted that our hypothesis goes beyond simply stating the fact that compartmentalization is important for piRNA processing. Compartmentalization plays an important role in almost all RNA processing pathways such as splicing, rRNA maturation etc. However, these processes still depend on the presence of sequence motifs in the RNA substrates (Henras, Plisson-Chastang, O'Donohue, Chakraborty, & Gleizes, 2015; G. Zhang, Taneja, Singer, & Green, 1994). In other words, in these pathways localization of RNA substrates to the processing compartments is necessary but not sufficient to trigger processing. We propose that the piRNA pathway operates differently and localization of RNA into nuage/Yb granules is both necessary and sufficient to initiate piRNA processing. Importantly, this hypothesis is testable as it suggests that piRNA biogenesis can be triggered in both cell types in a sequence-independent fashion by recruiting RNA into the processing compartment.

### **Acknowledgements**

This work was supported by grants from the National Institutes of Health R01 GM097363, Ministry of Education and Science of the Russian Federation and the Packard Fellowship Award.

## References

- Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., . . . Tuschl, T. (2006). A novel class of small RNAs bind to MILI protein in mouse testes. *Nature*, *442*(7099), 203-207. doi:10.1038/nature04916
- Aravin, A. A., Hannon, G. J., & Brennecke, J. (2007). The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science*, *318*(5851), 761-764. doi:10.1126/science.1146484
- Aravin, A. A., Naumova, N. M., Tulin, A. V., Vagin, V. V., Rozovsky, Y. M., & Gvozdev, V. A. (2001). Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Curr Biol*, *11*(13), 1017-1027. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11470406>
- Aravin, A. A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K. F., . . . Hannon, G. J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Molecular Cell*, *31*(6), 785-799. doi:10.1016/j.molcel.2008.09.003
- Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K., & Hannon, G. J. (2007). Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science*, *316*(5825), 744-747. doi:10.1126/science.1142612
- Bernstein, E., Caudy, A. A., Hammond, S. M., & Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, *409*(6818), 363-366. doi:Doi 10.1038/35053110
- Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., & Hannon, G. J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell*, *128*(6), 1089-1103. doi:10.1016/j.cell.2007.01.043
- Brennecke, J., Malone, C. D., Aravin, A. A., Sachidanandam, R., Stark, A., & Hannon, G. J. (2008). An Epigenetic Role for Maternally Inherited piRNAs in Transposon Silencing. *Science*, *322*(5906), 1387-1392. doi:10.1126/science.1165171
- Carmell, M. A., Girard, A., van de Kant, H. J., Bourc'his, D., Bestor, T. H., de Rooij, D. G., & Hannon, G. J. (2007). MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev Cell*, *12*(4), 503-514. doi:10.1016/j.devcel.2007.03.001
- Carmell, M. A., Xuan, Z., Zhang, M. Q., & Hannon, G. J. (2002). The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev*, *16*(21), 2733-2742. doi:10.1101/gad.1026102
- Chen, Y., Pane, A., & Schupbach, T. (2007). cutoff and aubergine mutations result in retrotransposon upregulation and checkpoint activation in *Drosophila*. *Current Biology*, *17*(7), 637-642. doi:10.1016/j.cub.2007.02.027
- Chen, Y. C. A., Stuwe, E., Luo, Y. C., Ninova, M., Le Thomas, A., Rozhavskaia, E., . . . Aravin, A. A. (2016). Cutoff Suppresses RNA Polymerase II Termination to Ensure Expression of piRNA Precursors. *Molecular Cell*, *63*(1), 97-109. doi:10.1016/j.molcel.2016.05.010
- Cox, D. N., Chao, A., Baker, J., Chang, L., Qiao, D., & Lin, H. (1998). A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-

- renewal. *Genes Dev*, *12*(23), 3715-3727. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9851978>
- Cox, D. N., Chao, A., & Lin, H. (2000). piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development*, *127*(3), 503-514. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10631171>
- Czech, B., & Hannon, G. J. (2016a). A Happy 3' Ending to the piRNA Maturation Story. *Cell*, *164*(5), 838-840. doi:10.1016/j.cell.2016.02.012
- Czech, B., & Hannon, G. J. (2016b). One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. *Trends in Biochemical Sciences*, *41*(4), 324-337. doi:10.1016/j.tibs.2015.12.008
- Czech, B., Preall, J. B., McGinn, J., & Hannon, G. J. (2013). A Transcriptome-wide RNAi Screen in the Drosophila Ovary Reveals Factors of the Germline piRNA Pathway. *Molecular Cell*, *50*(5), 749-761. doi:10.1016/j.molcel.2013.04.007
- Dehghani, M., & Lasko, P. (2015). In vivo mapping of the functional regions of the DEAD-box helicase Vasa. *Biology Open*, *4*(4), 450-U457. doi:10.1242/bio.201410579
- Feltzin, V. L., Khaladkar, M., Abe, M., Parisi, M., Hendriks, G. J., Kim, J., & Bonini, N. M. (2015). The exonuclease Nibbler regulates age-associated traits and modulates piRNA length in Drosophila. *Aging Cell*, *14*(3), 443-452. doi:10.1111/ace.12323
- Fire, A., Xu, S. Q., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, *391*(6669), 806-811. doi:10.1038/35888
- Girard, A., Sachidanandam, R., Hannon, G. J., & Carmell, M. A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature*, *442*(7099), 199-202. doi:10.1038/nature04917
- Goriaux, C., Desset, S., Renaud, Y., Vaury, C., & Brasset, E. (2014). Transcriptional properties and splicing of the flamenco piRNA cluster. *Embo Reports*, *15*(4), 411-418. doi:10.1002/embr.201337898
- Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B. J., Chiang, H. R., King, N., . . . Bartel, D. P. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature*, *455*(7217), 1193-U1115. doi:10.1038/nature07415
- Grivna, S. T., Beyret, E., Wang, Z., & Lin, H. (2006). A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev*, *20*(13), 1709-1714. doi:10.1101/gad.1434406
- Gunawardane, L. S., Saito, K., Nishida, K. M., Miyoshi, K., Kawamura, Y., Nagami, T., . . . Siomi, M. C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. *Science*, *315*(5818), 1587-1590. doi:10.1126/science.1140494
- Hamilton, A. J., & Baulcombe, D. C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science*, *286*(5441), 950-952. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10542148>
- Han, B. W., Wang, W., Li, C. J., Weng, Z. P., & Zamore, P. D. (2015). piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. *Science*, *348*(6236), 817-821. doi:10.1126/science.aaa1264
- Handler, D., Meixner, K., Pizka, M., Lauss, K., Schmied, C., Gruber, F. S., & Brennecke, J. (2013). The Genetic Makeup of the Drosophila piRNA Pathway. *Molecular Cell*, *50*(5), 762-777. doi:10.1016/j.molcel.2013.04.031

- Handler, D., Olivieri, D., Novatchkova, M., Gruber, F. S., Meixner, K., Mechtler, K., . . . 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 Journal*, *30*(19), 3977-3993. doi:10.1038/emboj.2011.308
- Hayashi, R., Schnabl, J., Handler, D., Mohn, F., Arneres, S. L., & Brennecke, J. (2016). Genetic and mechanistic diversity of piRNA 3'-end formation. *Nature*, *539*(7630), 588-+. doi:10.1038/nature20162
- Henras, A. K., Plisson-Chastang, C., O'Donohue, M. F., Chakraborty, A., & Gleizes, P. E. (2015). An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley Interdiscip Rev RNA*, *6*(2), 225-242. doi:10.1002/wrna.1269
- Hirano, T., Iwasaki, Y. W., Lin, Z. Y. C., Imamura, M., Seki, N. M., Sasaki, E., . . . Siomi, H. (2014). Small RNA profiling and characterization of piRNA clusters in the adult testes of the common marmoset, a model primate. *Rna*, *20*(8), 1223-1237. doi:10.1261/rna.045310.114
- Homolka, D., Pandey, R. R., Goriaux, C., Brassat, E., Vaury, C., Sachidanandam, R., . . . Pillai, R. S. (2015). PIWI Slicing and RNA Elements in Precursors Instruct Directional Primary piRNA Biogenesis. *Cell Reports*, *12*(3), 418-428. doi:10.1016/j.celrep.2015.06.030
- Horwich, M. D., Li, C. J., Matranga, C., Vagin, V., Farley, G., Wang, P., & Zamore, P. D. (2007). The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Current Biology*, *17*(14), 1265-1272. doi:10.1016/j.cub.2007.06.030
- Houwing, S., Kamminga, L. M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., . . . Ketting, R. F. (2007). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in zebrafish. *Cell*, *129*(1), 69-82. doi:10.1016/j.cell.2007.03.026
- Huang, H. D., Li, Y. J., Szulwach, K. E., Zhang, G. Q., Jin, P., & Chen, D. H. (2014). AGO3 Slicer activity regulates mitochondria-nucleus localization of Armitage and piRNA amplification. *Journal of Cell Biology*, *206*(2), 217-230. doi:10.1083/jcb.201401002
- Hur, J. K., Luo, Y. C., Moon, S., Ninova, M., Marinov, G. K., Chung, Y. D., & Aravin, A. A. (2016). Splicing-independent loading of TREX on nascent RNA is required for efficient expression of dual-strand piRNA clusters in *Drosophila*. *Genes & Development*, *30*(7), 840-855. doi:10.1101/gad.276030.115
- Hur, J. K., Olovnikov, I., & Aravin, A. A. (2014). Prokaryotic Argonautes defend genomes against invasive DNA. *Trends Biochem Sci*, *39*(6), 257-259. doi:10.1016/j.tibs.2014.04.006
- Ipsaro, J. J., Haase, A. D., Knott, S. R., Joshua-Tor, L., & Hannon, G. J. (2012). The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. *Nature*, *491*(7423), 279-U151. doi:10.1038/nature11502
- Ishizu, H., Iwasaki, Y. W., Hirakata, S., Ozaki, H., Iwasaki, W., Siomi, H., & Siomi, M. C. (2015). Somatic Primary piRNA Biogenesis Driven by cis-Acting RNA Elements and trans-Acting Yb. *Cell Reports*, *12*(3), 429-440. doi:10.1016/j.celrep.2015.06.035

- Iwasaki, Y. W., Siomi, M. C., & Siomi, H. (2015). PIWI-Interacting RNA: Its Biogenesis and Functions. *Annu Rev Biochem*, *84*, 405-433. doi:10.1146/annurev-biochem-060614-034258
- Jeske, M., Bordi, M., Glatt, S., Muller, S., Rybin, V., Muller, C. W., & Ephrussi, A. (2015). The Crystal Structure of the *Drosophila* Germline Inducer Oskar Identifies Two Domains with Distinct Vasa Helicase- and RNA-Binding Activities. *Cell Reports*, *12*(4), 587-598. doi:10.1016/j.celrep.2015.06.055
- Johnson, E., Wayne, S., & Nagoshi, R. (1995). Fs(1) Yb Is Required for Ovary Follicle Cell-Differentiation in *Drosophila-Melanogaster* and Has Genetic Interactions with the Notch Group of Neurogenic Genes. *Genetics*, *140*(1), 207-217. Retrieved from <Go to ISI>://WOS:A1995QV52100017
- Kalmykova, A. I., Klenov, M. S., & Gvozdev, V. A. (2005). Argonaute protein PIWI controls mobilization of retrotransposons in the *Drosophila* male germline. *Nucleic Acids Res*, *33*(6), 2052-2059. doi:10.1093/nar/gki323
- Katahira, J. (2012). mRNA export and the TREX complex. *Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms*, *1819*(6), 507-513. doi:10.1016/j.bbagr.2011.12.001
- Kawaoka, S., Izumi, N., Katsuma, S., & Tomari, Y. (2011). 3' End Formation of PIWI-Interacting RNAs In Vitro. *Molecular Cell*, *43*(6), 1015-1022. doi:10.1016/j.molcel.2011.07.029
- Kim, V. N., Han, J., & Siomi, M. C. (2009). Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol*, *10*(2), 126-139. doi:10.1038/nrm2632
- Klattenhoff, C., Bratu, D. P., McGinnis-Schultz, N., Koppetsch, B. S., Cook, H. A., & Theurkauf, W. E. (2007). *Drosophila* rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. *Dev Cell*, *12*(1), 45-55. doi:10.1016/j.devcel.2006.12.001
- Klattenhoff, C., & Theurkauf, W. (2008). Biogenesis and germline functions of piRNAs. *Development*, *135*(1), 3-9. doi:10.1242/dev.006486
- Klattenhoff, C., Xi, H. L., Li, C. J., Lee, S., Xu, J., Khurana, J. S., . . . Theurkauf, W. E. (2009). The *Drosophila* HP1 Homolog Rhino Is Required for Transposon Silencing and piRNA Production by Dual-Strand Clusters. *Cell*, *138*(6), 1137-1149. doi:10.1016/j.cell.2009.07.014
- Klenov, M. S., Lavrov, S. A., Korbut, A. P., Stolyarenko, A. D., Yakushev, E. Y., Reuter, M., . . . Gvozdev, V. A. (2014). Impact of nuclear Piwi elimination on chromatin state in *Drosophila melanogaster* ovaries. *Nucleic Acids Research*, *42*(10), 6208-6218. doi:10.1093/nar/gku268
- Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., . . . Nakano, T. (2008). DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev*, *22*(7), 908-917. doi:10.1101/gad.1640708
- Lau, N. C., Seto, A. G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D. P., & Kingston, R. E. (2006). Characterization of the piRNA complex from rat testes. *Science*, *313*(5785), 363-367. doi:10.1126/science.1130164

- Le Thomas, A., Marinov, G. K., & Aravin, A. A. (2014). A Transgenerational Process Defines piRNA Biogenesis in *Drosophila virilis*. *Cell Reports*, 8(6), 1617-1623. doi:10.1016/j.celrep.2014.08.013
- Le Thomas, A., Rogers, A. K., Webster, A., Marinov, G. K., Liao, S. E., Perkins, E. M., . . . Toth, K. F. (2013). Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes & Development*, 27(4), 390-399. doi:10.1101/gad.209841.112
- Le Thomas, A., Stuwe, E., Li, S. S., Du, J. M., Marinov, G., Rozhkov, N., . . . Aravin, A. A. (2014). Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. *Genes & Development*, 28(15), 1667-1680. doi:10.1101/gad.245514.114
- Le Thomas, A., Toth, K. F., & Aravin, A. A. (2014). To be or not to be a piRNA: genomic origin and processing of piRNAs. *Genome Biology*, 15(1). doi:Artn 204 10.1186/Gb4154
- Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., . . . Kim, V. N. (2003). The nuclear RNase III Drosha initiates microRNA processing. *Nature*, 425(6956), 415-419. doi:10.1038/nature01957
- Li, C., Vagin, V. V., Lee, S., Xu, J., Ma, S., Xi, H., . . . Zamore, P. D. (2009). Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell*, 137(3), 509-521. doi:10.1016/j.cell.2009.04.027
- Li, S. S., & Patel, D. J. (2016). Drosha and Dicer: Slicers cut from the same cloth. *Cell Research*, 26(5), 511-512. doi:10.1038/cr.2016.19
- Li, X. Z. G., Roy, C. K., Dong, X. J., Bolcun-Filas, E., Wang, J., Han, B. W., . . . Zamore, P. D. (2013). An Ancient Transcription Factor Initiates the Burst of piRNA Production during Early Meiosis in Mouse Testes. *Molecular Cell*, 50(1), 67-81. doi:10.1016/j.molcel.2013.02.016
- Lim, A. K., & Kai, T. (2007). Unique germ-line organelle, nuage, functions to repress selfish genetic elements in *Drosophila melanogaster* (vol 104, pg 6714, 2007). *Proceedings of the National Academy of Sciences of the United States of America*, 104(50), 20143-20143. doi:10.1073/pnas.0710102104
- Lim, R. S. M., Anand, A., Nishimiya-Fujisawa, C., Kobayashi, S., & Kai, T. (2014). Analysis of Hydra PIWI proteins and piRNAs uncover early evolutionary origins of the piRNA pathway. *Developmental Biology*, 386(1), 237-251. doi:10.1016/j.ydbio.2013.12.007
- Lund, E., & Dahlberg, J. E. (2006). Substrate selectivity of Exportin 5 and Dicer in the biogenesis of microRNAs. *Cold Spring Harbor Symposia on Quantitative Biology*, 71, 59-66. doi:DOI 10.1101/sqb.2006.71.050
- Ma, J. B., Yuan, Y. R., Meister, G., Pei, Y., Tuschl, T., & Patel, D. J. (2005). Structural basis for 5'-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. *Nature*, 434(7033), 666-670. doi:10.1038/nature03514
- Ma, L., Buchold, G. M., Greenbaum, M. P., Roy, A., Burns, K. H., Zhu, H. F., . . . Matzuk, M. M. (2009). GASZ Is Essential for Male Meiosis and Suppression of Retrotransposon Expression in the Male Germline. *Plos Genetics*, 5(9). doi:ARTN e1000635 10.1371/journal.pgen.1000635

- Malone, C. D., Brennecke, J., Dus, M., Stark, A., McCombie, W. R., Sachidanandam, R., & Hannon, G. J. (2009). Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell*, *137*(3), 522-535. doi:10.1016/j.cell.2009.03.040
- Masuda, S., Das, R., Cheng, H., Hurt, E., Dorman, N., & Reed, R. (2005). Recruitment of the human TREX complex to mRNA during splicing. *Genes & Development*, *19*(13), 1512-1517. doi:10.1101/gad.1302205
- Meister, G. (2013). Argonaute proteins: functional insights and emerging roles. *Nature Reviews Genetics*, *14*(7), 447-459. doi:10.1038/nrg3462
- Mohn, F., Handler, D., & Brennecke, J. (2015). piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. *Science*, *348*(6236), 812-817. doi:10.1126/science.aaa1039
- Mohn, F., Sienski, G., Handler, D., & Brennecke, J. (2014). The Rhino-Deadlock-Cutoff Complex Licenses Noncanonical Transcription of Dual-Strand piRNA Clusters in *Drosophila*. *Cell*, *157*(6), 1364-1379. doi:10.1016/j.cell.2014.04.031
- Muerdter, F., Guzzardo, P. M., Gillis, J., Luo, Y. C., Yu, Y., Chen, C. F., . . . Hannon, G. J. (2013). A Genome-wide RNAi Screen Draws a Genetic Framework for Transposon Control and Primary piRNA Biogenesis in *Drosophila*. *Molecular Cell*, *50*(5), 736-748. doi:10.1016/j.molcel.2013.04.006
- Muerdter, F., Olovnikov, I., Molaro, A., Rozhkov, N. V., Czech, B., Gordon, A., . . . Aravin, A. A. (2012). Production of artificial piRNAs in flies and mice. *Rna*, *18*(1), 42-52. doi:10.1261/rna.029769.111
- Murota, Y., Ishizu, H., Nakagawa, S., Iwasaki, Y. W., Shibata, S., Kamatani, M. K., . . . Siomi, M. C. (2014). Yb Integrates piRNA Intermediates and Processing Factors into Perinuclear Bodies to Enhance piRISC Assembly. *Cell Reports*, *8*(1), 103-113. doi:10.1016/j.celrep.2014.05.043
- Nishida, K. M., Iwasaki, Y. W., Murota, Y., Nagao, A., Mannen, T., Kato, Y., . . . Siomi, M. C. (2015). Respective Functions of Two Distinct Siwi Complexes Assembled during PIWI-Interacting RNA Biogenesis in *Bombyx* Germ Cells. *Cell Reports*, *10*(2), 193-203. doi:10.1016/j.celrep.2014.12.013
- Nishida, K. M., Saito, K., Mori, T., Kawamura, Y., Nagami-Okada, T., Inagaki, S., . . . Siomi, M. C. (2007). Gene silencing mechanisms mediated by Aubergine piRNA complexes in *Drosophila* male gonad. *RNA*, *13*(11), 1911-1922. doi:10.1261/rna.744307
- Nishimasu, H., Ishizu, H., Saito, K., Fukuhara, S., Kamatani, M. K., Bonnefond, L., . . . Nureki, O. (2012). Structure and function of Zucchini endoribonuclease in piRNA biogenesis. *Nature*, *491*(7423), 284-U157. doi:10.1038/nature11509
- Nudler, E. (2012). RNA Polymerase Backtracking in Gene Regulation and Genome Instability. *Cell*, *149*(7), 1438-1445. doi:10.1016/j.cell.2012.06.003
- Pane, A., Jiang, P., Zhao, D. Y., Singh, M., & Schupbach, T. (2011). The Cutoff protein regulates piRNA cluster expression and piRNA production in the *Drosophila* germline. *Embo Journal*, *30*(22), 4601-4615. doi:10.1038/emboj.2011.334
- Pane, A., Wehr, K., & Schupbach, T. (2007). zucchini and squash encode two putative nucleases required for rasiRNA production in the *Drosophila* germline. *Developmental Cell*, *12*(6), 851-862. doi:10.1016/j.devcel.2007.03.022

- Park, J. E., Heo, I., Tian, Y., Simanshu, D. K., Chang, H., Jee, D., . . . Kim, V. N. (2011). Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature*, 475(7355), 201-U107. doi:10.1038/nature10198
- Patel, A., Lee, H. O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M. Y., . . . Alberti, S. (2015). A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell*, 162(5), 1066-1077. doi:10.1016/j.cell.2015.07.047
- Pek, J. W., Patil, V. S., & Kai, T. (2012). piRNA pathway and the potential processing site, the nuage, in the *Drosophila* germline. *Development Growth & Differentiation*, 54(1), 66-77. doi:10.1111/j.1440-169X.2011.01316.x
- Peters, L., & Meister, G. (2007). Argonaute proteins: mediators of RNA silencing. *Mol Cell*, 26(5), 611-623. doi:10.1016/j.molcel.2007.05.001
- Qi, H., Watanabe, T., Ku, H. Y., Liu, N., Zhong, M., & Lin, H. F. (2011). The Yb Body, a Major Site for Piwi-associated RNA Biogenesis and a Gateway for Piwi Expression and Transport to the Nucleus in Somatic Cells. *Journal of Biological Chemistry*, 286(5), 3789-3797. doi:10.1074/jbc.M110.193888
- Rangan, P., Malone, C. D., Navarro, C., Newbold, S. P., Hayes, P. S., Sachidanandam, R., . . . Lehmann, R. (2011). piRNA Production Requires Heterochromatin Formation in *Drosophila*. *Current Biology*, 21(16), 1373-1379. doi:10.1016/j.cub.2011.06.057
- Reed, R. (2003). Coupling transcription, splicing and mRNA export. *Current Opinion in Cell Biology*, 15(3), 326-331. doi:10.1016/S0955-0674(03)00048-6
- Rivas, F. V., Tolia, N. H., Song, J. J., Aragon, J. P., Liu, J. D., Hannon, G. J., & Joshua-Tor, L. (2005). Purified Argonaute2 and an siRNA form recombinant human RISC. *Nature Structural & Molecular Biology*, 12(4), 340-349. doi:10.1038/nsmb918
- Robine, N., Lau, N. C., Balla, S., Jin, Z. G., Okamura, K., Kuramochi-Miyagawa, S., . . . Lai, E. C. (2009). A Broadly Conserved Pathway Generates 3' UTR-Directed Primary piRNAs. *Current Biology*, 19(24), 2066-2076. doi:10.1016/j.cub.2009.11.064
- Rozhkov, N. V., Hammell, M., & Hannon, G. J. (2013). Multiple roles for Piwi in silencing *Drosophila* transposons. *Genes & Development*, 27(4), 400-412. doi:10.1101/gad.209767.112
- Saito, K., Inagaki, S., Mituyama, T., Kawamura, Y., Ono, Y., Sakota, E., . . . Siomi, M. C. (2009). A regulatory circuit for piwi by the large Maf gene traffic jam in *Drosophila*. *Nature*, 461(7268), 1296-U1135. doi:10.1038/nature08501
- Saito, K., Ishizu, H., Komai, M., Kotani, H., Kawamura, Y., Nishida, K. M., . . . Siomi, M. C. (2010). Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in *Drosophila*. *Genes & Development*, 24(22), 2493-2498. doi:10.1101/gad.1989510
- Saito, K., Nishida, K. M., Mori, T., Kawamura, Y., Miyoshi, K., Nagami, T., . . . Siomi, M. C. (2006). Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev*, 20(16), 2214-2222. doi:10.1101/gad.1454806
- Saito, K., Sakaguchi, Y., Suzuki, T., Suzuki, T., Siomi, H., & Siomi, M. C. (2007). Pimet, the *Drosophila* homolog of HEN1, mediates 2'-O-methylation of PIWI-interacting RNAs at their 3' ends. *Genes & Development*, 21(13), 1603-1608. doi:10.1101/gad.1563607

- Sato, K., Iwasaki, Y. W., Shibuya, A., Carninci, P., Tsuchizawa, Y., Ishizu, H., . . . Siomi, H. (2015). Krimper Enforces an Antisense Bias on piRNA Pools by Binding AGO3 in the *Drosophila* Germline. *Molecular Cell*, 59(4), 553-563. doi:10.1016/j.molcel.2015.06.024
- Sengoku, T., Nureki, O., Nakamura, A., Satoru, K. I., & Yokoyama, S. (2006). Structural basis for RNA unwinding by the DEAD-box protein *Drosophila* vasa. *Cell*, 125(2), 287-300. doi:10.1016/j.cell.2006.01.054
- Shiromoto, Y., Kuramochi-Miyagawa, S., Daiba, A., Chuma, S., Katanaya, A., Katsumata, A., . . . Nakano, T. (2013). GPAT2, a mitochondrial outer membrane protein, in piRNA biogenesis in germline stem cells. *Rna-a Publication of the Rna Society*, 19(6), 803-810. doi:10.1261/rna.038521.113
- Sienski, G., Donertas, D., & Brennecke, J. (2012). Transcriptional Silencing of Transposons by Piwi and Maelstrom and Its Impact on Chromatin State and Gene Expression. *Cell*, 151(5), 964-980. doi:10.1016/j.cell.2012.10.040
- Siomi, H., & Siomi, M. C. (2015). Phased piRNAs tackle transposons. *Science*, 348(6236), 756-757. doi:10.1126/science.aab3004
- Song, J. J., Smith, S. K., Hannon, G. J., & Joshua-Tor, L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. *Science*, 305(5689), 1434-1437. doi:10.1126/science.1102514
- Strasser, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., . . . Hurt, E. (2002). TREX is a conserved complex coupling transcription with messenger RNA export. *Nature*, 417(6886), 304-308. doi:10.1038/nature746
- Swarts, D. C., Jore, M. M., Westra, E. R., Zhu, Y., Janssen, J. H., Snijders, A. P., . . . van der Oost, J. (2014). DNA-guided DNA interference by a prokaryotic Argonaute. *Nature*, 507(7491), 258-261. doi:10.1038/nature12971
- Thomson, T., & Lin, H. (2009). The biogenesis and function of PIWI proteins and piRNAs: progress and prospect. *Annu Rev Cell Dev Biol*, 25, 355-376. doi:10.1146/annurev.cellbio.24.110707.175327
- Vagin, V. V., Klenov, M. S., Kalmykova, A. I., Stolyarenko, A. D., Kotelnikov, R. N., & Gvozdev, V. A. (2004). The RNA interference proteins and vasa locus are involved in the silencing of retrotransposons in the female germline of *Drosophila melanogaster*. *RNA Biol*, 1(1), 54-58. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/17194939>
- Vagin, V. V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., & Zamore, P. D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. *Science*, 313(5785), 320-324. doi:10.1126/science.1129333
- Vagin, V. V., Yu, Y., Jankowska, A., Luo, Y. C., Wasik, K. A., Malone, C. D., . . . Hannon, G. J. (2013). Minotaur is critical for primary piRNA biogenesis. *Rna-a Publication of the Rna Society*, 19(8), 1064-1077. doi:10.1261/rna.039669.113
- Voigt, F., Reuter, M., Kasaruho, A., Schulz, E. C., Pillai, R. S., & Barabas, O. (2012). Crystal structure of the primary piRNA biogenesis factor Zucchini reveals similarity to the bacterial PLD endonuclease. *Nuc. Rna*, 18(12), 2128-2134. doi:10.1261/rna.034967.112

- Wang, H., Ma, Z. J., Niu, K. Y., Xiao, Y., Wu, X. F., Pan, C. Y., . . . Liu, N. (2016). Antagonistic roles of Nibbler and Hen1 in modulating piRNA 3' ends in *Drosophila*. *Development*, *143*(3), 530-539. doi:10.1242/dev.128116
- Wang, W., Han, B. W., Tipping, C., Ge, D. T., Zhang, Z., Weng, Z. P., & Zamore, P. D. (2015). Slicing and Binding by Ago3 or Aub Trigger Piwi-Bound piRNA Production by Distinct Mechanisms. *Molecular Cell*, *59*(5), 819-830. doi:10.1016/j.molcel.2015.08.007
- Watanabe, T., Takeda, A., Tsukiyama, T., Mise, K., Okuno, T., Sasaki, H., . . . Imai, H. (2006). Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev*, *20*(13), 1732-1743. doi:10.1101/gad.1425706
- Webster, A., Li, S., Hur, J. K., Wachsmuth, M., Bois, J. S., Perkins, E. M., . . . Aravin, A. A. (2015). Aub and Ago3 Are Recruited to Nuage through Two Mechanisms to Form a Ping-Pong Complex Assembled by Krimper. *Molecular Cell*, *59*(4), 564-575. doi:10.1016/j.molcel.2015.07.017
- Xiol, J., Spinelli, P., Laussmann, M. A., Homolka, D., Yang, Z., Cora, E., . . . Pillai, R. S. (2014). RNA Clamping by Vasa Assembles a piRNA Amplifier Complex on Transposon Transcripts. *Cell*, *157*(7), 1698-1711. doi:10.1016/j.cell.2014.05.018
- Zamore, P. D. (2010). Somatic piRNA biogenesis. *Embo Journal*, *29*(19), 3219-3221. doi:10.1038/emboj.2010.232
- Zamparini, A. L., Davis, M. Y., Malone, C. D., Vieira, E., Zavadil, J., Sachidanandam, R., . . . Lehmann, R. (2011). Vreteno, a gonad-specific protein, is essential for germline development and primary piRNA biogenesis in *Drosophila*. *Development*, *138*(18), 4039-4050. doi:10.1242/dev.069187
- Zanni, V., Eymery, A., Coiffet, M., Zytnicki, M., Luyten, I., Quesneville, H., . . . Jensen, S. (2013). Distribution, evolution, and diversity of retrotransposons at the flamenco locus reflect the regulatory properties of piRNA clusters. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(49), 19842-19847. doi:10.1073/pnas.1313677110
- Zeng, Y., Yi, R., & Cullen, B. R. (2005). Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *Embo Journal*, *24*(1), 138-148. doi:10.1038/sj.emboj.7600491
- Zhang, F., Wang, J., Xu, J., Zhang, Z., Koppetsch, B. S., Schultz, N., . . . Theurkauf, W. E. (2012). UAP56 Couples piRNA Clusters to the Perinuclear Transposon Silencing Machinery. *Cell*, *151*(4), 871-884. doi:10.1016/j.cell.2012.09.040
- Zhang, G., Taneja, K. L., Singer, R. H., & Green, M. R. (1994). Localization of pre-mRNA splicing in mammalian nuclei. *Nature*, *372*(6508), 809-812. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7997273>
- Zhang, Z., Koppetsch, B. S., Wang, J., Tipping, C., Weng, Z. P., Theurkauf, W. E., & Zamore, P. D. (2014). Antisense piRNA amplification, but not piRNA production or nuage assembly, requires the Tudor-domain protein Qin. *Embo Journal*, *33*(6), 536-539. doi:10.1002/emboj.201384895
- Zhang, Z., Wang, J., Schultz, N., Zhang, F., Parhad, S. S., Tu, S., . . . Theurkauf, W. E. (2014). The HP1 Homolog Rhino Anchors a Nuclear Complex that Suppresses piRNA Precursor Splicing. *Cell*, *157*(6), 1353-1363. doi:10.1016/j.cell.2014.04.030

Zhang, Z., Xu, J., Koppetsch, B. S., Wang, J., Tipping, C., Ma, S. M., . . . Zamore, P. D. (2011). Heterotypic piRNA Ping-Pong Requires Qin, a Protein with Both E3 Ligase and Tudor Domains. *Molecular Cell*, 44(4), 572-584. doi:10.1016/j.molcel.2011.10.011

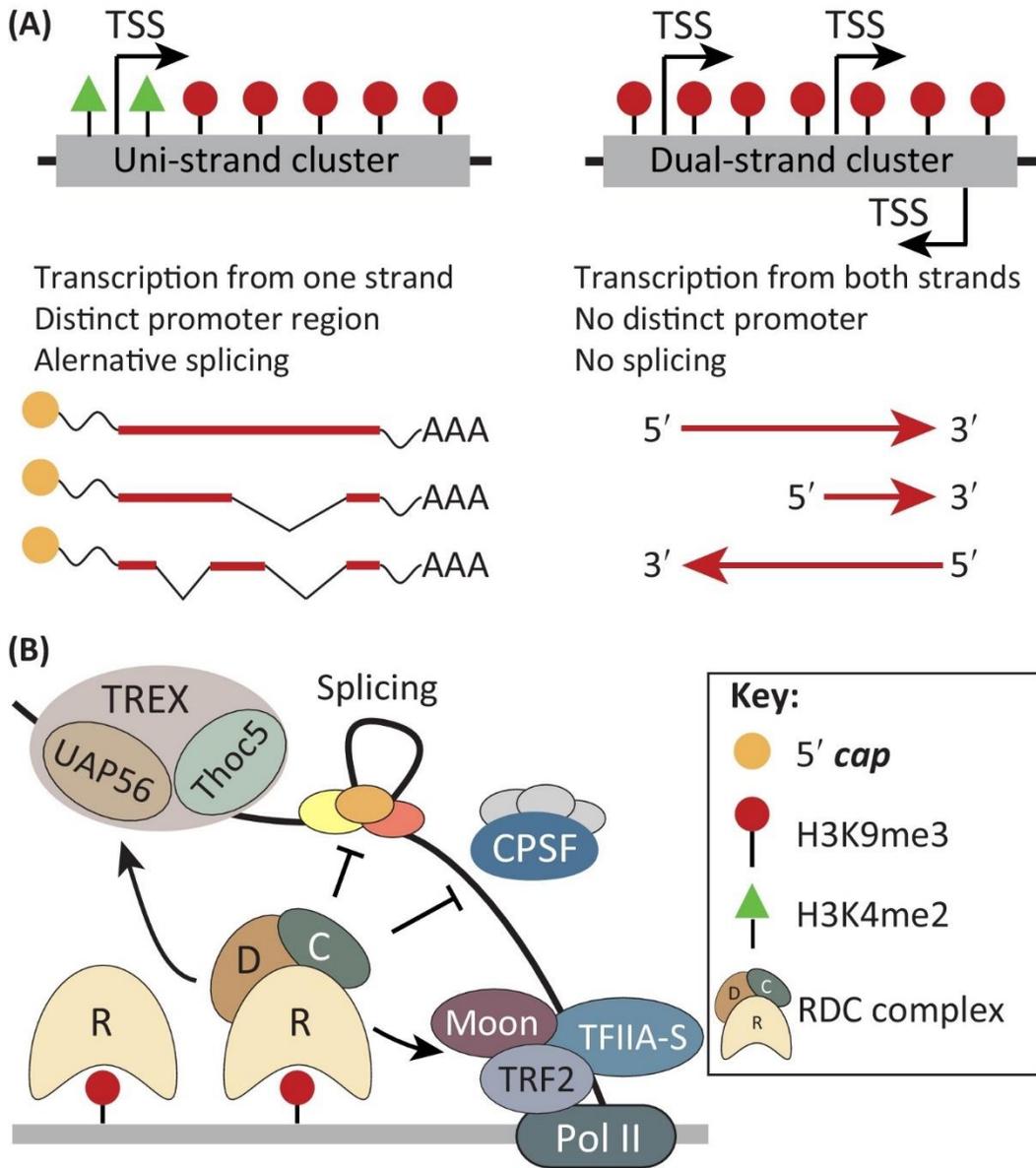
## Figures and Figure Legends

### Figure 1. Transcription of PIWI-Interacting RNA (piRNA) Clusters.

(A) Two types of piRNA cluster, uni-strand and dual-strand, differ in their transcription. Transcription of uni-strand clusters is initiated at distinct promoter regions that are marked by RNA polymerase II (Pol II) Ser5P and histone 3 lysine 4 di-methylation (H3K4me2) peaks, and transcription appears to be canonical: transcripts go through 5' end capping, 3' end polyadenylation, and sometimes alternative splicing. Dual-strand clusters are transcribed from both genomic strands and do not have defined promoters, suggesting that transcription is initiated at multiple sites. The nascent RNA produced by dual-strand clusters is not spliced and lacks a polyA tail. Both types of cluster are decorated with the heterochromatic histone 3 lysine 9 tri-methylation (H3K9me3) mark, except for the promoter regions of uni-strand clusters.

(B) The transcription of dual-strand clusters is regulated by the chromatin-associated Rhino-Deadlock-Cutoff (RDC) complex. The HP1 paralog Rhino directly binds to the H3K9me3 mark through its chromodomain. Rhino forms a complex with Deadlock and Cutoff. RDC promotes transcription initiation through interaction with Moonshiner and components of the transcription initiation complex, TRF2 and TFIIA-S. Cutoff suppresses transcription termination by two mechanisms: it prevents cleavage at poly(A) sites by the CPSF complex and interferes with termination if cleavage does occur. RDC also suppresses splicing of precursor transcripts. Finally, RDC is required for co-transcriptional loading of TREX on nascent RNA, which might promote export of piRNA precursors to the cytoplasm. Abbreviation: TSS, transcription start site.

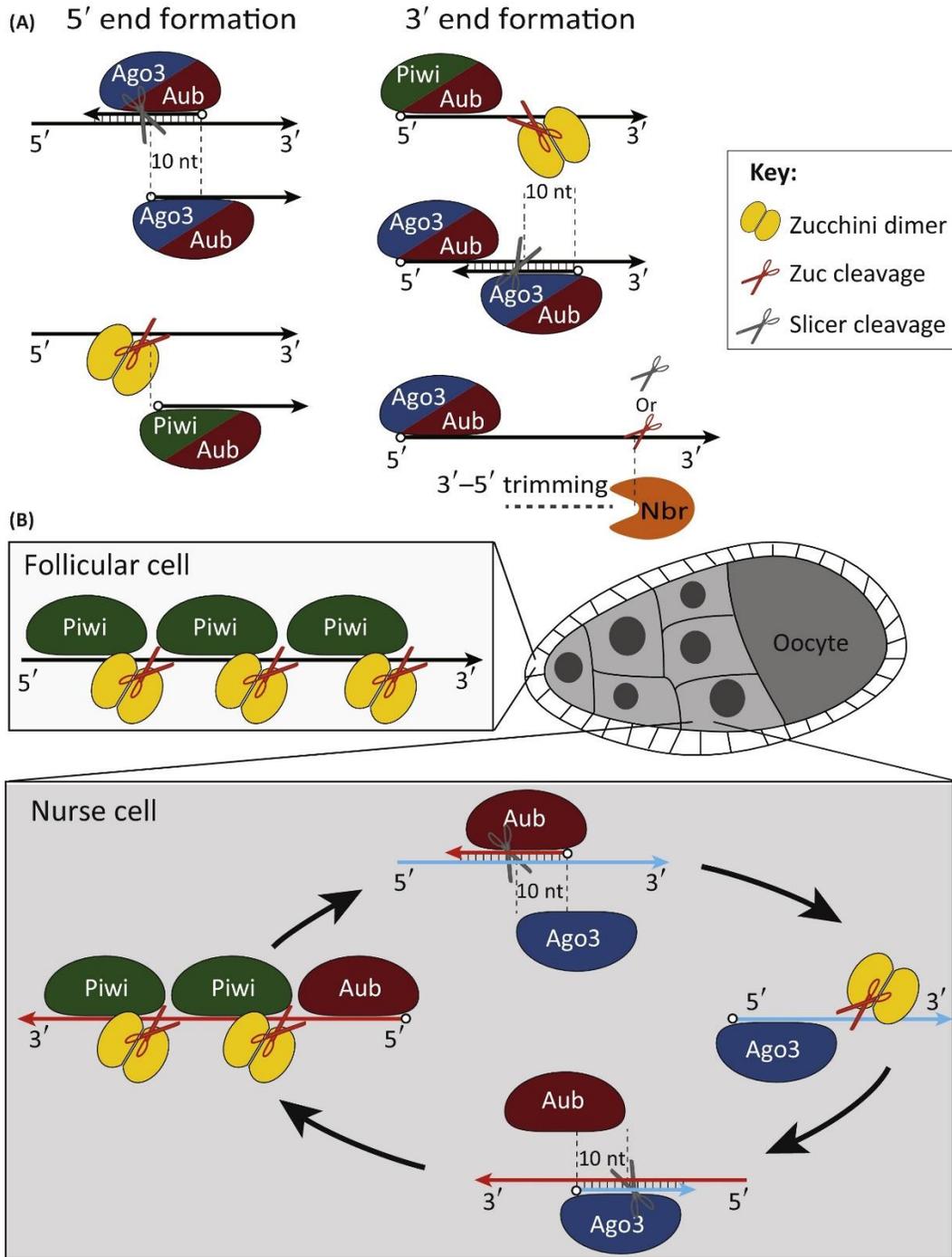
**Figure 1**



**Figure 2. Processing of the 5'- and the 3'-end of PIWI-Interacting RNA.**

(A) The 5' end of piRNAs can be formed either through the endonuclease (slicer) activity of the cytoplasmic Piwi proteins, Aubergine (Aub) and Argonaute (Ago)-3, or through cleavage by the endonuclease Zucchini (Zuc). Slicer cleavage of piRNA precursors is guided by complementary piRNA. The 5' end of a slicer product is shifted by exactly ten nucleotides (nt) relative to the 5' end of the guide piRNA. Cleavage mediated by Zuc is independent of guide piRNA. piRNAs formed through the slicer-dependent mechanism are loaded into Aub and Ago3, while piRNAs formed by Zuc are loaded into Piwi and Aub. The 3' end of piRNAs can be formed by three mechanisms: (i) through endonucleolytic cleavage by Zuc or (ii) slicer or (iii) by 3'-to-5' trimming of longer precursors by the exonuclease Nibbler. (B) piRNA processing in somatic follicular and germline cells of the *Drosophila* ovary. Only Piwi, but not Aub and Ago3, is expressed in follicular cells. Therefore, in these cells, both ends of mature piRNAs are formed exclusively through Zuc-mediated processing with possible contribution of a 3' end trimming activity. The single Zuc cleavage can simultaneously generate the 5' end of a downstream and the 3' end of an upstream RNA, resulting in a characteristic phased pattern of piRNAs. Slicer-dependent and Zuc-dependent processing coexists and the two pathways cooperate in germline nurse cells. When a new piRNA is formed by the slicer-dependent mechanism and loaded into Aub or Ago3, it can guide formation of the next piRNA, giving rise to the so-called 'ping-pong cycle'. Products of Aub-guided cleavage are predominantly loaded into Ago3, while products of Ago3-guided cleavage are loaded onto Aub. Slicer-dependent cleavage by Aub or Ago3 also directs Zuc-dependent substrate RNA processing.

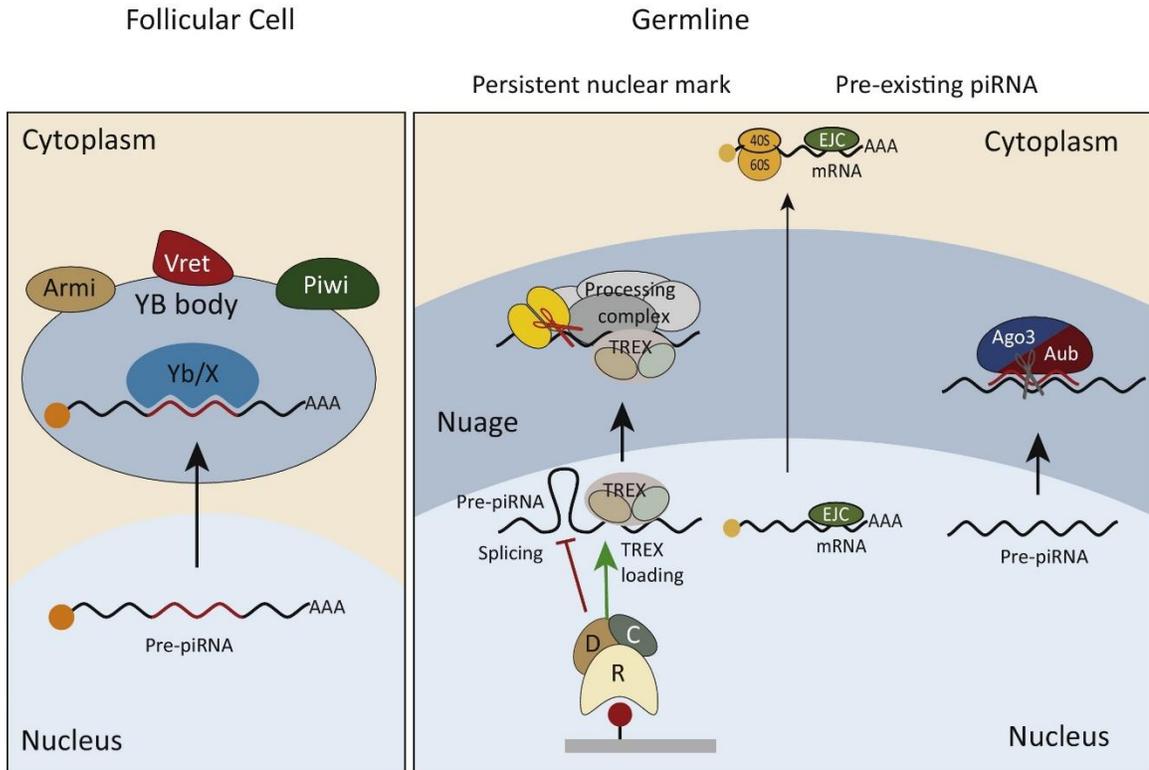
**Figure 2**



**Figure 3. Selection of PIWI-Interacting RNA precursors for processing.**

The selection mechanisms of piRNA precursors appear to differ between somatic (follicular) and germline cells. In follicular cells, specific sequences in piRNA precursors are recognized in the cytoplasm through binding by Yb or other yet-to-be-identified RNA-binding proteins (X). It was proposed that Yb, which forms cytoplasmic granules called Yb bodies, recruits other factors necessary for piRNA processing, such as Zucchini (Zuc), Vret and Armi. Two models were proposed for the selection of piRNA precursors in germline cells. According to the ‘persistent nuclear mark’ model, piRNA precursors are marked and licensed for processing in the nucleus. The mark (probably an RNA-binding protein) was proposed to shuttle with the piRNA precursor into the cytoplasm and activate the processing machinery. The identity of the mark is not known, although its deposition was proposed to depend on the RDC complex, which is present on chromatin of piRNA-generating loci. Normal mRNAs destined for translation are bound by the exon-junction complex (EJC) loaded on mRNA as a result of productive splicing. It was proposed that stalled splicing of piRNA precursors might induce marking of piRNA precursors and trigger their processing in the cytoplasm. The TREX complex, which is loaded on piRNA precursors in an RDC-dependent fashion, is a candidate for such a nuclear mark. According to the second model, selection of RNA for processing only occurs in the cytoplasm and is governed by complementary piRNAs that recognize potential precursors and trigger their processing through slicer- and Zuc-dependent mechanisms.

**Figure 3**



*Chapter 3***BINDING OF GUIDE PIRNA TRIGGERS METHYLATION OF THE UNSTRUCTURED N-TERMINAL REGION OF AUB LEADING TO ASSEMBLY OF THE PIRNA AMPLIFICATION COMPLEX**

Xiawei Huang<sup>1#</sup>, Hongmiao Hu<sup>2,3,4#</sup>, Alexandre Webster<sup>1</sup>, Fan Zou<sup>2</sup>, Jiamu Du<sup>2,3</sup>,  
Dinshaw J. Patel<sup>5</sup>, Ravi Sachidanandam<sup>6</sup>, Katalin Fejes Toth<sup>1</sup>, Alexei A. Aravin<sup>1&</sup>, Sisi  
Li<sup>2&</sup>

<sup>1</sup>California Institute of Technology, Division of Biology and Biological Engineering  
Pasadena, CA 91125, USA.

<sup>2</sup>Department of Biology, Southern University of Science and Technology of China,  
Shenzhen, Guangdong 518055, China.

<sup>3</sup>National Key Laboratory of Plant Molecular Genetics & Shanghai Center for Plant Stress  
Biology, Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences,  
Shanghai 201602, China.

<sup>4</sup>University of Chinese Academy of Sciences, Beijing 100049, China.

<sup>5</sup>Structural Biology Program, Memorial Sloan Kettering Cancer Center, New York, New  
York, USA

<sup>6</sup>Icahn School of Medicine at Mount Sinai, 5925, Department of Oncological Sciences, New  
York, New York, USA

#Those contributed equally

&To whom correspondence should be addressed:

Alexei Aravin [aaa@caltech.edu](mailto:aaa@caltech.edu); Sisi Li [liss6@sustech.edu.cn](mailto:liss6@sustech.edu.cn)

This chapter is a manuscript currently in submission for publication

bioRxiv: doi: <https://doi.org/10.1101/2020.07.14.203323>

**Abstract**

PIWI proteins use guide piRNAs to repress selfish genomic elements, protecting the genomic integrity of gametes and ensuring the fertility of animal species. Efficient transposon repression depends on amplification of piRNA guides in the ping-pong cycle, which in *Drosophila* entails tight cooperation between two PIWI proteins, Aub and Ago3. Here we show that post-translational modification, symmetric dimethylarginine (sDMA), of Aub is essential for piRNA biogenesis, transposon silencing and fertility. Methylation is triggered by loading of a piRNA guide into Aub, which exposes its unstructured N-terminal region to the PRMT5 methylosome complex. Thus, sDMA modification is a signal that Aub is loaded with piRNA guide. Amplification of piRNA in the ping-pong cycle requires assembly of a tertiary complex scaffolded by Krimper, which simultaneously binds the N-terminal regions of Aub and Ago3. To promote generation of new piRNA, Krimper uses its two Tudor domains to bind Aub and Ago3 in opposite modification and piRNA-loading states. Our results reveal that post-translational modifications in unstructured regions of PIWI proteins and their binding by Tudor domains that are capable of discriminating between modification states is essential for piRNA biogenesis and silencing.

## Introduction

The PIWI-interacting RNA (piRNA) pathway acts as a conserved defensive system that represses the proliferation of transposable elements (TEs) in the germline of sexually reproducing animals (Aravin et al., 2001; Brennecke et al., 2007; Gunawardane et al., 2007; Kalmykova, Klenov, & Gvozdev, 2005; Li et al., 2009; Nishida et al., 2007; Vagin et al., 2006). Loss of PIWI proteins causes derepression of transposons associated with gametogenesis failure and sterility in flies and mice (Carmell et al., 2007; Cox et al., 1998; Girard, Sachidanandam, Hannon, & Carmell, 2006; Khurana & Theurkauf, 2010). PIWI proteins recognize transposon targets with help of the associated small (23-30 nt) non-coding RNA guides, piwi-interacting RNAs.

PIWI proteins belong to the conserved Argonaute protein family present in all domains of life (Carmell, Xuan, Zhang, & Hannon, 2002; Cox et al., 1998; Cox, Chao, & Lin, 2000). Argonautes bind nucleic acid guides and share common domain architecture, all containing the conserved N, PAZ, MID and PIWI domains (Elkayam et al., 2012; Schirle & MacRae, 2012; Song, Smith, Hannon, & Joshua-Tor, 2004; Wei, Wu, Chen, Chen, & Xie, 2012). The MID and PAZ domains bind the 5' and 3' ends of the guide RNA, respectively (Boland, Huntzinger, Schmidt, Izaurralde, & Weichenrieder, 2011; Frank, Sonenberg, & Nagar, 2010; Ma, Ye, & Patel, 2004; Song et al., 2003; Song et al., 2004; Yan et al., 2003). The PIWI domain contains an RNase-H like fold with a conserved amino acid tetrad that endows Argonautes with endonuclease activity for precise cleavage of the target (Martinez & Tuschl, 2004; Parker, Roe, & Barford, 2004; Song et al., 2004). The degradation of complementary target mRNA by PIWI proteins can trigger the generation of new RNA guides in a process termed the ping-pong cycle. Ping-pong requires cooperativity between

two PIWI molecules as the product resulting from target cleavage by one protein is passed to the other and is converted to a new piRNA guide. In *Drosophila*, two distinct cytoplasmic PIWI proteins, Aub and Ago3, cooperate in the ping-pong cycle with each protein generating an RNA guide that is loaded into its partner. Amplification of piRNA guides through the ping-pong cycle is believed to be essential for efficient transposon repression as it allows the pathway to mount an adaptive response to actively transcribed transposons (Brennecke et al., 2007; Gunawardane et al., 2007; Huang, Fejes Toth, & Aravin, 2017).

In addition to four conserved domains, eukaryotic members of the Argonaute family, including PIWI proteins, contain a N-terminal extension region of various lengths with low sequence conservation. Structural studies of Agos suggest that the N-terminal regions adopt a disordered conformation (Faehnle, Elkayam, Haase, Hannon, & Joshua-Tor, 2013; Nakanishi, Weinberg, Bartel, & Patel, 2012; Park et al., 2019; Schirle & MacRae, 2012; Schirle, Sheu-Gruttadauria, & MacRae, 2014). Despite low overall conservation, the N-terminal region of the majority of PIWI proteins harbors arginine rich (A/G)R motifs. In both insects and mammals these motifs were shown to be substrates for post-translational modification by the PRMT5 methyltransferase, which produces symmetrically dimethylated arginine (sDMA) residues (Honda, Kirino, & Kirino, 2014; Kirino et al., 2009; Nishida et al., 2009; Reuter et al., 2009; Vagin, Hannon, & Aravin, 2009). Loss of *Prmt5* (encoded by the *Csul* and *Vls* genes) in *Drosophila* leads to reduced piRNA level and accumulation of transposon transcripts in germ cells, suggesting that sDMA modification of PIWIs plays an important role in the piRNA pathway.

Multiple Tudor domain-containing proteins can bind to sDMA modifications. Aromatic residues in binding pocket of Tudor domains form cation- $\pi$  interactions with sDMA (Chen, Nott, Jin, & Pawson, 2011; H. Liu et al., 2010; K. Liu et al., 2010; Tripsianes et al., 2011). Studies in *Drosophila* and mouse revealed that several Tudor domain-containing proteins (TDRDs) interact with PIWIs and are involved in piRNA-guided transposon repression, although their specific molecular functions remains poorly understood (Chen et al., 2011; Handler et al., 2011; Nishida et al., 2009; Sato, Iwasaki, Siomi, & Siomi, 2015; Siomi, Mannen, & Siomi, 2010). Previously we found that the Tudor-domain containing protein Krimper is required for ping-pong piRNA amplification and is capable of both self-interactions and binding of the two ping-pong partners, Aub and Ago3 (Webster et al., 2015). Krimper co-localizes with Aub and Ago3 in nuage, a membraneless perinuclear cytoplasmic compartment where piRNA-guided target degradation and ping-pong are proposed to take place. Ago3 requires Krimper for recruitment into this compartment, though Aub does not. These results led to proposal that Krimper assembles a complex that brings Ago3 to Aub and coordinates ping-pong in nuage (Webster et al., 2015). However, the architecture of the ping-pong piRNA processing (4P) complex and the extent to which Krimper regulates ping-pong remained unresolved.

Both the ping-pong cycle and sDMA modification of PIWI proteins are conserved features of the piRNA pathway, found in many organisms, suggesting that these processes are essential for pathway functions (Aravin et al., 2008; Brennecke et al., 2007; Gunawardane et al., 2007; Honda et al., 2014; Siomi et al., 2010). sDMA modification of PIWIs provides a binding platform for interactions with Tudor-domain proteins, however, its biological function and regulation are not known. Despite the essential role of ping-

pong in transposon repression, we similarly have little understanding of the molecular mechanisms that control this process. Here we revealed the biological function of Aub and Ago3 sDMA modifications and show that it plays an essential role in orchestrating assembly of the 4P complex in the ping-pong cycle. The modification signals whether PIWI proteins are loaded with guide piRNA, and this information is used to assemble a ping-pong complex that is receptive for directional transfer of RNA to an unloaded PIWI protein.

## **Results**

### **Krimper can simultaneously bind Aub and Ago3 with specificity for methylation state using two separate Tudor domains.**

The essential step in the ping-pong cycle is the transfer of RNA between two PIWI proteins, Aub and Ago3. This process requires that the two proteins are in physical proximity to each other. Previously we found that Krimper binds both Aub and Ago3 suggesting that it might bring them into physical proximity to allow transfer of the RNA during ping-pong (Webster et al., 2015). This might be possible through self-interaction of two Krimper molecules, each binding one PIWI protein, or by a single Krimper molecule simultaneously interacting with both Aub and Ago3. To discriminate between these two models, we employed two-step sequential immunoprecipitation to examine complex formation between Aub, Ago3 and a Krimper that lacks the N-terminal self-interaction region (residues 1-300aa) (Fig.1a). We immuno-precipitated tagged Aub followed by elution of purified complexes and another round of immunoprecipitation against the self-interaction-deficient Krimper. If a single Krimper molecule is capable of simultaneously interacting with Aub and Ago3, then the final immunoprecipitated complexes should contain Ago3 in addition to Krimper and

Aub. All three proteins were detected by Western blot after sequential purification (Fig.1b). This indicates that a single Krimper molecule can concomitantly interact with both Aub and Ago3.

To understand how Krimper is able to bind Aub and Ago3 simultaneously, we analyzed its domain architecture. In addition to its self-interacting N-terminal region, Krimper contains two extended Tudor domains: eTud1(272-512) and eTud2 (562-746) (Fig. 1c). Tudor domains were previously implicated in protein-protein interactions through specific binding of methylated Arginine or Lysine residues (H. Liu et al., 2010; K. Liu et al., 2010; Tripsianes et al., 2011). PIWI proteins, including Aub and Ago3, harbor conserved symmetrically dimethylated Arginine (sDMA) residues embedded into RA/RG-motifs in their N-terminal extended regions (Kirino et al., 2009; Kirino et al., 2010; Nishida et al., 2009) In agreement with published structures of other PIWI and Ago proteins, which suggest that N-terminal extensions of these proteins adopt an unstructured conformation (Elkayam et al., 2012; Faehnle et al., 2013; Matsumoto et al., 2016; Nakanishi et al., 2012; Schirle & MacRae, 2012; Yamaguchi et al., 2020), the N-terminal extended regions of both Aub and Ago3 have high levels of predicted disorder (Fig. 1d and 1e). Remarkably, RA/RG motifs, which are subject to sDMA modification are the only conserved sequences in the N-terminal regions of Aub and Ago3. The N-terminal regions are otherwise highly variable even between closely related *Drosophila* species (Fig. 1d and 1e). We previously found that the Krimper/Aub interaction depends on Aub methylation and found that the eTud2 domain of Krimper specifically binds sDMA-modified Aub (Webster et al., 2015). However, in contrast to Aub, Ago3 was shown to interact with Krimper in its unmethylated state (Sato, Iwasaki,

Shibuya, et al., 2015; Webster et al., 2015), raising the question of how these two proteins interact.

Previous studies revealed that the N-terminal (1-83aa) sequence of Ago3 is both necessary and sufficient for Krimper binding (Sato, Iwasaki, Shibuya, et al., 2015). However, the region of Krimper responsible for Ago3 binding remained unknown (Webster et al., 2015). To study the Ago3-Krimper interaction, we expressed and purified eTud1 and eTud2 domains of Krimper (Fig. 1c) and tested their binding to peptides derived from the N-terminal Ago3 sequence using isothermal titration calorimetry (ITC) (Fig. 1f and Extended Data Fig. 1a). We found that eTud1 binds to the Ago3-2 peptide (aa residues 63-78 of Ago3, which includes the conserved RA/AG motif) with a binding affinity of 62.8  $\mu$ M (Fig. 1f). Importantly, eTud1 does not detectably interact with the Aub-1 peptide that binds eTud2 (Fig. 1f). Thus, the two Tudor domains of Krimper have different binding preferences ensuring that Aub and Ago3 can bind Krimper simultaneously without direct competition for a binding site.

The Aub-Krimper (eTud2) interaction requires methylation of at least one of five Arg residues positioned within the RA/RG motif of the Aub N-terminal region (Webster et al., 2015). Remarkably, the Ago3-2 peptide that binds eTud1 domain also contains three Arg residues (R68, 70 and 72) within the RG/RA motif (Fig. 1d), which can be methylated *in vivo*, however, methylation is not required for Ago3 binding to Krimper (eTud1) (Fig. 1f). In fact, in agreement with *in vivo* experiments that indicate that Ago3 is bound to Krimper in its unmethylated state (Sato, Iwasaki, Shibuya, et al., 2015; Webster et al., 2015), methylation of any of the three Arg residues disrupts binding of the Ago3-2 peptide to eTud1 (Fig. 1f).

Our results indicate that sDMA-modified Aub interacts with eTud2, while unmodified Ago3 interacts with eTud1, however, we previously reported that Aub co-immunoprecipitates with both eTud2 and N+eTud1 fragments (Webster et al., 2015). To investigate this apparent discrepancy, we have repeated co-IP experiments and tested additional Krimp fragments. The new results confirmed our previous observation that Aub coimmunoprecipitates with the N+eTud1 fragment, however, we also found that the eTud1 fragment without N-terminal domain does not co-IP with Aub (Extended Data Fig. 1b). These results demonstrate that eTud1 does not bind Aub, confirming its preference toward Ago3, while they also show that the presence of the self-interacting N-terminal region is indispensable for co-immunoprecipitation of Aub with the N+eTud1 fragment. The most parsimonious explanation for these results is that formation of the complex between Aub and N+eTud1 observed in co-immunoprecipitation experiments is mediated by the full-length Krimper protein expressed in S2 cells that binds the N+eTud1 fragment through its self-interacting N-region.

Overall, our results indicate that binding of both Aub and Ago3 to Krimper is mediated by the RG/RA-motifs embedded in their unstructured N-terminal extended regions. However, the two proteins interact with Krimper in opposite modification states: Aub must be sDMA-modified, while Ago3 remains unmethylated. The distinct specificities of the two Tudor domains explain how Krimper can simultaneously bind Aub and Ago3 to bring them into physical proximity for ping-pong.

**Structural differences in the two Tudor domains of Krimper explain their different binding preference towards Arg-modified Aub and unmodified Ago3**

To gain a detailed understanding of the interaction between Krimper (eTud2) and Aub, we undertook structural studies. We solved the crystal structure of the Krimper (eTud2) domain in complex with the Aub-R15me2 peptide that showed highest binding affinity in ITC experiments (Webster et al., 2015). In the 2.7 Å structure the eTud2 domain adopted a typical Tudor subdomain and a staphylococcal nuclease (SN)-like subdomain (Tudor-SN) architecture (Fig. 2a, 2b and 2c). sDMA-modified Arg-containing peptide binds within a negatively-charged cleft between the Tudor and SN domains, and is specifically recognized by a conserved aromatic cage consisting of four aromatic amino acids Tyr624, Tyr630, Tyr646 and Tyr649 (Fig. 2a, 2d and 2e), stabilized further by hydrophobic and cation- $\pi$  interactions (Fig. 2e). In addition, eTud2 Asn651, which is conserved amongst other extended Tudor domains, forms a hydrogen bond with the guanidyl nitrogen of the methylated Arg15 (Fig. 2e and Extended Data Fig. 2a). Besides Arg15, other flanking arginine residues in the (G/A)R motif of Aub also contribute to protein-peptide binding, (Extended Data Fig. 2a). Together, these structural features of eTud2 explain its specific binding to sDMA-modified Aub.

Unlike eTud2, Krimper (eTud1) has an unusual binding specificity towards the unmethylated Ago3 sequence containing the (G/A)R motif. To elucidate the interaction between eTud1 and Ago3 we solved the crystal structure of eTud1 in both apo and peptide bound forms. In the 2.1 Å eTud1 apo structure, crystals belong to the  $P2_12_12_1$  space group with two monomers in the asymmetric unit (Extended Data Table 1). The overall structure is similar to eTud2 apart from the C-terminal region (484-511aa) that forms a long  $\pi$ -helical fold that inserts into the narrow interface between the Tudor and SN-like subdomains. We named this novel C-terminal helical topology the ‘latch helix’ (Fig. 2f and Extended Data

Fig. 2b). The latch helix is stabilized through formation of extensive hydrogen bonding and salt bridge interactions with its amphipathic concave binding site within eTud1. Detailed interaction patterns and paradigm are shown in Extended Data Fig. 2c.

Next, we determined the crystal structure of eTud1-Ago3-2 peptide complex at 2.4 Å resolution (Extended Data Table 1). The structure of eTud1 in the eTud1-Ago3-2 peptide complex exhibits almost an identical fold to the apo form, with an overall root mean square deviation of 0.60 Å (Extended Data Fig. 2d and 2e). However, the Ago3 peptide displaces the ‘latch helix’ and adopts an  $\alpha$ -helical structure to bind in the same cleft between the Tudor and SN-like subdomains, but with reverse directionality compared to the ‘latch helix’ (Fig. 2g and Extended Data Fig. 2d and 2e). These results suggest that the ‘latch helix’ might regulate binding of Ago3 to eTud1. Indeed, ITC results showed that eTud1 devoid of the latch helix (named eTud1  $\Delta$ C) has a higher binding affinity towards the unmethylated Ago3-2 peptide (Fig. 1f).

The Ago3 peptide binds both a hydrophilic concave and a hydrophobic concave region in the cleft between the Tudor and SN-like subdomains of eTud1 (Fig. 2h). However, instead of a conserved aromatic cage that interacts with sDMA in other canonical Tudor domains including eTud2, the binding of unmethylated Arg70 of Ago3 in eTud1 is provided by a mostly hydrophilic pocket composed of three non-aromatic amino acids (Asp373, Thr380 and Leu397) (Fig. 2h and 2i). Arg70 forms hydrogen bonds with Asp373 and Glu402 in eTud1, supplemented by a cation- $\pi$  interaction with Phe400 (Fig. 2i). Many extended Tudor domains adopts an incomplete aromatic cage, which may prefer unmethylated arginine peptides. In human TDRD2 protein, replacement of one of the four aromatic amino acids in the aromatic cage by a leucine residue, results in the preferential binding of unmethylated

(A/G)R repeat with higher affinity (H. Zhang et al., 2017). In addition, the backbone of Arg68 and the side chain of Arg72 form hydrogen bonds with Asn376 and Thr445, respectively (Extended Data Fig. 2f). Notably, the side chain of Phe400 is rotated by 90° to form a stronger cation- $\pi$  interaction with Arg70 when compared with the structure in the apo state (Extended Data Fig. 2g). Adjacent to the hydrophilic concave region there exists a hydrophobic concave region formed by Phe400, Trp339, and Tyr446. The Ago3 peptide forms a hydrophobic contact with the hydrophobic concave region utilizing Leu73 and Leu77 that strengthens its interactions (Fig. 2h and Extended Data Fig. 2f). The bound unmethylated arginine peptide in the Krimper eTud1 complex adopts the same directionality and similar localization as its counterpart in the TDRD2 complex (PDB code: 6B57) (Extended Data Fig. 3a, 3b and 3c). However, the peptide-binding interface of eTud1 is much narrower compared with the extended and negatively charged interface of TDRD2. Thus, eTud1 specifically recognizes unmodified R(A/G) motif in Ago3 using a unique hydrophilic concave and a hydrophobic concave region in the interface between the Tudor and SN-like subdomains. The N-terminal of Aub sequence with an additional RG motif appears to be more hydrophilic and may not readily fit into the narrow hydrophobic concave cleft of eTud1, while the peptide of Ago3-1 with only one RG motif appears to be unable to form strong interactions with eTud1.

Although the overall topology of eTud1 is similar to other extended Tudor domains, eTud1 has its own unique folding and recognition characteristics. First, the N-terminus of eTud1 contains a long-loop (spanning amino acids 272-300) connecting and interacting with both the Tudor and the SN-like subdomains (Extended Data Fig. 3d). Pro292 within this N-terminal loop forms a CH- $\pi$  interaction with Tyr365, Glu297 in the loop region forms a

hydrogen bonding with the backbone of Ser387, Arg300 in the loop forms a hydrogen bond and salt bridge interactions with Asp394, while Tyr304 in the loop forms cation- $\pi$  interaction with Arg428 in the SN-like domain. All these intramolecular interactions help to stabilize the overall structure of eTud1 (Extended Data Fig. 3d). Second, the linker helix  $\alpha 2$  is shorter and shifted towards the Tudor-SN interface compared with other extended Tudor domains (PDB code: 3OMC), facilitating the interactions with the C-terminal ‘latch helix’ or bound Ago3-2 peptide (Extended Data Fig. 3e).

Overall, our results reveal that the two Tudor domains of Krimper have distinct architectures that are responsible for differential binding of two PIWI proteins. Aub and Ago3 have similar organization, with N-terminal (G/A)R motifs that are subject to sDMA modification *in vivo*. However, while Aub interacts with the eTud2 domain in a methylated state through sDMA binding to a conserved aromatic cage, Ago3 binds the eTud1 domain in an unmethylated state employing a binding pocket that is distinct from other Tudor domain interactions. Furthermore, eTud1 possesses a ‘latch helix’ that must move out of the cleft to allow Ago3 binding providing a potential regulatory mechanism. These results highlight the crucial role that sDMA modifications of PIWI proteins play in regulating the formation of the tertiary ping-pong complex.

### **sDMA modification of Aub is required for piRNA biogenesis through the ping-pong cycle but is dispensable for loading of piRNA into Aub**

Alongside previous studies, our results demonstrate the critical function that the methylation states of Aub and Ago3 plays in their binding to Krimper. In order to investigate

the function of sDMA modification of Aub in the broader piRNA pathway *in vivo*, we generated transgenic flies expressing a methylation-deficient version of Aub (mdAub) by mutating five Arg into Lys in the N-terminal region (Fig. 1e). The methylation of wild-type Aub was readily detected using Western blotting with SYM11 antibody that recognizes sDMA residues (Fig. 3a). In contrast, mdAub produced no signal, indicating that it indeed lacks sDMA modifications. Transgenic mdAub and control wild-type protein were expressed in the ovary at a similar level (Extended Data Fig. 4a). Wild-type Aub localized to nuage, the membrane-less subcellular compartment that surrounds the nucleus of nurse cells and is believed to play an essential role in piRNA biogenesis and TE repression. mdAub is also recruited to nuage, although compared to the wild-type protein a higher fraction is dispersed in the cytoplasm (Fig. 3b, 3c and Extended Data Fig. 4b). FRAP experiments show a slightly increased mobility of mdAub in nuage compared to wild-type Aub (Fig. 3c and Extended Data Fig. 4c). Thus, the methylation status of Aub does not affect protein stability and has only a minor effect on subcellular protein localization.

To explore the function of sDMA modification of Aub *in vivo*, we studied the ability of mdAub to rescue defects observed in Aub mutant flies. We expressed transgenic wild-type and methylation-deficient versions of Aub under the control of the endogenous Aub promoter and used genetic crosses to introduce these transgenes into the *aub*<sup>HN/QC</sup> mutant background. *Aub*<sup>HN/QC</sup> females are completely sterile due to severe defects in oocyte axis specification and DNA damage, presumably caused by derepression of multiple transposon families (Cook, Koppetsch, Wu, & Theurkauf, 2004; Klattenhoff et al., 2007). While the wild-type Aub protein was able to rescue transposon derepression and sterility of the *aub* mutant background, methylation-deficient Aub failed to rescue both phenotypes (Fig. 3d and

3e). The inability of mdAub to rescue sterility and transposon activation suggests that sDMA modification is essential for Aub function *in vivo*.

Sequencing of total small RNA libraries from ovaries of *aub* mutant females that express wild-type or methylation-deficient variants of Aub revealed that piRNAs mapping to different transposon families and major piRNA clusters were severely reduced in the *aub*<sup>HN/QC</sup> mutant (Fig. 4a, 4b and 4d). While expression of methylation-deficient Aub increased piRNA levels, this increase was smaller than that observed in flies rescued with wild-type protein. In ovaries of control (*aub* heterozygotes) flies piRNA are twice as abundant as miRNA, while this ratio drops to only 0.38 (more than 5-fold) in *aub* mutants. Expression of wild-type Aub completely rescues piRNA expression, while expression of methylation-deficient Aub increase the ratio to 0.69, which is higher than the mutant, but still far from levels in the control (Fig. 4a). The analysis of piRNAs targeting individual TE families showed that expression of wild-type protein rescues piRNA loss for all families, while expression of mdAub provides intermediate piRNA level between the mutant and the control (Fig 4b). Both sense and antisense piRNAs to individual TE families are reduced in flies expressing mdAub. To comprehensively analyze the effect of sDMA of Aub on piRNA biogenesis, we analyzed piRNA generation throughout the genome in 5 kb windows. Almost all regions that generate piRNAs with notable exception of uni-strand piRNA clusters such as *flam* and *20A* show dramatic loss of piRNA in the *aub*<sup>HN/QC</sup> mutant. Unlike the wild-type protein, expression of mdAub fails to rescue piRNA generation throughout the genome indicating that sDMA is crucial for piRNA biogenesis (Fig 4b and 4d).

We previously proposed that binding of Aub and Ago3 to Krimper assembles a ping-pong piRNA processing (4P) complex that is essential for the ping-pong cycle (Webster et

al., 2015). To explore the role of Aub methylation in ping-pong we analyzed piRNA profiles of *aub* mutant flies rescued with either wild-type or mdAub to find signatures of ping-pong processing: pairs of piRNAs whose 5' ends are complementary across 10 nt. In the piRNA libraries isolated from ovaries of mdAub-rescued flies, piRNAs with ping-pong processing signatures were reduced for almost all major families of transposons compared to those from the wild type Aub rescue (Fig. 4b), indicating defects in the heterotypic ping-pong cycle. Another signature of ping-pong processing is the presence of sense piRNAs that lack U at the first position but have a bias for A in position 10 (10A bias). 10A bias in sense piRNAs was dramatically reduced for individual TE families in *aub* mutants. While expression of wtAub rescued the bias, mdAub expression led to only partial rescue (Fig. 4b). For all repeat-derived piRNAs, 10A bias among sense non-1U piRNA is 47% in heterozygous control and drops to 28% in *aub* mutants. Expression of wild-type Aub rescued the bias to 49%, while expression of mdAub led to an intermediate value of 38% (Fig. 4e). Mutation of Aub and its rescue by wtAub and mdAub do not change other nucleotide biases in piRNA (Extended Data Fig. 5a).

To explore if disruption of piRNA biogenesis might be caused by an inability of mdAub to be loaded with piRNAs, we expressed the mdAub transgene in wild-type flies that also express endogenous wild-type Aub. These flies have normal fertility and transposon expression (data not shown) indicating that expression of mdAub does not induce a dominant-negative effect. Next, we purified mdAub and control wild-type Aub protein, and radiolabeled isolated RNAs associated with both proteins. Quantification of the signal showed that small RNA loading of wild type and mdAub is similar (Fig. 4f). Small RNAs associated with the two proteins have a similar length distribution (Extended Data Fig. 5b).

To explore if small RNA loaded in mdAub might be different from sequences loaded into wild-type Aub protein we cloned and analyzed small RNAs residing in both complexes. Similar to wild-type Aub protein, the majority of sequences in mdAub are antisense to various transposons families (Fig. 4g). Overall, we did not find any abnormalities in the amount and composition of piRNA loaded into mdAub when it was expressed in the wild-type Aub background. Thus, the methylation status of Aub plays no direct role in piRNA loading.

Amplification of piRNAs in the ping-pong cycle is dependent on the slicer activity of Aub. To test if the mdAub mutation has an effect on its slicer activity, we have performed *in vitro* cleavage assays with proteins purified from S2 cells. When loaded with guide RNA both wild-type and mdAub cleave complementary RNA with similar efficiency and precision indicating that mdAub exhibits normal slicer activity (Extended Data Fig. 4d).

Taken together, our results indicate that sDMA modification of Aub is essential for piRNA biogenesis through piRNA amplification in the ping-pong cycle and hence the repression of transposons. However, mdAub has a normal slicer activity and sDMA modification is not required for loading of piRNA into mdAub if cells also express wild-type Aub that ensures functional piRNA biogenesis. This finding corroborates our earlier proposal that Aub methylation is required for the formation of the Krimper, Aub and Ago3 complex, which in turn is required for Ago3 loading (Webster et al., 2015). Overall, our results indicate that the sDMA modification of Aub has a crucial role in piRNA biogenesis and specifically in the ping-pong amplification cycle.

### **piRNA loading induces Aub methylation**

Our results demonstrate that sDMA modification of Aub is essential for an ability of the piRNA pathway to generate a repertoire of guide piRNAs that effectively target and repress transposons. To understand the molecular mechanism of sDMA modifications we decided to explore if loading of piRNAs into Aub affects its modification. We used flies that express a piRNA binding deficient version of Aub (pdAub) due to mutation in two conserved residues within the 5' end binding pocket located in the MID domain (Webster et al., 2015). We expressed pdAub on a wild-type background to preserve an intact piRNA biogenesis pathway and tested the methylation status of Aub using SYM11 antibody that recognizes methylated Arginine residues. Remarkably, we found that methylation of pdAub is strongly reduced compared to wild-type protein (Fig. 5a) suggesting that prior loading with piRNA is necessary for the post-translational sDMA modification of Aub. We employed a Tudor-Aub binding assay that strictly depends on the Aub methylation (Kirino et al., 2010) as another tool to assess the modification status of Aub. As expected, Tudor co-immunoprecipitated with wild-type Aub, but not mdAub, from ovary lysates (Fig. 5b). Similar to mdAub, pdAub did not bind Tudor protein confirming that Aub piRNA loading is a prerequisite for sDMA modification.

To further explore the link between piRNA loading and sDMA modification we tested Aub interaction with Tudor *in situ*, inside of the cells. In growing oocytes, Tudor protein localizes to the pole plasm, which is the posterior region of the oocyte essential for germ cell specification in the zygote (Arkov, Wang, Ramos, & Lehmann, 2006; Santos & Lehmann, 2004; Thomson & Lasko, 2005). Tudor binds sDMA-modified Aub leading to its accumulation in the pole plasm (Anne & Mechler, 2005). As expected, while wild-type Aub was concentrated in the pole plasm, mdAub failed to be properly localized because it could

not interact with Tudor protein (Fig. 5c). Similarly, to methylation-deficient Aub, pdAub had strong defects in pole plasm localization (Fig. 5d) confirming that piRNA-free Aub is unable to be methylated.

The experiments described above suggest that a defect in piRNA loading impairs Aub methylation, however it is possible that two point mutations introduced into pdAub to impair piRNA binding may cause additional structural changes that render it incapable of methylation. Therefore, we further explored the link between piRNA loading and sDMA modification using wild-type Aub protein. We analyzed the modification status of wild-type Aub in flies that have a low level of piRNAs due to knockdown of Zucchini (Zuc): a nuclease that is essential for piRNA processing (Han, Wang, Li, Weng, & Zamore, 2015; Malone et al., 2009; Mohn, Handler, & Brennecke, 2015). Depletion of Zuc in fly ovaries by RNAi caused a reduction of Aub methylation (Fig. 5e), suggesting that the modification status of wild-type Aub depends on its piRNA loading status.

To directly test if binding of guide piRNA induces Aub modification, we studied if the loading of a chemically synthesized RNA in cell lysate triggers Aub methylation. Argonaute proteins (including members of the PIWI clade) can be loaded with 5'-monophosphorylated short RNA when incubated together *in vitro* (Elkayam et al., 2012; Matsumoto et al., 2016; Yamaguchi et al., 2020). We expressed tagged Aub in S2 cells that lack an active piRNA pathway and do not express piRNA. We added different amounts of 5'-monophosphorylated 26 nt RNA into the lysate of S2 cells that expressed Aub protein and incubated lysate for one hour to allow Aub to be loaded with RNA and the methylation reaction to proceed. Next, we purified Aub and analyzed associated RNA and its modification status (Fig. 5f). The detection of bound RNA indicates that Aub binds exogenous short RNA during incubation

(Fig. 5g). Importantly, the addition of increasing amounts of short RNA led to a progressive increase in Aub methylation (Fig. 5h). Thus, loading of Aub with synthetic RNA triggers sDMA modification in cell lysate, indicating a direct link between RNA binding and modification.

PIWI proteins bind mature piRNAs by anchoring both of its ends: the 5' end is bound by the MID domain, while the 3' end is anchored by the PAZ domain (Simon et al., 2011; Tian, Simanshu, Ma, & Patel, 2011). However, loading and final processing of piRNA guides is coupled, so that longer piRNA precursors are first anchored at their 5' end, followed by trimming of their 3' end before they can be anchored by the PAZ domain (Izumi et al., 2016). To explore if sDMA modification is triggered by 5' end binding or requires complete binding of guide at both ends, we generated another Aub mutant unable to bind the 3' end, by mutating two conserved Tyr residues in its PAZ domain (Tyr345 and Tyr346) that were shown to be indispensable for this process (Yamashiro et al., 2020). Though this mutant has an intact 5' end binding pocket, it has miniscule level of sDMA modification upon addition of the 26nt synthetic RNA (Fig 5i), indicating that complete binding of the guide – both 5' and 3' ends – is required to induce modification.

**piRNA loading induces conformational change and promotes Aub modification by increased accessibility of the N-terminal sDMA motif to the PRMT5 methylosome complex**

How can piRNA loading into Aub induce its methylation? Binding of RNA might induce conformational changes in protein structure that expose previously hidden residues to the solvent and enhance access to enzyme that catalyzes modification. Indeed, structural

studies of Argonaute proteins demonstrated extensive conformational change in protein structure upon binding to guide RNAs (Elkayam et al., 2012; Yashiro et al., 2018). However, the structure of *Drosophila* Aub protein has yet to be solved and the N-terminal extension regions are not present in the available structures of PIWI proteins (Matsumoto et al., 2016; Yamaguchi et al., 2020). To study if loading of RNA into Aub induces its conformation change, we employed the RNA loading strategy described in the previous experiment (Fig. 5f) followed by a partial proteolysis assay. Aub protein with a FLAG tag introduced at its N-terminus was expressed in S2 cells and half of the lysate was incubated with 1  $\mu$ M of synthetic 26 nt RNA for one hour, while RNA was omitted for the other half. After incubation, protein was captured on anti-FLAG beads and subjected to serial dilution of chymotrypsin protease followed by SDS-PAGE gel electrophoresis (Extended Data Fig. 6a). Western blot with anti-FLAG antibody showed that the N-terminus of Aub has different accessibility to protease in free and RNA-bound forms (Extended Data Fig. 6a) suggesting that Aub undergoes a conformational change upon RNA loading.

To further explore how conformation affects Aub methylation we compared the modification status of full-length protein and N-terminal fragment that harbors the (G/A)R motif. Tagged full length Aub and the N-terminus (1-105 aa) fragment were expressed in S2 cells and the methylation status was probed by Western blot with SYM11 antibody. Similar to previous experiments, only a very weak modification was observed on full-length protein, however, we detected strong methylation of the N-terminal fragment (Fig. 6a). This result suggests that the N-terminal region of Aub protein is not easily accessible to modification if the protein is not bound to guide RNA, but the same sequence is methylated strongly if it is apart from the whole protein.

Next we tested if we can alter the accessibility of the N-terminal region to the methylation machinery by inserting an extra sequence between this region and the rest of the protein (named pdAub-EXT). While Aub deficient in piRNA binding was methylated only on a very weak level, insertion of a GFP sequence after AA105 caused robust methylation (Fig. 6b). Thus, insertion of an extended sequence between N-terminus and the rest of the protein converts the N-terminal (GA)R motif into a good substrate for modification even if the protein is not loaded with RNA. Overall, these experiments suggest that the N-terminal region is not readily accessible if Aub is not bound to RNA, however, it can be readily methylated upon solvent exposure.

We tested the biochemical interactions between Aub and the components of the methylosome complex responsible for sDMA modifications. sDMA modifications are catalyzed by conserved methylosome enzyme composed of *Drosophila* PRMT5 homolog Capsuleen (Csul) and MEP50 homolog Valois (Vls) (Friesen et al., 2001; Friesen et al., 2002; Meister et al., 2001) and previous genetic studies implicated both proteins are required for Aub sDMA modifications (Anne & Mechler, 2005; Anne, Olló, Ephrussi, & Mechler, 2007; Gonsalvez, Rajendra, Tian, & Matera, 2006). We found that Csul and Vls co-purifies with Aub in S2 cells (Fig. 6c). We further dissected Aub protein into individual domains and found that the N-terminal fragment (1-220aa) alone was sufficient for binding to Csul, while deletion of this region disrupted the interaction (Extended Data Fig. 6b). This indicated that the methylosome interacts with the N-terminal region of Aub harboring (G/A)R motif, while other regions are dispensable for binding.

To study if piRNA loading affects the interaction of Aub with the methylosome complex we tested binding of Csul with Aub mutants. Csul binds wild-type and mdAub indicating

that arginine residues that are targeted for methylation are dispensable for interaction with Csu1 (Fig. 6d). However, Csu1 did not interact with the piRNA-deficient pdAub mutant (Fig. 6d) suggesting that conformation change induced by binding to RNA is required for interactions with Csu1. Finally, we loaded wild-type Aub with synthesized 26 nt RNA and tested the effect of RNA binding on the interaction with Csu1. Addition of synthetic RNA increased interaction of Aub with Csu1 and subsequent Aub methylation (Fig. 6e). Thus, we determined that increased levels of sDMA modifications upon RNA binding appears to be a direct result of stronger binding of RNA-loaded Aub to the methylosome complex.

## **Discussion**

Ago proteins are present in all domains of life and use a conserved molecular architecture to bind guide nucleic acids and recognize and process complementary targets (Elkayam et al., 2012; Schirle & MacRae, 2012; Song et al., 2004; Wei et al., 2012). Despite the simplicity of the effector complex composed of one protein and one nucleic acid, Argonautes play crucial roles in the control of gene expression and have remarkably diverse sets of targets and functions. The diversity of post-translational modifications expands the regulation and function of Argonaute proteins (Johnston, Geoffroy, Sobala, Hay, & Hutvagner, 2010; Qi et al., 2008; Zeng, Sankala, Zhang, & Graves, 2008). We found that a post-translational modification specific to members of the PIWI clade of the Argonaute family, sDMA in the flexible N-terminal region, encodes information about guide RNA loading status and regulates interactions, cellular localization and function of PIWI proteins.

Although PIWI proteins and piRNAs share many similarities with other Agos and their RNA guides, the piRNA pathway has evolved unique features that are essential for its

function as an adaptive genome defense system. One such unique property is amplification of piRNAs that target active transposons in the ping-pong cycle (Brennecke et al., 2007; Gunawardane et al., 2007). Ping-pong employs the intrinsic RNA binding and processing capabilities of Ago proteins, however, it creates new functionality through the cooperation between two PIWI proteins. Our results indicate that the ping-pong cycle and sDMA-modification are tightly linked and that the modification status of PIWI proteins regulates assembly of the ping-pong processing complex.

### **Molecular mechanism of sDMA regulation**

Several lines of evidence suggest that sDMA modification of Aub is induced by binding of a piRNA guide. First, Aub mutants that are deficient in piRNA binding due to mutation in either the RNA 5' or the 3' end binding pocket have a decreased level of sDMA modification (Fig. 5a,b,i). Second, disruption of piRNA biogenesis diminishes methylation of wild-type Aub (Fig. 5e). Finally, loading of chemically synthesized RNA into Aub promotes its association with the methylosome complex (Fig. 6e) and sDMA modification (Fig. 5g-i, 6e). In contrast, sDMA modification of Aub is not required for its loading with piRNA (Fig. 4f, 4g and Extended Data Fig. 5b) and for its slicer activity (Extended Data Fig. 4d). Together, these results suggest that sDMA modification of Aub acts as a signal of its piRNA-bound state.

Our results suggest that piRNA loading induces sDMA methylation through a conformational change that makes the N-terminal sequence accessible to the methylation enzyme. While unloaded Aub is poorly methylated, the N-terminal sequence alone is a good substrate for methylation (Fig. 6a). Insertion of a sequence between the N-terminal region

and the rest of the protein also promotes methylation (despite the protein not being able to bind piRNA), suggesting that other parts of the protein inhibit modification (Fig. 6b). Finally, partial proteolysis indicates that Aub undergoes a conformation change upon piRNA loading (Extended Data Fig. 6a). Combined, these experiments suggest that the N-terminal sequence is poorly accessible to the modifying enzyme until Aub binds a guide RNA, inducing a conformation change that exposes its N-terminus (Fig. 7a).

Structural differences between empty and loaded states were reported for several prokaryotic and eukaryotic Agos (Doxzen & Doudna, 2017; Elkayam et al., 2012; Miyoshi, Ito, Murakami, & Uchiumi, 2016; Parker, Roe, & Barford, 2005; Rashid et al., 2007; Swarts et al., 2015; Wang, Sheng, Juranek, Tuschl, & Patel, 2008; Willkomm et al., 2017), corroborating the idea that binding to guide RNA induces conformational change. The PAZ domain of Agos exhibit high level of flexibility upon loading of guide RNA/DNA. During the recognition of target RNA, the PAZ domain undergoes a conformational transition that releases the 3' end of the guide and facilitates downstream guide-target base pairing. Our results indicate that binding of the 3' end by the PAZ domain is critical for sDMA modification of Aub's N-terminal region. Unfortunately, the N-terminal extension region was often truncated to facilitate Ago expression and crystallization and thus reported structures do not provide information about the N-terminal extension region (Matsumoto et al., 2016; Nakanishi et al., 2012; Yamaguchi et al., 2020). If the N-terminal region is preserved, it exists in an unstructured conformation that remains unresolved by crystallography (Faehnle et al., 2013; Nakanishi et al., 2013; Schirle & MacRae, 2012). However, piRNA loading of the nuclear PIWI protein in *Drosophila* was shown to induce a conformational change that exposes the nuclear localization sequence (NLS) located in its

N-terminus and to enable its binding to importin (Yashiro et al., 2018). Thus, two PIWI clade proteins, Aub and Piwi, harbor an N-terminal sequence that becomes accessible upon piRNA loading and its exposure promotes interactions with other factors and regulates protein function. Similar to Aub, the N-terminal extension region of Ago3 also harbors a (G/A)R motif that can be modified. Considering that piRNA binding triggers exposure of the N-terminus in Aub and Piwi, a similar process might occur in Ago3. Indeed, previous studies (Sato, Iwasaki, Shibuya, et al., 2015; Webster et al., 2015) and our results revealed that, unlike the bulk of the cellular Ago3 pool, Krimper-bound Ago3 is both unloaded and unmethylated, indicating that piRNA binding and modification are correlated for Ago3 as well as for Aub.

Our results demonstrate the importance of the N-terminal region in the function of PIWI proteins. Unlike other domains (PAZ, MID, PIWI) of Argonautes with well-characterized functions in RNA guide binding and target cleavage, the N-terminal region has received little attention due to its disordered conformation and its low conservation between different Agos (Kwak & Tomari, 2012; Meister et al., 2007; Ryazansky, Kulbachinskiy, & Aravin, 2018; Schirle & MacRae, 2012; Sheng et al., 2014; Wang et al., 2009). Our results suggest that the low conservation and absence of a fixed structure are in fact important features of the N-terminal region that are critical for PIWI proteins function. The flexible structure of this region might provide sensitivity to changes in overall protein conformation, such as the changes triggered by guide RNA binding. In Aub and Ago3, the modification and binding of sDMA sites to other proteins, as well as NLS-mediated interaction of Piwi require only a short linear motif, and thus the N-terminal region does not require a strongly conserved sequence or rigid folding. In agreement with this, the presence of a (G/A)R motif in Aub and

Ago3 proteins is conserved in other *Drosophila* species, however, the specific position and sequence context of the motif is diverse (Fig. 1d and 1e). Poor similarity between N-terminal sequences of different Agos might endow them with distinct functions. It might be worth exploring whether signaling of the guide-loading state through exposure of the N-terminal region is also conserved in Ago-clade proteins and whether it regulates their function.

### **Function of sDMA in the ping-pong cycle**

The central feature of ping-pong is that the cleavage of target RNA by one PIWI protein results in transfer of the cleaved product to another PIWI protein (Fig. 7b). Although the original model of ping-pong did not provide information on the molecular complex and interactions within the complex, ping-pong intuitively implies physical proximity between the two PIWI proteins followed by complex molecular rearrangements. We found that, although sDMA modification does not affect slicer activity of Aub (Extended Data Fig. 4d), information about the piRNA-loading state of PIWI proteins signaled by their sDMA modifications is used to assemble a complex that enables transfer of the processed RNA from Aub to Ago3.

While our previous findings strongly suggest that Krimper plays a role in assembly of the ping-pong piRNA processing (4P) complex in which Aub and Ago3 are brought into close physical proximity (Webster et al., 2015), the architecture of this complex and the extent to which Krimper regulates ping-pong remained unknown. Our results indicate that a single Krimper molecule interacts simultaneously with Aub and Ago3, suggesting that ping-pong takes place within a tertiary complex containing one molecule of each protein. Krimper actively selects the two ping-pong partners using the distinct specificities of its two Tudor

domains: eTud1 uniquely binds Ago3, while eTud2 recognizes modified Aub (Fig. 1f and Extended Data Fig. 1a). We found that *in vitro* the eTud2 domain is capable of binding both sDMA-modified Aub and Ago3 peptides (Extended Data Fig. 1a), however, *in vivo* Krimper complexes were reported to contain exclusively unmodified Ago3 (Sato, Iwasaki, Shibuya, et al., 2015), suggesting that in the proper cellular context eTud2 only binds sDMA-Aub. Thus, the domain architecture of Krimper ensures that tertiary complexes contain Aub-Ago3 partners rather than random pairs. This finding is in line with the observation that ping-pong occurs predominantly between Aub and Ago3 (Brennecke et al., 2007; Gunawardane et al., 2007), although, in principle ping-pong can take place between two identical proteins and a small level of homotypic Aub/Aub ping-pong was previously detected (Li et al., 2009; Z. Zhang et al., 2011). Thus, our results suggest that the propensity for heterotypic ping-pong is, at least in part, due to Krimper (Fig. 7a).

Ping-pong not only requires the physical proximity of two PIWI proteins but also that they have opposite piRNA-loading states: one protein induces piRNA-guided RNA cleavage (and therefore has to be loaded with a piRNA guide), while the other accepts the product of this reaction (and therefore has to be free of piRNA). Our results suggest that the opposite binding preference of the two Tudor domains towards sDMA ensures that the tertiary complex contains PIWI proteins in opposite RNA-loading states. While the overall fold structure of the two Tudor domains is similar, they have critical differences responsible for their distinct binding preferences. The binding pocket of eTud2 is similar to that of other Tudor domains and contains four aromatic residues that interact with sDMA (Fig. 2a and 2e). As sDMA modification of Aub signals its piRNA-binding status, the binding of eTud2 to modified Aub ensures that the complex contains Aub/piRNA. The structural studies and

*in vitro* binding assays revealed that Ago3 binds to eTud1 in its unmethylated state and sDMA modification of any of the Arg residues within its (A/G)R motif prevents this interaction. The unusual binding preference of eTud1 is reflected in its non-canonical binding pocket, which lacks three of the four conserved aromatic residues (Fig. 2i). Binding of methylated Aub and unmethylated Ago3 ensures that Aub has guide piRNA and Ago3 is free, thus enabling loading of Ago3 with RNA generated by Aub/piRNA-induced cleavage (Fig. 7a).

The architecture of the tertiary complex assembled by Krimper permits Aub-dependent generation and loading of RNA into Ago3. However, the ping-pong cycle also includes the opposite step, Ago3-dependent generation of Aub piRNA (henceforth we termed these steps ‘ping’ and ‘pong’). Our results suggest that the ping and pong steps require assembly of two distinct complexes discriminated by the modification status of Aub and Ago3.

### **Formation of a membraneless cellular compartment and its function in ping-pong**

As a single Krimper simultaneously binds Aub and Ago3, Krimper dimerization might be dispensable for ping-pong, raising the question what the function of Krimper self-interaction is. Previous findings suggest that Krimper forms a scaffold for assembly of nuage, a membraneless organelle (MLO) that surrounds nuclei of nurse cells and resembles other MLO possibly formed through liquid-liquid phase separation (Webster et al., 2015). Several lines of evidence point at Krimper as an essential component of nuage that acts as a scaffold for its assembly and the recruitment of client components. First, unlike other nuage components, FRAP measurements show very little Krimper exchange between nuage and the dispersed cytoplasmic compartment. Second, wild type, but not mutant Krimper that

lacks the self-interaction domain, forms cytoplasmic granules upon expression in heterologous cells that do not contain other nuage proteins. In contrast, other nuage components including Aub and Ago3 are dispersed in the cytoplasm when expressed in a similar setting, suggesting that they do not form condensates on their own and rely on other components for recruitment to nuage. Krimper recruits both Aub and Ago3 into MLO that it forms in heterologous cells. Combined, these data indicating that Krimper works as a scaffold and Ago3 and Aub as its clients for nuage assembly. Thus, the interactions between Krimper and the N-terminal regions of Aub and Ago3 is not only essential for the assembly of the tertiary molecular complex but is also responsible for recruitment of these proteins into membraneless cellular compartment (Fig. 7a). The high local concentration of proteins and RNA involved in the piRNA pathway in nuage might enhance the efficiency of ping-pong as well as recognition of RNA targets by Aub and Ago3.

### **Acknowledgements**

We thank members of the Aravin lab for discussion and comments. We thank the BL19U1 beamlines staff at the Shanghai Synchrotron Radiation Facility for assistance during data collection. We thank Igor Antoshechkin (Caltech) for help with sequencing. This work was supported by grants from the National Institutes of Health (R01 GM097363) and by the HHMI Faculty Scholar Award to A.A.A. and the National Natural Science Foundation of China (31870755) and the Guangdong Innovation Research Team Fund (2016ZT06S172) to S.L.

**Author Contributions**

S.L. and D.J.P. conceived and supervised ITC and structural work; A.A.A. conceived and supervised all other experiments. X.H., H.H., F.Z. and A.W. performed the experiments. R.S. developed tools for analysis of small RNA libraries. A.A.A., X.H., K.F.T. and S.L. wrote the manuscript.

## Materials and Methods

### *Fly stocks*

Short hairpin RNA (shRNA) lines used for knockdown including sh-White (BDSC #33623) and sh-Zuc (BDSC #35227), maternal alpha-Tubulin 67C-Gal4 drivers on chromosome two (BDSC #7062) or chromosome three (BDSC #7063) in addition to the Aub mutant stocks aub<sup>HN2</sup> cn<sup>1</sup> bw<sup>1</sup>/CyO (BDSC #8517) and aub<sup>QC42</sup> cn<sup>1</sup> bw<sup>1</sup>/CyO, (BDSC # 4968) were obtained from the Bloomington *Drosophila* Stock Center. Flies were kept on yeast for 2 days and ovaries dissected 5 days after hatching.

### *Generation of Transgenic Fly Lines*

Transgenic constructs for injection were generated using the Gateway cloning system (Life Technologies). cDNAs were obtained by RT-PCR from ovarian or testes RNA of adult *Drosophila melanogaster*, Oregon R strain. mdAub, pdAub, pdAub-EXT and AubPAZmut were generated by overlap PCR and inserted into the pENTR-D-TOPO directional cloning vector (Life Technologies). Transgenes were cloned into the pUASP-Gateway-phiC31 fly injection vector derived from the pCasPeR5-phiC31 vector containing GFP, mKate2 or Strep-FLAG tags using the Gateway cloning system (Life Technologies). The expression of each transgene was controlled using the yeast upstream activation sequence promoter (UASp) stably crossed with a maternal a-Tubulin67c-Gal4-VP16 (MaG4) driver. Transgenes were generated in flies by PhiC31-mediated transformation (BestGene) using PhiC31 landing pads on either chromosome two (BDSC #9736) or chromosome three (BDSC #9750). The GFP-wtAub and GFP-mdAub BAC line was generated by cloning of the *aub*

genomic locus from the BAC clone BACN04M10 into the pCasPeR4 vector using restriction sites XhoI and SpeI. Bacterial recombineering (Gene Bridges Counter Selection kit) was used to insert an in-frame GFP tag in the start site of Aub. GFP-wtAub and GFP-mdAub rescue lines were generated by crossing transgenic construct into the  $aub^{[HN]}/aub^{[QC]}$  background.

#### *Cell Culture, Immunoprecipitation and Western Blots*

Schneider S2 cells were cultured in complete Schneider medium (10% heat inactivated FBS; 100U penicillin [Life technologies]; 100 $\mu$ g streptomycin [Life technologies]). Plasmids were generated using Gateway cloning (Life technologies) using the *Drosophila* Gateway Vector Collection (DGVC) destination vectors, pAGW, pAFW and pAHW for GFP, 3xFLAG and 3XHA tags, respectively, driven by the Actin5C promoter. All constructs were N terminal tagged. Cells were transfected using TransIT-LT1 transfection reagent (Mirus biosciences) according to the manufacturer's recommendation using 3 $\mu$ g of total plasmid, with 1.5 $\mu$ g each for double transfections, or 1.0 $\mu$ g each for triple transfections. S2 cells were lysed in S2 lysis buffer (20mM Tris at pH7.4, 200mM KCl, 0.1% Tween-20, 0.1% Igepal, EDTA-free Complete Protease Inhibitor Cocktail [Roche], 100 $\mu$ g/mL RNase A). Supernatant was cleared by centrifugation at 4,000 x g for 20 minutes at 4°C. Input sample was collected from the supernatant at concentrations of 1-3 $\mu$ g/ $\mu$ L. Anti-FLAG M2 beads (Sigma Aldrich), anti-HA agarose beads (Thermo Fisher) and anti-GFP antibody (Covance) conjugated to Dynabeads (Thermo Fisher) were blocked in 5mg/ml BSA for 10 minutes at 4°C, followed by washing in S2 lysis buffer. Beads were added to the supernatant and rotated at 4°C for 4h,

washed three times in lysis buffer and eluted by boiling in reducing SDS loading buffer. In the two-step co-IP, FLAG beads were eluted using 3X FLAG peptide (Sigma- Aldrich). For co-IP from ovaries, dissected ovaries were lysed in NT2 buffer (50 mM Tris at pH 7.4, 150 mM NaCl, 1mM MgCl<sub>2</sub>, 0.05% Igepal, EDTA-free Complete Protease Inhibitor Cocktail), while for IP experiments RIPA buffer (25 mM Tris at pH 7.4, 150 mM NaCl, 1% NP-40, 0.5 sodium deoxycholate and 0.1% SDS, Igepal, EDTA-free Complete Protease Inhibitor Cocktail) was used for lysis. Lysate was incubated in the presence or absence of 100µg/mL RNase A and cleared by centrifugation. For co-IP experiments, lysates were incubated with anti-FLAG M2 beads (Sigma Aldrich) or with anti-GFP antibody (Covance) conjugated to Dynabeads (Thermo Fisher) at 4°C for 4h. For IP experiments, lysates were incubated with GFP nanobody (Chromo Tek), followed by washing and elution in reducing SDS buffer. Western blot was probed with rabbit anti-GFP (Covance) (1:3K), rabbit anti-Zuc (Endogenous) (1:1K), rabbit SYM11 antibody (Sigma Aldrich) (1:1K), mouse anti-GFP (Santa Cruz) (1:3K), anti-FLAG M2 (Sigma Aldrich) (1:10K), mouse anti-HA (Sigma Aldrich) (1:10K).

#### *piRNA isolation from immunopurified protein complexes*

Immunopurified protein-RNA complexes were spiked with 5pmol of 42-nt RNA oligonucleotide, followed by proteinase K digestion and phenol extraction. Isolated RNA was CIP treated, radiolabeled using PNK and gamma-P32 ATP, and run on a 15% urea-PAGE gel. Semi-Quantitative piRNA binding analysis and a detailed protocol can be found in a previous paper (Webster et al., 2015).

### *Small RNA-seq*

Small RNA libraries were cloned from total ovarian lysate of Aub heterozygous, *aub*<sup>[HN]</sup>/*aub*<sup>[QC]</sup> mutant and *aub*<sup>[HN]</sup>/*aub*<sup>[QC]</sup>, GFP-wtAub or GFP-mdAub rescue flies. 30 ovaries were dissected, and total RNA was isolated with Ribozol (Amresco, N580). Small RNAs within a 19- to 29-nt window were isolated from 15% polyacrylamide gels from 4 µg of ovarian total RNA. For samples that were NaIO<sub>4</sub>-treated, 5× borate buffer (pH 8.6; 150 mM borax, 150 mM boric acid) and 200 mM sodium periodate were added to the size-selected small RNA, and the samples were incubated for 30 min at 25°C. The NaIO<sub>4</sub>-treated small RNA was then ethanol-precipitated before proceeding to library construction. The small RNA libraries were constructed using the NEBNext small RNA library preparation set for Illumina (no. E7330S), according to the protocol, using NEBNext multiplex oligos for Illumina (no. E7335S). Libraries were sequenced on the Illumina HiSeq 2500 (SE 50-bp reads) platform.

### *Immunoprecipitation small RNA-seq*

Ovaries (~100 per immunoprecipitation) from flies expressing GFP-wtAub and GFP-mdAub under the control of endogenous promoter were dissected and lysed on ice in 250 µL of lysis buffer [30 mM Hepes-KOH at pH7.4, 100 mM KOAc, 2 mM Mg(OAc)<sub>2</sub>, 5 mM DTT, 0.5% (v/v) NP40, proteinase inhibitor (Roche, 11836170001), RNasin Plus (Promega, N2611)]. Lysate was dounced and clarified by centrifugation at maximum speed at 4°C. The supernatant was incubated with rabbit polyclonal anti-GFP (Covance) conjugated to Dynabeads (Thermo Fisher) for 4 h at 4°C. The immunoprecipitation and RNA isolation were carried out as described previously (Vagin et al., 2006). A fifth of the RNA was CIP-

treated (New England Biolabs, M0290S) in NEB buffer #3 (New England Biolabs, B7003S) for 30 min at 37°C and then ethanol-precipitated after phenol:chloroform and chloroform extraction. The CIP-treated RNA was then PNK-treated with 1 µL of 10× T4 polynucleotide kinase buffer (New England Biolabs, B0201S), 2 µL of [ $\gamma$ -P<sup>32</sup>] ATP (PerkinElmer, BLU502A250UC), and 1 µL of T4 polynucleotide kinase (New England Biolabs, M0201S) for 45 min at 37°C. The CIP- and PNK-treated RNA was added back to the remainder of the RNA isolated from the immunoprecipitation. Size selection, library preparation, and analysis were performed as described in small RNA-seq, except that fragments were gel-extracted based on labeled immunoprecipitation material.

#### *Sequence analysis of piRNA libraries*

After adapter trimming, reads were mapped to the dm3 assembly of the *Drosophila melanogaster* genome using Bowtie 0.12.7 (Langmead, Trapnell, Pop, & Salzberg, 2009). Only reads that map to the genome with no mismatches were considered for further analysis. piRNAs were defined as 23-29 nt reads that mapped to the repeatmasker track (<http://www.girinst.org/>), while miRNAs were defined based on their mapping to annotated miRNA genes

To analyze piRNA generation throughout the genome, the genome was split in 5 kb intervals and windows that produce at least 5 reads per million of miRNAs in control (*aub* heterozygous) and showed more than 80% reduction of piRNAs in the *aub* mutants were further analyzed. To compare piRNA abundance between the libraries piRNA reads were normalized to total miRNA reads. For detailed comparison of piRNA generation from 42AB, 38C, 80EF and *flamenco* clusters, 1kb genomic intervals were used.

To analyze piRNAs to individual TE families piRNA reads were mapped to TE consensus sequences ([http://www.fruitfly.org/p\\_disrupt/TE.html](http://www.fruitfly.org/p_disrupt/TE.html)) allowing for up to 2 mismatches. To compare TE piRNA abundance between the libraries piRNA reads were normalized to total miRNA reads. The fold-change in read count for each TE family was calculated by obtaining the base 2 logarithm of ratio of normalized reads in experimental libraries and heterozygous control libraries. The fraction of ping-pong piRNA pairs for each TE family was obtained using analysis described in a previous study (Webster et al., 2015). To compare abundance of TE piRNA associated with wtAub- and mdAub, reads mapping to each TE family were normalized to total library reads (RPM) after rRNA reads were discarded.

### *Microscopy*

Ovaries were dissected in PBS and fixed in 4% PFA in PBS for 20 minutes, washed 3X10 min at RT, permeabilized in 1% Triton-X100 in PBS for 10 minutes, and DAPI stained (Sigma-Aldrich). Ovaries were washed in PBS and mounted in Vectashield medium (Vector Labs). S2 cells were allowed to settle on coverslips treated with Poly-L-Lysine (Sigma-Aldrich). After gentle washing, cells were fixed in 0.5% PFA in PBS for 20 minutes followed by staining with DAPI (Sigma-Aldrich), washed and mounted in Vectashield medium (VectorLabs). Images were captured using an AxioImager microscope; an Apotome structured illumination system was used for optical sections (Carl Zeiss).

### *In vitro cleavage assay*

The *in vitro* assay for Aub slicer activity was adapted from a method described previously (Gunawardane et al., 2007). FLAG-tagged wild-type and mdAub proteins were

expressed and immunopurified from S2 cells. Proteins were eluted using 3X Flag peptide (Thermo Fisher). 50K MWCO protein concentrator (Thermo Fisher) was used to remove the 3X Flag peptide. 100 nM purified proteins were incubated with 100 nM of single-strand 26 nt guide RNA in cleavage buffer (25mM Hepes-KOH, pH 7.5, 50 mM KOAc, 5 mM Mg(OAc)<sub>2</sub>, 5 mM DTT) for 90 min at 25°C, and then <sup>32</sup>P-5'-labeled 29 nt complementary RNA (PerkinElmer, BLU502A250UC) was added and incubated for another 90 min. The cleaved products were analyzed on urea-containing PAGE.

#### *Limited protease assay*

10ng/ul chymotrypsin stock solution was prepared in chymotrypsin reaction buffer (10 mM Tris-HCl [pH 8.0], 2 mM CaCl<sub>2</sub>, and 5% glycerol). FLAG-wtAub was expressed in S2 cells. Lysates were divided into two equal fractions, one was incubated with 26nt ssRNA oligo for 1h at RT, another was incubated without oligos, followed by immunoprecipitation using anti-FLAG M2 beads. Beads were washed three times with lysis buffer and then incubated with a 1:2K, 1:4K and 1:8K serial dilution of the thermolysin protease for 30 min at 37°C. After extensive washing, beads were eluted in reducing SDS buffer. Samples were analyzed by western blot using anti-FLAG M2 antibody.

#### *Protein disorder prediction and conservation analysis*

Disorder predictions of full-length Krimper and the N terminal region of Aub and Ago3 were obtained using the PrDOS server based on a previously used algorithm (Ishida & Kinoshita, 2007). Conservation is measured as a numerical index reflecting the conservation of physico-chemical properties of amino acids in the alignment: Identities score highest, followed by

substitutions to amino acids lying in the same physico-chemical class. A detailed description of the algorithm can be found in (Livingstone & Barton, 1993).

#### *Hatching rate calculation*

3-day old mated adult female flies fed on yeast paste were transferred to fresh grape agar plates and allowed to lay eggs for 12 hours. Eggs were counted, and the hatching rate was determined over the following 36h hours. Counting was repeated on 10 consecutive days.

#### *RT-qPCR*

Total RNA was isolated from 20 ovaries with Ribozol (Amresco, N580). RNA was DNase I treated (Invitrogen, 18068-015) and reverse transcribed using SuperScript III (Invitrogen) with oligo d(T)<sub>15</sub> according to the manufacturer's recommendations. RT-qPCR was performed using Mastercycler ep Realplex PCR thermal cycler machine (Eppendorf). Target expression was normalized to *rp49*. Primers are shown in Table S2.

#### *Fluorescence Recovery After Photobleaching (FRAP)*

For each construct, at least 20 independent FRAP experiments were performed using ovaries expressing a single GFP-tagged transgene under the endogenous promoter. FRAP experiments were captured on a Zeiss LSM710 confocal microscope (Carl Zeiss AIM) equipped with a 25x/0.8 NA Imm. Corr. Multi-immersion objective lens using the Zeiss Zen Black software. Image acquisition for all experiments utilized an identical 488nm AOTF laser power setting of 7% to ensure laser power was not influencing measurements. PMT gain settings were variably set to accommodate the expression level of each GFP tagged

protein. Images were acquired at 256 x 256-pixel resolution at 0.07 $\mu$ m pixel size and scan speed of 614.4 ms per frame with 1.0  $\mu$ s pixel dwell time. A single bleach region was defined for each experiment, consisting of a region of 7x7 pixels equal to 0.49  $\mu$ m x 0.49  $\mu$ m and was bleached by a single iteration of 100% laser power from 488, 561 and 633nm wavelengths. Five initial pre-bleach images were captured prior to bleaching and 115 subsequent postbleach images were acquired every 614.4ms to assess fluorescence recovery in the bleach zone. FIJI (FIJI; <http://fiji.sc/>) software was used to analyze FRAP experiments. Detailed analysis of the mobile fraction and the nuage/cytoplasm ratio can be found in a previous study (Webster et al., 2015).

#### *RNA binding in cell lysates and measurement of methylation level*

S2 cells were transfected with 10 $\mu$ g of plasmid expressing N terminally Flag tagged wtAub driven by the Actin5C promoter. After 48h incubation, cells were lysed in S2 lysis buffer (20mM Tris at pH7.4, 200mM KCl, 0.1% Tween-20, 0.1% Igepal, EDTA-free Complete Protease Inhibitor Cocktail [Roche], 100 $\mu$ g/mL RNase A). Lysates were divided into four equal fractions. 26nt synthetic RNA oligo (IDT) labelled with [ $\gamma$ -P<sup>32</sup>] ATP (PerkinElmer, BLU502A250UC) as described above was added into lysate fractions to final concentration 1 $\mu$ M. Lysates were incubated at RT for 1h followed by IP with FLAG M2 beads at 4°C for 4h. Half of the beads were subject to RNA isolation, the other half were used for western blot. In the oligo binding concentration gradient experiment, lysates were divided into four equal fractions. non-radioactivity labelled 5'-end phosphorylated ssRNA oligo (IDT) was added into each fraction to a final concentration of 0, 0.1, 0.5 and 1 $\mu$ M, respectively. Lysates were incubated at RT for 1h followed by anti-FLAG IP at 4°C for 4h. Protein and methylation

levels were detected by Western blot and methylation level was estimated as the ratio of the background subtracted methylation signal intensity and the background subtracted protein signal intensity.

Relative methylation level was measured by normalizing to the methylation level of the no oligo control.

#### *Protein expression and purification*

Krimper eTud1 (residues 272-512) was cloned into a self-modified pSumo vector with 10xHis tag followed by a yeast sumo sequence. The plasmid was transformed into *E. coli* strain BL21(DE3) Rosseta and cultured at 37 °C in LB medium. The protein expression was induced by adding IPTG to a final concentration of 0.2 mM when the OD600 reached 0.7, and the cells were cooled to 16 °C. The recombinant expressed protein was purified using a HisTrap column (GE Healthcare). The hexahistidine plus yeast sumo tag was removed by ulp1 protease digestion followed by a second step HisTrap column (GE Healthcare). The target protein was further purified using MonoQ and Superdex G75 columns (GE Healthcare).

Krimper eTud2 (residues 562-746) was cloned into a self-modified His-MBP vector. Protein production procedure was to with Krimper eTud1. The recombinant expressed protein was purified using a HisTrap column (GE Healthcare). The hexahistidine plus MBP tag was removed by TEV protease digestion followed by a second step HisTrap column (GE Healthcare). Protein was further purified using Q and Superdex G75 columns (GE Healthcare).

*Crystallization and crystal growth*

Krimper eTud1 (272-512) was concentrated to 20 mg/ml and screened in sitting drop at 4 °C. Crystals were grown for 5 days in 0.1 M HEPES pH 7.5, 1.26 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>. For obtaining the protein-peptide complex, 20 mg/ml Krimper eTud1 was mixed with Ago3-2 peptide with a molar ratio of 1:4 and incubated for 1h before setting up sitting drop screening at 4 °C. Crystals were grown for 5 days in 0.1 M HEPES pH 7.5, 1.5 Li<sub>2</sub>SO<sub>4</sub>. 10 mg/ml Krimper eTud2 was mixed with AubR15me2 peptide with a molar ratio of 1:4 and incubated for 1h before sitting drop. Crystals appeared in 2 days in 16% PEG3350 and 0.1 M NH<sub>4</sub>Ac.

*Data collection and structure determination*

All data were collected at the Shanghai Synchrotron Radiation Facilities beamline 18U1 (SSRF-BL18U1) and beamline 19U1 (SSRF-BL19U1). Data were processed using the program HKL3000 (Minor, Cymborowski, Otwinowski, & Chruszcz, 2006). Structures of eTud1-Ago3 complex and eTud2-Aub complex were solved by SAD using the anomalous signals of SeMet using the program Phenix (Adams et al., 2010). eTud1 apo structure was solved by molecular replacement using the program Phenix (Adams et al., 2010) with eTud1-Ago3 complex as a search model. Models were refined in Phenix (Adams et al., 2010) and COOT (Emsley, Lohkamp, Scott, & Cowtan, 2010) iteratively and finally presented using Pymol (DeLano Scientific).

*ITC*

All ITC was performed using the Malvern PEAQ ITC instrument (Malvern) in a buffer of 50 mM NaCl, 20 mM HEPES pH 7.5 and 2 mM  $\beta$ -mercaptoethanol. Data analysis was performed using the Malvern data analysis software and Origin 7.0.

*Data availability*

Small RNA-seq data are available on the GEO database, GSE153156.

X-ray structures have been deposited in the RCSB Protein Data Bank with the accession codes: 7CFB for the eTud1 apo structure, 7CFC for the eTud1-Ago3 complex structure and 7CFD for the eTud2-AubR15me2 structure.

## References

- Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., . . . Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*, *66*(Pt 2), 213-221. doi:10.1107/S0907444909052925
- Anne, J., & Mechler, B. M. (2005). Valois, a component of the nuage and pole plasm, is involved in assembly of these structures, and binds to Tudor and the methyltransferase Capsuleen. *Development*, *132*(9), 2167-2177. Retrieved from <Go to ISI>://WOS:000229606100015
- Anne, J., Ollo, R., Ephrussi, A., & Mechler, B. M. (2007). Arginine methyltransferase Capsuleen is essential for methylation of spliceosomal Sm proteins and germ cell formation in *Drosophila*. *Development*, *134*(1), 137-146. doi:10.1242/dev.02687
- Aravin, A. A., Naumova, N. M., Tulin, A. V., Vagin, V. V., Rozovsky, Y. M., & Gvozdev, V. A. (2001). Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D-melanogaster* germline. *Current Biology*, *11*(13), 1017-1027. doi:Doi 10.1016/S0960-9822(01)00299-8
- Aravin, A. A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K. F., . . . Hannon, G. J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Molecular Cell*, *31*(6), 785-799. Retrieved from <Go to ISI>://WOS:000259721600004
- Arkov, A. L., Wang, J. Y., Ramos, A., & Lehmann, R. (2006). The role of Tudor domains in germline development and polar granule architecture. *Development*, *133*(20), 4053-4062. doi:10.1242/dev.02572
- Boland, A., Huntzinger, E., Schmidt, S., Izaurralde, E., & Weichenrieder, O. (2011). Crystal structure of the MID-PIWI lobe of a eukaryotic Argonaute protein. *Proc Natl Acad Sci U S A*, *108*(26), 10466-10471. doi:10.1073/pnas.1103946108
- Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., & Hannon, G. J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell*, *128*(6), 1089-1103. doi:10.1016/j.cell.2007.01.043
- Carmell, M. A., Girard, A., van de Kant, H. J. G., Bourc'his, D., Bestor, T. H., de Rooij, D. G., & Hannon, G. J. (2007). MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Developmental Cell*, *12*(4), 503-514. doi:10.1016/j.devcel.2007.03.001
- Carmell, M. A., Xuan, Z., Zhang, M. Q., & Hannon, G. J. (2002). The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev*, *16*(21), 2733-2742. doi:10.1101/gad.1026102
- Chen, C., Nott, T. J., Jin, J., & Pawson, T. (2011). Deciphering arginine methylation: Tudor tells the tale. *Nat Rev Mol Cell Biol*, *12*(10), 629-642. doi:10.1038/nrm3185
- Cook, H. A., Koppetsch, B. S., Wu, J., & Theurkauf, W. E. (2004). The *Drosophila* SDE3 homolog armitage is required for oskar mRNA silencing and embryonic axis specification. *Cell*, *116*(6), 817-829. Retrieved from <Go to ISI>://WOS:000221499900008
- Cox, D. N., Chao, A., Baker, J., Chang, L., Qiao, D., & Lin, H. F. (1998). A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-

- renewal. *Genes & Development*, 12(23), 3715-3727. doi:DOI 10.1101/gad.12.23.3715
- Cox, D. N., Chao, A., & Lin, H. (2000). piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development*, 127(3), 503-514. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/10631171>
- Doxzen, K. W., & Doudna, J. A. (2017). DNA recognition by an RNA-guided bacterial Argonaute. *Plos One*, 12(5). Retrieved from <Go to ISI>://WOS:000401487700036
- Elkayam, E., Kuhn, C. D., Tocilj, A., Haase, A. D., Greene, E. M., Hannon, G. J., & Joshua-Tor, L. (2012). The structure of human argonaute-2 in complex with miR-20a. *Cell*, 150(1), 100-110. doi:10.1016/j.cell.2012.05.017
- Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*, 66(Pt 4), 486-501. doi:10.1107/S0907444910007493
- Faehnle, C. R., Elkayam, E., Haase, A. D., Hannon, G. J., & Joshua-Tor, L. (2013). The making of a slicer: activation of human Argonaute-1. *Cell Reports*, 3(6), 1901-1909. doi:10.1016/j.celrep.2013.05.033
- Frank, F., Sonenberg, N., & Nagar, B. (2010). Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature*, 465(7299), 818-822. doi:10.1038/nature09039
- Friesen, W. J., Paushkin, S., Wyce, A., Massenet, S., Pesiridis, G. S., Van Duyne, G., . . . Dreyfuss, G. (2001). The methylosome, a 20S complex containing JBP1 and pICln, produces dimethylarginine-modified Sm proteins. *Molecular and Cellular Biology*, 21(24), 8289-8300. Retrieved from <Go to ISI>://WOS:000172315800005
- Friesen, W. J., Wyce, A., Paushkin, S., Abel, L., Rappsilber, J., Mann, M., & Dreyfuss, G. (2002). A novel WD repeat protein component of the methylosome binds Sm proteins. *Journal of Biological Chemistry*, 277(10), 8243-8247. Retrieved from <Go to ISI>://WOS:000174268000079
- Girard, A., Sachidanandam, R., Hannon, G. J., & Carmell, M. A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature*, 442(7099), 199-202. doi:10.1038/nature04917
- Gonsalvez, G. B., Rajendra, T. K., Tian, L. P., & Matera, A. G. (2006). The Sm-protein methyltransferase, dart5, is essential for germ-cell specification and maintenance. *Current Biology*, 16(11), 1077-1089. doi:10.1016/j.cub.2006.04.037
- Gunawardane, L. S., Saito, K., Nishida, K. M., Miyoshi, K., Kawamura, Y., Nagami, T., . . . Siomi, M. C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science*, 315(5818), 1587-1590. doi:10.1126/science.1140494
- Han, B. W., Wang, W., Li, C. J., Weng, Z. P., & Zamore, P. D. (2015). piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. *Science*, 348(6236), 817-821. doi:10.1126/science.aaa1264
- Handler, D., Olivieri, D., Novatchkova, M., Gruber, F. S., Meixner, K., Mechtler, K., . . . 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(19), 3977-3993. doi:10.1038/emboj.2011.308

- Honda, S., Kirino, Y., & Kirino, Y. (2014). Analysis of sDMA modifications of PIWI proteins. *Methods Mol Biol*, *1093*, 137-148. doi:10.1007/978-1-62703-694-8\_11
- Huang, X., Fejes Toth, K., & Aravin, A. A. (2017). piRNA Biogenesis in *Drosophila melanogaster*. *Trends Genet*, *33*(11), 882-894. doi:10.1016/j.tig.2017.09.002
- Ishida, T., & Kinoshita, K. (2007). PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Research*, *35*(Web Server issue), W460-464. doi:10.1093/nar/gkm363
- Izumi, N., Shoji, K., Sakaguchi, Y., Honda, S., Kirino, Y., Suzuki, T., . . . Tomari, Y. (2016). Identification and Functional Analysis of the Pre-piRNA 3' Trimmer in Silkworms. *Cell*, *164*(5), 962-973. doi:10.1016/j.cell.2016.01.008
- Johnston, M., Geoffroy, M. C., Sobala, A., Hay, R., & Hutvagner, G. (2010). HSP90 protein stabilizes unloaded argonaute complexes and microscopic P-bodies in human cells. *Mol Biol Cell*, *21*(9), 1462-1469. doi:10.1091/mbc.E09-10-0885
- Kalmykova, A. I., Klenov, M. S., & Gvozdev, V. A. (2005). Argonaute protein PIWI controls mobilization of retrotransposons in the *Drosophila* male germline. *Nucleic Acids Research*, *33*(6), 2052-2059. doi:10.1093/nar/gki323
- Khurana, J. S., & Theurkauf, W. (2010). piRNAs, transposon silencing, and *Drosophila* germline development. *Journal of Cell Biology*, *191*(5), 905-913. doi:10.1083/jcb.201006034
- Kirino, Y., Kim, N., de Planell-Sagner, M., Khandros, E., Chiorean, S., Klein, P. S., . . . Mourelatos, Z. (2009). Arginine methylation of Piwi proteins catalysed by dPRMT5 is required for Ago3 and Aub stability. *Nat Cell Biol*, *11*(5), 652-658. doi:10.1038/ncb1872
- Kirino, Y., Vourekas, A., Sayed, N., de Lima Alves, F., Thomson, T., Lasko, P., . . . Mourelatos, Z. (2010). Arginine methylation of Aubergine mediates Tudor binding and germ plasm localization. *RNA*, *16*(1), 70-78. doi:10.1261/rna.1869710
- Klattenhoff, C., Bratu, D. P., McGinnis-Schultz, N., Koppetsch, B. S., Cook, H. A., & Theurkauf, W. E. (2007). *Drosophila* rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. *Developmental Cell*, *12*(1), 45-55. doi:10.1016/j.devcel.2006.12.001
- Kwak, P. B., & Tomari, Y. (2012). The N domain of Argonaute drives duplex unwinding during RISC assembly. *Nature Structural & Molecular Biology*, *19*(2), 145-151. Retrieved from <Go to ISI>://WOS:000300029400005
- Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, *10*(3), R25. doi:10.1186/gb-2009-10-3-r25
- Li, C. J., Vagin, V. V., Lee, S. H., Xu, J., Ma, S. M., Xi, H. L., . . . Zamore, P. D. (2009). Collapse of Germline piRNAs in the Absence of Argonaute3 Reveals Somatic piRNAs in Flies. *Cell*, *137*(3), 509-521. Retrieved from <Go to ISI>://WOS:000265677200020
- Liu, H., Wang, J. Y., Huang, Y., Li, Z., Gong, W., Lehmann, R., & Xu, R. M. (2010). Structural basis for methylarginine-dependent recognition of Aubergine by Tudor. *Genes Dev*, *24*(17), 1876-1881. doi:10.1101/gad.1956010

- Liu, K., Chen, C., Guo, Y., Lam, R., Bian, C., Xu, C., . . . Min, J. (2010). Structural basis for recognition of arginine methylated Piwi proteins by the extended Tudor domain. *Proc Natl Acad Sci U S A*, *107*(43), 18398-18403. doi:10.1073/pnas.1013106107
- Livingstone, C. D., & Barton, G. J. (1993). Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Comput Appl Biosci*, *9*(6), 745-756. doi:10.1093/bioinformatics/9.6.745
- Ma, J. B., Ye, K., & Patel, D. J. (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature*, *429*(6989), 318-322. doi:10.1038/nature02519
- Malone, C. D., Brennecke, J., Dus, M., Stark, A., McCombie, W. R., Sachidanandam, R., & Hannon, G. J. (2009). Specialized piRNA Pathways Act in Germline and Somatic Tissues of the Drosophila Ovary. *Cell*, *137*(3), 522-535. doi:10.1016/j.cell.2009.03.040
- Martinez, J., & Tuschl, T. (2004). RISC is a 5' phosphomonoester-producing RNA endonuclease. *Genes Dev*, *18*(9), 975-980. doi:10.1101/gad.1187904
- Matsumoto, N., Nishimasu, H., Sakakibara, K., Nishida, K. M., Hirano, T., Ishitani, R., . . . Nureki, O. (2016). Crystal Structure of Silkworm PIWI-Clade Argonaute Siwi Bound to piRNA. *Cell*, *167*(2), 484-497 e489. doi:10.1016/j.cell.2016.09.002
- Meister, G., Eggert, C., Buhler, D., Brahms, H., Kambach, C., & Fischer, U. (2001). Methylation of Sm proteins by a complex containing PRMT5 and the putative U snRNP assembly factor pICln. *Current Biology*, *11*(24), 1990-1994. Retrieved from <Go to ISI>://WOS:000172754700030
- Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., & Tuschl, T. (2007). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs (Reprinted from Molecular Cell, vol 15, pg 185-197, 2004). *Cell*, *131*(4), 50-62. Retrieved from <Go to ISI>://WOS:000252438800006
- Minor, W., Cymborowski, M., Otwinowski, Z., & Chruszcz, M. (2006). HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes. *Acta Crystallogr D Biol Crystallogr*, *62*(Pt 8), 859-866. doi:10.1107/S0907444906019949
- Miyoshi, T., Ito, K., Murakami, R., & Uchiumi, T. (2016). Structural basis for the recognition of guide RNA and target DNA heteroduplex by Argonaute. *Nature Communications*, *7*. Retrieved from <Go to ISI>://WOS:000379075800001
- Mohn, F., Handler, D., & Brennecke, J. (2015). piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. *Science*, *348*(6236), 812-817. doi:10.1126/science.aaa1039
- Nakanishi, K., Ascano, M., Gogakos, T., Ishibe-Murakami, S., Serganov, A. A., Briskin, D., . . . Patel, D. J. (2013). Eukaryote-specific insertion elements control human ARGONAUTE slicer activity. *Cell Rep*, *3*(6), 1893-1900. doi:10.1016/j.celrep.2013.06.010
- Nakanishi, K., Weinberg, D. E., Bartel, D. P., & Patel, D. J. (2012). Structure of yeast Argonaute with guide RNA. *Nature*, *486*(7403), 368-374. doi:10.1038/nature11211
- Nishida, K. M., Okada, T. N., Kawamura, T., Mituyama, T., Kawamura, Y., Inagaki, S., . . . Siomi, M. C. (2009). Functional involvement of Tudor and dPRMT5 in the piRNA

- processing pathway in *Drosophila* germlines. *Embo Journal*, 28(24), 3820-3831. doi:10.1038/emboj.2009.365
- Nishida, K. M., Saito, K., Mori, T., Kawamura, Y., Nagami-Okada, T., Inagaki, S., . . . Siomi, M. C. (2007). Gene silencing mechanisms mediated by Aubergine-piRNA complexes in *Drosophila* male gonad. *Rna-a Publication of the Rna Society*, 13(11), 1911-1922. doi:10.1261/rna.744307
- Park, M. S., Araya-Secchi, R., Brackbill, J. A., Phan, H. D., Kehling, A. C., Abd El-Wahab, E. W., . . . Nakanishi, K. (2019). Multidomain Convergence of Argonaute during RISC Assembly Correlates with the Formation of Internal Water Clusters. *Mol Cell*, 75(4), 725-740 e726. doi:10.1016/j.molcel.2019.06.011
- Parker, J. S., Roe, S. M., & Barford, D. (2004). Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. *EMBO J*, 23(24), 4727-4737. doi:10.1038/sj.emboj.7600488
- Parker, J. S., Roe, S. M., & Barford, D. (2005). Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature*, 434(7033), 663-666. Retrieved from <Go to ISI>://WOS:000228011000049
- Qi, H. H., Ongusaha, P. P., Myllyharju, J., Cheng, D., Pakkanen, O., Shi, Y., . . . Shi, Y. (2008). Prolyl 4-hydroxylation regulates Argonaute 2 stability. *Nature*, 455(7211), 421-424. doi:10.1038/nature07186
- Rashid, U. J., Paterok, D., Koglin, A., Gohlke, H., Piehler, J., & Chen, J. C. H. (2007). Structure of *Aquifex aeolicus* Argonaute highlights conformational flexibility of the PAZ domain as a potential regulator of RNA-induced silencing complex function. *Journal of Biological Chemistry*, 282(18), 13824-13832. Retrieved from <Go to ISI>://WOS:000246060300072
- Reuter, M., Chuma, S., Tanaka, T., Franz, T., Stark, A., & Pillai, R. S. (2009). Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile. *Nat Struct Mol Biol*, 16(6), 639-646. doi:10.1038/nsmb.1615
- Ryazansky, S., Kulbachinskiy, A., & Aravin, A. A. (2018). The Expanded Universe of Prokaryotic Argonaute Proteins. *Mbio*, 9(6). Retrieved from <Go to ISI>://WOS:000454730100038
- Santos, A. C., & Lehmann, R. (2004). Germ cell specification and migration in *Drosophila* and beyond. *Current Biology*, 14(14), R578-589. doi:10.1016/j.cub.2004.07.018
- Sato, K., Iwasaki, Y. W., Shibuya, A., Carninci, P., Tsuchizawa, Y., Ishizu, H., . . . Siomi, H. (2015). Krimper Enforces an Antisense Bias on piRNA Pools by Binding AGO3 in the *Drosophila* Germline. *Molecular Cell*, 59(4), 553-563. doi:10.1016/j.molcel.2015.06.024
- Sato, K., Iwasaki, Y. W., Siomi, H., & Siomi, M. C. (2015). Tudor-domain containing proteins act to make the piRNA pathways more robust in *Drosophila*. *Fly (Austin)*, 9(2), 86-90. doi:10.1080/19336934.2015.1128599
- Schirle, N. T., & MacRae, I. J. (2012). The Crystal Structure of Human Argonaute2. *Science*, 336(6084), 1037-1040. Retrieved from <Go to ISI>://WOS:000304406800046
- Schirle, N. T., Sheu-Gruttadauria, J., & MacRae, I. J. (2014). Structural basis for microRNA targeting. *Science*, 346(6209), 608-613. doi:10.1126/science.1258040

- Sheng, G., Zhao, H. T., Wang, J. Y., Rao, Y., Tian, W. W., Swarts, D. C., . . . Wang, Y. L. (2014). Structure-based cleavage mechanism of *Thermus thermophilus* Argonaute DNA guide strand-mediated DNA target cleavage. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(2), 652-657. Retrieved from <Go to ISI>://WOS:000329614500031
- Simon, B., Kirkpatrick, J. P., Eckhardt, S., Reuter, M., Rocha, E. A., Andrade-Navarro, M. A., . . . Carlomagno, T. (2011). Recognition of 2'-O-methylated 3'-end of piRNA by the PAZ domain of a Piwi protein. *Structure*, *19*(2), 172-180. doi:10.1016/j.str.2010.11.015
- Siomi, M. C., Mannen, T., & Siomi, H. (2010). How does the royal family of Tudor rule the PIWI-interacting RNA pathway? *Genes Dev*, *24*(7), 636-646. doi:10.1101/gad.1899210
- Song, J. J., Liu, J., Tolia, N. H., Schneiderman, J., Smith, S. K., Martienssen, R. A., . . . Joshua-Tor, L. (2003). The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nat Struct Biol*, *10*(12), 1026-1032. doi:10.1038/nsb1016
- Song, J. J., Smith, S. K., Hannon, G. J., & Joshua-Tor, L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. *Science*, *305*(5689), 1434-1437. doi:10.1126/science.1102514
- Swarts, D. C., Hegge, J. W., Hinojo, I., Shiimori, M., Ellis, M. A., Dumrongkulraksa, J., . . . van der Oost, J. (2015). Argonaute of the archaeon *Pyrococcus furiosus* is a DNA-guided nuclease that targets cognate DNA. *Nucleic Acids Research*, *43*(10), 5120-5129. Retrieved from <Go to ISI>://WOS:000355929200032
- Thomson, T., & Lasko, P. (2005). Tudor and its domains: germ cell formation from a Tudor perspective. *Cell Res*, *15*(4), 281-291. doi:10.1038/sj.cr.7290297
- Tian, Y., Simanshu, D. K., Ma, J. B., & Patel, D. J. (2011). Structural basis for piRNA 2'-O-methylated 3'-end recognition by Piwi PAZ (Piwi/Argonaute/Zwille) domains. *Proc Natl Acad Sci U S A*, *108*(3), 903-910. doi:10.1073/pnas.1017762108
- Tripsianes, K., Madl, T., Machyna, M., Fessas, D., Englbrecht, C., Fischer, U., . . . Sattler, M. (2011). Structural basis for dimethylarginine recognition by the Tudor domains of human SMN and SPF30 proteins. *Nature Structural & Molecular Biology*, *18*(12), 1414-1420. doi:10.1038/nsmb.2185
- Vagin, V. V., Hannon, G. J., & Aravin, A. A. (2009). Arginine methylation as a molecular signature of the Piwi small RNA pathway. *Cell Cycle*, *8*(24), 4003-4004. doi:10.4161/cc.8.24.10146
- Vagin, V. V., Sigova, A., Li, C. J., Seitz, H., Gvozdev, V., & Zamore, P. D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. *Science*, *313*(5785), 320-324. doi:10.1126/science.1129333
- Wang, Y. L., Juranek, S., Li, H. T., Sheng, G., Wardle, G. S., Tuschl, T., & Patel, D. J. (2009). Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature*, *461*(7265), 754-U753. Retrieved from <Go to ISI>://WOS:000270547500028
- Wang, Y. L., Sheng, G., Juranek, S., Tuschl, T., & Patel, D. J. (2008). Structure of the guide-strand-containing argonaute silencing complex. *Nature*, *456*(7219), 209-U234. Retrieved from <Go to ISI>://WOS:000261039300033

- Webster, A., Li, S., Hur, J. K., Wachsmuth, M., Bois, J. S., Perkins, E. M., . . . Aravin, A. A. (2015). Aub and Ago3 Are Recruited to Nuage through Two Mechanisms to Form a Ping-Pong Complex Assembled by Krimper. *Molecular Cell*, *59*(4), 564-575. doi:10.1016/j.molcel.2015.07.017
- Wei, K. F., Wu, L. J., Chen, J., Chen, Y. F., & Xie, D. X. (2012). Structural evolution and functional diversification analyses of argonaute protein. *Journal of Cellular Biochemistry*, *113*(8), 2576-2585. Retrieved from <Go to ISI>://WOS:000305334900004
- Willkomm, S., Oellig, C. A., Zander, A., Restle, T., Keegan, R., Grohmann, D., & Schneider, S. (2017). Structural and mechanistic insights into an archaeal DNA-guided Argonaute protein. *Nature Microbiology*, *2*(6). Retrieved from <Go to ISI>://WOS:000406000900007
- Yamaguchi, S., Oe, A., Nishida, K. M., Yamashita, K., Kajiya, A., Hirano, S., . . . Nureki, O. (2020). Crystal structure of Drosophila Piwi. *Nature Communications*, *11*(1), 858. doi:10.1038/s41467-020-14687-1
- Yamashiro, H., Negishi, M., Kinoshita, T., Ishizu, H., Ohtani, H., & Siomi, M. C. (2020). Armitage determines Piwi-piRISC processing from precursor formation and quality control to inter-organelle translocation. *EMBO Rep*, *21*(2), e48769. doi:10.15252/embr.201948769
- Yan, K. S., Yan, S., Farooq, A., Han, A., Zeng, L., & Zhou, M. M. (2003). Structure and conserved RNA binding of the PAZ domain. *Nature*, *426*(6965), 468-474. doi:10.1038/nature02129
- Yashiro, R., Murota, Y., Nishida, K. M., Yamashiro, H., Fujii, K., Ogai, A., . . . Siomi, M. C. (2018). Piwi Nuclear Localization and Its Regulatory Mechanism in Drosophila Ovarian Somatic Cells. *Cell Rep*, *23*(12), 3647-3657. doi:10.1016/j.celrep.2018.05.051
- Zeng, Y., Sankala, H., Zhang, X., & Graves, P. R. (2008). Phosphorylation of Argonaute 2 at serine-387 facilitates its localization to processing bodies. *Biochem J*, *413*(3), 429-436. doi:10.1042/BJ20080599
- Zhang, H., Liu, K., Izumi, N., Huang, H., Ding, D., Ni, Z., . . . Min, J. (2017). Structural basis for arginine methylation-independent recognition of PIWIL1 by TDRD2. *Proc Natl Acad Sci U S A*, *114*(47), 12483-12488. doi:10.1073/pnas.1711486114
- Zhang, Z., Xu, J., Koppetsch, B. S., Wang, J., Tipping, C., Ma, S. M., . . . Zamore, P. D. (2011). Heterotypic piRNA Ping-Pong Requires Qin, a Protein with Both E3 Ligase and Tudor Domains (vol 44, pg 572, 2011). *Molecular Cell*, *44*(6), 1005-1005. Retrieved from <Go to ISI>://WOS:000298827200019

## Figures and Figure legends

### Figure 1. Krimper monomer simultaneously binds N-terminal regions of Aub and Ago3

(a) Scheme depicting two-step sequential immunoprecipitation. FLAG tagged Aub, HA tagged Ago3 and GFP tagged truncated Krimper lacking its N terminal dimerization domain were co-expressed in S2 cell. FLAG-immunoprecipitation was followed by elution and GFP IP.

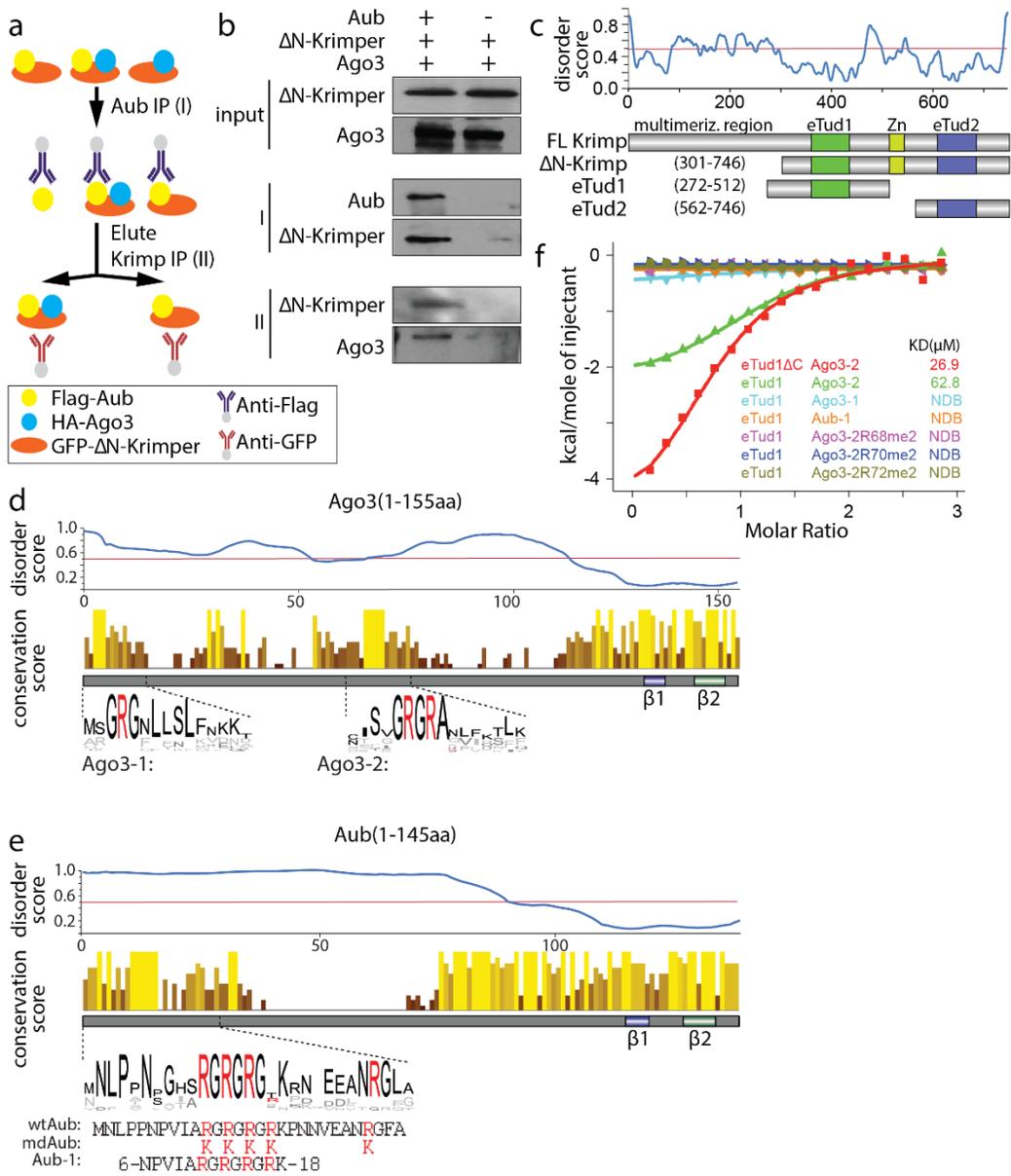
(b) Krimper monomer interacts with Aub and Ago3 simultaneously. Western blot detection of FLAG-Aub, HA-Ago3 and GFP-deltaN-Krimper in input and after 1<sup>st</sup> FLAG- and 2<sup>nd</sup> GFP-IP.

(c) Top: Protein disorder prediction of Krimper. Bottom panel: Domain architecture of Krimper and constructs used for ITC and structural analyses.

(d) Conservation analysis of Ago3 (1-155aa). Top: Disorder prediction of Ago3 (1-155aa). Middle: Conservation score of each amino acid within Ago3(1-155aa). Conserved RG-repeat regions are. Bottom: Peptides used for ITC and structural analysis.

(e) Conservation analysis of Aub (1-145aa). Top: Disorder prediction of Aub (1-145aa). Middle: Conservation score of each amino acid within Aub(1-145aa). Conserved RG-repeat region is enlarged. Bottom: Sequence difference between wtAub and mdAub. Five arginines within N terminus were replaced by lysines. Aub-1 peptide used for ITC and structural analysis.

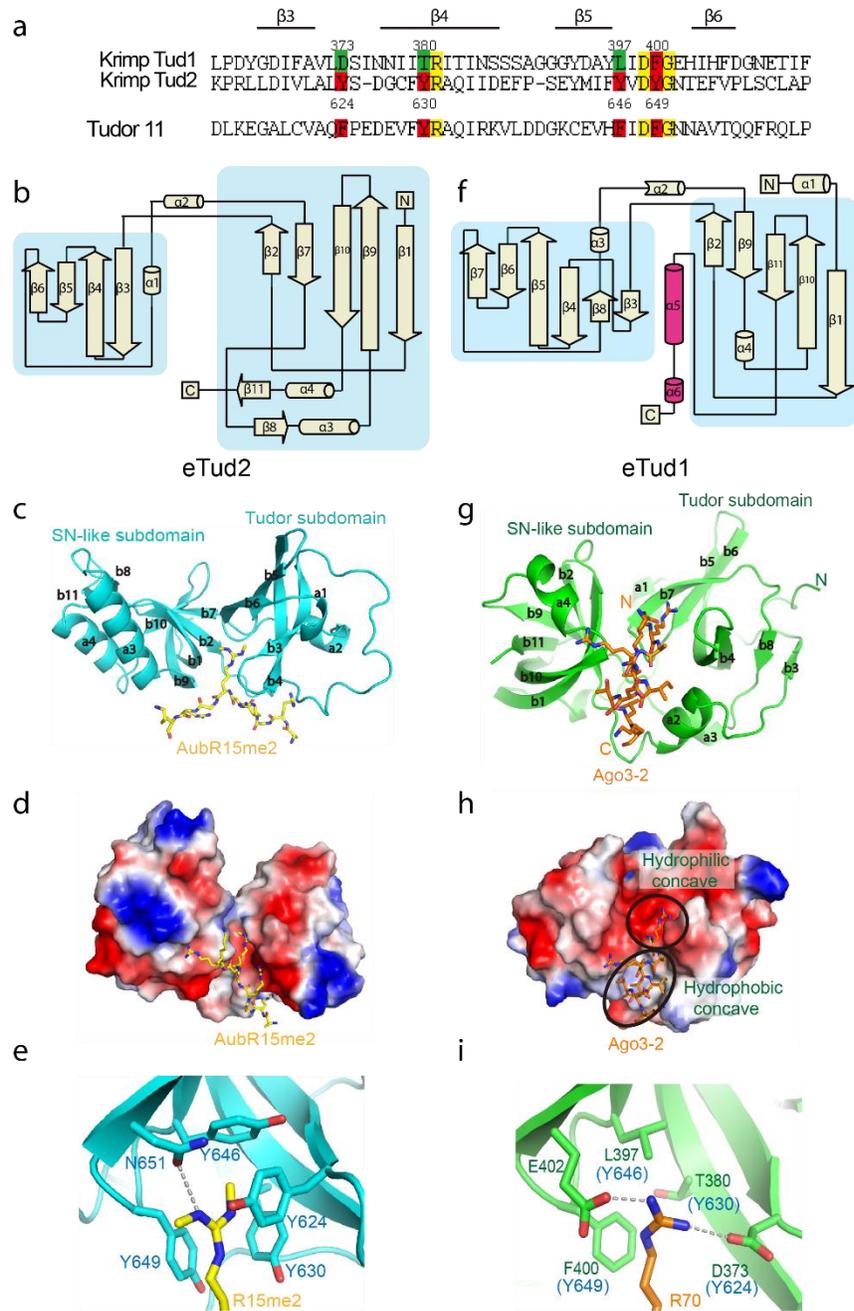
(f) eTud1 preferentially binds unmethylated Ago3. ITC analysis of Krimper eTud1 with different Ago3 or Aub peptides.



**Figure 2. Structure of Krimper eTud2-Aub and eTud1-Ago3 complex**

- (a) Krimper contains the canonical eTud2 domain and the non-canonical eTud1 domain. Alignments of Krimper Tudor domains with Tudor1 domain 11 are shown. Residues forming the me-Arg binding pocket (red) and corresponding residues in tud1 (green) are shown. Additional conserved residues marked yellow. Structural features are shown above.
- (b) Overall topology demonstration of Krimper eTud2 domain. The Tudor and SN-like subdomain are indicated.
- (c) Overall structure of the Krimper eTud2-Aub-R15me2 complex with the bound Aub-R15me2 peptide shown in yellow.
- (d) Enlarged view of the eTud2 aromatic cage.
- (e) Electrostatic surface of eTud2 with bound Aub-R15me2 peptide shown in yellow.
- (f) Overall topology demonstration of Krimper eTud1 domain. The Tudor and SN-like subdomain are indicated. Latch helix is marked magenta.
- (g) Overall structure of the Krimper eTud1-Ago3-2 complex with bound Ago3-2 peptide shown in orange.
- (h) Enlarged view of eTud1 binding pocket with inserted Ago3 R70 residue indicated in orange.
- (i) Electrostatic surface of eTud1 with bound Ago3-2 peptide shown in orange.

Huang et al., Figure 2



**Figure 3. Arginine methylation of Aub is required for fertility and TE repression**

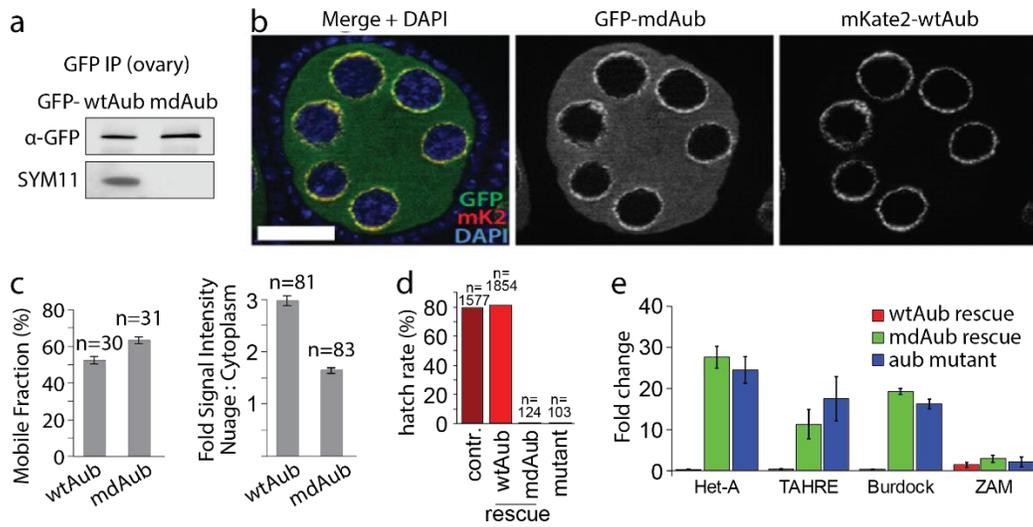
(a) Mutating conserved N-terminal Arginines to Lysines (mdAub) leads to loss of symmetric dimethylation of Aub. Similar amount of wtAub and mdAub was immunoprecipitated from ovaries. Methylation was detected by Western blot using the SYM11 antibody.

(b) In the wildtype background mdAub colocalizes with wildtype Aub in nuage. GFP-mdAub and mKate2-wtAub were co-expressed in ovaries under the control of the endogenous Aub promoter. Scale bar: 5 $\mu$ m.

(c) Arginine methylation deficiency slightly increases the fraction of mobile and dispersed cytoplasmic Aub protein. Left: the mobile fraction of nuage-localized GFP-wtAub and GFP-mdAub was determined in replicate FRAP experiments (constructs as in Extended Data Fig. 4c). Right: microscopic quantification of the ratio of GFP signal originating from nuage versus cytoplasm is shown. Error bars indicate st. dev. All transgenes are expressed under the control of the endogenous Aub promoter.

(d) Arginine methylation of Aub is required for fertility. Hatching rate of heterozygous control, Aub<sup>HN/QC</sup> mutant females and mutant females rescued with wt- or mdAub transgenes is indicated as % of eggs laid.

(e) Arginine methylation of Aub is required for TE suppression. Fold changes of different TE transcripts in ovaries of flies with the indicated genotypes compared to the heterozygous control as measured by RT-qPCR. n=3, error bars indicate st. dev.



**Figure 4. Arginine methylation of Aub is dispensable for piRNA binding but required for piRNA expression and ping-pong processing**

(a) Length distribution of small RNAs in *aub* mutants and upon rescue with wtAub and mdAub. Shown are the length distributions of reads annotated as miRNAs and repeats in ovarian small RNA libraries from heterozygous control, wtAub rescue, mdAub rescue and *aub* mutant flies. The ratio of repeat-derived piRNAs to miRNAs in each genotype is indicated above the graphs.

(b) Arginine methylation of Aub is required for piRNA expression and ping-pong processing of transposon piRNA. Abundance of sense and antisense piRNAs and the fraction of ping-pong pairs with 10nt overlap was analyzed in heterozygous control, wtAub rescue, mdAub rescue and *aub* mutant flies. For each library the abundance of transposon piRNAs was normalized to total miRNA read count; the heatmap shows the fold change compared to control (*aub* heterozygotes) for the top 20 most abundant TE families present in the control ovary. The last two heatmaps show changes in 10A bias for all sense piRNAs and non-1U sense piRNAs, respectively.

(c) Aub sDMA modification is required for generation of piRNAs from piRNA clusters. Reads from 5 kb genomic windows were normalized to total miRNA read count. Genomic windows with more than 5 RPM in libraries from heterozygous ovaries and more than 80% reduction in *aub* mutant were selected for further analysis. Heatmap shows piRNA fold change compared to control (*aub* heterozygotes). Note that expression of piRNAs from uni-strand clusters such as flam and 20A is not affected in *aub* mutants.

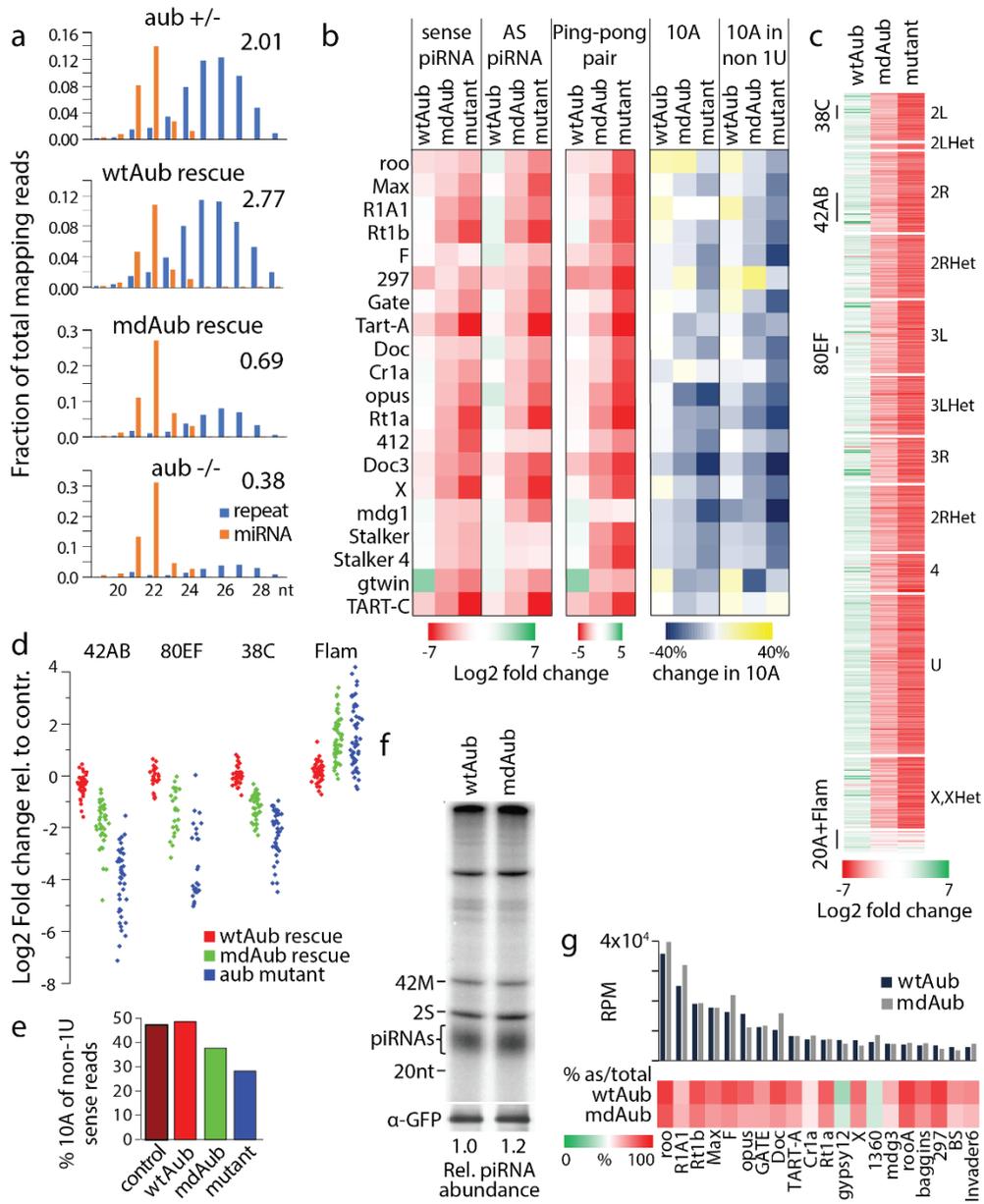
(d) Detailed analysis of piRNA generation from major piRNA clusters. Reads uniquely mapping to different piRNA clusters was determined in heterozygous control, wtAub rescue,

mdAub rescue and *aub* mutant flies and plotted as the fold change compared to the het control for four clusters (42AB, 38C, 80EF, Flam). Dot represents 1kb windows with uniquely mapping piRNAs within the clusters.

(e) The fraction of piRNAs that have signature of the ping-pong processing (piRNAs that map to TEs in sense orientation and have 10A, but not 1U bias) is reduced in *aub* mutants and only partially rescued by mdAub expression. Total small RNAs were cloned from ovaries of indicated genotypes and piRNAs were defined as 23 to 29 nt reads that mapped to the repeat track of the genome. The graphs show fraction of sense non-1U piRNAs that have 10A in libraries from control (*aub* heterozygous), wtAub and mdAub rescues and *aub* mutant ovaries.

(f) Aub Arginine methylation is not required for piRNA binding. GFP-tagged mdAub and wtAub were immunoprecipitated from ovaries where the transgene was expressed in the wildtype background. Protein level were detected by WB, associated RNAs were isolated, radiolabeled and run on a Urea PAGE gel. 42 nt ssRNA (42M) was spiked into each IP sample to control for labeling efficiency and RNA loss during isolation. Relative piRNA abundance normalized to IP-ed protein is estimated based on band intensity.

(g) Arginine methylation of Aub does not greatly affect the TE and antisense fractions of Aub-bound piRNAs. Bar chart shows normalized read counts (RPM) of wtAub and mdAub bound small RNAs mapped to the 20 TE families with post abundant piRNAs. Aub transgenes were expressed on the wildtype background. Heatmap shows the % of antisense reads mapping to each TE family.



**Figure 5. RNA binding promotes Aub arginine methylation**

(a) RNA binding is required for Aub arginine methylation *in vivo*. GFP tagged wtAub, mdAub and piRNA-binding deficient pdAub were immunoprecipitated from ovaries. Transgenes were expressed on the wildtype background. Total protein level and methylation was detected in Western blot using anti-GFP and SYM11 antibodies, respectively.

(b) RNA binding is required for Aub interaction with Tudor *in vivo*. HA tagged Tudor was co-expressed with GFP tagged wtAub, mdAub and pdAub. Aub was immunoprecipitated from ovarian lysate, followed by Western blot detection of the tagged transgenes.

(c) Arginine methylation is required for localization of Aub to the pole plasm. GFP tagged wtAub and mdAub were separately expressed in ovaries on the wild-type Aub background.

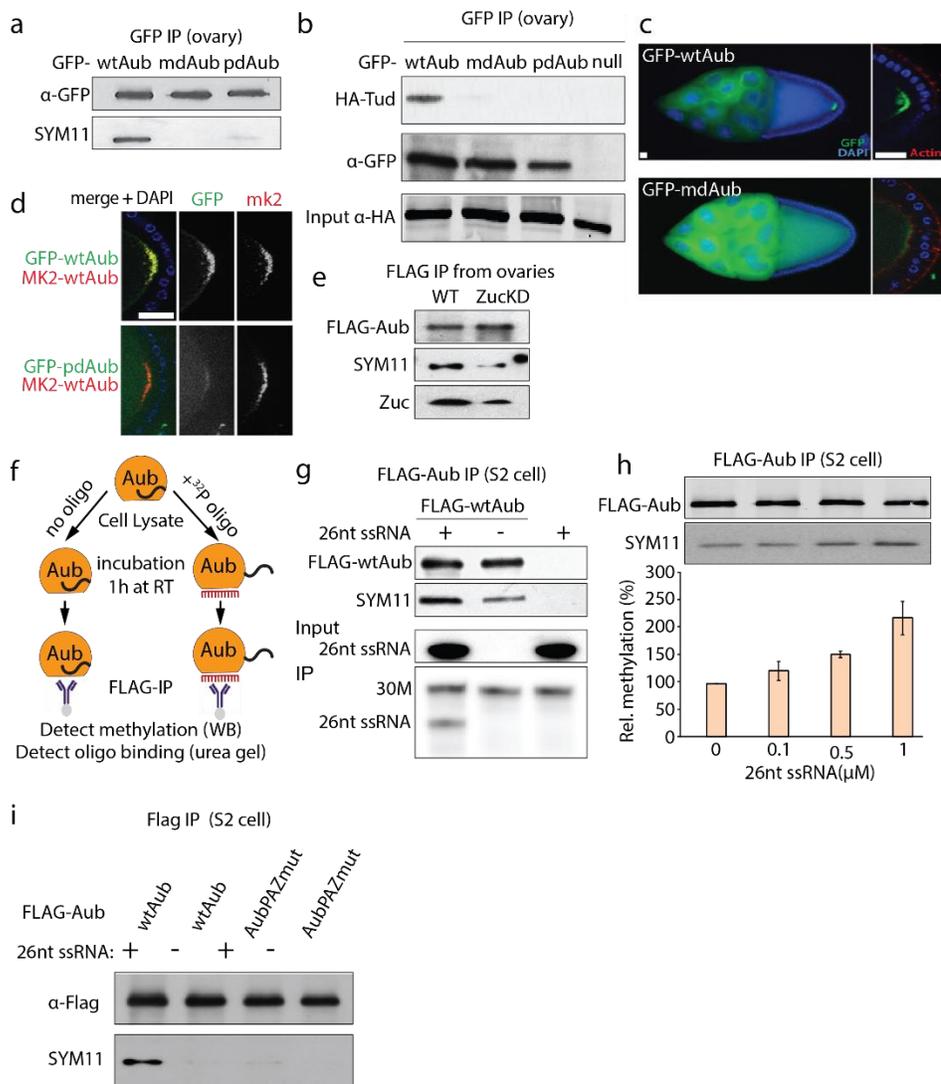
Scale bar: 5 $\mu$ m.

(d) RNA binding is required for Aub localization into the pole plasm. Upper panel: GFP-wtAub and mKate2-wtAub were co-expressed in ovaries in the wild-type Aub background. Bottom panel: GFP-pdAub and mKate2-wtAub were co-expressed in ovaries in the wild-type Aub background. Scale bar: 5 $\mu$ m.

(e) Zucchini KD leads to decreased Aub arginine methylation. FLAG-tagged wtAub was immunoprecipitated from control and Zuc KD ovaries in the wildtype Aub background. Methylation was detected by Western blot using the SYM11 antibody.

(f) Scheme of *in vitro* oligo-binding experiments. Lysate of S2 cells expressing FLAG-Aub was incubated with or without <sup>32</sup>P labelled 26nt RNA oligos followed by FLAG immunoprecipitation and detection of methylation by Western blot and oligo binding on urea gel.

- (g) RNA oligo binding promotes Aub sDMA modification. Aub expressed in S2 was loaded with 26nt ssRNA oligo in cell lysate, followed by immunoprecipitation. 30 nt (30M) ssRNA was spiked into each IP-ed sample to normalize for total IP-ed RNA amount.
- (h) Aub methylation correlates with synthetic ssRNA concentration added to lysates of S2 cells expressing FLAG-Aub. Methylation was quantified based on band intensity in FLAG and SYM11 Western blot and normalized to the no oligo control. Error bars indicate st. dev (n=2).
- (i) Binding of the RNA 3' end by Aub's PAZ domain is required for sDMA modification. Wild-type Aub and mutant protein lacking 3'-end piRNA binding due to mutation of two conserved residues in the PAZ domain (PAZmut) were expressed in S2 cells. Cell lysates were incubated with 26nt ssRNA oligos and sDMA modification was detected by western blot using SYM11 antibody.



**Figure 6. RNA binding triggers conformational change in Aub, exposing its N terminus to the methyltransferase CsuI/Vls complex**

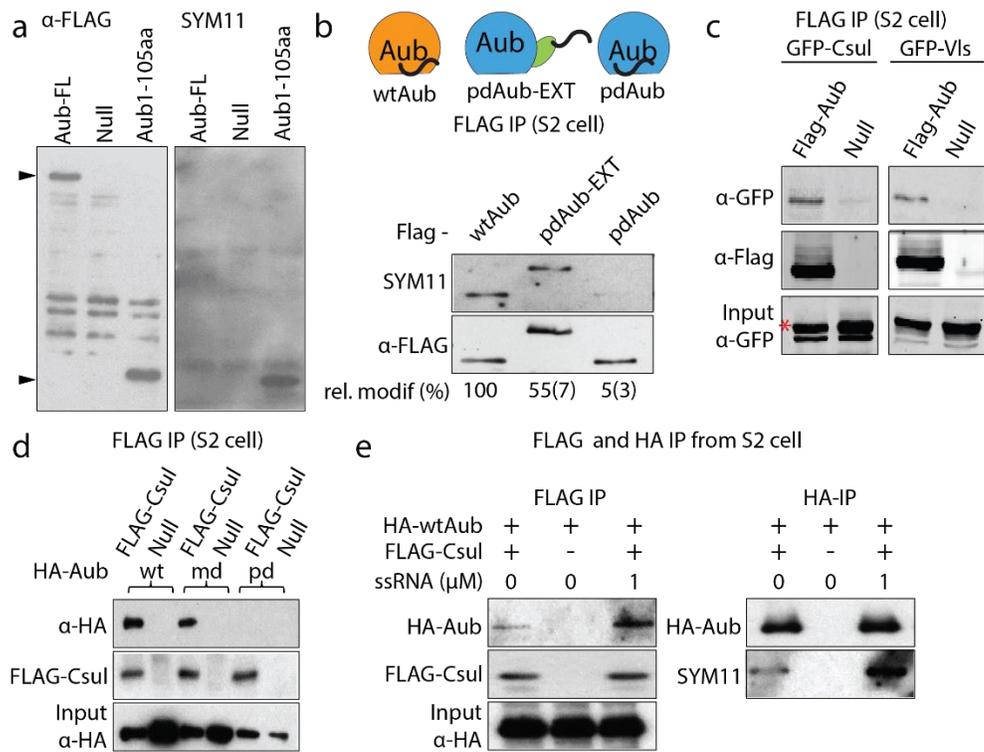
(a) N-terminal region within Aub protein is not easily accessible to methylation. FLAG tagged full length and N-terminal truncated (1-105aa) Aub were expressed in S2 cell and immunoprecipitated. Total protein and methylation level was assessed by Western blot. Arrowheads indicate correct size for full-length and N-terminal fragment, respectively.

(b) Top: Scheme showing architectures of different Aub constructs expressed and IP-ed from S2 cells. EGFP inserted between the N-terminal and PAZ domain (pdAub-EXT) artificially exposes N-terminus in absence of piRNA binding. Bottom: Western blot analysis of methylation states. Relative methylation level as estimated by ratio of SYM11/FLAG band intensities normalized to wildtype are listed, standard deviations are shown in brackets, n=2.

(c) Aub interacts with CsuI and Vls in S2 cells. Co-IP Western blot of tagged transgenes. Asterisk indicates band corresponding to GFP-CsuI in the INPUT.

(d) Aub binding to CsuI depends on RNA binding but not on Arg methylation. FLAG-CsuI and HA-Aub transgenes were expressed in S2 cells and coIP followed by Western detection.

(e) RNA loading of Aub leads to increased binding to CsuI and increased methylation of Aub. FLAG-CsuI and HA-Aub were expressed in S2 cells, lysates were incubated in the presence or absence of ssRNA oligo prior to coIP followed by Western detection.

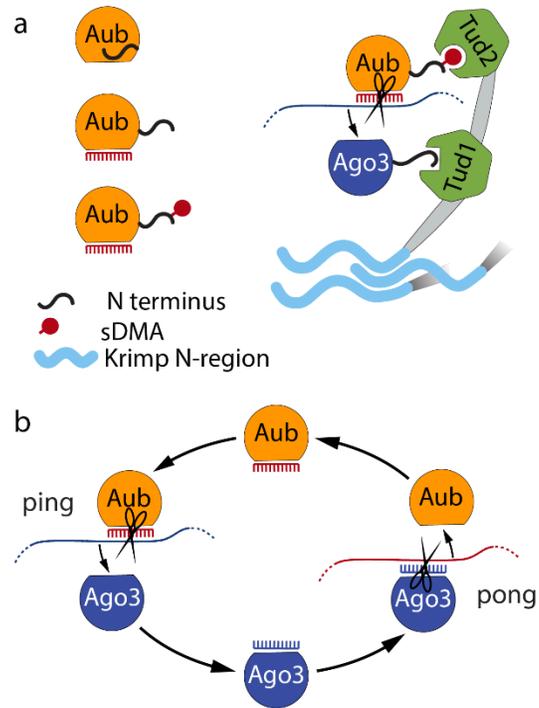


**Figure 7. Model for sDMA regulation and its function in ping-pong cycle**

(a) Model sDMA dependent assembly of the ping processing complex. The N-terminus of unloaded Aub is inaccessible. Binding to piRNA guide leads to conformational change of Aub exposing its N terminus and enabling methylation of its N-terminal arginines. Methylation thus serves as a signal of loading-state and enables Aub binding to the Tud2 domain of Krimper, which specifically recognizes methylated Arginines. The Tud1 domain of Krimper binds unmethylated, unloaded Ago3, enabling Ago3 loading with the newly processed piRNA. The N-terminal unstructured region of Krimper allows Krimper multimerization resulting in a Krimper scaffold that might assist in nuage assembly and ensuring high local concentration of Aub and Ago3.

(b) The Ping-Pong cycle consists of two distinct stages. In the ping stage, Aub with its piRNA guide targets piRNA precursors or TE transcripts. Cleaved, mature piRNA is loaded into Ago3. In the pong stage, piRNA-loaded Ago3 targets antisense piRNA precursors. Cleaved mature piRNA is loaded into Aub. The ping and pong processing could be accomplished by different complexes in nuage.

Huang et al., Figure 7



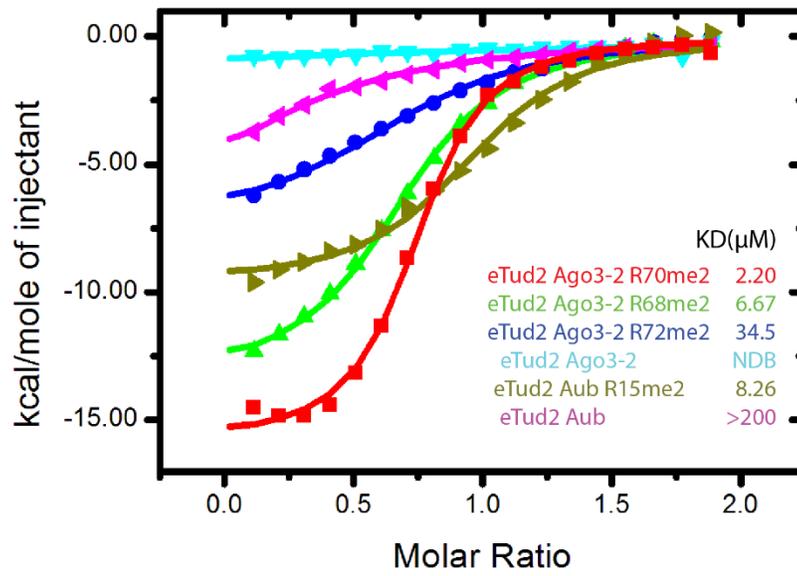
**Extended Data Figures****Extended Data Figure 1. Interaction between Krimp, Aub and Ago3**

(a) eTud2 preferentially binds methylated peptides. ITC analysis of Krimper eTud2 interaction with different Ago3 or Aub peptides.

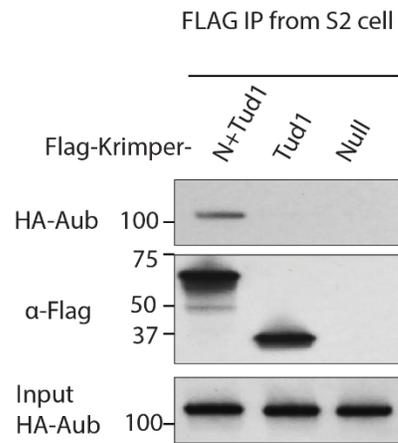
(b) Co-immunoprecipitation of Aub with Krimp fragments in S2 cells. FLAG-tagged eTud1 and N+eTud1 fragments were coexpressed with HA-Aub in S2 cells. Anti-FLAG immunoprecipitation was followed by Western blot. Unlike the N+eTud1 fragment, the eTud1 fragment without the N-terminal domain does not co-IP with Aub.

## Huang et al., Extended Data Fig. 1

a



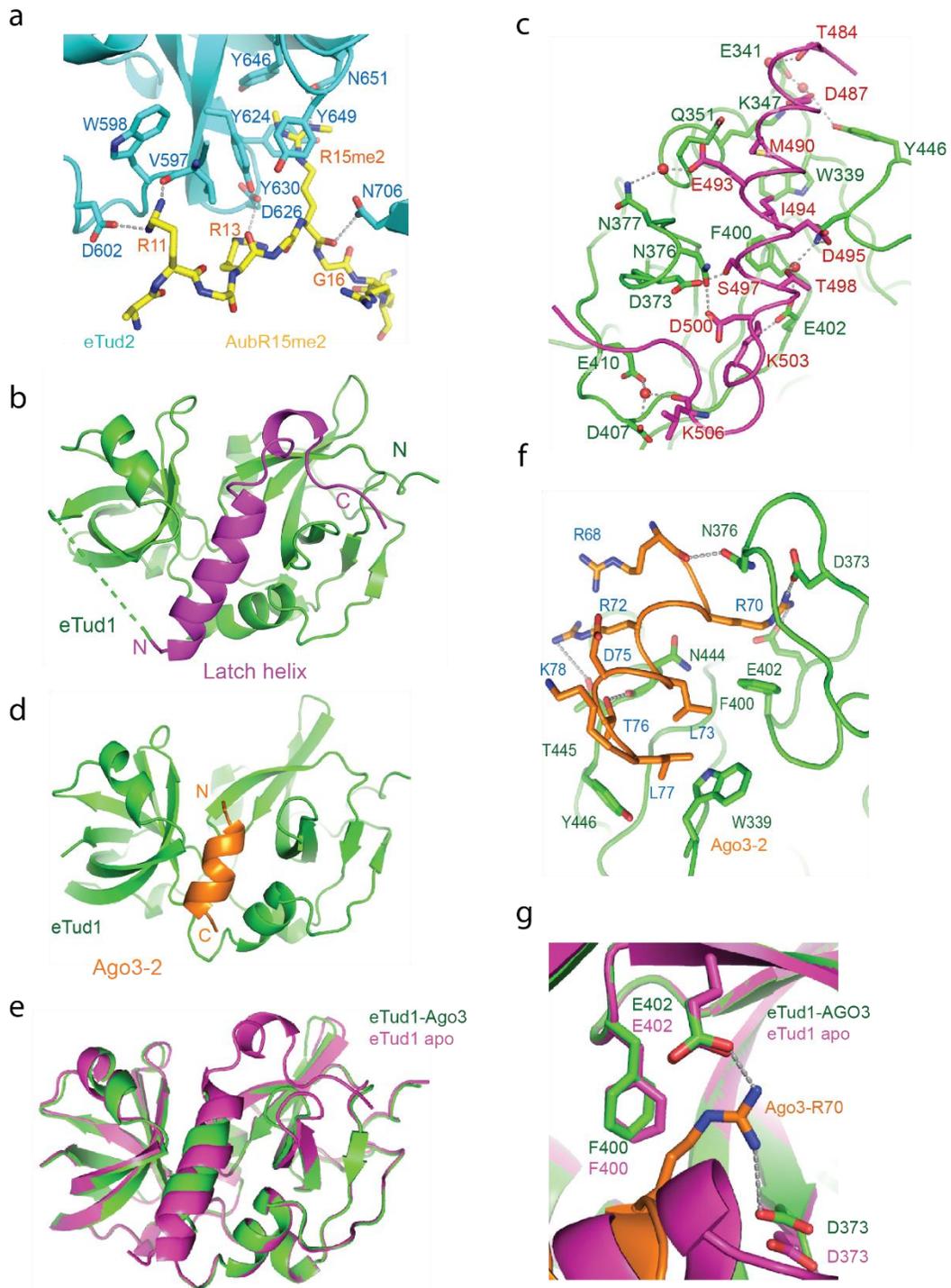
b



**Extended Data Figure 2. Detailed analysis of Krimper eTud1 apo structure**

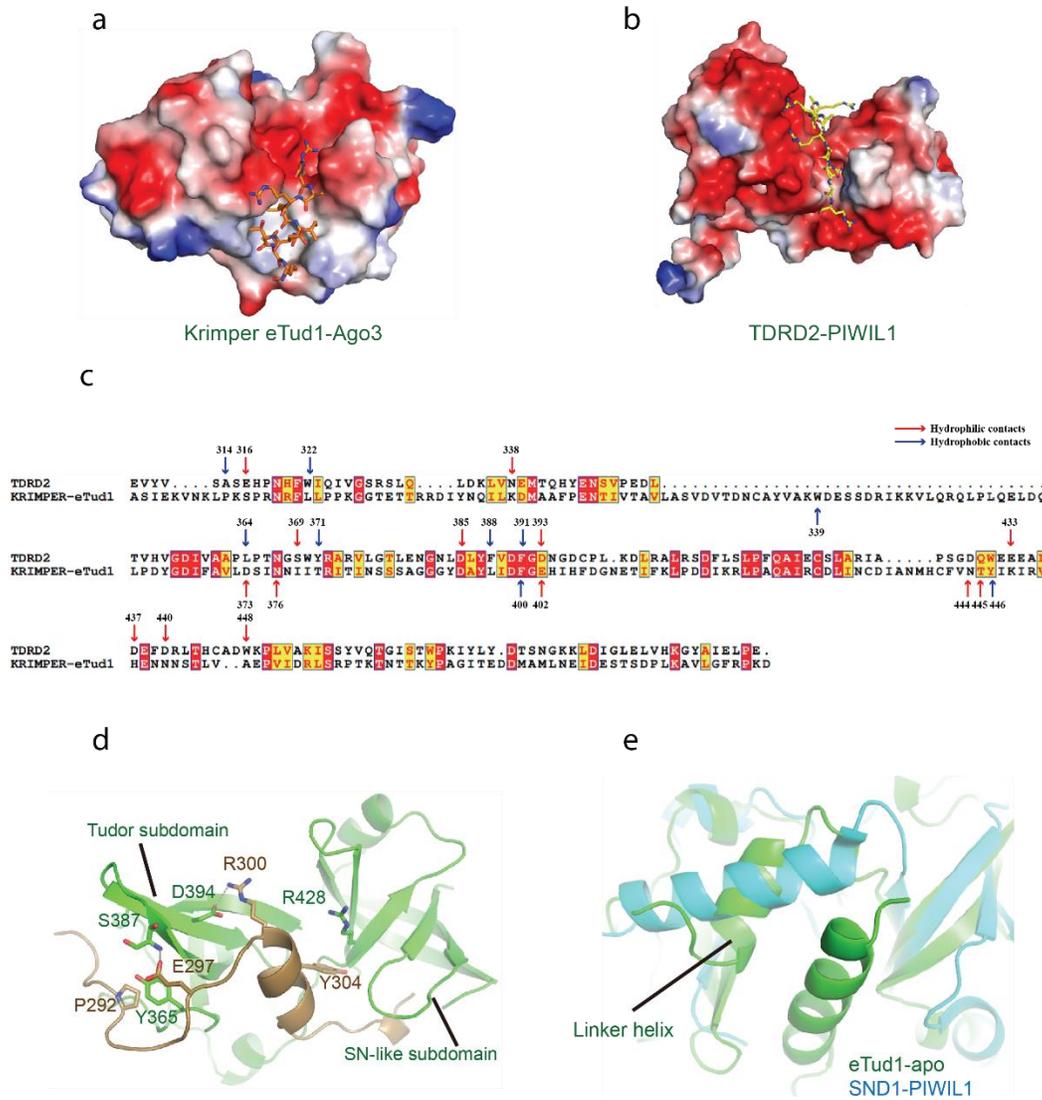
- (a) Detailed interactions of eTud2 with Aub peptide. Hydrogen bonds are shown in grey dashed lines.
- (b) Overall structure of the Krimper eTud1 domain. The ‘latch helix’ region is highlighted in magenta.
- (c) Overall structure of the Krimper eTud1-Ago3-2 complex with Ago3-2 shown in orange.
- (d) Structural superposition of Krimper eTud1 apo and Ago3 peptide-bound structure.
- (e) Detailed interaction networks of Krimper eTud1 with the Ago3-2 peptide.
- (f) Detailed interactions of the ‘latch helix’ with other parts of Krimper eTud1.
- (g) The Phe400 side chain in the Ago3 peptide bound structure (green) rotates 90° to accommodate Ago3-R70 when compared with the apo structure (magenta).

Huang et al., Extended Data Fig. 2



**Extended Data Figure 3. Structural comparison of Krimper eTud1-Ago3-2 complex with other extended Tudor domains**

- (a) Electrostatic surface of the eTud1-Ago3 complex binding cleft
- (b) Electrostatic surface of the TDRD2-PIWIL1 complex binding cleft (PDB code:6B57)
- (c) Sequence alignment between TDRD2 and eTud1 of Krimper. The position of amino acids involved in the hydrophilic contacts in the TDRD2-PIWIL1 structure are labeled with red arrows above the alignment. The position of amino acids involved in the hydrophobic contacts in the TDRD2-PIWIL1 structure are labeled with blue arrows above the alignment. The position of amino acids involved in the hydrophilic contacts in the eTud1-Ago3 structure are labeled with red arrows below the alignment. The position of amino acids involved in the hydrophobic contacts in the eTud1-Ago3 structure are labeled with blue arrows below the alignment.
- (d) eTud1 N-terminal structure is highlighted in brown.
- (e) Structural superposition of eTud1 apo structure with a canonical extended Tudor domain (SND1-PIWIL1, PDB code: 3OMC). The position of the linker helix is different in the two structures.



**Extended Data Figure 4. Localization, mobility and slicer activity of wtAub and mdAub proteins**

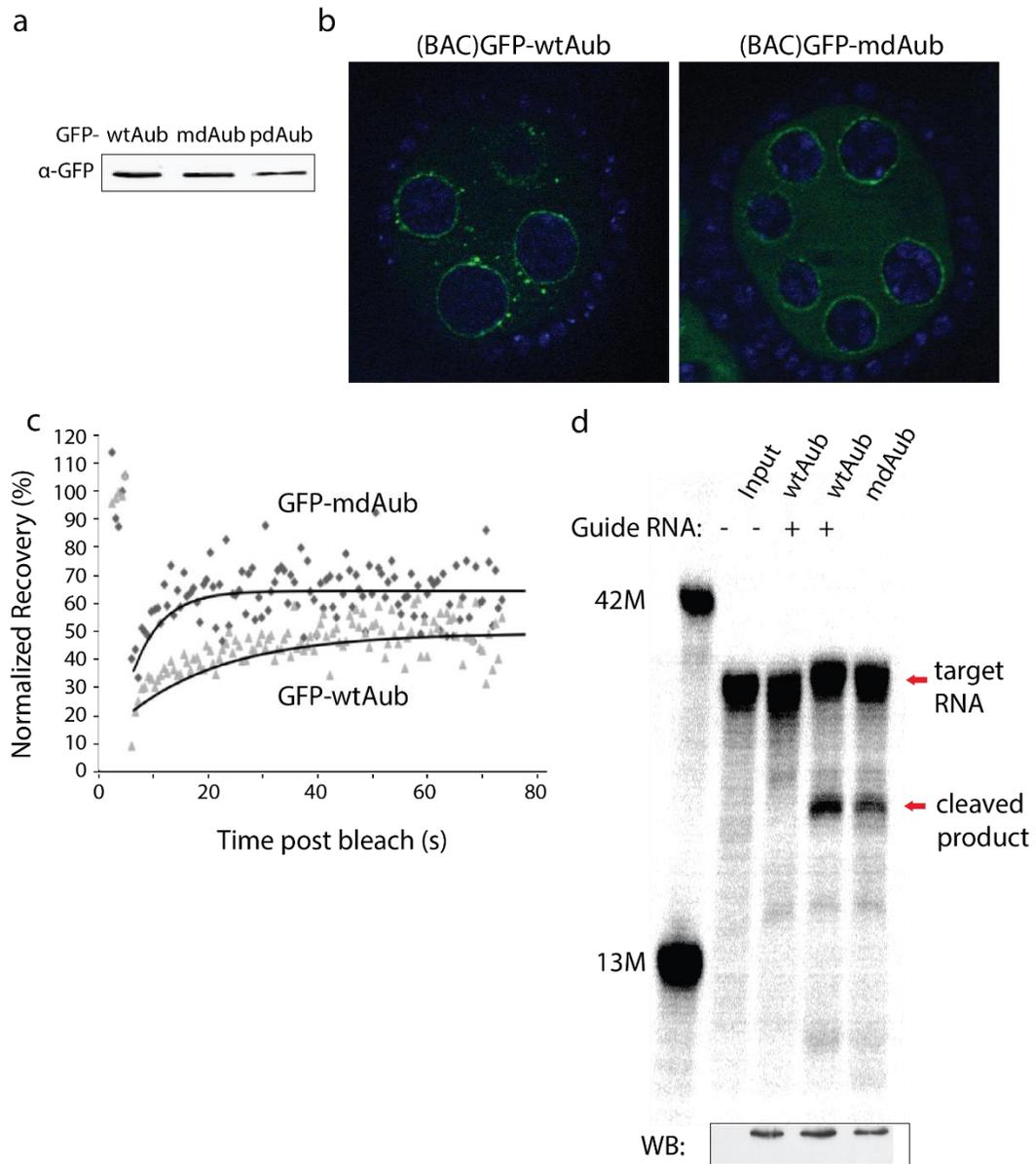
(a) Expression of EGFP-tagged wtAub, mdAub and pdAub proteins. Lysates were prepared from 100 ovaries dissected from flies expressing EGFP-tagged wtAub, mdAub and pdAub proteins and 3% of each lysate was loaded on the gel followed by Western blot analysis using anti-EGFP antibody.

(b) Localization of GFP-wtAub and GFP-mdAub expressed under the control of the endogenous Aub promoter in nurse cells at stage 6 of oogenesis.

(c) Methylation deficiency increases Aub mobile fraction. Representative FRAP experiments showing that the normalized recovery of GFP-mdAub is approximately 10% greater compared to GFP-wtAub. The mobile fraction was determined by modeling the recovery to an exponential recovery curve.

(d) mdAub has slicer activity. Purified wtAub and mdAub were used in a cleavage assay with 29 nt 5' end radiolabeled target containing a sequence complementary to the guide RNA. The products were resolved on a denaturing urea gel. Western blot shows the amounts of wtAub and mdAub used in the cleavage assay.

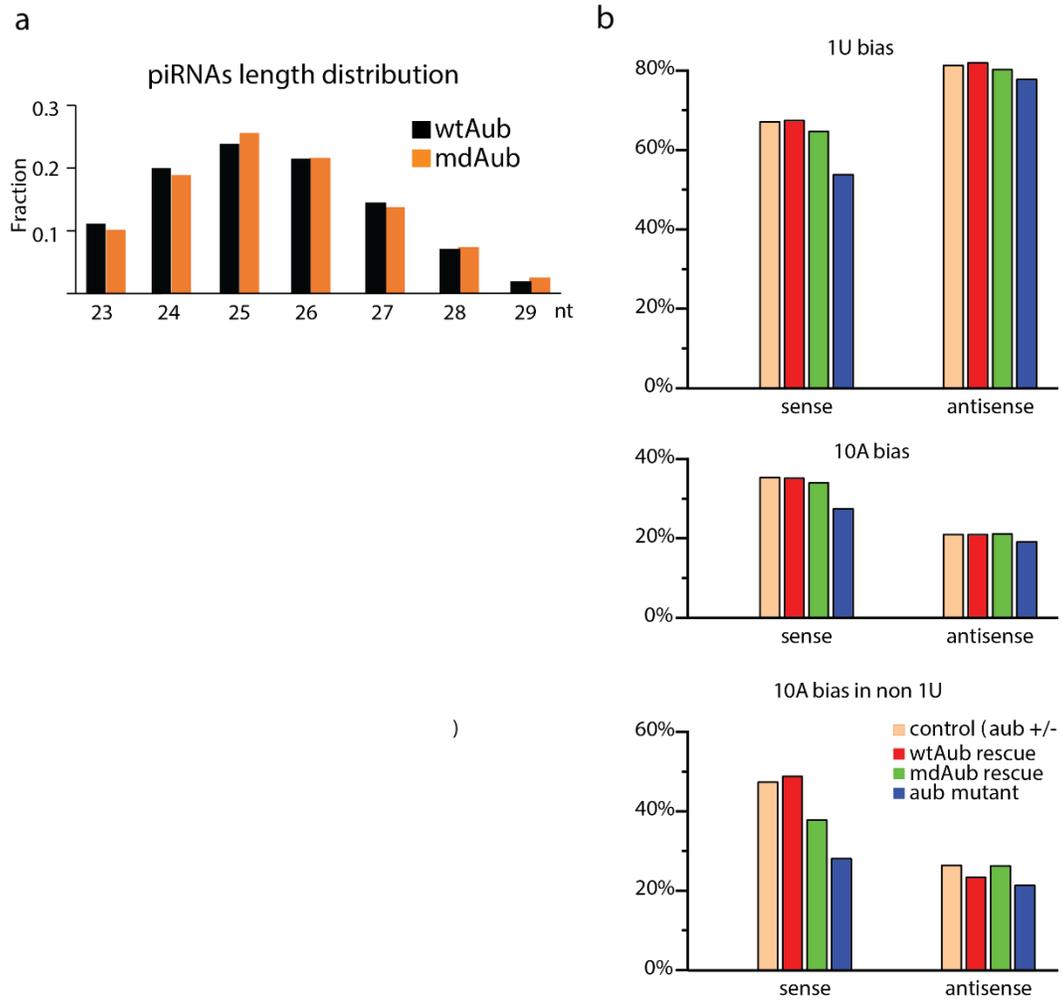
## Huang et al., Extended Data Fig. 4



**Extended Data Figure 5. Analysis of piRNAs in flies expressing wtAub and mdAub proteins**

- (a) Nucleotide biases in piRNA populations in ovaries of wtAub- and mdAub-rescue flies. Total 18 to 30 nt RNA was cloned from ovaries with the indicated genotypes and piRNAs were defined as 23 to 29 reads that mapped to the repeat track of the genome in sense and antisense orientation relative to the annotated TE sequence. The graphs show frequencies of 1U and 10A in the sense and anti-sense piRNAs in libraries cloned from control (*aub* heterozygous), wtAub and mdAub rescues and *aub* mutant ovaries.
- (b) Length distribution of small RNAs bound by GFP-wtAub and GFP-mdAub in the wildtype *aub* background.

## Huang et al., Extended Data Fig. 5



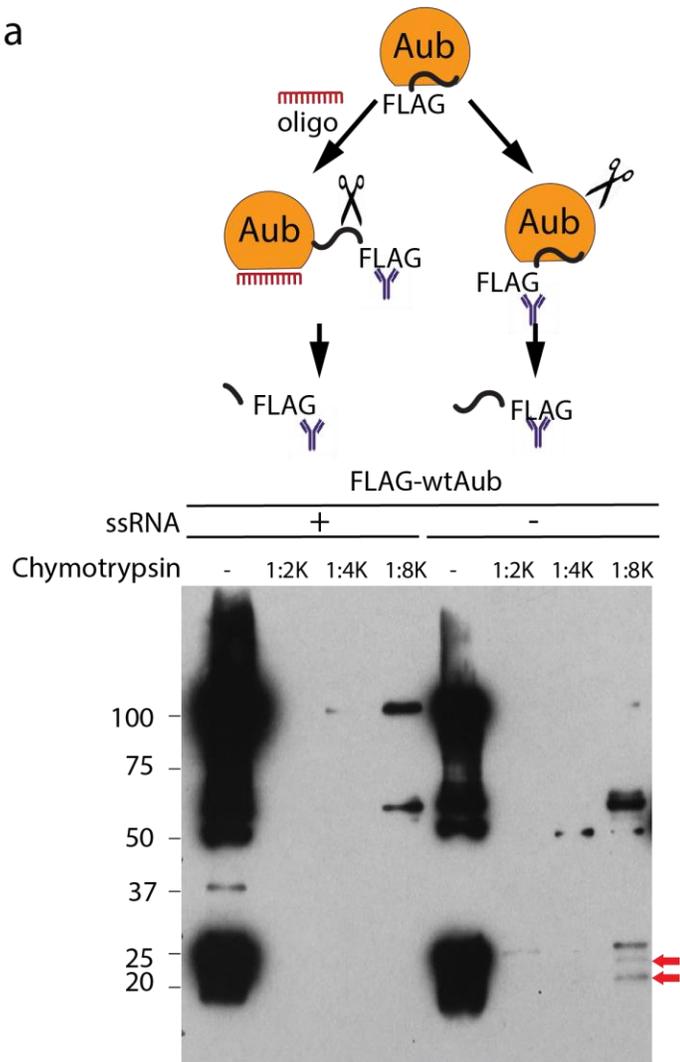
**Extended Data Figure 6. Oligo binding causes Aub conformational change**

(a) Scheme of limited protease assay (top). S2 cell lysate expressing FLAG-Aub was incubated with or without 1 $\mu$ M synthetic ssRNA oligos followed by FLAG IP. Bead-bound proteins were incubated with different concentration of chymotrypsin. Bead fraction was analyzed by Western blot (bottom) using FLAG detection. Red arrow indicates FLAG-tagged N-terminal fragment that is undigested in the absence of piRNA loading but digested upon small RNA loading.

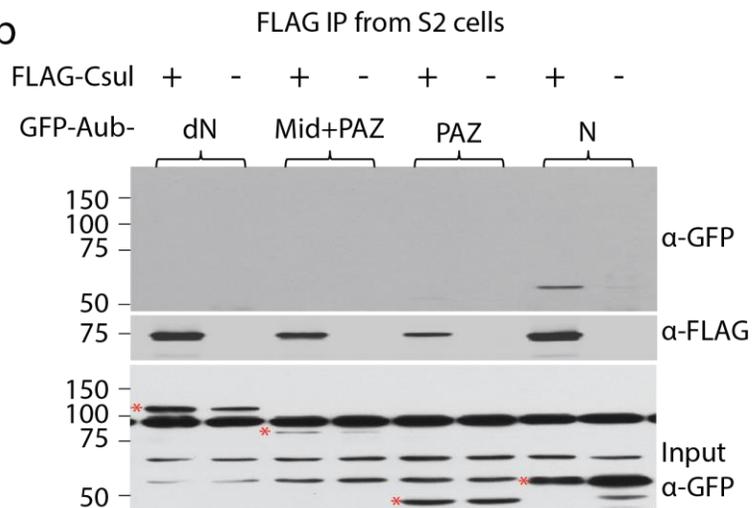
(b) Csu1 interacts with the N-terminus of Aub. FLAG tagged Csu1 was co-expressed in S2 cells with different GFP-tagged Aub fragments and coIP followed by Western detection of IP-ed proteins. Asterisk indicates bands corresponding to indicated GFP-tagged Aub protein fragments.

## Huang et al., Extended Data Fig. 6

a



b



## Extended Data Tables:

Extended Data Table 1. Data collection and refinement statistics

	Krimper Tud1 apo	Krimper Tud1- Ago3	Krimper Tud2- AubR15me2
<b>Data collection</b>			
Beamline	SSRF-BL19U1	SSRF-BL19U1	SSRF-BL19U1
Space group	$P2_12_12_1$	$P3_1$	$P2_12_12_1$
Wavelength (Å)	0.9792	0.9792	0.9792
Cell dimensions (Å)			
<i>a</i>	60.5	88.5	97.1
<i>b</i>	87.7	88.5	101.4
<i>c</i>	90.2	181.7	191.3
Resolution (Å)	50.0-2.1 (2.18-2.10) <sup>a</sup>	50.0-2.4 (2.49-2.40)	50.0-2.7 (2.80-2.70)
$R_{\text{merge}}$	0.134 (1.029)	0.094 (0.891)	0.114 (0.511)
$I / \sigma I$	12.5 (1.4)	9.0 (1.25)	12.5 (1.53)
Completeness (%)	99.8 (100.0)	99.3 (98.6)	95.5 (75.1)
Redundancy	5.4 (5.5)	4.3 (4.3)	5.1 (3.7)
CC1/2	0.727	0.872	0.897
<b>Refinement</b>			
$R_{\text{work}} / R_{\text{free}}$	0.204 / 0.222	0.234 / 0.259	0.243 / 0.291
No. reflections	28,403	61,721	49,822
No. atoms	3,274	7,519	12,195
Protein	3,062	6,899	11,794
Peptide	-	325	327
Solvent	212	295	74
$B$ -factors (Å <sup>2</sup> )	39.3	56.4	58.1
Protein	39.1	56.0	57.9
Peptide	-	67.6	69.2
Solvent	42.8	53.3	50.7
R.m.s. deviations			
Bond lengths (Å)	0.023	0.004	0.009
Bond angles (°)	1.771	0.743	1.237

<sup>a</sup>Highest-resolution shell is shown in parentheses.

Extended Data Table 2. Oligo sequences.

## qPCR primers

HeT-A-F	CGCGCGGAACCCATCTTCAGA
HeT-A-R	CGCCGCAGTCGTTTGGTGAGT
ZAM-F	ACTTGACCTGGATACACTCACAAC
ZAM-R	GAGTATTACGGCGACTAGGGATAC
Burdock-F	AGGGAAATATTTGGCCATCC
Burdock-R	TTTTGGCCCTGTAAACCTTG
TAHRE-F	CTGTTGCACAAAGCCAAGAA
TAHRE-R	GTTGGTAATGTTTCGCGTCCT
RP49-F	CCGCTTCAAGGGACAGTATCT
RP49-R	ATCTCGCCGCAGTAAACG

## RNA oligos

13nt ss RNA (size marker)	rCrCrArUrCrGrArUrArArArG
26nt ssRNA	rUrCrGrArArGrUrArUrUrCrCrGrCrGrUrArCrGrUrGrArUrGrUrU
29nt target ssRNA	rArCrCrArArCrArUrCrArCrGrUrArCrGrCrGrGrArArUrArCrUrUrCrGrA
30nt ssRNA (size marker)	rCrCrArUrCrGrArUrArArArGrUrUrArArArCrGrArGrCrUrUrCrCrCrG
42nt ssRNA (size marker)	rCrCrArUrCrCrArUrCrGrArUrArArArGrUrUrUrArArArCrGrArGrCrUrUrCrCrCrGrCrGrUrArCrGrGrA

*Chapter 4***IDENTIFICATION OF PIWI DIRECT INTERACTORS USING  
HETEROLOGOUS CELL CULTURE ASSAY**

This chapter was adapted from manuscript Rogers, A. K. et al. "Mago and Arp6 identified as novel *Drosophila* Piwi interactors using a heterologous two-hybrid system".

**Abstract**

The piRNA pathway acts as a safeguard against transposable elements for the germline genomes of all animals. In *Drosophila melanogaster*, the piRNA pathway is composed of three piwi clade Argonaute proteins – Aubergine (Aub), Argonaute3 (Ago3), and Piwi – and their associated piwi-interacting RNAs (piRNAs). Genetic screens have identified numerous factors involved in the piRNA pathway mediated transcriptional silencing and post-transcriptional silencing. An essential step to understanding the interaction network and the hierarchy among these factors and PIWI-clade proteins. While a complete list of all proteins that exhibit direct interaction has not been established. We developed a novel strategy to distinguish between direct and indirect interactions with Piwi and employed our strategy on a candidate list of factors. Our approach identifies Piwi direct interaction partners using a microscopy-based assay in a heterologous cell culture system in which Piwi is stably tethered to DNA. By performing our assay in a heterologous cell culture system, we limit the possibility of identifying false positives. Our assay allowed us to discern from previously identified Piwi interactors directly interacting with Piwi – thus further elucidating their roles within the piRNA pathway. We also utilized our approach to identify novel direct protein interactions with Piwi from a list of factors genetically identified to participate in the piRNA pathway.

## Introduction

In the germline of *Drosophila melanogaster*, the PIWI-clade Argonaute proteins associate with piRNAs to form effector complexes that recognize and silence selfish mobile genetic elements called transposable elements (TEs). The *Drosophila* genome encodes three PIWI-clade Argonaute proteins: Piwi, Aubergine (Aub), and Argonaute3 (Ago3). All three proteins are expressed in germline cells, while in the somatic follicular cells, only Piwi is present (Cox et al. 2000; Harris and Macdonald 2001; Saito et al. 2006; Brennecke et al. 2007; Gunawardane et al. 2007). Piwi is loaded with piRNA in cytoplasmic, perinuclear granules termed nuage, and then the Piwi-piRNA complex is translocated to the nucleus. In the nucleus, Piwi scans transcripts for a complementary target to its piRNA. Once a target has been identified, Piwi recruits the silencing machinery necessary to establish the H3K9me3 silencing mark at the site of the target (Saito et al. 2009; Olivieri et al. 2010; Handler et al. 2011; Ishizu et al. 2011; Klenov et al. 2011; Shpiz et al. 2011; Sienski et al. 2012; Gu and Elgin 2013; Le Thomas et al. 2013; Rozhkov et al. 2013). To accomplish its function, Piwi has to collaborate with at least two distinct sets of machinery. In the cytoplasm, Piwi must interact with the piRNA biogenesis machinery to be loaded. In contrast, in the nucleus, Piwi must interact with the silencing effector complex to regulate the chromatin structure. However, many additional factors have been genetically identified as part of the piRNA pathway. Only a few of the piRNA pathway factors have been biochemically dissected, with most interactions demonstrated predominantly through co-immunoprecipitation followed by western blotting. Further, the ability to dissect the interactome is complicated by the complexity of the interactions.

It is still unclear exactly whether Piwi engages in direct or indirect biochemical interactions with these factors. The ability to discern direct biochemical interactions with Piwi would provide important insights into understanding the hierarchy of events that occur for Piwi to be loaded in the cytoplasm, translocated to the nucleus, and accomplish its function in the nucleus. Two commonly used techniques are yeast two-hybrid screening and affinity purification of complexes in vitro followed by mass spectrometry to screen the candidates for the protein-protein interactions. Both techniques have limitations. Previously, we performed immunoprecipitation followed by mass spectrometry of tagged-Piwi in the presence and absence of RNase A treatment and found that Piwi predominantly interacts with RNA-binding proteins in an RNA-dependent manner (Le Thomas et al. 2013). However, due to the bias of affinity purification (AP) or co-immunoprecipitation (co-IP) towards high affinity or highly abundant proteins, we suspect that the list of interactors we obtained may be missing less abundant factors as well as any weak or transient interactions. Since we could not fix our samples prior to mass spectrometry, we also suspect that the list of interactors we obtained contains unspecific hits. Furthermore, we cannot discriminate between direct and indirect protein interactions based on the mass spectrometry data. Based on FLIP data, we expect that Piwi interacts very transiently with chromatin at the site of targets (Le Thomas et al. 2013). Piwi's transient interaction with chromatin suggests that factors downstream of Piwi recognition of a target could be interacting transiently as well, resulting in the inability to capture these interactions with co-IP followed by mass spectrometry. Yeast two-hybrid (Y2H) system, which identifies defined, direct binary interactions, are relatively inexpensive and do not rely on specialized or large equipment. When looking for direct interactions with Piwi, both AP/MS using *Drosophila* ovary lysate

and traditional Y2H systems are plagued with false positives – indirect interactions – generated by interactions mediated by other piRNA pathway proteins. We developed a fluorescence-based optimized two-hybrid (2H) approach to identify biochemical interactions of Piwi. Our method can detect transient or weak protein-protein interactions in vivo and discriminate between direct and indirect interactions with *D. melanogaster* Piwi. Our work establishes a technique that allows us to take a candidate list of factors and distinguish between factors that directly interact with Piwi and those that indirectly interact with Piwi (possibly mediated by RNA or other *D. melanogaster* proteins). Co-IP experiments in S2 cells are further applied to confirm the direct interactors.

## **Results**

### **A heterologous cell culture assay utilizing lacO tethering to detect direct protein interactions**

To determine if factors previously identified to interact with Piwi were directly or indirectly interacting with Piwi, we employed a new method capable of detecting direct protein-protein interactions in vivo. Our approach utilizes the fusion of a bait protein to the DNA-binding domain of lacI and a fluorescent tag mKate2. The lacI-mKate2 fused bait protein is stably expressed in cells that contain three genomic loci with arrays of the lacO sequence (Jegou et al. 2009), resulting in the recruitment of the bait protein to the lacO loci, which can be visualized using fluorescent microscopy as three distinct nuclear dots. Cells stably expressing the tethered bait protein are transiently transfected with prey proteins fused with a different fluorescent tag (CFP). If the bait and prey proteins interact, the prey is recruited to the nuclear lacO loci, resulting in co-localization of the mKate2 and CFP signals (Fig 1A). Tethering of the bait protein to lacO foci allows for weak or transient interactions,

which may be missed by other techniques, to be identified. We utilized our system with Piwi as bait in human U2OS cells, which lack other *Drosophila* germline proteins – thus favoring detection of direct interactions between Piwi and the prey proteins. When we stably express lacI-mKate2-Piwi in U2OS cells containing lacO arrays, Piwi localizes to nuclear dots, whereas transiently transfected CFP (control) is evenly distributed throughout the cell and is not enriched at lacO arrays (p-value=0.234; std=0.193; n=5) (Fig 1B). We composed a list of candidate interaction partners from factors previously identified to interact with Piwi (Fig 1C) and the top hits identified from at least two out of three RNAi screens aimed to identify proteins that are involved in transposon repression (Czech et al. 2013; Handler et al. 2013; Muerdter et al. 2013). We used our heterologous cell culture system to determine whether factors on our candidate list engaged in direct interactions with Piwi. We quantified our images using a homemade script that employs the Costes method on the nuclei, deemed regions of interest (ROI), of cells expressing both the lacI- mKate2-Piwi dots and CFP-tagged candidates where a p-value greater than or equal to 0.95 indicates co-localization. (Costes et al. 2004; Schindelin et al. 2012).

### **Identification of Piwi Direct Interactors in Nuage**

Piwi is believed to be loaded with piRNAs in the nuage. Using our heterologous cell culture system, we tested for biochemical interactions between Piwi and several well-known nuage components. CFP-tagged Hen1 (p-value=0.096; std=0.124; n=5), Gasz (p-value=0.154; std=0.344; n=5), Nibbler (Nbr) (p-value=0.00; std=0.00; n=5), Minotaur (Mino) (p-value= 0.018; std=0.040; n=5), Vreteno (Vret) (p-value=0.212; std=0.172; n=5), and BoYb (p-value=0.062; std=0.139; n=5) were not recruited to lacImKate2-Piwi dots (Fig S1). Furthermore, the core cytoplasmic components of the primary processing machinery,

including the helicase Armitage (Armi) (p-value=0.296; std=0.362; n=5) and the endonuclease Zucchini (Zuc) (p-value=0.196; std = 0.210; n=5) were not recruited by Piwi to the lacO foci (Fig 2A, S1). In steady-state, Zuc is localized to the mitochondrial membrane (Malone et al. 2009; Haase et al. 2010). We generated Zuc with a mutated predicted mitochondrial localization signal ( $\Delta$ MLS-Zuc) to test whether the subcellular localization of Zuc affected the result of our assay. We found that  $\Delta$ MLS-Zuc was also not recruited to the lacI-mKate2-Piwi foci (p-value=0.098; std=0.197; n=5) (Fig 2A). Other labs concluded that Armi interacts with Piwi in an RNA-independent manner (Olivieri et al. 2010; Ilyin et al. 2017). We performed co-IP of ArmiHA and FLAG-Piwi from S2 cells in the presence and absence of RNase A treatment. While Armi and Piwi are pulled down together, the majority of the interaction between Piwi and Armi is lost upon RNase A treatment (Fig 2B). Based on this result, we believe most stable interactions between Armi and Piwi are mediated by RNA. Previously identified as Piwi interactors, CFP-tagged Vasa (p-value= 0.996; std=0.009; n=5), Shutdown (Shu) (p-value=0.994; std=0.013; n=5) and Papi (p-value=1.00; std=0.00; n=5) were recruited by Piwi to lacO foci, indicating a direct biochemical interaction (Fig 2C). Squash (Squ) (p-value=0.968; std=0.066;n=5), an RNase H II-related protein involved in the effector step of the piRNA pathway (Pane et al. 2007; Haase et al. 2010), and Qin (p-value=0.986; std=0.031; n=5), which is required for heterotypic ping-pong amplification (Zhang et al. 2011; Zhang et al. 2014a) were also found to directly interact with Piwi (Fig 2C). We identified the RNA helicase Spindle-E (Spn-E) (p-value=0.996; std=0.005; n=5), but not the Vasa paralog Belle (Bel) (p-value=0.028; std=0.063; n=5), to be a novel interactor with Piwi (Fig 2C, S1). We used co-immunoprecipitation of tagged Piwi and Spn-E proteins in S2 cells to confirm this interaction (Fig 2D). Immunoprecipitation of

FLAG-tagged Piwi readily pulled down HA-tagged Spn-E in an RNA-independent manner. We tested for recruitment of components of the ping-pong piRNA processing (4P) complex to lacO foci and identified Aubergine (Aub) (p-value=0.990; std=0.010; n=5), but not Argonaute3 (Ago3) (p-value=0.010; std=0.017; n=5) or Krimper (p-value=0.00; std=0.00; n=5), as a novel interactor with Piwi (Fig 2C, S1). We also tested whether CFPtagged Piwi could be recruited to the lacI-mKate2-Piwi foci and found that Piwi (p-value= 0.162; std=0.362) could not form dimers with itself in our assay (Fig S1). We used co-immunoprecipitation of tagged Piwi, Aub, and Ago3 proteins in S2 cells to confirm their interaction (Fig 2E). Immunoprecipitation of FLAG-tagged Piwi readily pulled down HA-tagged Aub but not HA-tagged Ago3; as a control, immunoprecipitated FLAG-tagged Ago3 pulled down HA-tagged Aub in a predominantly RNA-dependent manner. Importantly, the interaction between Piwi and Aub is not mediated by RNA as it was also observed upon treatment of the lysate with RNase A. Overall, the heterologous cell culture assay, with novel interactions validated by co-IP, suggests that Piwi directly interacts with nuage components – Vasa, Papi, Shu, Squash, Qin, Spn-E, and Aub.

### **Piwi Directly Interacts with the Exon Junction Complex (EJC) Component Mago**

The mechanism by which RNA transcripts are translocated from the nucleus to the piRNA processing machinery remains elusive. Within the list of potential direct Piwi-interacting partners, some are core components of the exon junction complex (EJC) and EJC accessory proteins. Nxt1 (p-value=0.024; std=0.054; n=5), RnpS1 (pvalue=0.268; std=0.343; n=5), and Acinus (Acn) (p-value=0.122; std=0.267; n=5) have previously been implicated in the piRNA pathway but did not colocalize with Piwi (Fig 3A). Mago (p-value=1.00; std=0.00; n=5) was identified as a novel, direct Piwi-interacting factor (Fig 3A).

We confirmed the interaction of Mago and Piwi by co-IP followed by western blot using overexpression of FLAG-tagged Piwi and HA-tagged Mago in S2 cells. We found that the interaction is not affected upon RNase A treatment, indicating it is not mediated by RNA (Fig 3B). Interestingly, we were also able to co-IP FLAG-tagged Piwi and HA-tagged Nxt1 in the presence and absence of RNase A treatment (Fig 3C). These results suggest that Piwi interacts with multiple components of the EJC, but of the factors tested, only directly interacts with Mago. We tested other factors that affect the translation, surveillance, and localization of mRNAs. Of these factors, we found that Fmr1 (p-value=0.00; std=0.00; n=5), Karyβ3 (pvalue=0.154; std=0.238; n=5) and Zn72D (p-value=0.00; std=0.00; n=5) were not recruited to lacI-mKate2-Piwi foci (Fig S2). Furthermore, Piwi failed to recruit the components of the transcription export (TREX) complex, UAP56 (p-value=0.002; std=0.004; n=5) and THOC5 (p-value=0.030; std=0.051; n=5) (Fig S2).

**Exclusive recruitment of the Rhino-Deadlock-Cutoff (RDC) complex to piRNA clusters may be mediated by Piwi's direct interactions with Cutoff and a novel interactor – Arp6**

We tested for direct interactions between Piwi and the factors that comprise the Rhino-Deadlock-Cutoff (RDC) complex, which exclusively binds H3K9me3 marks along with piRNA clusters and is required for pre-piRNA transcription (Klattenhoff et al. 2009; Pane et al. 2011; Le Thomas et al. 2014; Mohn et al. 2014; Zhang et al. 2014b; Chen et al. 2016). Rhino was not recruited to lacO foci (p-value=0.018; std=0.040; n=5) (Fig 4A). However, another RDC component, Cutoff (Cuff) (p-value=1.00; std=0.00; n=5), was recruited by Piwi (Fig 4A). Not much is known about the actin-related protein 6 (Arp6), which colocalizes with HP1 on centric heterochromatin (Kato et al. 2001). However, our assay identified Arp6 as a novel direct interactor with Piwi (p-value=1.00; std=0.00; n=5) (Fig 4A). We were able

to confirm this interaction by co-IP of tagged Arp6 and Piwi expressed in S2 cells in the presence and absence of RNase A treatment (Fig 4B). Furthermore, we were able to show in S2 cells Arp6 RNA-independently co-IPs with Rhino but does not co-IP with HP1a (Fig 4C). This data suggests that direct interactions mediate the exclusive recruitment of the RDC complex to piRNA clusters between Piwi and the factors Arp6 and Cuff. Our lab showed that Cuff plays an important role in protecting pre-piRNA transcripts from degradation by the exonuclease Rat1 (Chen et al. 2016). We found that Piwi directly interacts with Rat1 (p-value=0.998; std=0.004; n=5) but not with its binding partner, Rai1 (p-value=0.438; std=0.267; n=5) (Fig 4D). This data suggests that Piwi may play a role in determining whether piRNA cluster-derived transcripts will be protected for transport to nuage or degraded by exonucleases.

## **Discussion**

In this study, we focused on elucidating the interactome of Piwi, which is responsible for identifying targets and recruiting the silencing machinery necessary to establish a repressive chromatin state at target loci (Saito et al. 2009; Olivieri et al. 2010; Handler et al. 2011; Ishizu et al. 2011; Klenov et al. 2011; Shpiz et al. 2011; Sienski et al. 2012; Gu and Elgin 2013; Le Thomas et al. 2013; Rozhkov et al. 2013). We employed an optimized two-hybrid (2H) system to discriminate between direct and indirect interactions with Piwi.

A heterologous cell culture assay discriminates between direct and indirect interactions with Piwi

Previously, many factors were identified genetically to be involved in the piRNA pathway. Some of these factors were determined to engage in biochemical interactions with Piwi via co-IP followed by western blot. However, such experiments can't distinguish

between direct and indirect interactions. Besides, traditional techniques such as affinity purification followed by mass spectrometry (AP/MS) can be plagued with unspecific hits and miss detections of transient or weak interactions. Our study developed a heterologous cell culture system capable of detecting transient or weak direct protein-protein interactions. The read-out of our assay is observed using co-localization of fluorescently labeled candidate proteins and lacI-mKate2-Piwi dots (Fig 1A-B). At the time of our study, there were 24 factors previously identified to interact with Piwi (Fig 1C). We generated a list of candidates from these factors, as well as factors previously identified as potential piRNA pathway components in genetic screens (Czech et al. 2013; Handler et al. 2013; Muerdter et al. 2013). Only small fractions of candidate factors are recruited to the lacO foci, suggesting only a small number of proteins directly interact with Piwi. In contrast, other factors' interactions with Piwi are mediated by a complex or RNA. It should be noted that while our heterologous cell culture assay does not produce false-positive results due to the low likelihood of mammalian somatic proteins mediating interactions with Piwi, we acknowledge that our assay is susceptible to false negatives. False-negative results could be due to (1) the missing modification made on the prey protein in *D. melanogaster* to interact with Piwi, (2) a conformational change that is made when Piwi is loaded with piRNAs, or (3) the prey protein is bound to a structure that cannot enter the nucleus.

Mago, a component of the exon junction complex (EJC), provides the connection between Piwi and the EJC to facilitate the export of piRNA precursors to nuage

How piRNA precursors are exported into the nuage (Pandey et al. 2017; Rogers et al. 2017) for further processing into mature piRNAs are remained unclear. Previously, it had been proposed that the transcript export (TREX) complex component UAP56, along with

Rhino, may bind piRNA precursor transcripts and facilitate their release into nuage by the nuclear pore complex (Zhang et al. 2012; Zhang et al. 2014b). When we tested the THO/TREX complex components UAP56 and THOC5, we did not find any direct interactions with Piwi (Fig S2). Our candidate list contained several components of the exon junction complex (EJC) and EJC accessory proteins. Despite previously being shown to be important for proper localization of flam piRNA precursors outside of the nucleus (Dennis et al. 2016), Nxt1 did not directly interact with Piwi in our assay (Fig 3A). However, Nxt1 can be pulled down with Piwi in an RNA-independent manner (Fig 3C), suggesting that Piwi interacts with Nxt1 via interactions with other EJC components. While RnpS1 and Acinus (Acn) have been implicated in mRNA quality control, pre-mRNA splicing, and transcriptional regulation (Schwerk et al. 2003), they only affect flam transcript intranuclear localization (Dennis et al. 2016) and do not directly interact with Piwi (Fig 3A). We did, however, identify the EJC component Mago as a novel direct Piwi interactor (Fig 3A-B). Mago was also previously shown to affect flam transcript export from the nucleus, but not flam piRNA production (Dennis et al. 2016). Overall, our data suggest that Mago directly interacts with Piwi, allowing the EJC to associate with bound target transcripts to be channeled into nuage for piRNA processing.

Piwi interacts with the Rhino-Deadlock-Cutoff (RDC) complex component Cutoff and a novel interactor – Arp6

Transcription of piRNA clusters requires RDC complex binding to the H3K9me3 marks. (Klattenhoff et al. 2009; Pane et al. 2011; Le Thomas et al. 2014; Mohn et al. 2014; Zhang et al. 2014b; Chen et al. 2016). How the RDC complex exclusively binds the H3K9me3 marks located within piRNA clusters is unclear. We found that Piwi directly interacts with

the RDC component Cuff (Cuff) but not Rhino (Fig 4A). In addition, we identified Arp6 as a novel direct interactor of Piwi in our system (Fig 4A-B). We showed that Arp6 interacts with Piwi in an RNA-independent manner and can be pulled down with Rhino in an RNA-independent manner (Fig 4B-C). It was previously shown that Arp6 colocalized with HP1a (Kato et al. 2001); however, we were able to pull down Arp6 specifically with Rhino (HP1d) and not with HP1a (Fig 4C). This data suggests a direct biochemical “bridge” between Piwi and the RDC complex at piRNA clusters mediated by interactions with Cuff and Arp6. We propose that when Piwi/piRNA complexes recognize nascent pre-piRNA transcripts, Piwi biochemically interacts with Cuff and Arp6 resulting in the exclusive recruitment and binding of RDC complexes to the H3K9me3 deposited on piRNA clusters. This observation would expand the RDC complex to the PARDC complex, where Arp6 and Rhino interact and Piwi guides the ARDC complex to target binding sites via its interactions with Cuff and Arp6. Cuff was shown to be essential in the protection of piRNA precursor transcripts from degradation before their transportation to nuage (Chen et al. 2016). Chen et al. showed that Cuff and the exonuclease Rat1, and its binding partner Rai1, play antagonistic roles in processing piRNA precursor transcripts. We found that Piwi directly interacts with Rat1, but not Rai1 (Fig 4D). We previously showed that Piwi could sequester a transcript in nuage, resulting in the processing of the transcript by the piRNA processing machinery (Rogers et al. 2017). It is tempting to suggest that Piwi’s interaction with Rat1 implies that after directing the RDC complex to piRNA clusters, Piwi is “handed” nascent precursor transcripts and ultimately determines the transcript’s fate. Piwi’s interactions with Rat1 and the EJC means it could channel the transcript for degradation by Rat1 or protect the transcript and channel it ultimately to be exported to nuage for processing into piRNAs.

Our optimized 2H system in mammalian cell culture allows for detecting weak or transient direct protein-protein interactions *in vivo*. This approach utilizes fluorescent detection of recruitment of the prey protein to the lacO tethered prey protein. Our approach can be employed in a modular fashion, switching the cell type containing the lacO arrays (as necessary) and the lacI-tagged bait protein to identify direct protein-protein interactions for any factor. We utilized our technique in two ways: (1) determining whether previously identified interactions with Piwi are direct and (2) screening candidates for potential novel direct interactions with Piwi. Our approach has allowed for elaborating the Piwi interactome by discerning direct interactions, proving it to be a useful tool for extending our understanding of the piRNA pathway. Furthermore, we believe this approach will prove useful for a wide audience screening for and establishing verified direct protein-protein interactions for factors in a variety of complex pathways.

## **Methods and Materials**

### *Piwi-tethering assay in U2OS cells*

The lacO U2OS cells were a generous gift from the Rippe lab (Jegou et al. 2009) and were grown in complete DMEM media (DMEM high glucose GlutaMAX supplement pyruvate (ThermoFisher, 10569010), 10% heat-inactivated FBS (Gemini Bio-Products, 100-106), 1% Penicillin-Streptomycin (ThermoFisher, 15140122)). The lacI-mKate2-Piwi vector was created using pcDNA6.2/N-EmGFP-DEST (Invitrogen, V32620) and lacI and mKate2 overlap PCR products. lacO U2OS cells expressing lacI-mKate2-Piwi were selected using 5µg/mL blasticidin (Invitrogen, A11139- 03). A stable cell line was established by colony picking, and cells positive for mKate2 expression were sorted by the Caltech Sorting Facility, using their recommended protocol based on mKate2 expression. The CFP-destination vector

was constructed by ligation of the CFP sequence from the pPCW vector (DGRC#1085) into the XbaI and NotI restriction sites of the backbone of pcDNA6.2/N-EmGFP-DEST (Invitrogen, V32620) using T4 Ligase (NEB, M0202) according to the manual. Entry vectors containing Asterix, RnpS1, Mago, Arp6, Nxt1, His2AV, Panoramix, Caf1-105, H1, Acn, and YL-1 were generated by PCR amplification of the respective cDNA and Topo cloning into the pENTR-D-TOPO vector according to the manufacturer's suggestion (Invitrogen K240020). Expression vectors for CFP-tagged candidates were generated from the CFP-destination vector and entry vectors using LR clonase (Invitrogen, 11791100) according to the manual. LacImKate2-Piwi LacO U2OS cells were transfected with 2.5 $\mu$ g of the CFP or CFP-tagged constructs using TransIT-LT1 transfection reagent (Mirus, MIR 2305) according to the manual. 48 hours after transfection, cells were fixed with 4% paraformaldehyde in PBS at room temperature (RT) for 20 minutes then washed 3x in PBS. Cells were incubated in 1% Triton X-100 for 10 minutes at RT then stained with 1:1000 diluted 5mg/ml DAPI at RT for 5 minutes. Then the cells were washed 3x in PBS and mounted with a coverslip using Prolong Gold antifade reagent before imaging on a confocal microscope. Quantification of co-localization was done with the in-house script to determine the Costes p-value (Costes et al. 2004). A region of interest (ROI) for each cell was defined by where DAPI signal overlapped with CFP and/or mKate2 signals. The Costes p-value is based off 10,000 iterations of scrambling the ROI for the channels, where a Costes p-value  $\geq 0.95$  indicates positive co-localization (Costes et al. 2004). For each factor, an N of 5 images was used, and the average p-value and standard deviation were calculated.

*Co-immunoprecipitation with Piwi from S2 cells*

Schneider S2 cells were cultured in complete Schneider's *Drosophila* medium (Schneider's *Drosophila* medium (ThermoFisher, 21720024), 10% heat inactivated FBS (Gemini Bio-Products, 100-106), 1% Penicillin-Streptomycin (ThermoFisher, 15140122)). Plasmids were generated using LR clonase (Invitrogen, 11791100) according to the manual using entry vectors created in the lab (see above) and *Drosophila* Gateway Vector Collection (DGVC) destination vectors – pAGW (DGRC#1071), pAFW (DGRC#1111), pAHW (DGRC#1095), pAWH (DGRC#1096) – expressed by the Actin5c promoter. Cells were transfected using TransIT-LT1 transfection reagent (Mirus, MIR 2305) according to the manual with 3 $\mu$ g of total plasmid. For co-immunoprecipitation of proteins expressed in S2 cells, a 3.5mm culture plate of transfected S2 cells was mechanically lysed and incubated for 20min on ice in 150 $\mu$ L S2 lysis buffer (20mM Tris pH7.4, 150mM KCL, 0.1% Tween-20, 0.1% NP-40 Igepal, EDTA-free Complete Protease Inhibitor Cocktail (Roche, 11836170001), 100 $\mu$ g/mL RNase A). The supernatant was cleared by centrifugation at 4,000g for 20 minutes at 4°C. 10% Input sample was collected from the supernatant at concentrations of 1-3 $\mu$ g/ $\mu$ L. Anti-FLAG M2 beads (Sigma Aldrich, M8823) were blocked in 5mg/ml BSA for 10 minutes at 4°C, followed by washing in S2 lysis buffer. Blocked beads were added to the supernatant and rotated at 4°C overnight. Beads were washed three times in PBS + 0.05% Tween-20 and eluted by boiling in reducing SDS loading buffer. Antibodies used for western blots were mouse monoclonal anti-FLAG M2 (Sigma Aldrich, F3165) and mouse monoclonal anti-HA (Sigma Aldrich, H3663) at 1:10000 in 5% milk in PBST. Homemade Anti-GFP rabbit polyclonal antibody was used at 1:2000 in 5% milk in PBST.

## References

- Anand, A., & Kai, T. (2012). The tudor domain protein kumo is required to assemble the nuage and to generate germline piRNAs in *Drosophila*. *EMBO J*, *31*(4), 870-882. doi:10.1038/emboj.2011.449
- Anand, A., & Kai, T. (2014). Antisense piRNA amplification, but not piRNA production or nuage assembly, requires the Tudor-domain protein Qin Response. *Embo Journal*, *33*(6), 540-541. doi:10.1002/emboj.201387548
- Andress, A., Bei, Y., Fonslow, B. R., Giri, R., Wu, Y., Yates, J. R., 3rd, & Carthew, R. W. (2016). Spindle-E cycling between nuage and cytoplasm is controlled by Qin and PIWI proteins. *J Cell Biol*, *213*(2), 201-211. doi:10.1083/jcb.201411076
- Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., & Hannon, G. J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell*, *128*(6), 1089-1103. doi:10.1016/j.cell.2007.01.043
- Brower-Toland, B., Findley, S. D., Jiang, L., Liu, L., Yin, H., Dus, M., . . . Lin, H. (2007). *Drosophila* PIWI associates with chromatin and interacts directly with HP1a. *Genes Dev*, *21*(18), 2300-2311. doi:10.1101/gad.1564307
- Chen, Y. A., Stuwe, E., Luo, Y., Ninova, M., Le Thomas, A., Rozhavskaia, E., . . . Aravin, A. A. (2016). Cutoff Suppresses RNA Polymerase II Termination to Ensure Expression of piRNA Precursors. *Mol Cell*, *63*(1), 97-109. doi:10.1016/j.molcel.2016.05.010
- Cook, H. A., Koppetsch, B. S., Wu, J., & Theurkauf, W. E. (2004). The *Drosophila* SDE3 homolog armitage is required for oskar mRNA silencing and embryonic axis specification. *Cell*, *116*(6), 817-829. doi:10.1016/s0092-8674(04)00250-8
- Costes, S. V., Daelemans, D., Cho, E. H., Dobbin, Z., Pavlakis, G., & Lockett, S. (2004). Automatic and quantitative measurement of protein-protein colocalization in live cells. *Biophys J*, *86*(6), 3993-4003. doi:10.1529/biophysj.103.038422
- Cox, D. N., Chao, A., & Lin, H. (2000). piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development*, *127*(3), 503-514. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/10631171>
- Czech, B., Preall, J. B., McGinn, J., & Hannon, G. J. (2013). A transcriptome-wide RNAi screen in the *Drosophila* ovary reveals factors of the germline piRNA pathway. *Mol Cell*, *50*(5), 749-761. doi:10.1016/j.molcel.2013.04.007
- Dennis, C., Brasslet, E., Sarkar, A., & Vauray, C. (2016). Export of piRNA precursors by EJC triggers assembly of cytoplasmic Yb-body in *Drosophila*. *Nat Commun*, *7*, 13739. doi:10.1038/ncomms13739
- Donertas, D., Sienski, G., & Brennecke, J. (2013). *Drosophila* Gtsf1 is an essential component of the Piwi-mediated transcriptional silencing complex. *Genes Dev*, *27*(15), 1693-1705. doi:10.1101/gad.221150.113
- Gangaraju, V. K., Yin, H., Weiner, M. M., Wang, J., Huang, X. A., & Lin, H. (2011). *Drosophila* Piwi functions in Hsp90-mediated suppression of phenotypic variation. *Nat Genet*, *43*(2), 153-158. doi:10.1038/ng.743
- Gu, T., & Elgin, S. C. (2013). Maternal depletion of Piwi, a component of the RNAi system, impacts heterochromatin formation in *Drosophila*. *PLoS Genet*, *9*(9), e1003780. doi:10.1371/journal.pgen.1003780

- Gunawardane, L. S., Saito, K., Nishida, K. M., Miyoshi, K., Kawamura, Y., Nagami, T., . . . Siomi, M. C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science*, *315*(5818), 1587-1590. doi:10.1126/science.1140494
- Haase, A. D., Fenoglio, S., Muerdter, F., Guzzardo, P. M., Czech, B., Pappin, D. J., . . . Hannon, G. J. (2010). Probing the initiation and effector phases of the somatic piRNA pathway in *Drosophila*. *Genes Dev*, *24*(22), 2499-2504. doi:10.1101/gad.1968110
- Han, B. W., Wang, W., Li, C., Weng, Z., & Zamore, P. D. (2015). Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. *Science*, *348*(6236), 817-821. doi:10.1126/science.aaa1264
- Handler, D., Meixner, K., Pizka, M., Lauss, K., Schmied, C., Gruber, F. S., & Brennecke, J. (2013). The genetic makeup of the *Drosophila* piRNA pathway. *Mol Cell*, *50*(5), 762-777. doi:10.1016/j.molcel.2013.04.031
- Harris, A. N., & Macdonald, P. M. (2001). Aubergine encodes a *Drosophila* polar granule component required for pole cell formation and related to eIF2C. *Development*, *128*(14), 2823-2832. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/11526087>
- Hayashi, R., Schnabl, J., Handler, D., Mohn, F., Ameres, S. L., & Brennecke, J. (2016). Genetic and mechanistic diversity of piRNA 3'-end formation. *Nature*, *539*(7630), 588-592. doi:10.1038/nature20162
- Horwich, M. D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., & Zamore, P. D. (2007). The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr Biol*, *17*(14), 1265-1272. doi:10.1016/j.cub.2007.06.030
- Ilyin, A. A., Ryazansky, S. S., Doronin, S. A., Olenkina, O. M., Mikhaleva, E. A., Yakushev, E. Y., . . . Shevelyov, Y. Y. (2017). Piwi interacts with chromatin at nuclear pores and promiscuously binds nuclear transcripts in *Drosophila* ovarian somatic cells. *Nucleic Acids Res*, *45*(13), 7666-7680. doi:10.1093/nar/gkx355
- Ishizu, H., Nagao, A., & Siomi, H. (2011). Gatekeepers for Piwi-piRNA complexes to enter the nucleus. *Curr Opin Genet Dev*, *21*(4), 484-490. doi:10.1016/j.gde.2011.05.001
- Iwasaki, Y. W., Murano, K., Ishizu, H., Shibuya, A., Iyoda, Y., Siomi, M. C., . . . Saito, K. (2016). Piwi Modulates Chromatin Accessibility by Regulating Multiple Factors Including Histone H1 to Repress Transposons. *Mol Cell*, *63*(3), 408-419. doi:10.1016/j.molcel.2016.06.008
- Jegou, T., Chung, I., Heuvelman, G., Wachsmuth, M., Gorisch, S. M., Greulich-Bode, K. M., . . . Rippe, K. (2009). Dynamics of telomeres and promyelocytic leukemia nuclear bodies in a telomerase-negative human cell line. *Mol Biol Cell*, *20*(7), 2070-2082. doi:10.1091/mbc.E08-02-0108
- Kato, M., Sasaki, M., Mizuno, S., & Harata, M. (2001). Novel actin-related proteins in vertebrates: similarities of structure and expression pattern to Arp6 localized on *Drosophila* heterochromatin. *Gene*, *268*(1-2), 133-140. doi:10.1016/s0378-1119(01)00420-6
- Kibanov, M. V., Egorova, K. S., Ryazansky, S. S., Sokolova, O. A., Kotov, A. A., Olenkina, O. M., . . . Olenina, L. V. (2011). A novel organelle, the piNG-body, in the nuage of

- Drosophila male germ cells is associated with piRNA-mediated gene silencing. *Mol Biol Cell*, 22(18), 3410-3419. doi:10.1091/mbc.E11-02-0168
- Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Khurana, J. S., . . . Theurkauf, W. E. (2009). The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell*, 138(6), 1137-1149. doi:10.1016/j.cell.2009.07.014
- Klenov, M. S., Sokolova, O. A., Yakushev, E. Y., Stolyarenko, A. D., Mikhaleva, E. A., Lavrov, S. A., & Gvozdev, V. A. (2011). Separation of stem cell maintenance and transposon silencing functions of Piwi protein. *Proc Natl Acad Sci U S A*, 108(46), 18760-18765. doi:10.1073/pnas.1106676108
- Kodjabachian, L., Delaage, M., Maurel, C., Miassod, R., Jacq, B., & Rosset, R. (1998). Mutations in ccf, a novel Drosophila gene encoding a chromosomal factor, affect progression through mitosis and interact with Pc-G mutations. *EMBO J*, 17(4), 1063-1075. doi:10.1093/emboj/17.4.1063
- Ku, H. Y., Gangaraju, V. K., Qi, H., Liu, N., & Lin, H. (2016). Tudor-SN Interacts with Piwi Antagonistically in Regulating Spermatogenesis but Synergistically in Silencing Transposons in Drosophila. *PLoS Genet*, 12(1), e1005813. doi:10.1371/journal.pgen.1005813
- Le Thomas, A., Rogers, A. K., Webster, A., Marinov, G. K., Liao, S. E., Perkins, E. M., . . . Toth, K. F. (2013). Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev*, 27(4), 390-399. doi:10.1101/gad.209841.112
- Le Thomas, A., Stuwe, E., Li, S., Du, J., Marinov, G., Rozhkov, N., . . . Aravin, A. A. (2014). Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. *Genes Dev*, 28(15), 1667-1680. doi:10.1101/gad.245514.114
- Li, C., Vagin, V. V., Lee, S., Xu, J., Ma, S., Xi, H., . . . Zamore, P. D. (2009). Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell*, 137(3), 509-521. doi:10.1016/j.cell.2009.04.027
- Liang, X., Shan, S., Pan, L., Zhao, J., Ranjan, A., Wang, F., . . . Zhou, Z. (2016). Structural basis of H2A.Z recognition by SRCAP chromatin-remodeling subunit YL1. *Nat Struct Mol Biol*, 23(4), 317-323. doi:10.1038/nsmb.3190
- Lim, A. K., & 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(16), 6714-6719. doi:10.1073/pnas.0701920104
- Lo, P. K., Huang, Y. C., Poulton, J. S., Leake, N., Palmer, W. H., Vera, D., . . . Deng, W. M. (2016). RNA helicase Belle/DDX3 regulates transgene expression in Drosophila. *Dev Biol*, 412(1), 57-70. doi:10.1016/j.ydbio.2016.02.014
- Lopez, A., Higuete, D., Rosset, R., Deutsch, J., & Peronnet, F. (2001a). corto genetically interacts with Pc-G and trx-G genes and maintains the anterior boundary of Ultrabithorax expression in Drosophila larvae. *Mol Genet Genomics*, 266(4), 572-583. doi:10.1007/s004380100572
- Lopez, A., Higuete, D., Rosset, R., Deutsch, J., & Peronnet, F. (2001b). corto genetically interacts with Pc-G and trx-G genes and maintains the anterior boundary of

- Ultrabithorax expression in *Drosophila* larvae. *Molecular Genetics and Genomics*, 266(4), 572-583. doi:10.1007/s004380100572
- Ma, L., Buchold, G. M., Greenbaum, M. P., Roy, A., Burns, K. H., Zhu, H. F., . . . Matzuk, M. M. (2009). GASZ Is Essential for Male Meiosis and Suppression of Retrotransposon Expression in the Male Germline. *Plos Genetics*, 5(9). doi:ARTN e1000635  
10.1371/journal.pgen.1000635
- Malone, C. D., Brennecke, J., Dus, M., Stark, A., McCombie, W. R., Sachidanandam, R., & Hannon, G. J. (2009). Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell*, 137(3), 522-535. doi:10.1016/j.cell.2009.03.040
- Mendez, D. L., Kim, D., Chruszcz, M., Stephens, G. E., Minor, W., Khorasanizadeh, S., & Elgin, S. C. (2011). The HP1a disordered C terminus and chromo shadow domain cooperate to select target peptide partners. *Chembiochem*, 12(7), 1084-1096. doi:10.1002/cbic.201000598
- Mendez, D. L., Mandt, R. E., & Elgin, S. C. (2013). Heterochromatin Protein 1a (HP1a) partner specificity is determined by critical amino acids in the chromo shadow domain and C-terminal extension. *J Biol Chem*, 288(31), 22315-22323. doi:10.1074/jbc.M113.468413
- Mohn, F., Handler, D., & Brennecke, J. (2015). Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. *Science*, 348(6236), 812-817. doi:10.1126/science.aaa1039
- Mohn, F., Sienski, G., Handler, D., & Brennecke, J. (2014). The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in *Drosophila*. *Cell*, 157(6), 1364-1379. doi:10.1016/j.cell.2014.04.031
- Moshkovich, N., Nisha, P., Boyle, P. J., Thompson, B. A., Dale, R. K., & Lei, E. P. (2011). RNAi-independent role for Argonaute2 in CTCF/CP190 chromatin insulator function. *Genes Dev*, 25(16), 1686-1701. doi:10.1101/gad.16651211
- Muerdter, F., Guzzardo, P. M., Gillis, J., Luo, Y., Yu, Y., Chen, C., . . . 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(5), 736-748. doi:10.1016/j.molcel.2013.04.006
- Ohtani, H., Iwasaki, Y. W., Shibuya, A., Siomi, H., Siomi, M. C., & Saito, K. (2013). DmGTSF1 is necessary for Piwi-piRISC-mediated transcriptional transposon silencing in the *Drosophila* ovary. *Genes Dev*, 27(15), 1656-1661. doi:10.1101/gad.221515.113
- Olivieri, D., Senti, K. A., Subramanian, S., Sachidanandam, R., & Brennecke, J. (2012). The Cochaperone Shutdown Defines a Group of Biogenesis Factors Essential for All piRNA Populations in *Drosophila*. *Molecular Cell*, 47(6), 954-969. doi:10.1016/j.molcel.2012.07.021
- Olivieri, D., Sykora, M. M., Sachidanandam, R., Mechtler, K., & Brennecke, J. (2010). An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in *Drosophila*. *Embo Journal*, 29(19), 3301-3317. doi:10.1038/emboj.2010.212

- Pandey, R. R., Homolka, D., Chen, K. M., Sachidanandam, R., Fauvarque, M. O., & Pillai, R. S. (2017). Recruitment of Armitage and Yb to a transcript triggers its phased processing into primary piRNAs in *Drosophila* ovaries. *Plos Genetics*, *13*(8). doi:ARTN e1006956  
10.1371/journal.pgen.1006956
- Pane, A., Jiang, P., Zhao, D. Y., Singh, M., & Schupbach, T. (2011). The Cutoff protein regulates piRNA cluster expression and piRNA production in the *Drosophila* germline. *Embo Journal*, *30*(22), 4601-4615. doi:10.1038/emboj.2011.334
- Pane, A., Wehr, K., & Schupbach, T. (2007). zucchini and squash encode two putative nucleases required for rasiRNA production in the *Drosophila* germline. *Developmental Cell*, *12*(6), 851-862. doi:10.1016/j.devcel.2007.03.022
- Patil, V. S., & Kai, T. (2010). Repression of Retroelements in *Drosophila* Germline via piRNA Pathway by the Tudor Domain Protein Tejas. *Current Biology*, *20*(8), 724-730. doi:10.1016/j.cub.2010.02.046
- Peng, J. C., Valouev, A., Liu, N., & Lin, H. F. (2016). Piwi maintains germline stem cells and oogenesis in *Drosophila* through negative regulation of Polycomb group proteins. *Nature Genetics*, *48*(3), 283-291. doi:10.1038/ng.3486
- Qi, H., Watanabe, T., Ku, H. Y., Liu, N., Zhong, M., & Lin, H. F. (2011). The Yb Body, a Major Site for Piwi-associated RNA Biogenesis and a Gateway for Piwi Expression and Transport to the Nucleus in Somatic Cells. *Journal of Biological Chemistry*, *286*(5), 3789-3797. doi:10.1074/jbc.M110.193888
- Roelens, B., Clemot, M., Leroux-Coyau, M., Klapholz, B., & Dostatni, N. (2017). Maintenance of Heterochromatin by the Large Subunit of the CAF-1 Replication-Coupled Histone Chaperone Requires Its Interaction with HP1a Through a Conserved Motif. *Genetics*, *205*(1), 125-+. doi:10.1534/genetics.116.190785
- Rogers, A. K., Situ, K., Perkins, E. M., & Toth, K. F. (2017). Zucchini-dependent piRNA processing is triggered by recruitment to the cytoplasmic processing machinery. *Genes & Development*, *31*(18), 1858-1869. doi:10.1101/gad.303214.117
- Rozhkov, N. V., Hammell, M., & Hannon, G. J. (2013). Multiple roles for Piwi in silencing *Drosophila* transposons. *Genes & Development*, *27*(4), 400-412. doi:10.1101/gad.209767.112
- Saito, K., Inagaki, S., Mituyama, T., Kawamura, Y., Ono, Y., Sakota, E., . . . Siomi, M. C. (2009). A regulatory circuit for piwi by the large Maf gene traffic jam in *Drosophila*. *Nature*, *461*(7268), 1296-U1135. doi:10.1038/nature08501
- Saito, K., Ishizu, H., Komai, M., Kotani, H., Kawamura, Y., Nishida, K. M., . . . Siomi, M. C. (2010). Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in *Drosophila*. *Genes & Development*, *24*(22), 2493-2498. doi:10.1101/gad.1989510
- Saito, K., Nishida, K. M., Mori, T., Kawamura, Y., Miyoshi, K., Nagami, T., . . . Siomi, M. C. (2006). Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes & Development*, *20*(16), 2214-2222. doi:10.1101/gad.1454806
- Saito, K., Sakaguchi, Y., Suzuki, T., Suzuki, T., Siomi, H., & Siomi, M. C. (2007). Pimet, the *Drosophila* homolog of HEN1, mediates 2'-O-methylation of PIWI-interacting

- RNAs at their 3' ends. *Genes & Development*, 21(13), 1603-1608. doi:10.1101/gad.1563607
- Salvaing, J., Lopez, A., Boivin, A., Deutsch, J. S., & Peronnet, F. (2003). The Drosophila Corto protein interacts with Polycomb-group proteins and the GAGA factor. *Nucleic Acids Research*, 31(11), 2873-2882. doi:10.1093/nar/gkg381
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., . . . Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9(7), 676-682. doi:10.1038/Nmeth.2019
- Schwerk, C., Prasad, J., Degenhardt, K., Erdjument-Bromage, H., White, E., Tempst, P., . . . Reinberg, D. (2003). ASAP, a novel protein complex involved in RNA processing and apoptosis. *Molecular and Cellular Biology*, 23(8), 2981-2990. doi:10.1128/Mcb.23.8.2981-2990.2003
- Shpiz, S., Olovnikov, I., Sergeeva, A., Lavrov, S., Abramov, Y., Savitsky, M., & Kalmykova, A. (2011). Mechanism of the piRNA-mediated silencing of Drosophila telomeric retrotransposons. *Nucleic Acids Research*, 39(20), 8703-8711. doi:10.1093/nar/gkr552
- Sienski, G., Batki, J., Senti, K. A., Donertas, D., Tirian, L., Meixner, K., & Brennecke, J. (2015). Silencio/CG9754 connects the Piwi-piRNA complex to the cellular heterochromatin machinery. *Genes & Development*, 29(21), 2258-2271. doi:10.1101/gad.271908.115
- Sienski, G., Donertas, D., & Brennecke, J. (2012). Transcriptional Silencing of Transposons by Piwi and Maelstrom and Its Impact on Chromatin State and Gene Expression. *Cell*, 151(5), 964-980. doi:10.1016/j.cell.2012.10.040
- Verni, F., & Cenci, G. (2015). The Drosophila histone variant H2A.V works in concert with HP1 to promote kinetochore-driven microtubule formation. *Cell Cycle*, 14(4), 577-588. doi:10.4161/15384101.2014.991176
- Vrettos, N., Maragkakis, M., Alexiou, P., & Mourelatos, Z. (2017). Kc167, a widely used Drosophila cell line, contains an active primary piRNA pathway. *Rna*, 23(1), 108-118. doi:10.1261/rna.059139.116
- Wang, W., Han, B. W., Tipping, C., Ge, D. T., Zhang, Z., Weng, Z. P., & Zamore, P. D. (2015). Slicing and Binding by Ago3 or Aub Trigger Piwi-Bound piRNA Production by Distinct Mechanisms. *Molecular Cell*, 59(5), 819-830. doi:10.1016/j.molcel.2015.08.007
- Yu, Y., Gu, J., Jin, Y., Luo, Y. C., Preall, J. B., Ma, J. B. A., . . . Hannon, G. J. (2015). Panoramix enforces piRNA-dependent cotranscriptional silencing. *Science*, 350(6258), 339-+. doi:10.1126/science.aab0700
- Zamparini, A. L., Davis, M. Y., Malone, C. D., Vieira, E., Zavadil, J., Sachidanandam, R., . . . Lehmann, R. (2011). Vreteno, a gonad-specific protein, is essential for germline development and primary piRNA biogenesis in Drosophila. *Development*, 138(18), 4039-4050. doi:10.1242/dev.069187
- Zhang, F., Wang, J., Xu, J., Zhang, Z., Koppetsch, B. S., Schultz, N., . . . Theurkauf, W. E. (2012). UAP56 Couples piRNA Clusters to the Perinuclear Transposon Silencing Machinery. *Cell*, 151(4), 871-884. doi:10.1016/j.cell.2012.09.040

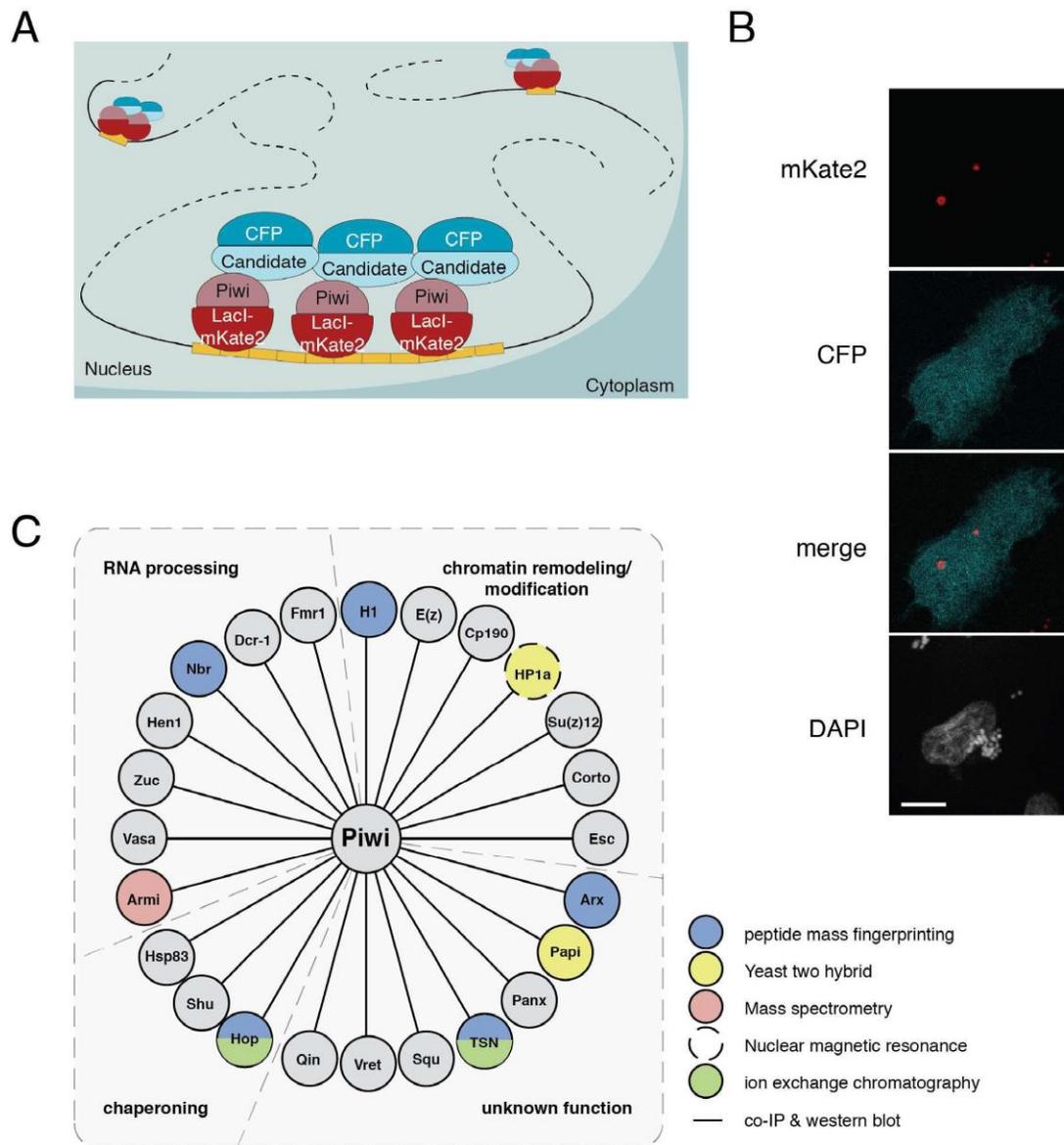
- Zhang, Z., Wang, J., Schultz, N., Zhang, F., Parhad, S. S., Tu, S., . . . Theurkauf, W. E. (2014). The HP1 Homolog Rhino Anchors a Nuclear Complex that Suppresses piRNA Precursor Splicing. *Cell*, *157*(6), 1353-1363. doi:10.1016/j.cell.2014.04.030
- Zhang, Z., Xu, J., Koppetsch, B. S., Wang, J., Tipping, C., Ma, S. M., . . . Zamore, P. D. (2011). Heterotypic piRNA Ping-Pong Requires Qin, a Protein with Both E3 Ligase and Tudor Domains (vol 44, pg 572, 2011). *Molecular Cell*, *44*(6), 1005-1005. doi:10.1016/j.molcel.2011.12.002

## Figures and Figure legends

### **Figure 1. A heterologous cell culture assay to distinguish between direct and indirect protein-protein interactions with *Drosophila melanogaster* Piwi.**

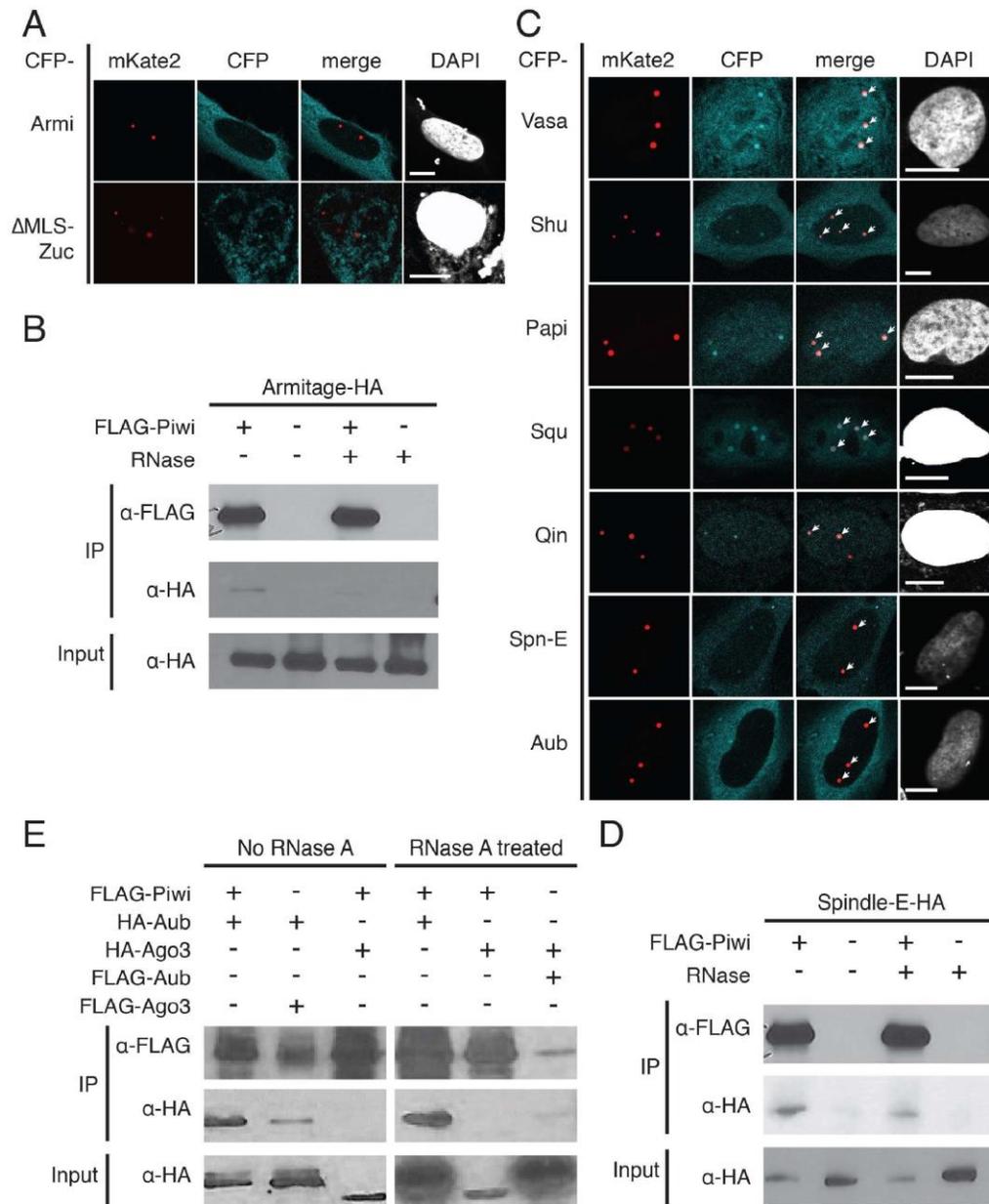
A) Schema of a heterologous cell culture assay to identify weak or transient interactions with Piwi. Human U2OS cells have three arrays composed of lacO repeats inserted in their genome. The bait protein, a fusion of the DNA binding domain of lacI with a fluorescent protein (mKate2) and Piwi, is tethered to the lacO arrays, resulting in distinct nuclear dots. The cells are transfected with CFP-tagged prey proteins. Colocalization of the CFP and mKate2 signals indicates a direct interaction between the bait and prey proteins. B) CFP alone does not colocalize with lacI-mKate2-Piwi. Representative confocal images of lacI-mKate2-Piwi U2OS cells transfected with CFP (p-value=0.234; std=0.193; n=5). Scale bar represents 10 $\mu$ m. C) A map of factors previously identified to interact with *D. melanogaster* Piwi in the piRNA pathway. It is not known which of these interactions with Piwi are direct or indirect (mediated by RNA or other proteins). The techniques by which these interactions have been determined are marked according to the legend.

**Figure 1**



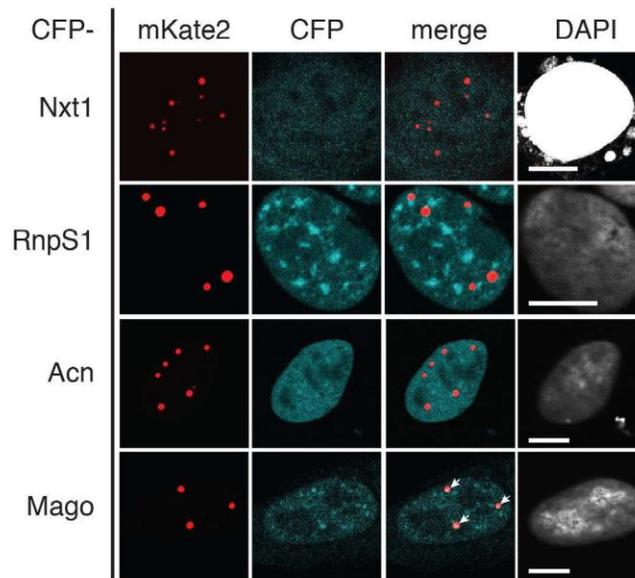
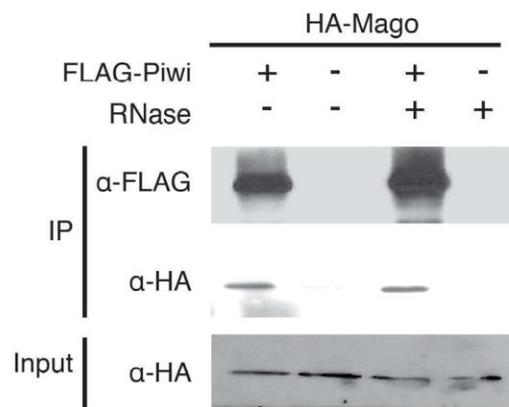
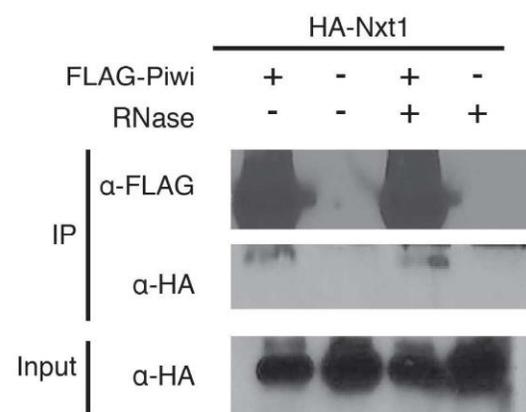
**Figure 2. Piwi directly interacts with nuage components Vasa, Shutdown, Papi, Squash, Qin, Spindle-E, and Aubergine.**

A) Piwi does not directly interact with core components of the primary piRNA processing machinery. Armitage (Armi) is not recruited by lacI-mKate2-Piwi (p-value=0.296; std=0.362; n=5). In steady state, Zucchini (Zuc) is localized to the mitochondrial membrane. Zuc with a mutated predicted mitochondrial localization signal ( $\Delta$ MLS-Zuc) is not recruited to lacO foci by Piwi (p-value=0.098; std=0.197; n=5). Representative images are shown. Scale bar represents 10 $\mu$ m. B) Piwi and Armi interactions are RNA-dependent. Co-immunoprecipitation (Co-IP) of HA-tagged Armi and FLAG-tagged Piwi in S2 cells indicates that the Piwi/Armi interaction is depleted upon RNase A treatment. C) Piwi does interact with some nuage localized factors. Vasa (p-value= 0.996; std=0.009; n=5), Shutdown (Shu) (p-value=0.994; std=0.013; n=5), Papi (p-value=1.00; std=0.00; n=5), Squash (Squ) (p-value=0.968; std=0.066; n=5), Qin (p-value=0.986; std=0.031; n=5), Spindle-E (Spn-E) (p-value=0.996; std=0.005; n=5), and Aubergine (Aub) (p-value=0.990; std=0.01; n=5) were recruited by lacI-mKate2-Piwi. Representative images are shown. White arrows indicate sites of recruitment of CFP-tagged candidates to lacI-mKate2-Piwi foci. Scale bars represent 10 $\mu$ m. D) Piwi and Spn-E interactions are RNA-independent. CoIP of HA-tagged Spn-E and FLAG-tagged Piwi in S2 cells indicates that the Piwi/Spn-E interaction is unaffected by RNase A treatment. E) Co-IP of Aub with Piwi from S2 cells is RNA-independent. The third *D.melanogaster* piwi clade Argonaute protein, Argonaute3 (Ago3), does not co-IP with Piwi, but does co-IP with Aub in an RNA-dependent manner.

**Figure 2**

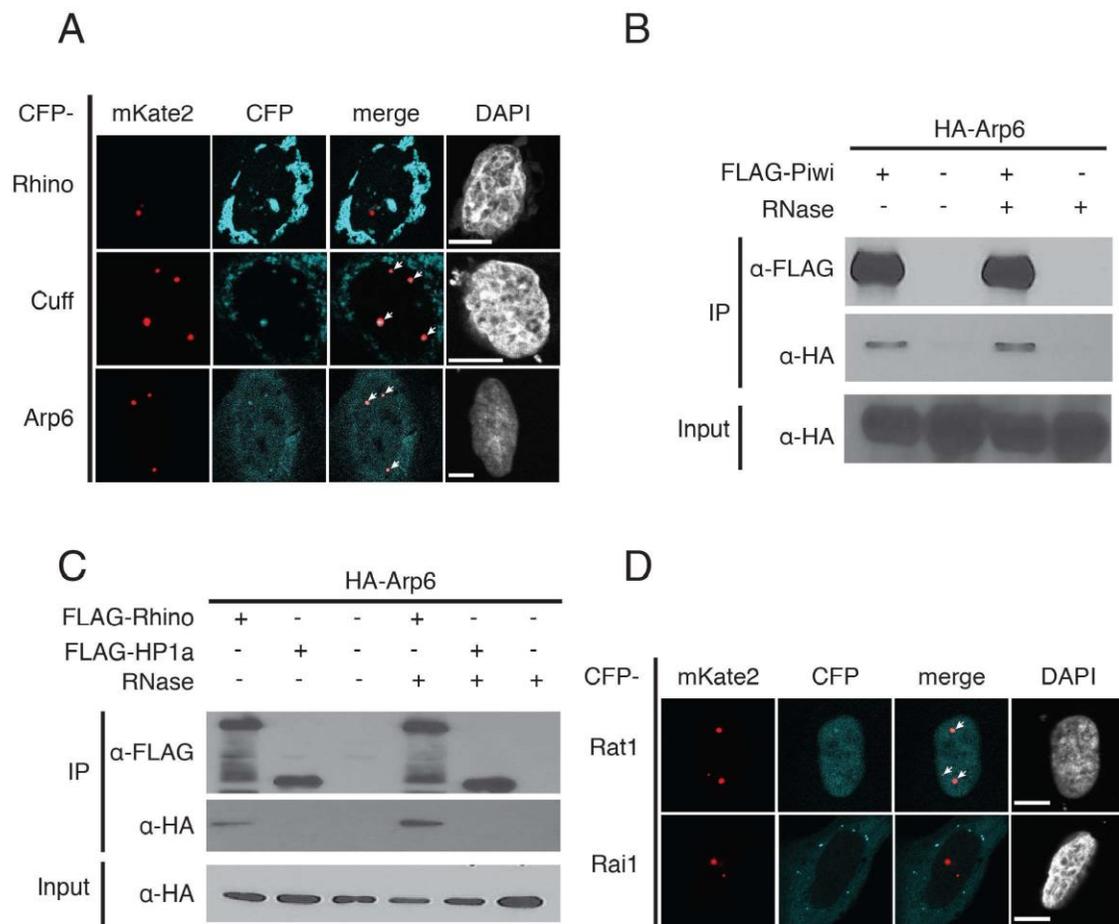
**Figure 3. Piwi directly interacts with the exon junction complex (EJC) component****Mago.**

A) Exportin Nxt1 (p-value=0.024 ; std=0.054 ; n=5) and EJC components RnpS1 (pvalue=0.268 ; std=0.343 ; n=5) and Acinus (Acn) (p-value=0.122 ; std=0.267 ; n=5) were not recruited to lacI-mKate2-Piwi foci. Piwi interacts with the EJC component Mago (pvalue=1.00; std=0.00 ; n=5). Representative confocal images are shown. White arrows indicate colocalization. Scale bars represent 10 $\mu$ m. B) Piwi and Mago interactions are RNA-independent. Co-IP of HA-tagged Mago and FLAG-tagged Piwi in S2 cells indicates that the Piwi/Mago interaction is unaffected by RNase A treatment. C) Piwi's interaction with Nxt1 is likely mediated by a protein complex. Piwi and Nxt1 interactions are RNA independent. Co-IP of HA-tagged Nxt1 and FLAG-tagged Piwi in S2 cells indicates that the Piwi/Nxt1 interaction is unaffected by RNase A treatment.

**Figure 3****A****B****C**

**Figure 4. Piwi directly interacts with the Rhino-Deadlock-Cutoff (RDC) complex component Cutoff and the novel interactor Arp6.**

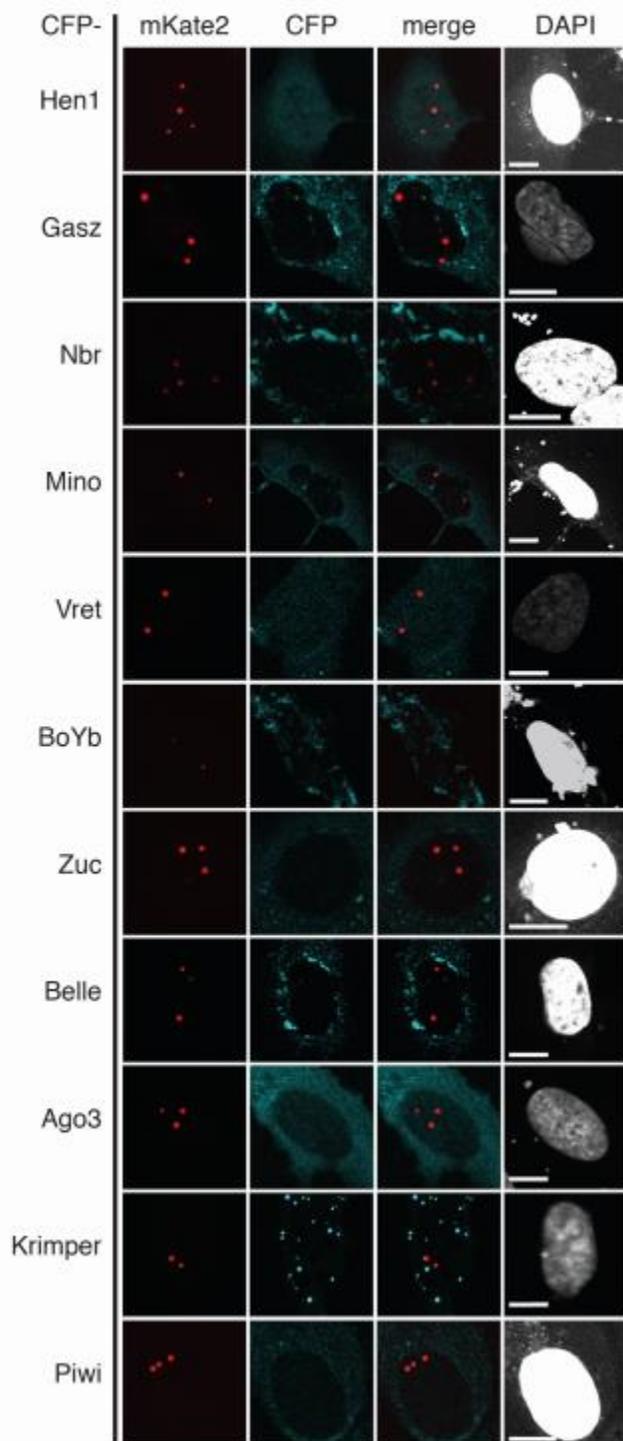
A) Piwi directly interacts with the RDC component Cutoff (Cuff) (p-value=1.00; std=0.00; n=5), but not with Rhino (p-value=0.018; std=0.040; n=5). Arp6 (p-value=1.00; std=0.00; n=5) was identified as a novel interactor of Piwi. Representative images are shown. White arrows indicate sites of colocalization. Scale bars represent 10 $\mu$ m. B) Piwi and Arp6 interactions are RNA-independent. Co-IP of HA-tagged Arp6 and FLAG-tagged Piwi in S2 cells indicates that the Piwi/Arp6 interaction is unaffected by RNase A treatment. C) Arp6 can be pulled down with Rhino, but not with HP1a. Co-IP of HA-tagged Arp6 and FLAGtagged Rhino in S2 cells indicates that the Arp6/Rhino interaction is unaffected by RNase A treatment. D) Piwi directly interacts with the exonuclease Rat1 (p-value=0.998; std=0.004; n=5), but not with its binding partner Rai1 (p-value=0.438; std=0.267; n=5).

**Figure 4**

**Supplemental Material****Supplementary Figure S1. Piwi does not directly interact with many nuage localized factors involved in piRNA biogenesis**

LacI-mKate2-Piwi foci did not recruit CFP-tagged Hen1 (p-value=0.096; std=0.124; n=5), Gasz (p-value=0.154; std=0.344; n=5), Nibbler (Nbr) (p-value=0.00; std=0.00; n=5), Minotaur (Mino) (p-value=0.018; std=0.040; n=5), Vreteno (Vret) (p-value=0.212; std=0.172; n=5), BoYb (p-value=0.062; std=0.139; n=5), Zucchini (Zuc) (p-value=0.196; std=0.210; n=5), Belle (p-value=0.028; std=0.063; n=5), Argonaute3 (Ago3) (p-value=0.010; std=0.017; n=5), Krimper (p-value=0.00; std=0.00; n=5), or Piwi (p-value=0.162; std=0.362; n=5). Representative images are shown. Scale bar represents 10 $\mu$ m.

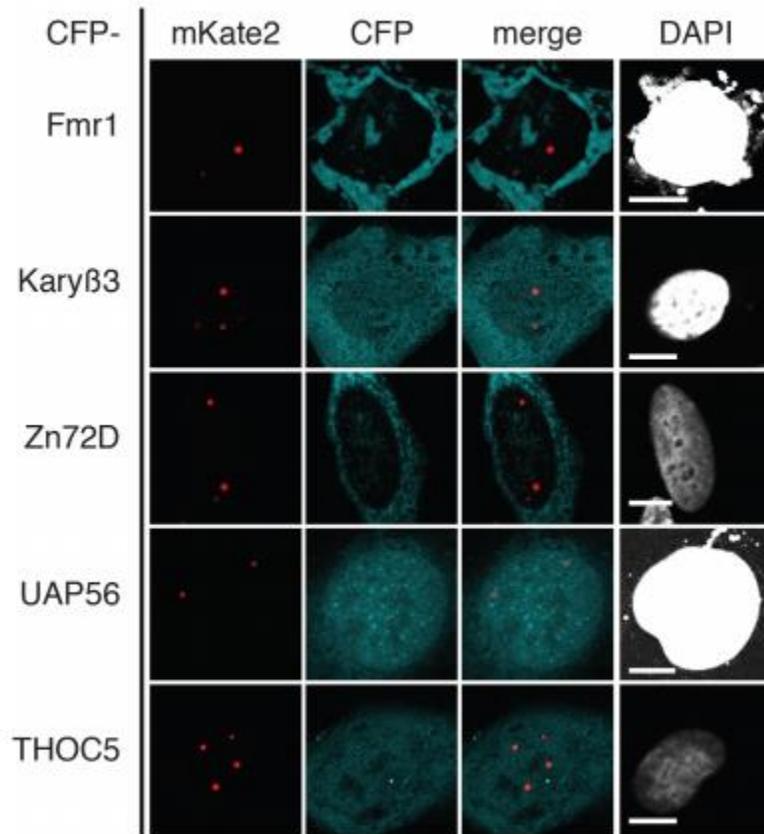
Supplemental Figure S1



**Supplementary Figure S2. Piwi does not directly interact with several factors involved in transcript export from the nucleus.**

LacI-mKate2-Piwi foci did not recruit CFP-tagged Fmr1 (p-value=0.00; std=0.00; n=5), Kary $\beta$ 3 (p-value=0.154; std=0.238; n=5), Zn72D (p-value=0.00; std=0.00; n=5), UAP56 (p-value=0.002; std=0.004; n=5), or THOC5 (p-value=0.030; std=0.051; n=5). Representative images are shown. Scale bar represents 10 $\mu$ m.

Supplemental Figure S2



*Chapter 5*

## CONCLUDING REMARKS AND FUTURE WORK

The PIWI-interacting RNA (piRNA) pathway is a conserved defense mechanism that protects the genetic information of animal germ cells from the deleterious effects of molecular parasites, such as transposons. This small RNA-based silencing system comprises PIWI-clade Argonaute proteins and their associated RNA-binding partners, the piRNAs. The projects present in the thesis were designed to advance our understand of piRNA biogenesis and decipher its regulation mechanisms.

In chapter 2, we focus on the piRNA biogenesis in *Drosophila* somatic and germline cells. We elaborate their processing phase in the nucleus and the cytoplasmic processing site ‘nuage’ separately. We focus on how piRNA precursor transcription is regulated and discuss current models for piRNA maturation. And highlight recent work that advanced our understanding of piRNA precursor processing to mature piRNAs. The major findings in the past years that significantly elucidate the piRNA biogenesis are: (1) The chromatin structure of genomic regions giving rise to piRNAs is crucial for piRNA biogenesis. (2) piRNA biogenesis depends on the presence of the allegedly repressive histone 3 lysine 9 trimethylation (H3K9me3) mark on dual-strand piRNA clusters. (3)The establishment of proper chromatin on dual-strand clusters requires trans-generational inheritance of homologous piRNAs. (4)The function of dual-strand clusters as memory banks of previous transposon invasions requires the presence of chromatin-bound proteins, which modulate transcription initiation and termination from these loci and splicing of, and TREX loading on, nascent piRNA precursor transcripts (Chen et al., 2016; Mohn, Sienski, Handler, &

Brennecke, 2014). (5) piRNAs can be processed from precursors by the endonuclease Zucchini (Zuc) or by Piwi proteins guided by complementary piRNAs (ping-pong). Zuc-dependent processing, which was considered the primary mechanism, can be triggered by ping-pong (Han, Wang, Li, Weng, & Zamore, 2015; Mohn, Handler, & Brennecke, 2015a, 2015b).

Two PIWI-clade proteins, Aub and Ago3, are essential for the piRNA post-transcriptional silencing. The silencing is accomplished via the processing so-called ping-pong cycle, which occurs in the nuage (Brennecke et al., 2007; Gunawardane et al., 2007; Vagin et al., 2006). Chapter 3 focused on one of the important post-translational modifications in the piRNA pathway, Aub symmetric dimethylarginine modification within its N terminal region. We are trying to understand the molecular function of sDMA for TE suppression and the regulation mechanism behind it. We reported that the loading of a piRNA guide triggers methylation of Aub, which exposes its unstructured N-terminal region to the PRMT5 methylosome complex. Thus, sDMA modification is a signal that Aub is loaded with a piRNA guide. Amplification of piRNA in the ping-pong cycle requires the assembly of a tertiary complex scaffolded by Krimper, which simultaneously binds the N-terminal regions of Aub and Ago3. Krimp uses its two Tudor domains to bind Aub and Ago3 in opposite modification and piRNA-loading states to promote the generation of new piRNA. Our results reveal that post-translational modifications in unstructured regions of Piwi proteins and their binding by Tudor domains that are capable of discriminating between modification states are essential for piRNA biogenesis and silencing.

Numerous piRNA pathway factors are identified via genetic screens (Czech, Preall, McGinn, & Hannon, 2013; Muerdter et al., 2013). To discern the direct and indirect

interactors of PIWI-clade proteins help us better understand the biological events that occur in order during piRNA biogenesis. The traditional yeast two-hybrid screening and affinity purification of complexes in vitro followed by mass spectrometry have their limitations. The results of which can be plagued with unspecific hits and miss detections of transient or weak interactions. In the chapter 4, we designed and utilized a heterologous cell culture assay to distinguish between direct and indirect protein-protein interactions with *Drosophila melanogaster* Piwi. We also identified novel Piwi interactor Arp6, which may play a role as a ‘bridge’ that physically interacts with Piwi and the RDC complex in the piRNA cluster loci. Arp6 may be the key factor that guides Piwi-RDC complexes exclusively binding to the H3K9me3 marks to regulate the transcription of piRNA precursors.

Although recent studies and our findings help us uncover molecular mechanisms of several key biological steps of the piRNA pathway, many outstanding questions remained. By what mechanism are some coding mRNAs selected for processing into piRNA? Do genic piRNAs generated by this process have a biological function, or are they by-products of promiscuous processing by an apparatus with low specificity? How is the processing of piRNA precursors initiated in the absence of complementary piRNAs? Particularly, we are extremely eager to know what drives the formation of ‘nuage’ – the essential processing site for piRISC assemblies, piRNAs maturation, TEs suppression. Cells organize many of their biochemical reactions in non-membrane compartments. Lines of evidence imply that Krimp acts as an essential component of nuage that acts as a scaffold for its assembly and the recruitment of client components. First, unlike other nuage components, FRAP measurements show very little Krimp exchange between nuage and the dispersed cytoplasmic compartment. Second, wild-type, but not mutant Krimp lacking the self-

interaction domain forms cytoplasmic granules upon expression in heterologous cells that do not contain other nuage proteins. In contrast, other nuage components, including Aub and Ago3, are dispersed in the cytoplasm when expressed in a similar setting, suggesting that they do not form condensates on their own and rely on other components recruitment to nuage.

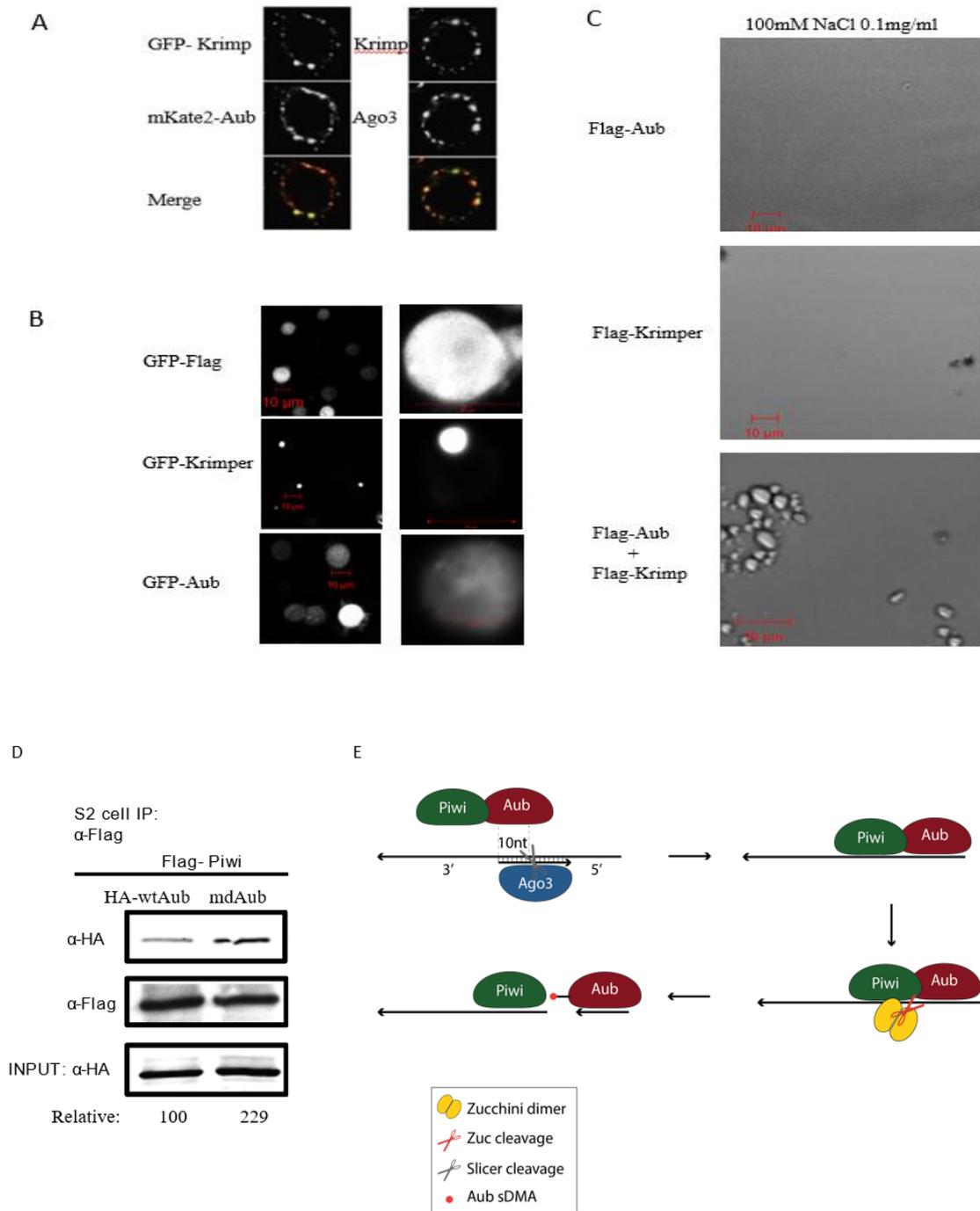
Recent evidence has shown that many of these compartments are liquids that form by phase separation from the cytoplasm. It's possible that scaffold Krimp, together with its interaction partner, drives the formation of nuage via LLPS. Our preliminary data support such hypothesis: (A) Krimp co-localizes with Aub and Ago3 in nuage, exhibiting droplet morphology. (B) Krimp but not Aub forms condensed granule in the heterologous culture cells. (C) Krimp, together with Aub, undergoes LLPS in vitro while Krimp or Aub alone will not triggers this process. These in vivo and vitro observations strongly imply the likelihood of the nature of nuage assembly- A Krimp mediated multivalent interaction-based liquid-liquid phase separations. It's reported that the N terminal region of Krimp is required for its self-interaction. Identify if the self-interaction ability is essential for the Krimp LLPS, and the role of Krimp self-interaction in the piRNA biogenesis should be elucidated in the future.

Our heterologous cell culture assay identified Aub but not Ago3 directly interacts with Piwi. What's the biological function of such biochemical interaction? Does it play a role in the piRNA loading? Based on the current piRNA phasing model. The first piRNA precursors derived from the 3' end product of Ago3 mediated cleavage should be loaded onto Aub. Then Aub bound piRNA precursors undergo the next several rounds of Zuc-mediate cleavage, and the 3'end cleaved products will be loaded onto Piwi. These consecutive processes generate phased Aub/Piwi bound piRNAs. Our preliminary results showed that

(D) Piwi predominantly interacts with unmethylated Aub, sDMA of Aub will dissociate the Piwi from the Aub. These observations suggest another role of Aub sDMA in the piRNA phasing, also explain why piRNA phasing can only be detected between Aub/Piwi bound piRNA but not Ago3/Piwi. We propose our sDMA-dependent piRNA phasing model. (E) Unmethylated piRNA-free Aub directly interacts with Piwi. Upon Ago3-piRISCs target and cleave piRNA precursors, the 3' end cleaved products are loaded onto Aub to form Piwi-Aub-RISCs. Zuc is then recruited to the Piwi-Aub-RISCs to define the 3' end of Aub bound piRNAs. Mature piRNA loading of Aub triggers its N terminal conformational change, dissociating piRNA-free Piwi from the complex. The released Piwi then receive newly generated piRNA precursors from Zuc mediated cleavage to accomplish the 'phasing'. In the future, it's important to dissect the domains required for the Piwi-Aub interaction. If the elimination of Aub-Piwi interaction will cause a defect in the phasing pathway will be an important question waiting to be solved.

.

## Preliminary Data



**(A)(B)(C) Krimper undergoes LLPS in vivo and in vitro.**

(A) Krimper co-localizes with Aub and Ago3 in the nuage. (B) GFP-Krimper forms a certain granule in the S2 cell cytoplasm while GFP-Aub and control GFP-Flag are dispersed evenly in the cytoplasm. (C) Binding to Aub triggers the Krimper undergoing phase separation in vitro.

**(D)(E) Aub sDMA guides phased piRNA biogenesis.**

(D) Flag-Piwi Co-IP with HA-wtAub and mdAub. Piwi prefers to interact with sDMA-free Aub. (E) Model for the sDMA mediated phased piRNA biogenesis.

**Methods and Materials***Cell Culture*

Schneider S2 cells were cultured in complete Schneider medium (10% heat inactivated FBS; 100U penicillin [Life technologies]; 100µg streptomycin [Life technologies]). Plasmids were generated using Gateway cloning (Life technologies) using the *Drosophila* Gateway Vector Collection (DGVC) destination vectors pAFW for 3xFLAG tag, driven by the Actin5C promoter. All constructs were N terminal tagged. Cells were transfected using TransIT-LT1 transfection reagent (Mirus biosciences) according to the manufacturer's recommendation using 3µg for cells in 10mm dish. S2 cells were lysed in S2 lysis buffer (20mM Tris at pH7.4, 200mM KCl, 0.1% Tween-20, 0.1% Igepal, EDTA-free Complete Protease Inhibitor Cocktail [Roche], 100µg/mL RNase A). Supernatant was cleared by centrifugation at 4,000 x g for 20 minutes at 4°C. Anti-FLAG M2 beads (Sigma Aldrich) were blocked in 5mg/ml BSA for 10 minutes at 4°C, followed by washing in S2 lysis buffer. Beads were added to the

supernatant and rotated at 4°C for 4h, washed three times in lysis buffer and eluted by 3X FLAG peptide (Sigma- Aldrich) in the elution buffer 25 mM HEPES, pH 7.6, 100 mM NaCl, 0.1 mM EDTA, 10% glycerol, 0.01% NP-40, 1 mM DT, 1X protease inhibitor. 3X Flag peptides were removed using MWCO 50 protein concentrator (ThermoFisher)

*In vitro liquid-liquid phase separation assay*

Flag tagged protein was stored in elution buffer. 0.1mg/ml – 0.5mg/ml Flag-Krimper was add to same concentration Flag-Aub. Protein mixture was incubated at 37 °C for 5 min then returned to RT. Place protein on the glass bottom dish for imaging with differential interference contrast microscopy at 63x on a Zeiss LSM710 confocal microscope.

## References

- Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., & Hannon, G. J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell*, *128*(6), 1089-1103. doi:10.1016/j.cell.2007.01.043
- Chen, Y. A., Stuwe, E., Luo, Y., Ninova, M., Le Thomas, A., Rozhavskaya, E., . . . Aravin, A. A. (2016). Cutoff Suppresses RNA Polymerase II Termination to Ensure Expression of piRNA Precursors. *Mol Cell*, *63*(1), 97-109. doi:10.1016/j.molcel.2016.05.010
- Czech, B., Preall, J. B., McGinn, J., & Hannon, G. J. (2013). A transcriptome-wide RNAi screen in the *Drosophila* ovary reveals factors of the germline piRNA pathway. *Mol Cell*, *50*(5), 749-761. doi:10.1016/j.molcel.2013.04.007
- Gunawardane, L. S., Saito, K., Nishida, K. M., Miyoshi, K., Kawamura, Y., Nagami, T., . . . Siomi, M. C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science*, *315*(5818), 1587-1590. doi:10.1126/science.1140494
- Han, B. W., Wang, W., Li, C. J., Weng, Z. P., & Zamore, P. D. (2015). piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. *Science*, *348*(6236), 817-821. doi:10.1126/science.aaa1264
- Mohn, F., Handler, D., & Brennecke, J. (2015a). Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. *Science*, *348*(6236), 812-817. doi:10.1126/science.aaa1039
- Mohn, F., Handler, D., & Brennecke, J. (2015b). piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. *Science*, *348*(6236), 812-817. doi:10.1126/science.aaa1039
- Mohn, F., Sienski, G., Handler, D., & Brennecke, J. (2014). The Rhino-Deadlock-Cutoff Complex Licenses Noncanonical Transcription of Dual-Strand piRNA Clusters in *Drosophila*. *Cell*, *157*(6), 1364-1379. doi:10.1016/j.cell.2014.04.031
- Muerdter, F., Guzzardo, P. M., Gillis, J., Luo, Y., Yu, Y., Chen, C., . . . 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*(5), 736-748. doi:10.1016/j.molcel.2013.04.006
- Vagin, V. V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., & Zamore, P. D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. *Science*, *313*(5785), 320-324. doi:10.1126/science.1129333