A Novel Role of UAP56 in piRNA Mediated Transposon Silencing: A Dissertation

Fan Zhang

University of Massachusetts Medical School

Let us know how access to this document benefits you.
Follow this and additional works at: https://escholarship.umassmed.edu/gsbs_diss

Part of the Biochemistry Commons, Genetics Commons, Genomics Commons, Molecular Biology Commons, and the Molecular Genetics Commons

Repository Citation

This material is brought to you by eScholarship@UMassChan. It has been accepted for inclusion in Morningside Graduate School of Biomedical Sciences Dissertations and Theses by an authorized administrator of eScholarship@UMassChan. For more information, please contact Lisa.Palmer@umassmed.edu.
A Novel Role of UAP56 in piRNA Mediated Transposon Silencing

A Dissertation Presented

By

Fan Zhang

Submitted to the Faculty of the

University of Massachusetts Graduate School of Biomedical Sciences, Worcester

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2, 2013

MOLECULAR MEDICINE
A NOVEL ROLE OF UAP56 IN piRNA MEDIATED TRANSPSON SILENCING

A Dissertation Presented

By

Fan Zhang

The signatures of the Dissertation Defense Committee signify completion and approval as to style and content of the Dissertation

______________________________________
William Theurkauf, Ph.D., Thesis Advisor

______________________________________
Nelson Lau, Ph.D., Member of Committee

______________________________________
Melissa Moore, Ph.D., Member of Committee

______________________________________
Craig Peterson, Ph.D., Member of Committee

______________________________________
Oliver Rando, Ph.D., Member of Committee

The signature of the Chair of the Committee signifies that the written dissertation meets the requirements of the Dissertation Committee

______________________________________
Victor Ambros, Ph.D., Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies that the student has met all graduation requirements of the school

______________________________________
Anthony Carruthers, Ph.D.,
Dean of the Graduate School of Biomedical Sciences
Interdisciplinary Graduate Program
August 2, 2013
Dedications

To my dear parents, who always encourage me to reach high;

To my wonderful husband, Hanhui, who always backs me up without a word;

To my sweet boys, Tian and Hai, who are my real motivations to move forward.
Acknowledgements

Seven years ago, when I first knocked on his office door, my thesis advisor, Dr. William Theurkauf, welcomed me with his profound scientific insight and bright personality. I was fortunate to join his lab and involved in this exciting project. I truly enjoyed my graduate study. With all sorts of great ideas and discussions, Bill kept on inspiring me and encouraging me to pursue something new. He taught me not to afraid of challenges and conquers myself. Sometimes when I felt stuck, his optimism kept my spirit up. With all sorts of collaborations drawn in by his open mind and wide connections, we were able to moving beyond our abilities. Whenever there was a little progress, his great appreciation made me feel fulfilled. As a daughter, a wife and a mother, I sometimes struggled between the balance of my family life and my professional career. Bill always very understood and his support really empowered me to overcome all the difficulties and finally finished my Ph.D. I wish I could carry all the great things he has taught me through all these years and pass it on to my new career and new life ahead.

Meanwhile, I was really lucky to have four great professors, Dr. Victor Ambro, Dr. Kirsten Hagstrom, Dr. Malissa Moore and Dr. Oliver Rando, as my graduate research committee. Their continuous scientific input and encouragement guided me through each little step of my graduate study.

Second, I would like to thank all the great people I worked with in Theurkauf lab. Beatrice Benoit was the first postdoc I worked with. She taught me the fly genetics from the very beginning by holding my hand. And Nadine Schultz was like a live dictionary of fly genetics and she magically solved almost all my fly issues. She was also a faithful
listener of all my happiness and worries and never hesitated to offer help. Birgit Koppetsch, with her specialties of confocal microscopy and image analysis, her valuable contribution to this project is indispensable. I am thankful to my fellow graduate student, Zhao Zhang, for preparing small RNA library and all the technical support of RNA-Seq library preparation. And I deeply appreciate Travis Thomson for comments and helping revise my thesis.

It was a great pleasure to collaborate with Dr. Zhiping Weng and her group on this project, especially Jie Wang and Thom Vreven. With their specialties of bioinformatics data analysis and protein structure analysis, they pushed the study of my project out of limit. And all the great discussion at the joined group meeting was inspiring and fruitful.

Third, I would like to thank Dr. Thoru Pederson for directing me into the graduate program at UMass Medical School, closely following my career development and all the great advises along the way. Without him, I did not even know where to start.

Last but least, I would like to thank all my families. My parents made a big sacrifice to let me pursue my graduated degree abroad. Being a scientist too, my dear husband, Hanhui, shared all my tears and joy. He was the first person I ran my scientific idea by over the kitchen table, and was also the one looking after the kids while I was spending the long night in the lab. And my two lovely boys, Tian and Hai, gave me extra challenges for my Ph.D. study. Meanwhile, their smiles always reset my mind back to peace no matter how hard the situation is.
ABSTRACT

Transposon silencing is required to maintain genome stability. The non-coding piRNAs effectively suppress of transposon activity during germline development. In the \textit{Drosophila} female germline, long precursors of piRNAs are transcribed from discrete heterochromatic clusters and then processed into primary piRNAs in the perinuclear nuage. However, the detailed mechanism of piRNA biogenesis, specifically how the nuclear and cytoplasmic processes are connected, is not well understood. The nuclear DEAD box protein UAP56 has been previously implicated in protein-coding gene transcript splicing and export. I have identified a novel function of UAP56 in piRNA biogenesis. In \textit{Drosophila} egg chambers, UAP56 co-localizes with the cluster-associated HP1 variant Rhino. Nuage is a germline-specific perinuclear structure rich in piRNA biogenesis proteins, including Vasa, a DEAD box with an established role in piRNA production. Vasa-containing nuage granules localize directly across the nuclear envelope from cluster foci containing UAP56 and Rhino, and cluster transcripts immunoprecipitate with both Vasa and UAP56. Significantly, a charge-substitution mutation that alters a conserved surface residue in UAP56 disrupts co-localization with Rhino, germline piRNA production, transposon silencing, and perinuclear localization of Vasa. I therefore propose that UAP56 and Vasa function in a piRNA-processing compartment that spans the nuclear envelope.
## Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>i</td>
</tr>
<tr>
<td>Signature page</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>iv</td>
</tr>
<tr>
<td>Abstract</td>
<td>vi</td>
</tr>
<tr>
<td>List of tables</td>
<td>x</td>
</tr>
<tr>
<td>List of figures</td>
<td>xi</td>
</tr>
<tr>
<td>Copyright notice</td>
<td>xiii</td>
</tr>
<tr>
<td>Contributions</td>
<td>xiv</td>
</tr>
</tbody>
</table>

Chapter 1: General Introductions

1.1. Genome stability and piRNA mediated transposon silencing 1

1.2. piRNAs biogenesis 2

1.2.1. Genomic Origins of piRNAs 3

1.2.2. Transcription of piRNA precursors 5

1.2.3. Processing of piRNA precursors 6

1.2.3.1. Primary piRNA biogenesis 6

1.2.3.2. Ping-Pong amplification 8

1.3. Compartmentalization of piRNA pathway machinery 13

1.4. piRNA mediated TE silencing 15

1.4.1. Cytoplasmic TE silencing 16
Chapter 2: UAP56 Couples piRNA Clusters to the Perinuclear Transposon Silencing Machinery

2.1. Summary

2.2. Introduction

2.3. Results

• UAP56 co-localization with the piRNA biogenesis machinery
• UAP56 is required for nuage localization of piRNA pathway proteins
• Germline DNA damage
• Gene and transposon expression
• piRNA production
• piRNA precursor binding
• The E245K substitution disrupts a conserved surface residue

2.4. Discussion

2.5. Experimental Procedures
Chapter 3: UAP56 suppresses piRNA precursor splicing in the Drosophila germline

3.1. Introduction 106
3.2. Results 107
  - UAP56 suppress piRNA precursors splicing 107
  - The E245K substitution disrupts UAP56 binding to the spliced piRNA precursors 114
3.3. Discussion 120
3.4. Experimental Procedures 121

Chapter 4: General Discussion 123

4.1 Introduction 123
4.2 piRNA cluster transcription 124
4.3 Co-transcriptional regulation of piRNA production 127
4.4 piRNA transcription and nuclear export 128
4.5 piRNA production and protein coding gene expression 129
4.6 Open Questions 133

References 135
List of Tables

Table 2.1. Egg production, D-V patterning, and hatch rates in \textit{uap56} mutants.

Table 2.2. Primer sequences.

Table 3.1. Primer sequences.
List of Figures

Figure 1.1. piRNA biogenesis in *Drosophila* female ovaries

Figure 2.1. Organization of the piRNA biogenesis machinery

Figure 2.2. UAP56 does not co-localize with Rhi in oocyte.

Figure 2.3. UAP56 expression in mutants and transgenic lines.

Figure 2.4. The E245K substitution and *rhi* mutations disrupt UAP56 localization to nuclear foci.

Figure 2.5. Rhi and UAP56 localization in later stage *uap56* mutant egg chambers.

Figure 2.6. Localization of transgenic wild type UAP56-Venus and sz-Venus fusion proteins.

Figure 2.7. Rhi and Vasa associate with the nuclear envelope.

Figure 2.8. *uap56* mutants disrupt nuage accumulation of Vasa.

Figure 2.9. Rhi and UAP56 localization in *vas* mutant.

Figure 2.10. UAP56 is required for PIWI protein localization to nuage.

Figure 2.11. Posterior localization of Aub and Vasa.

Figure 2.12. DNA damage in *uap56* mutant.

Figure 2.13. *uap56*, *rhi* and *vasa* mutations disrupt transposon Blood silencing but do not alter Ent1 gene expression.

Figure 2.14. *uap56*, *rhi* and *vasa* mutations disrupt transposon silencing but do not alter gene expression.

Figure 2.15. Transposon over-expression in *uap56* mutant.

Figure 2.16. piRNA expression in wt and *uap56*sz15'/*uap56*28 mutant ovaries.
Figure 2.17. Transposon mapping piRNA expression in \textit{uap56}, \textit{rhi} and \textit{vasa} mutant ovaries.

Figure 2.18. Cluster mapping piRNA expression in \textit{uap56}, \textit{rhi} and \textit{vasa} mutant ovaries.

Figure 2.19. Genic piRNA expression in \textit{uap56}, \textit{rhi} and \textit{vasa} mutant ovaries.

Figure 2.20. piRNA precursor expression in \textit{uap56} and \textit{vasa} mutant ovaries.

Figure 2.21. piRNA expression and precursor binding to UAP56.

Figure 2.22. piRNA precursors bind to UAP56 and Vasa

Figure 2.23. UAP56 structure.

Figure 2.24. UAP56 sequence conservation.

Figure 2.25. UAP56 function in a piRNA processing compartment.

Figure 3.1. \textit{uap56} mutations lead to piRNA precursors accumulation at 42AB cluster.

Figure 3.2. \textit{uap56} mutations disrupt piRNA production and cause precursor accumulation at dual strand cluster 3.

Figure 3.3. Cluster 3 piRNA precursors bind to UAP56.

Figure 3.4. Spliced piRNA precursor transcripts from cluster 3 bind to UAP56.

Figure 4.1. Nuclear piRNA pathway
Copyright Notice

Chapter 2 is published in Cell.

List of Contributions

Figure 2.7 was analyzed by Birgit Koppetsch.

Figure 2.13 and 2.14 was prepared by Jie Wang and Jia Xu.

Figure 2.15 A was prepared by Birgit Koppetsch.

Figure 2.16-2.19 was prepared by Jie Wang. Small RNA library was prepared by Zhao Zhang.

Figure 2.21 and 2.22 A and B was prepared by Jie Wang.

Figure 2.23 was prepared by Thom Vreven.
1.1. Genome stability and piRNA mediated transposon silencing

Higher eukaryotes have large genomes dominated by non-coding sequences, including transposable elements (TEs) and their remnants. These mobile elements occupy approximately 45% of human genome and 15-20% of Drosophila genome (Biemont and Vieira, 2006; Kapitonov and Jurka, 2003; Lander et al., 2001). Transposons can move from one location to another in the genome by a cut-paste mechanism (DNA elements) or by inserting a copy into a new site through an RNA intermediate (retroelements) (Gogvadze and Buzdin, 2009). Mobilization of TEs can rearrange the genome, altering gene sequences and their expression patterns. This genomic dynamism could provide organisms with the ability to adapt to evolutionary challenges and give rise to biological diversity, but most transposition events are considered deleterious as they can cause chromosomal breaks, insertional mutations (of protein-coding genes) and illegitimate recombination (Levin and Moran, 2011). To ensure that genetic information passes faithfully from generation to generation, strong selective pressure limits TE activity in the germline. Repressive protein complexes, DNA methylation and small inhibiting non-coding RNAs, are important epigenetic mechanisms to silence TEs in mammals, worms, plants and fungus (Levin and Moran, 2011; Slotkin and Martienssen, 2007).

The Piwi-interacting RNAs (piRNAs) are a distinct group of small RNAs associated with the Piwi-clade of Argonaute proteins (Malone and Hannon, 2009; Siomi
et al., 2011). piRNAs share sequence homology with TEs and have been implicated in silencing TEs and other repetitive sequences by directing heterochromatin formation and post-transcriptional silencing (Brennecke et al., 2007; Rouget et al., 2010; Saito et al., 2006). The mechanisms driving piRNA biogenesis and TE silencing are poorly understood, but Next-Generation Sequencing technology and genetic approaches have provided insight into the molecular machinery of the piRNA pathway.

1.2. piRNAs biogenesis

piRNAs are endogenous 23-29nt RNAs that interact with Piwi clade Argonaute proteins. As the predominant expression of PIWI proteins is in the animal gonad and mutations in pathway components often disrupt germline maintenance and cause sterility, piRNAs appear to function primarily in the germline (Klattenhoff and Theurkauf, 2007). The first piRNAs was discovered in Drosophila where small RNAs encoded by the Suppressor of Stellate (Su(Ste)) locus suppress testis-expressed Stellate gene repeats (Aravin et al., 2001). Subsequent studies in fly testes and early embryos revealed a set of related small RNAs derived from repetitive genomic elements and these RNAs were initially named repeat-associated small interfering RNAs (rasiRNAs) (Aravin et al., 2003) . Later, similar kinds of RNAs were identified in zebrafish, mouse and rat and were shown to associate with PIWI clade Argonaute proteins. The rasiRNAs were then recognized as a subgroup of PIWI-interacting RNAs (piRNAs) (Aravin et al., 2006; Chen et al., 2005; Girard et al., 2006; Grivna et al., 2006; Lau et al., 2006; Watanabe et al., 2006). In C. elegans, Piwi interacting RNAs are shorter, and are known as 21U RNAs.
However, due to their conserved role in transposon silencing, 21U RNAs are still considered bona fide piRNAs (Batista et al., 2008; Ruby et al., 2006).

Compared to siRNAs and miRNAs, piRNAs are a relatively new member of small silencing RNA family and their biogenesis is not well understood. Here I will summarize the current understanding of piRNA production, primarily using findings from the *Drosophila* system, with highlights from other organisms.

1.2.1. Genomic Origins of piRNAs

PiRNAs distribute unevenly in the genome, and are concentrated at discrete loci called clusters (Brennecke et al., 2007). Most of the clusters are located in peri-centromeric and sub-telomeric regions with a high density of repetitive sequences, although some of piRNAs map to untranslated regions of protein coding genes (Brennecke et al., 2007; Ro et al., 2007; Robine et al., 2009; Saito et al., 2009). In zebrafish and fruit flies, most piRNAs are derived from TEs and TE remnants, providing sequence homology for active transposon recognition (Brennecke et al., 2007; Houwing et al., 2007). However, this is not always true. For example, the majority of piRNAs in mammalian testis are derived from unique intergenic regions devoid of transposons and repetitive elements (Aravin and A, 2006; Girard et al., 2006; Lau et al., 2006). And in *C. elegans*, piRNAs target only one transposon family (Batista et al., 2008; Das et al., 2008). Broad piRNA functions in addition to transposon targeting have been reported in multiple systems, but the biological significance of these interactions have not been firmly established (Ashe et al., 2012; Lee et al., 2012; Rajasethupathy et al., 2012; Rouget et al., 2010; Shirayama et al., 2012; Vourekas et al., 2012).
In *Drosophila*, there are over 100 piRNA clusters localized in pericentromeric and subtelomeric heterochromatin regions. The distribution of piRNAs to either both or only one of the genomic strands define two classes of clusters, designated dual-strand or uni-strand (Brennecke et al., 2007). The 42AB Cluster is the largest in the fly genome, and is estimated to produce as much of 30% of germline piRNAs (Brennecke et al., 2007; Li et al., 2009). The *flamenco* locus is a uni-strand cluster with unique piRNAs mapping only to the plus genomic strand with TE sequences that comprise the *flamenco* transcript on the minus strand. piRNAs from this locus are therefore anti-sense to transposons. This locus appears to function primarily in ovarian somatic follicle cells (Brennecke et al., 2007; Malone et al., 2009).

In *Drosophila*, heterochromatic regions contain a high density of TE and TE remnants, but the piRNA clusters only occupy a subset of these transposon-rich domains. It is still unclear how the clusters became specific “hot spots” for piRNA production, but Rhino (Rhi), a germline-specific Heterochromatin Protein 1 (HP1) homolog, binds to dual-strand clusters and is required for piRNA production (Klattenhoff et al., 2009). HP1a, the founding member of the HP1 family, binds to H3K9m3 and recruits methyltransferase and that generates this mark, and thus triggers the spread of this heterochromatin modification to the neighboring Histone (Bannister et al., 2001; Lachner et al., 2001; Nakayama et al., 2001). This suggests piRNA clusters are epigenetically defined. This idea is supported by a recent study in mouse which shows that piRNAs can be made from a foreign DNA sequence genetically introduced into a piRNA cluster (Yamamoto et al., 2013).
1.2.2. Transcription of piRNA precursors

The nature of cluster transcripts is the most mysterious aspects of the piRNA pathway. Unlike siRNAs and miRNAs, which are generated from double stranded precursors by Dicer, piRNAs appear to be derived from single stranded RNAs. In Drosophila, a single P element insertion in the 5’ end (presumably the promoter) of the \textit{flamenco} locus disrupts piRNA production from the downstream 160kb genomic region providing genetic evidence for a single long transcription unit. Little is known about transcription of dual strand clusters, but the asymmetrical mapping of piRNAs to the two genetic strands of these clusters suggests that the transcripts are processed separately (Brennecke et al., 2007). By high-throughput RNA sequencing, the long single stranded piRNA precursors were identified in silkworm ovary-derived BmN4 cells and mouse testis (Kawaoka et al., 2013; Li et al., 2013). In \textit{C. elegans}, each of the short 21U piRNAs is produced from a \textasciitilde 26nt precursor (Gu et al., 2012).

The piRNA cluster transcription machinery is largely uncharacterized. Recently, CAGE-seq (cap analysis of gene expression followed by sequencing) and PAS-seq (polyadenylation site sequencing) uncovered 5’-cap and 3’-poly A-tails of piRNA precursors in BmN4 cells, mouse testis and worms (Gu et al., 2012; Kawaoka et al., 2013; Li et al., 2013). And by ChIP-seq (chromatin immunoprecipitation followed by sequencing) with antibodies against RNA polymerase II (Pol II), Pol II was identified across the clusters, and A-MYB was the first transcription factor identified that initiates the transcription of piRNAs clusters in mouse testis (Li et al., 2013). These observations indicate that Pol II transcribes clusters.
In *Drosophila*, Cutoff (cuff), a protein related to the de-capping factor Rai1, co-immunoprecipitates with Rhi and is required for localization of Rhi to nuclear foci, and germline piRNA production (Pane et al., 2011). The biochemical function for Cuff in piRNA biogenesis has not been defined, but it appears to function with Rhi to epigenetically define piRNA clusters.

### 1.2.3. Processing of piRNA precursors

piRNAs, unlike miRNAs and siRNAs, are generated from single stranded precursors by a Dicer-independent mechanism (Houwing et al., 2007; Vagin et al., 2006). Current models propose two distinct mechanisms for piRNA production: primary piRNA biogenesis and ping-pong amplification (Malone and Hannon, 2009; Siomi et al., 2011). The primary pathway initiates piRNA production and is the main mechanism used at the uni-strand clusters that produce somatic piRNAs (Ishizu et al., 2012; Zamore, 2010). The dual-strand clusters that dominate in the germline cells are also proposed to generate primary piRNAs, which then are amplified by an Aub/Ago3 dependent Ping-Pong cycle (Brennecke et al., 2007; Gunawardane et al., 2007).

### 1.2.3.1. Primary piRNA biogenesis

Initial piRNAs produced from both dual-strand or uni-strand clusters must be generated independent of pre-existing piRNAs. In *Drosophila* female ovaries, oocytes and nurse cells are the germline cells, and the surrounding follicle cells are somatic. Both tissues utilize the piRNA pathway to suppress TE activities, but the uni-strand clusters, e.g. *flamenco*, produce the vast majority of piRNAs in the somatic follicle cells. These
cells express only Piwi, the founding member of the PIWI clade Argonautes (Brennecke et al., 2007). In *Drosophila* ovaries, by contrast, three PIWI proteins, Piwi, Aubergine (Aub) and Ago3 play essential roles in piRNA processing (Cox et al., 2000; Li et al., 2009; Lin and Spradling, 1997; Schmidt et al., 1999). These three proteins localize to different subcellular compartments and show different tissue specificity (Brennecke et al., 2007). Piwi is predominantly nuclear and is expressed in both germline cells and the surrounding somatic follicle cells, while Aub and Ago3 are restricted to the cytoplasm of germline cells (Brennecke et al., 2007). This suggests two distinct pathways for piRNA processing and transposon silencing in the soma and germline, and that the somatic function of Piwi is independent of Aub and Ago3 (Li et al., 2009; Malone et al., 2009; Saito et al., 2009).

By utilizing the ovarian somatic sheet cell (OSCC) line and *in vivo* RNAi to specifically knockdown somatic pathway components, progress has been made in exploring the mechanism of primary piRNA production (Olivieri et al., 2010; Saito et al., 2009). The nuclease Zucchini (Zuc), the RNA helicases Armitage (Armi) and the Tud domain-containing (Tdrd) RNA helicase Yb were found to be the essential for primary piRNA production and TE silencing (Olivieri et al., 2010; Saito et al., 2010). In follicle cells, these proteins accumulate in cytoplasmic foci called Yb bodies (Szakmary et al., 2009). Recently, biochemical and structural analyses showed that Zuc can function as a single-strand-specific endonuclease, and may cleave nascent cluster transcripts and generate 5’ ends of piRNAs (Ipsaro et al., 2012; Nishimasu et al., 2012). The resulting piRNA intermediates are then appear to be loaded into Piwi and trimmed to form mature 3’ ends (Kawaoka et al., 2011) (Figure 1.1). Armi and Yb are in the same complex with
Piwi and required for Piwi loading. A recent study revealed a new piRNA pathway player, Shutdown (Shu), which interacts with Hsp83/HSP90 through its tetratrico-peptide repeat (TPR) domain and loads piRNA intermediates into PIWI proteins (Olivieri et al., 2012; Preall et al., 2012). Finally, Vreteno (Vret), a newly identified Tdrd protein, is also required for Piwi loading in the primary pathway (Zamparini et al., 2011).

The last step in piRNA biogenesis is 3’ end modification. In Drosophila, both endo-siRNAs and piRNAs have a 2’-O-methylation at the 3’ end, which is generated by the methyltransferase HEN1 (Horwich et al., 2007; Ohara et al., 2007; Vagin et al., 2006). This modification appears to take place after RNA binding to PIWI, and increases piRNA stability in vivo (Horwich et al., 2007; Kamminga et al., 2010; Kirino and Mourelatos, 2007; Kurth and Mochizuki, 2009; Saito et al., 2007). Piwi binds to the mature piRNAs then is transported into nucleus where it plays roles in transcriptional silencing (Ishizu et al., 2011). In germline cells, mature primary piRNAs are also loaded into Aub and Ago3 complexes, which feed into the Ping-Pong amplification cycle.

1.2.3.2. Ping-Pong amplification

PIWI proteins are the direct binding partners of piRNAs. They play essential roles in both piRNA processing and transposon silencing (Brennecke et al., 2007; Cox et al., 2000; Harris and Macdonald, 2001; Li et al., 2009; Vagin et al., 2006). Like other Argonaute proteins, they feature a PAZ domain with RNA binding property and a PIWI domain with endonuclease activity (Jinek and Doudna, 2009; Sashital and Doudna, 2010). piRNAs mediate target recognition, and the three PIWI proteins bind to piRNAs that differ in strand bias and therefore target specificity. Piwi and Aub bind mainly to
piRNAs antisense to TEs, while Ago3 interacts primarily with sense piRNAs. The antisense piRNAs bound to Aub and Piwi are proposed to target TEs for silencing and produce the precursors of sense piRNAs, which are bound by Ago3. Aub and Ago3 associating piRNAs show sequence complementary between their first 10 nucleotides (nt), and in vitro experiments confirmed the endonuclease activity of these PIWI proteins (Brennecke et al., 2007; Gunawardane et al., 2007). Based on these findings, a Ping-Pong amplification model for piRNA production in germline cells has been proposed. In this model, primary piRNAs antisense to TEs are loaded into Piwi or Aub and direct cleavage of TE mRNAs, generating the precursors of sense piRNAs bound to Ago3. Through the endoribonuclease activity of Ago3 and Aub, these sense strand piRNAs direct cleavage of anti-sense RNAs derived from piRNA clusters, producing the precursor of anti-sense piRNAs bound to Aub and Piwi. PIWI proteins cleave between 10th and 11th nucleotide of the “guide piRNA”, which have a strong bias for U at the first nucleotide. As a result, piRNAs from opposite strands that are produced by the Ping-Pong amplification show a bias toward both a 10nt overlap, and a U at position 1 and an A at position 10 (Brennecke et al., 2007) (Figure 1.1).

piRNA maturation appears to be completed after long precursors are bound to PIWI proteins, when a 3’-5’ exonuclease trims the bound RNA to final length. The three PIWI proteins appear to have different “footprints” on the precursors, such that the mature piRNAs bound by Piwi, Aub and Ago3 show length distribution peaks at 25nt, 24nt and 23nt respectively (Brennecke et al., 2007). A recent study in BmN4 cell extracts confirmed this 3’-5’ exonucleolytic activity in vitro (Kawaoka et al., 2011). After trimming, the 3’ ends of mature piRNAs are methylated by HEN1, which appears
to increase piRNA stability (Horwich et al., 2007; Kamminga et al., 2010; Kirino and Mourelatos, 2007; Kurth and Mochizuki, 2009; Saito et al., 2007).
Figure 1.1. piRNA biogenesis in *Drosophila* female ovaries

In *Drosophila* female ovaries, oocytes and nurse cells are the germline cells, and the surrounding follicle cells are somatic. Both tissues utilize the piRNA pathway to suppress TE activities, however the machinery and processing is different. In somatic follicle cells, the long single stranded piRNA precursors are transcribed from somatic uni-strand clusters and then exported to cytoplasmic perinuclear structure, Yb bodies. Yb bodies accommodate Zuc, Armi and other primary piRNA processing machineries. It has been proposed that Zuc cleaves the nascent cluster transcripts and generates 5’ ends of piRNAs with a U bias. Armi, Yb, Shu and Vret facilitate the loading of piRNA intermediates into Piwi. An unknown exonuclease then trims the 3’ ends of Piwi bound piRNA intermediates to form mature piRNAs, which will be transported into nucleus and play roles in transcriptional silencing TEs. In germline cells, piRNAs are produced from both dual-strand clusters and uni-strand clusters by both primary processing and secondary Ping-Pong amplification. Nuclear proteins Rhi and Cuff interact with dual-strand clusters and required for piRNA production. Similar to somatic piRNA biogenesis, the long piRNA precursors are transcribed and exported to germline specific perinuclear structure, nuage. The antisense piRNAs generated by primary processing bind to Aub and target TE mRNAs by sequence complementary. Aub cleaves TE mRNAs to generate the precursors of sense piRNAs bound to Ago3. Through the endoribonuclease activity of Ago3 and Aub, these sense strand piRNAs direct cleavage of anti-sense RNAs derived from piRNA clusters, producing the precursor of anti-sense piRNAs bound to Aub and Piwi and perpetuate the cycle. The Piwi bound antisense piRNAs are relocated into nucleus and suppress TE at transcriptional level.
Figure 1.1
1.3. Compartmentalization of piRNA pathway machinery

piRNA production starts with transcription of piRNA clusters. So far, Piwi, Rhi and Cuff are the only piRNA pathway components that localize in the nucleus, and both Rhi and Cuff associate with dual-strand piRNA clusters and epigenetically regulate piRNA production from these genomic loci. These two proteins co-localize to heterochromatic foci in centromeric or pericentromeric regions in nurse cell nuclei, and co-immunoprecipitate from ovary extracts (Klattenhoff et al., 2009; Pane et al., 2011). Piwi is the only PIWI protein in *Drosophila* that shows steady state localization in the nucleus. Proper nuclear localization of Piwi requires Ago3 and Aub, suggesting Piwi functions downstream of piRNA processing and that its import requires loading of piRNAs in cytoplasm (Li et al., 2009; Wang and Elgin, 2011). Piwi protein with a N-terminal deletion that removes the nuclear localization signal (NLS) abolishes its nuclear import and disrupts its transposon silencing function (Klenov et al., 2011). Since Piwi is expressed in both somatic and germline cells, nuclear localization appears to be essential to Piwi-mediated transcriptional TE silencing in both tissues.

The majority of piRNA pathway components are in the cytoplasm, which suggests a post-transcriptional TE silencing mechanism. Germline cells and somatic cells organize their piRNA machineries in two different kinds of cytoplasmic structures, nuage and Yb bodies. These two subcellular structures share some components, but are differentiated by a number of unique features.

Nuage (the French word for “cloud”) is a germline specific electron-dense perinuclear structure located at the cytoplasmic face of nurse cell nuclear envelope (Mahowald, 1972). Similar granules have been identified in *C. elegans, Xenopus,*
zebrafish and mouse (Clark and Eddy, 1975; Ikenishi and Tanaka, 1997; Knaut et al., 2000; Strome and Wood, 1983). In *Drosophila*, most of the well characterized piRNA proteins localized to nuage, including Aub and Ago3, which are core components of the Ping-Pong amplification machinery. In addition, the helicases Armi, Spindle E and Vasa, the nucleases Zucchini and Squash, and the Tudor-domain protein Krimper localize to nuage. Mutations in genes encoding any of these proteins disrupt piRNA production and elevate TE expression (Brennecke et al., 2007; Cook et al., 2004; Lim and Kai, 2007; Malone et al., 2009; Pane et al., 2007). Vasa, a germline specific DEAD-box RNA-dependent helicase, is one of the founding components of nuage (Hay et al., 1990; Lasko and Ashburner, 1990). Vasa is proposed to recruit Tudor and Tdrd, which in turn form a platform for assembly of Ping-Pong amplification machinery by recruiting Aub and Ago3. Aub and Ago3 interact with these Tudor domain proteins through symmetric dimethyl-arginines (sDMAs) that are modified by dPRMT5 (Nishida et al., 2009). Recently, Qin/Kumo, a newly characterized Tdrd protein, was also shown to be required for the heterotopic Ping-Pong between Aub and Ago3 (Zhang et al., 2011).

Ultrastructure studies have shown that nuage is closely associated with nuclear pores, which are the main path for nuclear export of mRNAs. This implicates nuage as a surveillance structure for post-transcriptional TE silencing (Chuma et al., 2009; Mahowald, 1972). Recently, a corresponding organelle, the piRNA nuage giant body (piNG-body), was found in the *Drosophila* male germ cells. This structure contains known nuage proteins and appears to mediate silencing of Stellate via the piRNA pathway (Kibanov et al., 2011).
The Yb body is an RNA-rich structure found in the somatic cells of *Drosophila* ovaries and testis, and is often associated with mitochondria. It is named after its marker component Yb protein, which is only expressed in somatic cells (Szakmary et al., 2009). Yb bodies accommodate Zuc, Armi and other piRNA precessing proteins (Olivieri et al., 2010; Saito et al., 2010). Yb body assembly appears to be important for primary piRNA maturation and Piwi loading. Most of the piRNA pathway proteins that localize to Yb bodies are also expressed in germline cells, suggesting that piRNA production in somatic follicle cells is the model for the primary piRNA processing in germline and soma.

### 1.4. piRNA mediated TE silencing

Since piRNA sequences are homologous to TEs and pathway mutations lead to highly elevated TE levels, piRNAs have been suggested to act as small silencing RNAs and their main targets are TEs. However, unlike siRNAs and miRNAs, piRNA sequences are highly heterogeneous with no “seed” sequences for target recognition. It is still an open question that how piRNAs recognize their targets. Most piRNA sequences perfectly match TE transcripts, thus targeting by perfect base pairing has been proposed. A recent study in mice shows piRNAs generated from a genetic knock-in sequence will suppress the complementary reporter gene (Yamamoto et al., 2013). In *Drosophila*, a similar study shows the targeting of a Piwi-binding piRNA is highly sequence specific, which further supports the complete base pairing theory (Huang et al., 2013). However, in other systems, for example *C. elegans*, mismatches are allowed for target recognition (Lee et al., 2012)
Although piRNAs originate from heterochromatic clusters, their putative target, TEs, have a much broader distribution throughout the genome. Disrupting piRNA production from the uni-strand *flamenco* clusters, for example, activates TEs elsewhere in the genome, suggesting that piRNAs can target TEs in *trans* (Brennecke et al., 2007).

As discussed below, different properties and functions of piRNA pathway proteins suggest that silencing involves post-transcriptional cleavage, heterochromatin formation, and transcriptional suppression.

### 1.4.1. Cytoplasmic TE silencing

Due to the endonuclease activity of nuage located PIWI proteins, Aub and Ago3, and their proposed role in the Ping-Pong cycle, the post-transcriptional destruction of mature TE transcripts has been proposed as a primary mechanism for piRNA-mediated silencing (Brennecke et al., 2007). The model suggests after TE transcripts are transcribed and exported from the nucleus, where they are recognized by antisense piRNAs associated with Aub in nuage. Aub cleaves TE transcripts and generates sense piRNA precursors, which will bind to Ago3 and perpetuate the cycle. The perinuclear localization of the machineries to nuage facilitates the scanning of transcripts as they come out of the nuclear pore. In the *Drosophila* genome, many germline expressed genes, including Ago3, have TE sequences embedded in their introns but are not silenced. Presumably, these genes bypass piRNA-mediated silencing because splicing removes the intronic piRNA matching sequences.

In mammalian systems, the association of PIWI proteins with the translational machinery suggests a translational suppression mechanism (Grivna et al., 2006).
Drosophila embryo, piRNAs have been proposed to specifically target nos 3’UTR and promote deadenylation and translational repression (Rouget et al., 2010). But these findings require relaxed stringency in target recognition by piRNAs, and remain to be confirmed.

1.4.2. Nuclear TE silencing

In both somatic and germline cells, Piwi shows a steady-state accumulation in the nucleus. Nuclear localization of Piwi appears depend on loading with piRNAs antisense to TEs, and transcriptional silencing of some TEs in insect somatic cells is linked to Piwi-dependent modification of histones (Cox et al., 2000; Huang et al., 2013; Le Thomas et al., 2013; Olivieri et al., 2010; Olivieri et al., 2012; Rozhkov et al., 2013; Saito et al., 2010; Shpiz et al., 2011; Wang and Elgin, 2011). Drosophila do not control transcription through DNA methylation, but mouse PIWI proteins appear to direct DNA methylation at piRNA target loci (Aravin et al., 2008; Kuramochi-Miyagawa et al., 2008). piRNAs thus appear to have a conserved role in TE silencing at transcription level, but the ultimate mechanism of silencing is not conserved.

One report suggests that Piwi associates with chromatin and may directly interact with heterochromatin protein 1a (HP1a), but these studies have not been confirmed (Brower-Toland et al., 2007). However, HP1a and H3K9me2 at transposons appear to be reduced in piwi knockdown (Wang and Elgin, 2011). And recent cell type specific knockdowns and genome-wide analysis of chromatin and transcripts by ChIP-Seq and RNA-Seq are consistent with Piwi functioning in heterochromatic silencing of TE transcription in somatic tissue. Two studies have shown that loss of Piwi reduces
H3K9me2/3 modifications and increases Pol II occupancy at target TEs, which is linked to overexpression of TE transcripts (Huang et al., 2013; Le Thomas et al., 2013; Rozhkov et al., 2013; Sienski et al., 2012). One study showed that Piwi and Pol II colocalize with each other on polytene chromosomes from nurse cells (Le Thomas et al., 2013). An interesting model was proposed based on data from Maelstrom (Mael) mutants (Sienski et al., 2012). Mael was first identified as a piRNA pathway component in mouse (Aravin et al., 2009). In mael mutants, although TE levels are elevated, piRNAs and Piwi loading are not affected and H3K9me3 did not increase. However, Pol II occupancy did increase, suggesting that Mael functions downstream of H3K9me3 in the silencing pathway (Sienski et al., 2012).

1.5. piRNA functions other than TE silencing

In Drosophila, most piRNAs map to TE and TE related sequences, but a subset of piRNAs map to the 3’UTRs of protein coding genes, which suggests that piRNAs could regulate gene expression (Aravin et al., 2006; Robine et al., 2009; Saito et al., 2009). For example, the traffic jam (tj) locus encodes a transcription factor, and piRNAs from the tj 3’UTR associate with Piwi and depend on Zuc for production. These sense strand tj piRNAs are proposed to target the FasIII gene, which controls follicle cell organization (Saito et al., 2009). In addition, the first piRNAs to be identified target the stellate (ste) gene, and prevent Stellate protein over-expression in Drosophila testis (Aravin et al., 2001). Finally, a population of piRNAs derived from the X chromosome TAS are antisense to the vasa gene, and aub and ago3 mutants that disrupt expression of these piRNAs lead to Vasa protein over-expression (Li et al., 2009). piRNAs targeting both ste
and *vasa* associate with Aub, suggesting a post-transcriptional silencing mechanism (Nishida et al., 2007).

In mouse and *C. elegans*, only a relatively small fraction of piRNAs are derived from TE and TE related sequences, and most map to unannotated intergenic regions (Aravin et al., 2006; Batista et al., 2008; Ruby et al., 2006). The function for these non-TE regulating piRNAs is yet to be determined. However, recent studies in *C. elegans* indicates that piRNAs serve a surveillance function in the genome, and may help identify protein coding genes as “self” and introduced transgenes as “non-self”, which generates an epigenetic memory that lasts for multiple generations (Ashe et al., 2012; Bagijn et al., 2012; Lee et al., 2012; Shirayama et al., 2012). In addition, recent studies suggest that Piwi/piRNA complexes may function in the soma to mediate methylation of the CREB2 promoter and epigenetically regulate synaptic plasticity in neurons (Rajaseethupathy et al., 2012).

1.6. piRNA pathway and *Drosophila* germline development

piRNAs and piRNA biogenesis proteins are predominantly expressed and function in the animal gonad. In *Drosophila* and other animals, disruption of the piRNA pathway perturbs germline development and leads to animal sterility. Better understanding the phenotypes caused by piRNA pathway mutations and the linkage between TE silencing and germline maintenance will help further explore the pathway and its function.

In *Drosophila melanogaster*, oogenesis starts with the division of germ line stem cells (GSCs) localized at the tip of the gerarium adjacent to somatic niche cells.
Asymmetric cleavage generates one stem cell and one cystoblast, the progenitor of the oocyte and nurse cells (Spradling, 1993). The initial identification of piwi was through a genetic screen for genes controlling asymmetric GSC division in flies (Lin and Spradling, 1997). Genetic mosaic and niche cell-specific expression studies indicate that Piwi functions in somatic niche cells to maintain GSCs and in GSCs to promote the division. piwi mutants show reduced GSC content and the ovaries are rudimentary (Cox et al., 1998; Cox et al., 2000). The mechanism of GSC maintenance by Piwi/piRNAs is largely unknown, but may be distinct from TE control, as recent studies show that a mutation that disrupts the Piwi nuclear localization signal blocks transposon silencing but does not prevent stem cell maintenance or division (Klenov et al., 2011).

After the asymmetric division of GSC, the cystoblasts undergo four rounds of incomplete division to produce a cyst of 16 interconnected cells. One of the 16 cells will migrate to the posterior end and develop into the oocyte, and the other 15 cells will take the fate of nurse cells. Later, the 16-cell cysts are surrounded by a layer of somatic follicle cells and together will bud off the germarium to form an egg chamber (Spradling, 1993). In wild type ovaries, meiotic DNA breaks are present only in the cysts at Region 2a and are repaired as the cysts progress to region 2b and 3. Region 2a cysts in piRNA pathway mutants accumulate massive foci of phosphorylated Histone H2av, which are marks of DNA double-strand breaks, and these foci persist until late in oogenesis. These foci are presumably linked to breaks caused by transposition of TEs (Klattenhoff et al., 2007).

During most of Drosophila oogenesis, the oocyte is transcriptionally silent and the nurse cells supply the RNAs, proteins and the other nutrients that drive growth. The
nurse cell deposited materials including morphogenetic RNAs and proteins that are localized to the pole of the oocyte through interactions with the microtubule cytoskeleton. These asymmetrically localized factors determine the body axis of the embryo. The *osk*, *bicoid* and *grk* gene encode key patterning factors, and their mRNAs/proteins localized to the posterior, anterior and anterior-dorsal cortex of oocyte (Grunert and St Johnston, 1996).

piRNA pathway mutations disrupt microtubule organization, leading to mis-localization of patterning mRNAs in oocyte (Chen et al., 2007; Cook et al., 2004; Klattenhoff et al., 2007; Pane et al., 2007). The axis-specific defects of animals carrying piRNA pathway mutations resemble mutants that fail to repair meiotic DNA breaks, and the patterning defects associated with these mutations are suppressed by mutations in *mei-41* and *mnk* (fly homologs of ATR and Chk2), which are required for DNA damage signaling (Abdu et al., 2002; Ghabrial and Schupbach, 1999). In many piRNA mutants, *mei-41* and *mnk* mutations also suppress ventralization of the egg (Klattenhoff et al., 2007). Since piRNA pathway mutants show a significant increase in the activity of transposable elements, increased transposition is likely to lead to DNA breaks, which activate the ATR/Chk2 pathway and thus disrupt axis specification.

1.7. UAP56 and its pre-identified functions

UAP56 is a member of ATP-dependent DEAD-box RNA-binding proteins. The first human UAP56 was identified as a nuclear spliceosome component, which is recruited by U2AF65 and stabilizes the interaction of U2 snRNA and pre-mRNA branch point, thus named after U2AF65 associated protein the size of 56kD (Fleckner et al.,
UAP56 protein is highly conserved through species with 8 motifs commonly found in DEAD box protein family. UAP56 has been implicated functions in many aspects of RNA metabolism including linking pre-mRNA splicing and mRNA export from the nucleus (Shen, 2009). In human cells, UAP56 is mainly associated with spliced mRNA and play an essential role in bulk mRNA export (Gatfield, 2001). *Xenopus* UAP56 and its yeast homologue Sub2p are required for co-transcriptionally recruiting the mRNA export factor Aly/REF to mRNA (Luo et al., 2001). Furthermore, in Drosophila cells, it has been shown that UAP56 is indispensable for the export of both spliced and un-spliced intronless mRNAs (Gatfield et al., 2001). It is such an essential gene that depletion of UAP56 protein leads to animal lethality (Eberl et al., 1997; Shen, 2009). At subcellular level, UAP56 accumulates at nuclear speckles involves in protein coding transcripts slicing and nuclear export. Decrease UAP56 level or disrupting UAP56 functions by blocking ATP binding leads to nuclear retain of mRNAs. The detailed protein structure analysis showed human UAP56 protein has two globule domains connected by a flexible linker, and ATP-binding pocket and RNA-binding motifs are located at the interface of the two globule domains. By hydrolyzing ATP, UAP56 change its confirmation clamping on an off the RNA transcripts. The *Drosophila* homolog Hel25E was identified almost at the same time the human UAP56 was identified (Eberl et al., 1997). Besides its conserve function in RNA processing, the *Drosophila* UAP56 is also an enhancer of position-effect variegation (PEV). PEV is the suppression of a euchromatin gene when it is relocated near heterochromatin. This observation suggests UAP56 may have a role in epigenetic regulation (Eberl et al., 1997). It was proposed that UAP56 could promote an open chromatin structure by
unwinding or releasing the mRNA from the site of transcription. However, the putative RNA helicase activity of UAP56 has only been shown by in vitro experiments. The details need to be examined by an in vivo system. A recent study of UAP56 function in Drosophila germline revealed a uap56 point mutation causing RNA localization defect and body-axis defect resembles piRNA pathway mutants. This suggests the potential function of UAP56 in piRNA pathway.

1.8. Conclusions

The gonad specific non-coding small RNAs, piRNAs are essential for germline function, and piRNA mutant animals show dramatically increased TE activity and are sterile. Tiling array and RNA sequencing experiments indicate that piRNA mutants do not disrupt gene expression in the ovary. It is therefore likely that sterility is the result of double strand breaks caused by transpositions, which overwhelm the DNA repair machinery. In every system studied to date, at least a subset of piRNAs share high sequence homology with TEs, and piRNA mutants lead to transposon mRNA over-expression (Aravin et al., 2006; Batista et al., 2008; Brennecke et al., 2007; Girard et al., 2006; Houwing et al., 2007; Ruby et al., 2006; Saito et al., 2006). piRNAs thus appear to have a conserved role in transposon silencing, but appear to have additional targets in other systems and perhaps somatic cells. A number of mechanisms have been proposed for piRNA mediated TE silencing. The Ping-Pong cycle in the perinuclear nuage leads to piRNA amplification and post-transcriptional silencing. Recent studies on the primary piRNA pathway in gonad somatic tissue and cell lines derived from them indicate that also mediate transcriptional silencing through a mechanism related to heterochromatin
formation (Huang et al., 2013; Le Thomas et al., 2013; Olivieri et al., 2010; Olivieri et al., 2012; Rozhkov et al., 2013; Saito et al., 2010; Shpiz et al., 2011; Sienski et al., 2012; Wang and Elgin, 2011). The spatial organization of the piRNA machinery, which is partitioned into different cellular and subcellular compartments, may add dynamism to the pathway and enhance efficient silencing in specific cellular contexts. In my studies, I have investigated each compartment, and the linkage between them, with the aim of building a more comprehensive understanding of piRNA biogenesis and transposon silencing.
Chapter 2

UAP56 Couples piRNA Clusters to the Perinuclear Transposon Silencing Machinery

This chapter is published in Cell.


2.1. Summary

piRNAs silence transposons during germline development. In Drosophila, transcripts from heterochromatic clusters are processed into primary piRNAs in the perinuclear nuage. The nuclear DEAD box protein UAP56 has been previously implicated in mRNA splicing and export, while the DEAD box protein Vasa has an established role in piRNA production and localizes to nuage with the piRNA binding PIWI proteins Ago3 and Aub. We show that UAP56 co-localizes with the cluster-associated HP1 variant Rhino, that nuage granules containing Vasa localize directly across the nuclear envelope from cluster foci containing UAP56 and Rhino, and that cluster transcripts immunoprecipitate with both Vasa and UAP56. Significantly, a charge-substitution mutation that alters a conserved surface residue in UAP56 disrupts co-localization with Rhino, germline piRNA production, transposon silencing, and
perinuclear localization of Vasa. We therefore propose that UAP56 and Vasa function in a piRNA-processing compartment that spans the nuclear envelope.

2.2. Introduction

Transposons are ubiquitous genome pathogens that can mobilize and induce mutations that alter gene expression, cause disease and drive evolution (Bennetzen, 2000; Britten, 2010; Hedges and Belancio, 2011). The 23-30 nucleotide (nt) long piRNAs, which guide sequence specific RNA cleavage by PIWI clade Argonaute proteins in vitro, silence transposons during germline development (Aravin et al., 2007; Ghildiyal and Zamore, 2009; Gunawardane et al., 2007; Khurana and Theurkauf, 2010; Malone and Hannon, 2009; Siomi et al., 2010). In *Drosophila*, the primary piRNAs that initiate transposon silencing are derived from peri-centromeric and sub-telomeric chromatin domains composed of complex arrays of nested transposon fragments, termed piRNA clusters (Bergman et al., 2006; Brennecke et al., 2007). The *Drosophila flamenco (flam)* cluster, for example, spans approximately 180kb of heterochromatin on the X chromosome (Brennecke et al., 2007; Mével-Ninio et al., 2007; Sarot et al., 2004). *P*-element insertion mutations at the 5' end of this locus reduce expression of mature piRNAs and putative precursor transcripts from across the cluster, and increase expression of the transposons represented by fragments in *flam* (Brennecke et al., 2007). piRNAs from *flam* are therefore proposed to trans-silence homologous elements that are scattered throughout the genome.

The piRNA pool appears to be amplified by a ping-pong cleavage cycle in which anti-sense primary piRNAs from heterochromatic clusters, bound by the PIWI protein
Aubergine (Aub), direct cleavage of transposon transcripts. This silences transposon expression and generates precursors for sense strand piRNAs that associate with Ago3. Ago3-piRNA complexes then cleave cluster transcripts to produce anti-sense piRNA precursors, completing the cycle (Brennecke et al., 2007; Gunawardane et al., 2007). Most of the proteins required for ping-pong amplification, including Ago3 and Aub, localize to the perinuclear nuage (Klattenhoff and Theurkauf, 2007; Malone et al., 2009). This germline-specific electron dense material is closely associated with the cytoplasmic face of nuclear pores (Eddy, 1974), but it is unclear how piRNA precursors are directed from clusters to the perinuclear processing machinery. The DEAD box is a conserved ATP-dependent RNA binding motif (Linder and Jankowsky, 2011). The DEAD box protein Vasa (Vas) is required for germline development and piRNA production, and was the first molecularly defined nuage component (Hay et al., 1990; Lasko and Ashburner, 1990; Liang et al., 1994; Malone et al., 2009). UAP56, by contrast, is a ubiquitously expressed DEAD box protein previously implicated in splicing and RNA export (Shen, 2009). Here we present evidence that UAP56 and Vas are cluster transcript binding proteins that organize a piRNA-processing compartment that spans the nuclear envelope.

2.3. Results

UAP56 co-localization with the piRNA biogenesis machinery

Drosophila piRNA pathway mutations lead to germline DNA damage and activation of a signaling cascade that blocks asymmetric RNA localization in the developing oocyte (reviewed by (Khurana and Theurkauf, 2010). UAP56 has a conserved function in splicing and mRNA export, and strong hypomorphic mutations are
lethal (Gatfield et al., 2001; MacMorris et al., 2003; Shen, 2009). However, Drosophila females carrying the uap56<sup>α15</sup> allele (an E245K substitution) in trans to hypomorphic 5’UTR deletions are sterile and show remarkably specific defects in asymmetric RNA localization (Eberl et al., 1997; Meignin and Davis, 2008), raising the possibility that UAP56 has a germline function in piRNA production and transposon silencing.

To explore a potential link to the piRNA pathway, we determined the subcellular localization of UAP56 relative to the piRNA cluster associated HP1 variant Rhino (Rhi) and the nuage component Vas (Figure 2.1). Immunofluorescence labeling and confocal imaging revealed UAP56 diffusely localized throughout germline and somatic nuclei, but also concentrated in distinct nuclear foci in the nurse cells, which produce most of the maternal mRNAs that are deposited in the oocyte (Figure 2.1A). Strikingly, 98.4% of these foci, defined by UAP56 signal above background, co-localized with Rhi foci. Furthermore, 99.1% of Rhi signal above background co-localized with UAP56 foci (Figure 2.1A and B; Pearson’s correlation coefficient R = 0.99). In the transcriptionally silent oocyte nucleus, however, Rhi localized to foci but UAP56 was dispersed (Figure 2.2). In <i>rhi</i><sup>2</sup>]<i>rhi</i><sup>KG</sup> mutant ovaries, which do not express detectable Rhi protein, UAP56 was dispersed throughout the nucleus and did not form foci. Rhi is therefore required for UAP localization to nuclear foci, which may also require transcription.
Figure 2.1. Organization of the piRNA biogenesis machinery

A, B. UAP56, Rhi and Vasa localization in a stage 8 nurse cell. Single channel images are shown on left and a 3-color merged image is show on the right (color assignments as indicated). The line in the merged image indicates the position of a line scan for fluorescence intensity (B). The nuclear foci contain overlapping peaks of UAP56 and Rhi, and Vasa-GFP signal accumulates across the nuclear envelope from foci that are closely associated with the periphery. Scale bars = 2 mm C. Scatter plot comparing Rhi signal in foci at the nuclear periphery with Vasa signal in adjacent nuage foci. The diagonal indicates identical signal levels. $R$ is the Pearson correlation coefficient, and the $p$ value is based on Pearson’s product moment correlation coefficient and follows a $t$ distribution. 135 pairs of foci from 25 nuclei were quantified.
Figure 2.1
Figure 2.2. UAP56 does not co-localize with Rhi in oocyte.

At the posterior of the *Drosophila* egg chamber, follicle cells surround transcriptional silenced oocyte. Rhi in green localizes to the oocyte nuclear foci, while UAP56 in red localizes to follicle cell nuclei but does not co-localize with Rhi in oocyte. Rhi and UAP56 are shown separately on the left, and a merged image with Rhi in green and UAP56 in red is on the right. Scale bars = 5 mm.
We next assayed Rhi localization to nuclear foci in \textit{uap56} mutants. The \textit{uap56}^{sx15} allele, when combined with hypomorphic 5’ UTR deletion alleles, is viable but sterile, and leads to spindle class oocyte patterning defects that are characteristic of piRNA pathway mutations (Meignin and Davis, 2008). The point mutation in the \textit{uap56}^{sx15} allele does not alter protein expression (Meignin and Davis, 2008), and the \textit{uap56}^{28} 5’ UTR deletion allele does not change protein structure, but appears to reduce expression (Figure 2.3; (Meignin and Davis, 2008). The \textit{uap56}^{sx15}/\textit{uap56}^{28} combination thus generates mutant ovaries in which most of the UAP56 is likely to carry the E245K substitution. In these ovaries, UAP56 was dispersed throughout the nurse cell nuclei. By contrast, Rhi was present in distinct foci during early oogenesis (Figure 2.4). In later stage egg chambers, however, Rhi was dispersed in the nurse cell nuclei (Figure 2.5). Low levels of wild type UAP56, produced by the \textit{uap56}^{28} allele, could support Rhi localization during early oogenesis. Alternatively, UAP56 could be required to maintain Rhi localization during later stages of oogenesis, but dispensable for localization early. We favor the former possibility, and speculate that UAP56 and Rhi are co-dependent for localization to nuclear foci.
Figure 2.3. UAP56 expression in mutants and transgenic lines.

Western blots for UAP56 in 2-4 day old ovaries. Genotypes are indicated above each lane. Endogenous UAP56 is approximately 50KD. The UAP-56-Venus fusion proteins migrate at 80KD. Actin was used as a loading control, and intensities were quantified using an Odyssey Imager. Bar graph in the lower panel shows UAP56 expression normalized to Actin. UAP56 is reduced by approximately 30% in \textit{uap56}^{28} heterozygous ovaries relative to \textit{uap56}^{\texttt{sz15r}} heterozygous ovaries, and by approximately 50% in \textit{uap56}^{\texttt{sz15r}}/\textit{uap56}^{28} transheterozygous ovaries relative to \textit{uap56}^{\texttt{sz15r}} heterozygotes. The endogenous UAP56, UAP56-Venus and sz-Venus fusions were expressed at comparable levels.
Figure 2.3

Quantification of Western Blot

Levels of proteins (normalized to Actin)

KD Marker uap56+ Cyo uap56−/Cyo uap56+/Cyo uap56+ uap56−/uap56+ UAP56-Venus sz15'-Venus

UAP56-Venus or sz15'-Venus

UAP56

Actin

Markers:
- UAP56
- UAP56-Venus
- sz15'-Venus

Levels of proteins normalized to Actin.
Figure 2.4. The E245K substitution and rhi mutations disrupt UAP56 localization to nuclear foci.

UAP56 (red), Rhi (green), and nuclear pores (cyan) in wild type, $uap56$ and $rhi$ mutant nurse cell nuclei. UAP56 and Rhi are shown with nuclear pores (NP) in the first two images in each row. A merged image of Rhi, UAP56 and nuclear pores is show in the last panel of each row. In stage 4 $uap56^{sz15}/uap56^{28}$ egg chambers, Rhi localizes to foci but UAP56 is dispersed in the nucleoplasm. In $rhi$ mutations, Rhi protein is not detected and UAP56 fails to localize to foci. Scale bars are 2µm.
Figure 2.4

wt

Rhi/NP  UAP56/NP  Rhi/UAP56/NP

uap56sz15/uap5628

rh2/rhikG
Figure 2.5. Rhi and UAP56 localization in later stage $uap56$ mutant egg chambers.

Localization of Rhi (green), UAP56 (red) and the nuclear pore complex (cyan) in stage 8 wild type and $uap56^{sz15}/uap56^{28}$ egg chambers. In later stage mutant egg chambers, both Rhi and UAP56 show little signal in foci, suggesting that $uap56$ is required for Rhi localization to clusters. Scale bars = 2$\mu$m.
Figure 2.5

wt

Rhi/NP  UAP56/NP  Rhi/UAP56/NP

uap5628

uap5628
To directly determine if the E245K substitution alters the function and subcellular localization of UAP56, we generated transgenes expressing wild type \textit{uap56} (UAP56-Venus) or the \textit{uap56^{sz15}} mutant allele (sz-Venus) fused to the fluorescent protein Venus. Both transgenes were driven by the \textit{uap56} promoter and integrated into the same chromosomal location, and western blots showed that both fusion proteins were expressed at comparable levels (Figure 2.3). The wild type UAP56-Venus fusion rescued the lethality associated with both \textit{uap56^{28}/Df} and \textit{uap56^{sz15}/Df} combinations, and 75\% of the embryos produced by \textit{uap56^{sz15}/Df} females expressing this fusion showed normal dorsal ventral patterning, and a small fraction of these embryos hatched (Table 2.1). The failure to restore full fertility is likely due to the Venus fusion, as a genomic fragment expressing untagged protein fully rescues the mutant phenotype (Meignin and Davis, 2008). By contrast, the sz-Venus fusion failed to rescue viability in either mutant background (Table 2.1). To assay fusion protein localization in the ovary, we expressed both transgenes in the viable \textit{uap56^{sz15}/uap56^{28}} trans-heterozygous background and labeled for Rhi. Wild type UAP56-Venus precisely co-localized with Rhi foci in early egg chambers (Figure 2.6), and this transgene rescued Rhi localization to foci during later stages (data not shown). By contrast, the sz-Venus fusion was dispersed in the germline nuclei, did not co-localize with Rhi (Figure 2.6), and failed to rescue Rhi foci in late stage egg chambers. The E245K substitution thus disrupts UAP56 co-localization with Rhi and Rhi focus stability.
Table 2.1. Egg production, D-V patterning, and hatch rates in *uap56* mutants.

Genotypes are indicated in the first column. Zygotic lethality is indicated in the second column (L). Virgin females were raised at 25°C with wild type Oregon-R strain males. Eggs were collected on grape juice plates and scored for total production (egg/female/day), fraction with wild type dorsal appendages, and fraction of eggs that hatch (%). N indicates total egg scored for each genotype.
<table>
<thead>
<tr>
<th>Genotype</th>
<th>Lethality (L)</th>
<th>egg/female/day</th>
<th>eggs with normal appendages (%)</th>
<th>Hatch (%)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>uap56</td>
<td>L</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>uap56 /uap56</td>
<td>L</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>uap56</td>
<td>L</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>uap56 /uap56</td>
<td>L</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>uap56 /uap56</td>
<td>76</td>
<td>0</td>
<td>0</td>
<td>2274</td>
<td></td>
</tr>
<tr>
<td>uap56 /uap56 : UAP56-Venus</td>
<td>36</td>
<td>75</td>
<td>1</td>
<td>2023</td>
<td></td>
</tr>
<tr>
<td>uap56 /uap56 : UAP56-Venus</td>
<td>13</td>
<td>3</td>
<td>1</td>
<td>549</td>
<td></td>
</tr>
<tr>
<td>uap56 /uap56 : UAP56-Venus</td>
<td>59</td>
<td>26</td>
<td>0</td>
<td>4316</td>
<td></td>
</tr>
<tr>
<td>uap56 /uap56 : sz-Venus</td>
<td>L</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>uap56 /uap56 : sz-Venus</td>
<td>L</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>uap56 /uap56 : sz-Venus</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>716</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2.6. Localization of transgenic wild type UAP56-Venus and sz-Venus fusion proteins.

Egg chambers expressing the transgenes were Immunolabeled for of Rhi (left panel, green in right panel). Fusion protein distribution is show in the middle panel and in red in the right panel. sz-Venus, which carries the E245K substitution found in the uap56\textsuperscript{e215} allele, fails to co-localize with Rhi foci. Scale bars are 2µm.
Figure 2.6

Rhi UAP56-Venus/sz-Venus Rhi/UAP56

uap56sz15'/uap5628; UAP56-Venus

uap56sz15'/uap5628; sz-Venus
UAP56 is required for nuage localization of piRNA pathway proteins

Our subcellular localization studies also showed that prominent UAP56-Rhi foci were closely associated with the nuclear periphery, and that nuage foci containing a functional GFP-Vasa fusion appeared to localize directly across the nuclear envelope (Figure 2.1A). Labeling for Rhi, Vas and nucleoporins (Figure 2.7A) showed that the pores lie between the nuclear foci containing Rhi and the nuage foci containing Vasa, and that a significant fraction of Vasa signal overlaps with signal for the nuclear pores (Figure 2.7B). To quantify the relationship between Rhi foci in the nucleus and Vasa foci in nuage, we identified all Rhi foci that were above threshold and within the confocal resolution limit of nuclear pores (the signals overlapped), and for each of these foci, determined if a focus of GFP-Vas was present at the adjacent nuclear periphery. 98.5% of Rhi foci near the pores had an associated focus of Vasa at the cytoplasmic face of the nucleus. We then quantified Rhi in the nuclear foci and Vasa levels in the adjacent nuage foci. Strikingly, the level of Vas was highly correlated with the level of Rhi (Figure 2.1C; R=0.72, p<2.2e-16). Furthermore, Vasa localization to nuage was disrupted in uap56ez15/uap5628 mutants (Figure 2.8). By contrast, and Rhi and UAP56 co-localized to distinct nuclear foci in vas mutants (Figure 2.9). Rhi and UAP56 thus appear to function at clusters, in a process that is upstream of Vas localization to nuage.
Figure 2.7. Rhi and Vasa associate with the nuclear envelope.

A. Triple label immunofluorescence and confocal imaging of Rhi, nuclear pores (NP) and Vasa. Rhi localizes to nuclear foci that are often closely associated with nuage foci containing Vas, and the nuclear pores lie between these foci, with significant overlap between the signal for Vas and the pores. Single channel images are shown on left and a 3-color merged image is shown on the right (color assignments as indicated). B. 3D intensity profile of the three fluorescent signals (Rhi in red, NP in blue and Vas in green). The left panel shows the profile of the whole nucleus. The middle and right panels are higher magnification views of the area indicated in (A), which suggest that Vas may be partially overlapped with the nuclear pores. Scale bar = 2μm.
Figure 2.7

A

Rhi

NP

Vas-GFP

2µm

Rhi/NP/Vas-GFP

B

Graphs showing data analysis.

47
Figure 2.8. *uap56* mutants disrupt nuage accumulation of Vasa.

Vasa (green), Actin (red) and DNA (blue) in *wt* and *uap56* mutant nurse cells. In wild type nurse cells, Vas localized to the perinuclear nuage, but this localization was absent in *uap56* mutants. Egg chambers were labeled with anti-Vas antibody, rhodamine phalloidin (F-Actin), and with TOTO3 (DNA). Imaging was performed with a Leica laser scanning confocal microscope. Scale bars = 2μm.
Figure 2.8
Figure 2.9. Rhi and UAP56 localization in vas mutant.

Localization of Rhi (green), UAP56 (red) and the nuclear pore complex (cyan) in wild type and vas<sup>D5</sup>/vas<sup>PH</sup> egg chambers. Rhi and UAP56 colocalize in nuclear foci in both genotypes. Scale bars = 2μm.
Figure 2.9
The PIWI proteins Aub and Ago3 localize with Vas in nuage, where they appear to drive piRNA amplification. The founding member of the PIWI family, Piwi, is concentrated in nuclei (reviewed by (Klattenhoff and Theurkauf, 2007)). The uap56<sup>ez15</sup>/uap56<sup>28</sup> transheterozygous combination did not alter Piwi localization to nuclei, but disrupted Aub and Ago3 localization to nuage (Figure 2.10A). Expression of wild type UAP56-Venus in the uap56<sup>ez15</sup>/uap56<sup>28</sup> background partially restored nuage localization of Aub and fully restored nuage localization of Ago3 (Figure 2.10B). By contrast, transgenic expression of sz-Venus failed to restore nuage localization of either protein (Figure 2.10B). Aub and Vas localize to the posterior pole of the oocyte, where they are incorporated into pole plasm. Posterior localization of both proteins is disrupted in uap56<sup>ez15</sup>/uap56<sup>28</sup> mutants (Meignin and Davis, 2008) (Figure 2.11A and B). Posterior localization of Aub was restored in mutants expressing the wild type UAP56-Venus transgene, but not in mutants expressing the sz-Venus fusion (Figure 2.11B). We did not assay Vas localization in these transgenic ovaries, but Aub localization to pole plasm requires Vas (Harris and Macdonald, 2001), suggesting that Vas localization is also restored. The E245K substitution thus disrupts the organization of nuclear and cytoplasmic components of the piRNA biogenesis machinery.
Figure 2.10. **UAP56 is required for PIWI protein localization to nuage**

A. Aub, Ago3 and Piwi localization in *Drosophila* nurse cells. Aub, Ago3, and Piwi are in green and DNA is in blue. Aub and Ago3 localized at perinuclear nuage in wild type nurse cells, while Piwi is in nuclei of both nurse cells and follicle cells. In *uap56^{sz15}/uap56^{28}* mutants, Aub and Ago3 localization to nuage is disrupted, but Piwi localizes to the nucleus. B. Aub and Ago3 nuage localization is rescued by the wild type UAP56-Venus transgene, but not the not by the mutant sz-Venus transgene. Gray scale image show Aub and Ago3 and merged images show Aub and Ago3 in green and DNA in blue. Scale bars are 2µm. Also see Figure S3.
Figure 2.10

A

<table>
<thead>
<tr>
<th></th>
<th>Aub/DNA</th>
<th>Ago3/DNA</th>
<th>Piwi/DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt</td>
<td><img src="image1.jpg" alt="Image" /></td>
<td><img src="image2.jpg" alt="Image" /></td>
<td><img src="image3.jpg" alt="Image" /></td>
</tr>
<tr>
<td>uap56sz15'/uap5628</td>
<td><img src="image4.jpg" alt="Image" /></td>
<td><img src="image5.jpg" alt="Image" /></td>
<td><img src="image6.jpg" alt="Image" /></td>
</tr>
</tbody>
</table>

B

<table>
<thead>
<tr>
<th></th>
<th>Aub</th>
<th>Aub/DNA</th>
<th>Ago3</th>
<th>Ago3/DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>uap56sz15'/uap5628</td>
<td><img src="image7.jpg" alt="Image" /></td>
<td><img src="image8.jpg" alt="Image" /></td>
<td><img src="image9.jpg" alt="Image" /></td>
<td><img src="image10.jpg" alt="Image" /></td>
</tr>
<tr>
<td>UAP56-Venus</td>
<td><img src="image11.jpg" alt="Image" /></td>
<td><img src="image12.jpg" alt="Image" /></td>
<td><img src="image13.jpg" alt="Image" /></td>
<td><img src="image14.jpg" alt="Image" /></td>
</tr>
<tr>
<td>sz-Venus</td>
<td><img src="image15.jpg" alt="Image" /></td>
<td><img src="image16.jpg" alt="Image" /></td>
<td><img src="image17.jpg" alt="Image" /></td>
<td><img src="image18.jpg" alt="Image" /></td>
</tr>
</tbody>
</table>
Figure 2.11. Posterior localization of Aub and Vasa.

A. B. Immunolocalization of Vasa (A, green) and Aub (B, green) in stage 9/10 oocytes. DNA is in blue. In wild type oocytes, Aub and Vasa localize to the posterior pole. In \textit{uap56^{sz155}/uap56^{28}} mutants, both proteins are dispersed. Posterior localization of Aub is restored in \textit{uap56^{sz155}/uap56^{28}} mutants expressing the wild type UAP56-Venus fusion, but not in mutants expressing the sz-Venus fusion. Scale bars = 2\,\mu m.
Figure 2.11

![Image showing genetic crosses involving wild type (wt) and uap56sz15/uap5628 alleles in Aub and Vas contexts.](image_url)

- **Aub**
  - wt
  - uap56sz15/uap5628
  - uap56sz15/uap5628; UAP56-Venus
  - sz15'-Venus

- **Vas**
  - wt
  - uap56sz15/uap5628
  - uap56sz15/uap5628; UAP56-Venus
  - sz15'-Venus
**Germline DNA damage**

piRNA pathway mutations lead to germline DNA damage, which is proposed to result from transposon mobilization (reviewed by Klattenhoff and Theurkauf, 2008). γH2Av is a phosphorylated histone variant that accumulates at DNA break sites (Madigan et al., 2002). In wild type egg chambers, γH2Av foci are present in region 2 of the germarium, where meiotic breaks form, and in later nurse cell nuclei undergoing endoreduplication (Jang et al., 2003). Stage 2 and later oocyte nuclei, by contrast, are consistently negative for γH2Av foci. In uap56ε15/uap56ε8 mutants, γH2Av foci persist in oocyte nuclei and appear to be enhanced in the nurse cells (Figure 2.12). In uap56ε15/uap56ε8 mutants expressing the wild type UAP56-Venus transgene, by contrast, γH2Av foci were not detected in later stage oocytes (Figure 2.12). UAP56 is therefore required to maintain germline genome integrity.
Figure 2.12. DNA damage in uap56 mutant.

Accumulation of DNA double strand breaks in uap56 mutant egg chambers. 2-4 day old egg chambers were labeled with antibody to γH2Av (green), a phosphorylated histone variant associated with DNA double strand break sites. DNA, labeled by TOTO3, is in red. In wild type, minimal γH2Av signal is detected. In uap56ε15r/uap5628 transheterozygotes, by contrast, γH2Av signal is increased in the nurse cells and oocyte (arrow). Transgenic expression of wild type UAP56-Venus in the uap56ε15r/uap5628 background reduced γH2Av labeling, confirming that germline DNA damage is due to the uap56 mutant alleles, and not background mutations. Scale bars = 5μm
Figure 2.12

![Image of immunofluorescence staining](image.png)

- **wt**
- **uap56sz15'/uap5628**
- **uap56sz15'/uap5628; UAP56-Venus**

Legend:
- γ-H2Av: Green
- DNA: Red
- γ-H2Av/DNA/Actin: Blue

Note: The image shows immunofluorescence staining for γ-H2Av, DNA, and γ-H2Av/DNA/Actin in different genotypes, with arrows indicating specific areas of interest.
Gene and transposon expression

UAP56 is required for splicing and export of a broad spectrum of mRNAs (Gatfield et al., 2001; Shen, 2009), but our phenotypic and localization data raised the possibility that UAP56 also functions with Vas and Rhi to control transposon activity. We therefore used whole genome tiling arrays to assay gene and transposon expression in \( uap56^{ce15} / uap56^{28} \) and vas mutant ovaries, and compared these data with an earlier analysis of rhi mutants (Klattenhoff et al., 2009). The \( \text{vasa intronic gene (vig)} \) is contained in a vas intron, and Vig co-purifies with FMRP and components of the siRNA machinery (Caudy et al., 2002), suggesting that it functions in RNA silencing. To specifically disrupt vas, we therefore analyzed a null deletion allele that removes both vas and vig in trans to a point mutation in vas that does not disrupt vig. The Genome Browser screen shot in Figure 2.13 shows expression of the retrotransposon Blood and neighboring Ent1 gene, which contains three introns, in control, \( uap56 \), rhi and vas mutants. In all three mutants, Blood is over-expressed and Ent1 is expressed at levels comparable to the control (Figure 2.13). This pattern extends across the genome, as no protein-coding gene showed a statistically significant difference from wild type in \( uap56 \), vas or rhi mutants (FDR<0.05) (Figure 2.14B). By contrast, all three mutations disrupted transposon silencing, with 11 transposon families showing significant over-expression in \( uap56 \), rhi, and vas (FDR<0.05; Figure 2.14A, red points; Venn diagram, Figure2.14C). RT-qPCR and northern blotting confirmed that expression of a control protein coding gene (RP49) did not change and that two transposons (Het-A and Blood) were over-expressed (Figure 2.15A and B). The \( uap56^{ce15}/uap56^{28} \) allelic combination, and mutations in vas and rhi, thus disrupt transposon silencing but do not alter gene expression. Although the wild
type UAP56-Venus transgene restored the proper nuclear UAP56 localization in \textit{uap56} mutants, Het-A and Blood transcripts levels were still elevated which is consistent with the failure to fully rescue the female fertility (Figure 2.15C and Table 2.1). This suggests the Venus tag may perturb UAP56 function.
Figure 2.13. *uap56*, *rhi* and *vasa* mutations disrupt transposon Blood silencing but do not alter Ent1 gene expression.

Whole genome tiling arrays were used to assay gene and transposon expression in wild type, *uap56*^{es15}/*uap56*^{28}, *vas*^{D5}/*vas*^{PH}, and *rhi*^{02086}/*rhi*^{KG00910} ovaries. A Genome Brower screen shot showing a region on chromosome 2L containing the nature transposon, Blood and Ent1 gene. In *uap56*, *vas* and *rhi* mutants, Blood expression is increased dramatically, while the Ent1 levels are comparable to control.
Figure 2.13

![Graph showing gene expression levels for uap56, rhi, vasa, and wt.](image)
Figure 2.14. *uap56*, *rhi* and *vasa* mutations disrupt transposon silencing but do not alter gene expression.

Whole genome tiling arrays were used to assay gene and transposon expression in wild type, *uap56<sup>sz15</sup>*/uap56<sup>28</sup>, *vas<sup>D5</sup>/vas<sup>PH</sup>*, and *rhi<sup>02086</sup>/rhi<sup>KG00910</sup>* ovaries. A. Scatter plots comparing the expression of transposons in mutants relative to wild type. The diagonal indicates identical expression levels. All 3 mutants show significant (FDR<0.05) over-expression of a subset of transposon families. Over-expression of selected transposon families was confirmed by northern blotting and qPCR (Figure 2.15). B. Scatter plots comparing the expression of protein-coding genes mutant and wild type ovaries. None of the mutations lead to a significant change (FDR<0.05) in protein coding genes expression. The *uap56, rhi* and *vas* transcripts are highlighted, as indicated by the legend. C. Venn diagram showing the overlap in transposon families with significant up-regulation (FDR<0.01) in *uap56, rhi* and *vas* mutants.
Figure 2.14

A

B

C

UAP56 RA; UAP56 RC
UAP56 RB
Rhi
Vas
Figure 2.15. Transposon over-expression in *uap56* mutant.

A. Northern blotting for Blood and HeT-A transposon transcripts. *OreR* and *armi* serve as negative and positive controls. The bar graph shows quantification of the northern blot signal. Consistent with tiling array data, HeT-A expression was increased by more than 10 fold and Blood increased approximately 5 fold. B. RT-qPCR quantification of RP49 (control mRNA) and HeT-A transcripts in wild type (*OreR*) and *uap56* mutant ovaries. 18s rRNAs was used as an internal control and no RT-primer background signal was subtracted. RP49 is expressed at similar levels in wild type and mutant ovaries. Two sets of PCR primers, amplifying different regions of HeT-A, confirm a 10-fold increase of transcript levels. C. RT-qPCR quantification of RP49 (control mRNA) and HeT-A, Blood transcripts fold changes in *sz15'/uap56*; UAP56-Venus ovaries compared with wild type (*OreR*). UAP56-Venus did not restore transposon silencing in *uap56* mutants.
Figure 2.15

A

Quantification of northern blot

B

Transcripts level quantified by qRT-PCR

C

Transcripts level quantified by qRT-PCR
piRNA production

To determine if the \textit{uap56}^{\text{sz15}}/\textit{uap56}^{28} combination blocks production of piRNAs, we deeply sequenced small RNAs from mutant ovaries, which revealed a significant reduction in total piRNAs (Figure 2.16A and B). piRNAs from opposite strands that overlap by 10nt are characteristic of ping-pong amplification. The bias toward 10nt overlap was reduced in \textit{uap56}^{\text{sz15}}/\textit{uap56}^{28}, \textit{rhi} and \textit{vas} mutant ovaries (Figure 2.16C, D and E). The scatter plots in Figure 2.17 show piRNAs mapping to group 1, 2 and 3 transposons in \textit{uap56}^{\text{sz15}}/\textit{uap56}^{28}, \textit{vas} and \textit{rhi} mutants compared with wild type controls. In all thee genotypes, piRNAs specific to group 1 transposable elements, which appear to be expressed primarily in the germline, are significantly reduced (Figure 2.17, black points). By contrast, the levels of piRNAs linked to group 3 transposons, which are expressed primarily in the somatic follicle cells, were comparable to wild type (Figure 2.17, red points). Vas and Rhi both appear to be specific to the germline, consistent with changes in germline piRNAs. However, UAP56 is expressed in both the germline and somatic follicle cells. The \textit{uap56}^{28}/\textit{uap56}^{\text{sz15}} allelic combination thus appears to disrupt a germline specific branch of the piRNA biogenesis pathway.
Figure 2.16. piRNA expression in wt and uap56<sup>515</sup>/uap56<sup>28</sup> mutant ovaries.

A and B. The length histogram of ovarian piRNAs from wild type (A) and uap56 mutants (B). The blue and red bars show the abundance of sense and antisense piRNAs, respectively.  C and D. Histograms showing the frequency of overlap between piRNAs from opposite stands in wild type (C) and uap56 mutant (D). Wild type piRNAs show a highly significant enrichment for piRNA from opposite strands that overlap by 10 nt, which is characteristic of ping-pong amplification (C). The uap56<sup>515</sup>/uap56<sup>28</sup> combination significantly reduced transposon and cluster piRNA abundance, and reduced piRNAs from opposite strands that overlap by 10 nt (D).  E. Distribution of pingping Z scores for transposon mapping piRNAs in w1118, uap, rhi and vasa. The p-values are calculated by the paired t-test.
Figure 2.16

A Transposon mapping small RNAs in wt uox

B Transposon mapping small RNAs in uap uox

C wt piRNA pingpong zscore=61.06

D uap piRNA pingpong zscore=20.71

E
Figure 2.17. Transposon mapping piRNA expression in *uap56, rhi* and *vasa* mutant ovaries.

Scatter plots comparing the abundance of piRNAs mapping to germline enriched group 1 transposons (black), soma enriched group 3 transposons (red) and class 2 transposons, which show a sense strand bias (green). All three mutations reduced piRNA from germline specific Group 1 transposon families.
Figure 2.17

- uap vs. wt
- rhi vs. wt
- vasa vs. wt
piRNA clusters that produce uniquely mapping piRNAs from both genomic strands dominate in the germline, while the major somatic piRNA cluster, *flam*, and cluster 2 produce unique piRNAs almost exclusively from one genomic strand (Brennecke et al., 2007; Gunawardane et al., 2007). Mutations in *rhi* nearly eliminate piRNAs from dual-strand clusters, but do not block piRNA production by *flam* or cluster 2 (Klattenhoff et al., 2009). The *uap56^{28}/uap56^{ex15}* allelic combination produces an essentially identical reduction in dual strand cluster expression (Figure 2.18A). Mutations in *vas* also block expression of dual-strand clusters, and do not alter *flam* expression. However, *vas* mutations also reduce piRNAs linked to cluster 2, which appears to be expressed in the germline and soma (Figure 2.18A; (Malone et al., 2009). To directly compare cluster expression in the three mutant backgrounds, we generated pair-wise scatter plots covering the top 50 clusters (Figure 2.18B). This analysis showed that *uap56^{ex15}/uap56^{28}* and *rhi^{2}/rhi^{KG}* produce nearly identical changes in cluster expression (R=0.96), and that very similar changes are produced by *vas* mutants (R=0.83 for both *vas-uap56* and *vas-rhi* comparisons; Figure 2.18B). UAP56, Rhi and Vas thus represent nuclear and nuage components of a dual-strand cluster expression system.
Figure 2.18. Cluster mapping piRNA expression in *uap56*, *rhi* and *vasa* mutant ovaries.

A. Scatter plots comparing the abundance of cluster piRNAs in 3 mutants and wild type ovaries. Each point represents piRNAs from a single cluster. All three mutations significantly reduced piRNAs to the germline clusters, including the major dual strand cluster at 42AB (red). By contrast, *uap56* and *rhi* mutants do not reduced piRNA to unistrand cluster 2 (green) or *flam* (blue). Mutations in *vas* reduce piRNAs linked to cluster 2, but not *flam*. B. Pair-wise comparison of cluster piRNAs in mutant ovaries. Cluster expression is highly correlated in all three mutants, with *rhi* and *uap56* showing almost identical patterns of cluster piRNA expression (R=0.96).
Figure 2.18

A

\[ uap \text{ vs. } \text{wt} \]

\[ rhi \text{ vs. } \text{wt} \]

\[ vasa \text{ vs. } \text{wt} \]

B

\[ uap \text{ vs. } \text{rhi} \]

\[ vasa \text{ vs. } \text{uap} \]

\[ vasa \text{ vs. } \text{rhi} \]
The majority of piRNAs in the *Drosophila* germline are derived from transposons and other repeats, but a subset map to gene transcripts. Chromosomal profiles of uniquely mapping 23 to 29 nt RNAs in *uap56sz15*/*uap56*28 mutants revealed an increase in small RNAs from euchromatic sites. Consistent with this observation, the scatter plots in Figure 2.19A show that 23 to 30 nt genic RNAs increase significantly in both *rhi* and *uap56* mutants. By contrast, *vas* produced a modest decrease in these genic RNAs. These data were normalized to sequencing depth, and reduced transposon mapping piRNAs could therefore artificially inflate the genic piRNA pool. We therefore normalized to total miRNAs and repeated the analysis, which confirmed the increase in genic 23 to 30 nt RNAs in *uap56* and *rhi* mutants (Figure 2.19E-G). Most of these species are derived from 3’ UTRs (Figure 2.19H). This 3’UTR bias is also present in wild type and *vas* mutants, but the high signal in *uap56* and *rhi* compresses profiles and obscures this bias.

In wild type ovaries, genic small RNAs show a peak at 21nt, which is characteristic of endo-siRNAs, and a peak at 26nt, which is consistent with piRNAs bound to Piwi (Brennecke et al., 2007; Figure 2.19B). In *uap56sz15*/*uap56*28 mutants, by contrast, the length distribution was much broader, with a shoulder between 20-22nt (Figure 2.19C). Some of these species could be non-specific mRNA breakdown products. Following binding to Piwi proteins, piRNAs are modified by the addition of a 2’-O-methyl group at the 3’ terminus (Horwich et al., 2007; Saito et al., 2007), which renders these species resistant to oxidation. To estimate the fraction of genic small RNAs that are bound to Piwi proteins, we therefore deep sequenced small RNAs after oxidation. These studies revealed a significant increase in oxidation-resistant genic small RNAs in
uap56 mutants, with a length distribution characteristic of Piwi binding (Figure 2.19D).

Mutations in uap56 thus reduce cluster piRNAs and increase genic piRNAs, suggesting that the specificity of processing has been compromised.
Figure 2.19. Genic piRNA expression in uap56, rhi and vasa mutant ovaries.

A. Genic piRNA expression in mutant ovaries. Scatter plots compare the abundance of piRNAs linked to protein coding genes in mutant ovaries relative to wild type controls. Both uap56 and rhi mutants lead to a significant increase in ectopic piRNAs from protein coding genes. R-values are shown at upper left corner. RNA sequencing data for vas and rhi are from Malone et al. (2009) and Klattenhoff et al. (2009). B-D. Length distribution of genic small RNAs in wild type (B) and uap56 mutants (C). D. Oxidation-resistant genic small RNAs in uap56 normalized by sequencing depth. The blue and red bars show sense and antisense small RNAs. E-G. Scatter plots showing genic piRNAs(≥23nt) in uap, rhi and vasa mutants relative to wild type control, with data normalized to miRNA abundance. Mutations in uap56 and rhi increase genic piRNAs. H. Distribution of genic piRNAs on mature transcripts. Transcripts were divided into 20 equal length bins, from the 5′ end to the 3′ end. Genic piRNAs were then normalized by transcript length plotted as a function of bin number. Both uap56 and rhi lead to an increase in piRNAs mapping to the 3′ ends of mature transcripts. Genic piRNAs in wild type and vas mutants show a 3′ bias in piRNA distribution, but this is obscured due to low signal relative to uap56 and rhi mutants.
Figure 2.19

genic small RNAs in uap unox (normalized to sequencing depth) (rpkm) (Log2)
genic small RNAs in rhi unox (normalized to sequencing depth) (rpkm) (Log2)
genic small RNAs in vasa unox (normalized to sequencing depth) (rpkm) (Log2)

p=2.81e-06  p=4.16e-50  p=1.65e-02

genetic small RNAs in wt unox (normalized to sequencing depth) (rpkm) (Log2)
piRNA precursor binding

Clusters are proposed to produce long precursor RNAs that are processed into primary piRNAs in the nuage. Strand specific RT-qPCR showed that the uap56<sup>sz15</sup>/uap56<sup>28</sup> combination significantly reduced the steady state level of plus and minus strand RNAs from two regions of the major 42AB cluster (Figure 2.20A), but did not reduce precursor RNAs from uni-strand cluster 2 or flam (Figure 2.20B). And the reduction of piRNA precursor transcripts is partially rescued by UAP56-Venus transgene (Figure 2.20E). Mutations in rhi produce similar changes in cluster transcript abundance (Klattenhoff et al., 2009). DEAD box proteins can function as ATP-dependent RNA clamps (Linder and Jankowsky, 2011), and we speculated that UAP56 binds and stabilizes transcripts derived from germline clusters. To test this hypothesis, we immunoprecipitated wild type UAP56-Venus and sz-Venus fusion proteins from ovary extracts and assayed RNA binding by random primed strand-specific deep sequencing. To quantify enrichment, genome mapping fragments in the input and precipitated fractions were normalized to noncoding RNAs (mainly ribosomal RNAs), which show only background binding to the beads.
Figure 2.20. piRNA precursor expression in *uap56* and *vasa* mutant ovaries.

Strand specific RT-qPCR quantification of piRNA precursor transcripts from cluster 1/42AB, cluster 2 and *flam*. (A and C) Two sets of RT-qPCR primers were used to quantify the abundance of the pre-piRNAs from dual-strand cluster 1/42AB. Both show a strong reduction in precursor levels in *uap56* mutants (A), and no significant change in *vas* mutants (C). (B and D) Quantification of precursor transcripts from uni-strand cluster 2 and *flam*. Transcript levels in both *uap56* and *vas* are not significantly different from controls. E. Quantification of precursor transcripts from dual-strand cluster 1 in wild type (OreR), *uap56* mutants (*sz15'/uap56*28) and *sz15'/uap56*28; UAP56-Venus. UAP56-Venus transgene partially rescued piRNA precursor production in *uap56* mutants. Error bars represent SE from 3 biological replicates.
Figure 2.20

A) pre-piRNAs from cluster 1/42AB

B) pre-piRNAs from cluster 2 and flam

C) pre-piRNAs from cluster 1/42AB

D) pre-piRNAs from cluster 2 and flam

E) pre-piRNAs from cluster 1/42AB
Strikingly, RNAs from the major germline piRNA cluster at 42AB were the most highly enriched species in the UAP56-Venus precipitated fraction (34 fold: Figure 2.21 and Figure 2.22A, orange point), and piRNAs from 42AB are nearly eliminated by the uap56 mutation (Figure 2.21). Transcripts from flanking protein-coding gene show relatively modest enrichment, and transcripts from the somatic flamenco cluster were not enriched (Figure S2.22A, blue point). RT-qPCR confirmed that transcripts from cluster 1/42AB are enriched in the precipitated fraction, while transcripts from cluster 2, flam are not (Figure 2.22C). Across the transcriptome, RNAs from the top 40 piRNA producing clusters showed a 7.9-fold enrichment, and species from all 142 clusters identified by Brennecke et al. (2007) showed an average 5.2 fold enrichment. Protein coding genes, by contrast, showed no average enrichment (Figure 2.22B). The uap56ε15 mutation caused a modest, but statistically significant, reduction in RNA binding for both cluster transcripts and mRNAs (Figure 2.21, 2.22A, B and C). By contrast, this substitution does not significantly alter gene expression (Figure 2.14B), but nearly eliminates piRNA production by germline clusters (Figure 2.18A) and blocks UAP56 co-localization with Rhi (Figure 2.4). The E245K substitution thus appears to specifically block an interaction between UAP56 and the piRNA biogenesis machinery.

Primary piRNAs show a significant strand bias relative to the embedded transposon fragments that make up clusters, with 60% +/-24% accumulating anti-sense to the mobile elements (Brennecke et al., 2007). The longer RNAs that immunoprecipitate with UAP56, by contrast, do not show strand bias relative to the transposons (47% +/-23% antisense). For example, the region of the 42AB cluster composed of GATE transposon fragments in Figure 2.21 shows typical anti-sense piRNA strand bias, but no
clear bias in UAP56-bound RNAs. This difference in strand bias is highly significant (p-value = 1.123e-06 by Wilcoxon signed rank test (paired)) (Figure 2.22B). Strand bias in the piRNA pool is proposed to originate from post-transcriptional ping-pong amplification of primary piRNAs (Brennecke et al., 2007). We therefore speculate that UAP56 binds to precursor RNAs that have not entered the piRNA amplification cycle.

The DEAD box protein Vas localizes across the nuclear envelope from UAP56 and is required for production of piRNAs from germline clusters. To determine if cluster transcripts associate with Vas, we immunoprecipitated a functional Vas-GFP fusion from wild type ovaries (Johnstone and Lasko, 2004) and assayed for cluster transcripts by RT-qPCR (Brennecke et al., 2007; Klattenhoff et al., 2009). Anti-GFP immunoprecipitation from the parental strain was used as a negative control. As shown in Figure 2.22D, RNAs from the 42AB cluster, but not from flam or uni-strand cluster 2, were significantly enriched by Vas-GFP immunoprecipitation. The DEAD box is an ATP dependent RNA binding domain, suggesting that this interaction is direct. Steady state cluster transcript levels were not reduced in vas mutants, and may be somewhat elevated (Figure 2.20C, D). These observations suggest that Vas binds cluster transcripts as they exit the nucleus, and deliver these RNAs to the piRNA biogenesis machinery.
Figure 2.21. piRNA expression and precursor binding to UAP56.

A. piRNAs and UAP56-associated RNAs mapping to the 1/42AB on chromosome 2R. The locus has many embedded transposable elements and is flanked by protein coding genes, Pld and Jing. In wild type, piRNAs map to both the plus and minus strands of the 42AB cluster (pink tracks), and are dramatically reduced in uap56<sup>2215</sup>/uap56<sup>28</sup> mutants (purple tracks). Wild type UAP56-Venus and sz-Venus transgenes were expressed in w<sup>1</sup> ovaries and immunoprecipitated with anti-FLAG. Bound RNAs were quantified by strand-specific RNA-Seq. The UAP-Venus input signal is in light brown, the UAP-Venus IP signal is in orange, the sz-Venus input is in light green, and the sz-Venus IP signal is dark green. All signals are normalized to ribosomal RNAs. Cluster transcripts were highly enriched over input by both wild type UAP56-Venus and sz-Venus IP. By contrast, the neighboring gene, Pld, showed a modest enrichment. The sz-Venus mutant showed somewhat diminished enrichment relative to wild type, for both cluster and gene transcripts.

B. Cluster specific piRNAs (pink) accumulate anti-sense to the embedded transposon fragments (brown, direction of transcription indicated by arrowheads). UAP56 bound precursor RNAs are derived from the same regions, but do not show clear strand bias (UAP-Venus RIP).
Figure 2.21
Figure 2.22. piRNA precursors bind to UAP56 and Vasa

The transcripts immunoprecipitated with UAP56-Venus or sz-Venus were analyzed genome-wildly by RNA-seq. A. Scatter plot comparing fold-enrichment of transcripts from UAP56-Venus IP versus sz-Venus IP. Each black point represents a protein-coding gene, and the red points represent piRNA clusters. The majority of the coding genes showed no enrichment (most of the black points are clustered around 0), while most of cluster transcripts were enriched in IP sample (most of the red points are above 0). Transcripts from cluster 1/42AB (Orange point) showed the highest fold enrichment in the transcriptome. By contrast, cluster 2 (green point) and flamenco (blue point) are not enriched in the IP samples. B. Box plots comparing fold change of cluster transcripts and coding gene transcripts with UAP56-Venus IP and sz-Venus IP. IP of both fusions produced a significant enrichment of cluster transcripts relative to genes. The sz-Venus fusion showed reduced binding to cluster transcripts and mRNAs. C. Cluster transcript immunoprecipitataion with wild type UAP56-Venus or sz-Venus fusion proteins. The fusions were expressed in w1 ovaries and immunoprecipitated with anti-FLAG. Anti-FLAG immunoprecipitation from parental w1 ovaries was used as a control. Transcripts were assayed by RT-qPCR. Transcripts from cluster 1/42AB were enriched 40-160 fold with the wild type fusion. Transcripts from flam and rp49 were not enriched. A relatively modest but statistically significant reduction in cluster transcript binding was observed with the uap56-sz15' point mutation. D. Cluster transcript immunoprecipitation with Vas. Vas-GFP was driven by the vas promoter in the w1 background. GFP fusion protein was precipitated with GFP-Trap beads and bound RNA was assayed by RT-qPCR. GFP-Trap precipitation from parental w1 ovaries expressing GFP was used as a
control. The precursor transcripts of cluster 1/42AB were enriched 10-60 fold in Vasa-GFP complex, but *flam* transcripts showed minimal enrichment. Cluster 2 transcripts showed a modest, but statistically insignificant, enrichment. Error bars represent SE from 3 biological samples. *** P<0.001; ** P<0.01; *P<0.05.
Figure 2.22

Fold enrichment of pre-piRNAs

RNA IP/qRT-PCR

GFP

VASA-GFP

Fold enrichment of pre-piRNAs

C

D

RNA IP/qRT-PCR

Fold enrichment of pre-piRNAs

cl1-A
c1-32
c2-A
flam
RP49

c1-A
c1-32
c2-A
flam

dual-strand cluster 1/42AB
uni-strand cluster 2 and flam

UAP56-Venus
sz-Venus
wt

GFP
VASA-GFP

***
**
*

***
*

+ -

+ -

+ -

+ -

+ -
The E245K substitution disrupts a conserved surface residue.

DEAD box proteins are composed of N- and C-terminal domains joined by a linker, and the ATP and RNA binding sites span these domains (reviewed by (Linder and Jankowsky, 2011). In the ATP-bound “closed” state, these domains form an RNA clamp. The studies described here indicate that the E245K substitution in uap56sz15 defines a domain that is critical to piRNA production and transposon silencing. To define the location of this substitution on UAP56, we threaded the Drosophila sequence into the crystal structure of the ATP bound form of human UAP56 (Shi et al., 2004; Zhao et al., 2004). This modeling places E245 at the protein surface, near the linker between the two UAP56 domains (Figure 2.23). This residue appears to form a salt bridge between the end of an a-helix and b-sheet, and a structurally analogous salt bridge appears to stabilize the closed ATP and RNA binding conformation in the related DEAD box protein Dpb5(Montpetit et al., 2011). This bridge is disrupted in the open form, which is engaged with the nuclear pore and the Gle1 activator and does not bind RNA (Montpetit et al., 2011). The E245K substitution in uap56sz15 may therefore contribute to the observed reduction in mRNA and cluster transcript binding (Figure 2.22A, B and C). By contrast, this substitution nearly eliminates piRNA production and co-localization with Rhi, but does not disrupt steady state mRNA expression (Figure 2.4, 2.14B and 2.16B). Reduced RNA binding cannot explain this specificity. We therefore propose that this surface residue lies in a domain of UAP56 that interacts with the piRNA machinery, but is dispensable for mRNA export and processing. This residue is conserved from humans to fission yeast (Figure 2.24), raising the possibility that this domain mediates interactions with the small RNA silencing machinery in other systems.
**Figure 2.23. UAP56 structure.**

*Drosophila* UAP56 amino acid sequence treded into a high-resolution crystal structure of human UAP56. The N-terminal and C-terminal domains are connected by a flexible linker (green), and ATP and RNA binding motifs (orange) are located at the interface between the 2 domains. Glutamate 245 (red), which is changed to a lysine in *uap56sz*¹⁵, is located in a β sheet in the N-terminal domain that is attached to the linker. This negatively charged residue also appears to form a salt bridge with Arg236 (blue), which is located in an α-helix in the N-terminal domain.
Figure 2.23
Figure 2.24. UAP56 sequence conservation.

The amino acid sequences of human, fly, worm, fission yeast and budding yeast UAP56 are aligned. E245 (red), which is changed to lysine in *uap56^{c15}*, is conserved in all four species.
Figure 2.24

<table>
<thead>
<tr>
<th>Species</th>
<th>Amino Acid Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>H. sapiens</td>
<td>MAENVDNELLDYDDEVETA--AGGDAEAPAKKDKVGSYVI</td>
<td>42</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>---MADNUDLLDYDDEEQET---TAVENQ--EAPKDVQINTVSI</td>
<td>38</td>
</tr>
<tr>
<td>C. elegans</td>
<td>------MeEQLDLYEVEEQEIQ------DQPQEVGGGDARKTGTKYASI</td>
<td>38</td>
</tr>
<tr>
<td>S. pombe</td>
<td>--MAGAQEDLIDYEEELVQDQPA---ETTPAADTNENEGSDKGKSYYVI</td>
<td>48</td>
</tr>
<tr>
<td>S. cerevisiae</td>
<td>--MSHEGAEELVEYDSNEQIQDIAEAGETGATSANGDNHTLTAAAGDKGKSYYVI</td>
<td>59</td>
</tr>
<tr>
<td>H. sapiens</td>
<td>HSSGFRLKFLKPELLRAIDVCPEHSEPVQHECIPQAILGDVLQARSGMKATFVFLA</td>
<td>102</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>HSSGFRLKFLKPELLRAIDVCPEHSEPVQHECIPQAILGDVLQARSGMKATFVFLA</td>
<td>98</td>
</tr>
<tr>
<td>C. elegans</td>
<td>HSSGFRLKFLKPELRAIGDCPEHSEPVQHECIPQAILGDVLQARSGMKATFVFLA</td>
<td>98</td>
</tr>
<tr>
<td>S. pombe</td>
<td>HSTGFRLKFLKPELLRAITDSGCEHSEPVQHECIPQALIGDVLQARSGMKATFVFLA</td>
<td>108</td>
</tr>
<tr>
<td>S. cerevisiae</td>
<td>HSTGFRLKFLKPELRAIDCPEHSEPVQHECIPQATLGDVLQARSGMKATFVFLA</td>
<td>119</td>
</tr>
<tr>
<td>H. sapiens</td>
<td>TLQQLEPVTG--QSVSLVMCHTRELAFQISKEYERSKFNPKVAVFFGGLSKEDEVL</td>
<td>161</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>TLQQLEPDSNNCHLMVNLVCHTRELAFQISKEYERSKYNFTKAVVFFGMAIQLDEET</td>
<td>158</td>
</tr>
<tr>
<td>C. elegans</td>
<td>TLQQLEPVGD--EVSVLMCHTRELAFQISKEYERSKYLPKLVKAVVFFGMAIJKDEE</td>
<td>157</td>
</tr>
<tr>
<td>S. pombe</td>
<td>TLQQLEPVGD--EVSVLMCMELQAFQISKEYERSKYNPKVAVFFGMAIQLDEE</td>
<td>157</td>
</tr>
<tr>
<td>S. cerevisiae</td>
<td>TLQQLEPVGD--EVAVVNLVMCHTRELAFQISKEYERSKYNPKVAVFFGMAIQLDEE</td>
<td>157</td>
</tr>
<tr>
<td>H. sapiens</td>
<td>KK-----------NCPHIVVGTPGRILALARMKSLNLKHFILDECKLMEQCLMRDVQEIFRMTS</td>
<td>219</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>KS-----------GTPHIVVGTPGRILALIRNMKLNLKHLDECKMLDQRDVQEIFRMTS</td>
<td>216</td>
</tr>
<tr>
<td>C. elegans</td>
<td>AN-----------DCPHIVVGTPGRMLALARSGKLKLDKVKYFVLDECKMDQDVQEIFRMTS</td>
<td>215</td>
</tr>
<tr>
<td>S. pombe</td>
<td>KDKSRSHPHIVATPGNLVREKILVNSKVHLVEDCKMSVMSDVQIFRMTS</td>
<td>215</td>
</tr>
<tr>
<td>S. cerevisiae</td>
<td>KNKDTAPHIVATPGNKALVREKILVNSKVHLVEDCKMSVMSDVQIFRMTS</td>
<td>215</td>
</tr>
<tr>
<td>H. sapiens</td>
<td>HEKQVMFSATLSKIEIRFVKCPEQDPMETFYVDEETKTLHGQLQYYVKKEDNKNKLF</td>
<td>279</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>HGKQVMFSATLSDIRVPCKKMFQDPMETFYVDEETKTLHGQLQYYVKKEDNKNKLF</td>
<td>276</td>
</tr>
<tr>
<td>C. elegans</td>
<td>QPKQVMFSATLSKIEIRFVKCPEQDPMETFYVDEETKTLHGQLQYYVKKEDNKNKLF</td>
<td>273</td>
</tr>
<tr>
<td>S. pombe</td>
<td>PDQVSMFSATLSKIEIRFVKCPEQDPMETFYVDEETKTLHGQLQYYVKKEDNKNKLF</td>
<td>273</td>
</tr>
<tr>
<td>S. cerevisiae</td>
<td>RDQVSMFSATLSKIEIRFVKCPEQDPMETFYVDEETKTLHGQLQYYVKKEDNKNKLF</td>
<td>273</td>
</tr>
<tr>
<td>H. sapiens</td>
<td>LNDVQDRFVNVISLDFDE----IDISSYIQTR</td>
<td>428</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>LNDVQDRFVNVISLDFDE----IDISSYIQTR</td>
<td>428</td>
</tr>
<tr>
<td>C. elegans</td>
<td>LNSVQDRFDSITLEFPK--IDVSTYQIRG</td>
<td>245</td>
</tr>
<tr>
<td>S. pombe</td>
<td>LDRIQERFVNVISLDFDE----IDVSTYQIRG</td>
<td>245</td>
</tr>
<tr>
<td>S. cerevisiae</td>
<td>LDRIQERFVNVISLDFDE----IDVSTYQIRG</td>
<td>245</td>
</tr>
<tr>
<td>S. cerevisiae</td>
<td>LDRIQERFVNVISLDFDE----IDVSTYQIRG</td>
<td>245</td>
</tr>
</tbody>
</table>

sz15'

<table>
<thead>
<tr>
<th>Species</th>
<th>Amino Acid Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>H. sapiens</td>
<td>HSSGFRDFLLKPELRAIDVCPEHSEPVQHECIPQAILGDVLQARSGMKATFVFLA</td>
<td>102</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>HSSGFRDFLLKPELRAIDVCPEHSEPVQHECIPQAILGDVLQARSGMKATFVFLA</td>
<td>98</td>
</tr>
<tr>
<td>C. elegans</td>
<td>HSSGFRDFLLKPELRAIDVCPEHSEPVQHECIPQAILGDVLQARSGMKATFVFLA</td>
<td>98</td>
</tr>
<tr>
<td>S. pombe</td>
<td>HSTGFDFNLKPELRAIDDSGCEHSEPVQHECIPQAIIGDVLQARSGMKATFVFLA</td>
<td>108</td>
</tr>
<tr>
<td>S. cerevisiae</td>
<td>HSTGFDFNLKPELRAIDCPEHSEPVQHECIPQAIIGDVLQARSGMKATFVFLA</td>
<td>119</td>
</tr>
<tr>
<td>H. sapiens</td>
<td>TLQQLEPVGT--QSVSLVMCHTRELAFQISKEYERSKFNPKVAVFFGGLSKEDEVL</td>
<td>161</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>TLQQLEPDSNNCHLMVNLVCHTRELAFQISKEYERSKYNFTKAVVFFGMAIQLDEET</td>
<td>158</td>
</tr>
<tr>
<td>C. elegans</td>
<td>TLQQLEPVGD--EVSVLMCHTRELAFQISKEYERSKYLPKLVKAVVFFGMAIJKDEE</td>
<td>157</td>
</tr>
<tr>
<td>S. pombe</td>
<td>TLQQLEPVGD--EVSVLMCMELQAFQISKEYERSKYNPKVAVFFGMAIQLDEE</td>
<td>157</td>
</tr>
<tr>
<td>S. cerevisiae</td>
<td>TLQQLEPVGD--EVAVVNLVMCHTRELAFQISKEYERSKYNPKVAVFFGMAIQLDEE</td>
<td>157</td>
</tr>
<tr>
<td>H. sapiens</td>
<td>KK-----------NCPHIVVGTPGRILALARMKSLNLKHFILDECKLMEQCLMRDVQEIFRMTS</td>
<td>219</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>KS-----------GTPHIVVGTPGRILALIRNMKLNLKHLDECKMLDQRDVQEIFRMTS</td>
<td>216</td>
</tr>
<tr>
<td>C. elegans</td>
<td>AN-----------DCPHIVVGTPGRMLALARSGKLKLDKVKYFVLDECKMDQDVQEIFRMTS</td>
<td>215</td>
</tr>
<tr>
<td>S. pombe</td>
<td>KDKSRSHPHIVATPGNLVREKILVNSKVHLVEDCKMSVMSDVQIFRMTS</td>
<td>215</td>
</tr>
<tr>
<td>S. cerevisiae</td>
<td>KNKDTAPHIVATPGNKALVREKILVNSKVHLVEDCKMSVMSDVQIFRMTS</td>
<td>215</td>
</tr>
<tr>
<td>H. sapiens</td>
<td>HEKQVMFSATLSKIEIRFVKCPEQDPMETFYVDEETKTLHGQLQYYVKKEDNKNKLF</td>
<td>279</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>HGKQVMFSATLSDIRVPCKKMFQDPMETFYVDEETKTLHGQLQYYVKKEDNKNKLF</td>
<td>276</td>
</tr>
<tr>
<td>C. elegans</td>
<td>QPKQVMFSATLSKIEIRFVKCPEQDPMETFYVDEETKTLHGQLQYYVKKEDNKNKLF</td>
<td>273</td>
</tr>
<tr>
<td>S. pombe</td>
<td>PDQVSMFSATLSKIEIRFVKCPEQDPMETFYVDEETKTLHGQLQYYVKKEDNKNKLF</td>
<td>273</td>
</tr>
<tr>
<td>S. cerevisiae</td>
<td>RDQVSMFSATLSKIEIRFVKCPEQDPMETFYVDEETKTLHGQLQYYVKKEDNKNKLF</td>
<td>273</td>
</tr>
<tr>
<td>H. sapiens</td>
<td>LNDVQDRFVNVISLDFDE----IDISSYIQTR</td>
<td>428</td>
</tr>
<tr>
<td>D. melanogaster</td>
<td>LNDVQDRFVNVISLDFDE----IDISSYIQTR</td>
<td>428</td>
</tr>
<tr>
<td>C. elegans</td>
<td>LNSVQDRFDSITLEFPK--IDVSTYQIRG</td>
<td>245</td>
</tr>
<tr>
<td>S. pombe</td>
<td>LDRIQERFVNVISLDFDE----IDVSTYQIRG</td>
<td>245</td>
</tr>
<tr>
<td>S. cerevisiae</td>
<td>LDRIQERFVNVISLDFDE----IDVSTYQIRG</td>
<td>245</td>
</tr>
</tbody>
</table>

E245K
2.4. Discussion

piRNAs potently repress transposon activity during germline development, and thus play a critical role in maintaining the inherited genome complement (Malone and Hannon, 2009; Siomi et al., 2010). Heterochromatic clusters encode primary piRNAs, but most of the piRNA biogenesis machinery is concentrated in the perinuclear nuage, which is closely associated with the cytoplasmic face of nuclear pores (Eddy, 1974; Eddy, 1975; Klattenhoff and Theurkauf, 2007; Lim and Kai, 2007). How transcripts are directed from clusters to the perinuclear piRNA biogenesis machinery, and how gene transcripts are excluded from this machinery, are not understood. We show that the DEAD box protein UAP56 co-localizes with the cluster-associated HP1 homolog Rhi, and that prominent foci containing both Rhi and UAP56 are closely associated with the nuclear periphery, directly opposite nuage foci containing Vas (Figure 2.1). Strikingly, an E245K mutation that prevents UAP56 co-localization with Rhi (Figure 2.4) also disrupts nuage localization of piRNA biogenesis proteins (Figure 2.10), transposon silencing (Figure 2.14), and piRNA production by the dual strand clusters (Figures 2.16 and 2.18). This mutation also increases ectopic piRNAs from protein coding genes (Figure 2.19). UAP56 thus has a previously unrecognized role in germline piRNA biogenesis and transposon silencing. Intriguingly, studies in C. elegans indicate that Vas-containing p-granules, which appear to be equivalent to nuage, form a size exclusion zone that extends the nuclear pore (Updike et al., 2011). We therefore propose that UAP56 functions with Vas to organize a piRNA processing compartment that spans the nuclear pore and increases the efficiency and specificity of piRNA production by coordinately directing
cluster transcripts to processing factors in the nuage, and excluding gene transcripts from these factors (Figure 2.25).

DEAD box proteins appear to function as ATP-dependent RNA clamps (Linder and Jankowsky, 2011), and transcripts from the major dual strand cluster immunoprecipitate with both UAP56 and Vas (Figure 2.22C and D). Intriguingly, UAP56 and Vas are closely associated with opposite faces of nuclear pore complexes (Figure 2.1, Figure 2.7), and nucleoporins stimulate ATP hydrolysis and RNA release by related DEAD box proteins (Montpetit et al., 2011). UAP56 binding to nuclear pores could therefore trigger release of cluster transcripts, which are bound by the DEAD box protein Vas in the nuage. We speculate that Vas direct these RNAs to the primary piRNA processing machinery and to Aub, which binds primary piRNAs and co-purifies with Vas (Lim and Kai, 2007; Malone et al., 2009).

The majority of piRNA pathway proteins are germline specific, but UAP56 is an ubiquitously expressed essential gene. The uap56\textsuperscript{sz15} allele thus reveals a link between germline-specific piRNA pathway components and the general RNA splicing and export machinery. It is interesting to note that mutations in the Drosophila mago nashi and tsunagi/Y14 genes, which encode conserved exon junction components, lead to axis specification defects that are strikingly similar to the defects associated with the uap56\textsuperscript{sz15} allele and other piRNA pathway mutants (Micklem et al., 1997; Mohr et al., 2001; Newmark and Boswell, 1994; Newmark et al., 1997). This may reflect a function for the splice junction complex, with UAP56, in piRNA production and transposon silencing.
**Figure 2.25. UAP56 function in a piRNA processing compartment.**

A model for UAP56 function within piRNA processing compartment that spans the nuclear envelope. (1) UAP56 associates with nascent transcripts from dual-strand piRNA clusters, which bind the HP1 homolog Rhi. We speculate that UAP56 interacts with clusters through a domain defined by residue 245E. (2) UAP56-cluster transcript complexes interact with the nuclear pore, which triggers RNA release and export. (3) Vas binds cluster transcripts within the pore, or as they emerge from the pore and enter nuage, and then delivers these RNAs to the processing machinery.
Figure 2.25

Nucleus dual-strand piRNA cluster

Nuage
2.5. Experimental Procedures

Fly stocks and transgenic lines

All the stocks and crosses were raised at 25˚ C on cornmeal medium using standard conditions. OregonR (OreR) and w^{1118} were used as control strains. The uap56^{sz15}/CyO, uap56^{28}/CyO and UASp-UAP56-GFP stocks were obtained from Ilan Davis (University of Oxford). uap56^{Df}/CyO, vas^{D5}/CyO, vas^{PH}/CyO, rhi^{2} (rhi^{02086})/CyO and rhi^{KG} (rhiKG00910)/CyO were obtained from Bloomington Stock Center.

The UASp-UAP56-GFP transgene was constructed using a full length UAP56 cDNA from the Gold collection (BDGP23644). Coding sequences were recovered by PCR amplification using a sense primer (5′-GGTACCATGGCCGACAATGACGATC-3′) that spanned the translation initiation codon, and an antisense primer that introduced a unique Nhe1 restriction site upstream of the stop codon (5′-GCTAGCGCGTCCCTCAATGAATGTAG-3′). The PCR product was cloned into pGemT (Promega) and sequenced. 2 silent mutations at position 22 (GAA to GAG) and 87 (TCC to TCT) of the protein were detected. The UAP56 cDNA was fused to the N-terminus of GFP (from vector KS-GFP kindly provided by A. Vincent). A KpnI-BamHI fragment encoding the UAP56-GFP fusion was cloned downstream of the promoter in the UASp transformation vector (Rorth, 1998). Transgenic lines were obtained in w^{1} background, and transgene expression was induced using the germline-specific nanos Gal4 vp16 driver.

The UAP56-Venus transgene was engineered as follows: The UAP56 promoter and 5′ UTR (695bp) were cloned upstream of a full-length cDNA, which was fused in frame to a multifunction affinity purification tag followed by Venus (Ma et al., 2012).
The fusion was followed by \textit{uap56} 3’ UTR (1255bp). Site directed mutagenesis was used to generate a G to A substitution at position 733 of the \textit{uap56} coding sequence, producing the E245K substitution found in the \textit{uap56}^{sz15^-} allele. The wild type and \textit{uap56}^{sz15^-} fusion genes were introduced into the attB vector, and independent transformants were obtained by site-direct phiC31 integration into chromosomal locus 68A4 (Ni et al., 2008).

**Immunohistochemistry**

Ovaries were immunolabeled using the Buffer A staining protocol, as described earlier (Klattenhoff et al., 2009). The following primary antibodies and dilutions were used: rabbit anti-Aub (1:1000) (Brennecke et al., 2007), rabbit anti-Ago3 (1:250) (Li et al., 2009), rabbit anti-Piwi (1:1000), rabbit anti-UAP56 (1:1000) (Eberl et al., 1997), guinea pig anti-Rhi (1:500) (Klattenhoff et al., 2009), and rabbit anti-Vas (1:5000) (Liang et al., 1994). Anti-UAP56 and anti-Vas were gifts from P. Lasko. A rabbit polyclonal antibody against γ-H2Av (1:500, Rockland) was used to detect DNA DSBs. A mouse monoclonal antibody against nuclear pore complex (1:1000, Covance) was used to label nuclear envelope. TOTO-3 dye (Molecular Probes) was used at 1:500 to label DNA after RNase One (1:500, Promega) treatment.

**Total RNA isolation and array analyses**

Total RNA was extracted from 2-4 day old ovaries from OreR, \textit{uap56}^{sz15^-}/\textit{uap56}^{28}, \textit{rhi}^2/\textit{rhi}^{KG} and \textit{vas}^{D5}/\textit{vas}^{PH} females using the RNeasy Kit (Qiagen) and the manufacturer’s
instructions. Tilling array analysis was performed as described previously (Klattenhoff et al., 2009). Data can be accessed through GEO reference series GSE35638.

**Small RNA sequencing**

Total RNA was extracted from ovaries dissected from 2-4 day old females using MirVana kit (Ambion). 18-29 nt small RNAs were gel purified following 2S rRNA depletion and libraries were prepared as described previously (Klattenhoff et al., 2009). Sequencing was performed using a Solexa Genome Analyzer (Illumina, San Diego, CA). Small RNA sequence analysis was performed as described previously (Klattenhoff et al., 2009; Li et al., 2009). Data can be accessed through GEO reference series GSE35638.

**Strand-specific Reverse Transcription qPCR**

Strand-specific RT-qPCR for cluster transcripts was performed as described previously (Klattenhoff et al., 2009). Signals were normalized to 18s rRNA after subtracting no RT primer background.

**RNA Immunoprecipitation and Strand-specific RNA Sequencing**

Whole ovaries were dissected from 2-4 day old flies in 1×PBS Buffer and homogenized on ice in Lysis Buffer (50mM Tris-HCl, pH7.5, 150mM NaCl, 1mM EDTA, 0.5% NP-40) with 40U/ml RNasin Plus (Promega N261), 1× Proteinase Inhibitor Cocktail (Sigma) and 1mM PMSF. The lysate was sonicated three times for 5 seconds each at 35% amplitude using a Diagenode Bioruptor, and then centrifuged at 14,000rpm for 15min at 4°C in a table top microfuge. 10% of the supernatants were saved as input.
samples and the rest were incubated with GFP Trap-A beads (Chromotek) for 2hr at 4°C. The beads were separated from the supernatants after incubation and washed three times with Lysis Buffer at room temperature. UAP56-Venus and sz-Venus were immunoprecipitated by ANTI-FLAG® M2 Affinity Gel (Sigma A2220) and eluted with 200µg/ml FLAG® Peptide (Sigma F3290). Three biological replicates were performed for each IP. Total RNA was extracted from the input lysate and the beads using RNeasy Micro Kit (Qiagen) using the manufacturer’s protocol. Strand-specific RNA sequencing was performed after depletion of ribosomal RNAs using Ribo-Zero (Epicenter Inc) by a modification of the procedure of Parkhomchuk et al. (Parkhomchuk et al., 2009) (Parkhomchuk et al., 2009), described in detail elsewhere (Zhang et al., 2012). Data can be accessed through GEO reference series GSE35638.

Quantification of transposon transcripts levels by RT-qPCR

Total RNA was prepared from whole ovaries using RNeasy Mini Kit (Qiagen#74104). Oligo(dT)$_{20}$ primer was mixed with an 18s rRNA-specific control primer and first strand cDNAs were synthesized using SuperScript™III reverse transcriptase (Invitrogen#18080-093) following the manufacture's protocol. The resulting cDNAs were used as templates for quantitative real-time PCR using the primers indicated (Table 2.2). qPCR reactions were performed using a StepOnePlus™ System (Applied Biosystems) and SYBR Green I (Qiagen#204145). The expression level of transcripts was measured relative to the 18s rRNA internal control. Background obtained with no RT primer reactions was subtracted. Three technical replicates were performed for each RT primer. Graphs show the average and standard deviation (SD).
Table 2.2 Primer sequences.

<table>
<thead>
<tr>
<th>RT primers</th>
<th>name</th>
<th>sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>18s rRNA-rt</td>
<td>GATACTACCATCAAAAGTTGATAGG</td>
</tr>
<tr>
<td></td>
<td>rp49-rt</td>
<td>GGAGGAGACGCCG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>qPCR primers</th>
<th>name</th>
<th>sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>18s rRNA-f</td>
<td>CACACGAACATGAACCTATGGGACGTGTG</td>
</tr>
<tr>
<td></td>
<td>18s rRNA-r</td>
<td>TCGGTACAAGACCATAACGATCTGC</td>
</tr>
<tr>
<td></td>
<td>pre-rp49-f</td>
<td>GCCAATTGTACCCGTGTTTT</td>
</tr>
<tr>
<td></td>
<td>pre-rp49-r</td>
<td>GTTCGATCCGTAACCGATGT</td>
</tr>
<tr>
<td></td>
<td>HeT-A-1-f</td>
<td>CAGCCCAAATCTGAGAGAGG</td>
</tr>
<tr>
<td></td>
<td>HeT-A-1-r</td>
<td>CCACCTGTAGCTGTTACG</td>
</tr>
<tr>
<td></td>
<td>HeT-A-2-f</td>
<td>GTGATGCGGAACAGAATTT</td>
</tr>
<tr>
<td></td>
<td>HeT-A-2-r</td>
<td>TGCTCGTAAATAATCCCGC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Primers for northern probes</th>
<th>name</th>
<th>sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HeT-A-f</td>
<td>gatttaggtgacactatatag GACACGCGAAAGCGAAGA</td>
</tr>
<tr>
<td></td>
<td>HeT-A-r</td>
<td>cgtaatagctcactatatagg GGAGAAGATCGCTGTTCTG</td>
</tr>
<tr>
<td></td>
<td>Blood-f</td>
<td>gatttaggtgacactatatag CGCAGATGACAACAGCAGAATGTGTCT</td>
</tr>
<tr>
<td></td>
<td>Blood-r</td>
<td>cgtaatagctcactatatagg GCAGAACCACGCGACACAGA</td>
</tr>
<tr>
<td></td>
<td>Rp49-f</td>
<td>CCAAGCAGTCATCCGACGCCAGAGTC</td>
</tr>
<tr>
<td></td>
<td>Rp49-r</td>
<td>cgtaatagctcactatatagg TCCGACCACGTTACAAGAACTCCTCA</td>
</tr>
</tbody>
</table>
Image analysis

The ImageJ WCIF plugin (http://imagej.nih.gov/ij/, 1997-2011) was used to quantify UAP56 and Rhi co-localization, following sequential imaging of double-labeled egg chambers. The threshold of each channel was determined by background analysis based on the approach of Costes et al. (Costes et al., 2004). The zero-zero pixels are included in calculated thresholds. $R$ is the Returns Pearson’s correlation coefficient for pixels, where both channels are above their respective threshold. The percentage of colocalization was calculated for each protein by dividing the colocalizing pixel intensities by the total pixel intensity above threshold. Foci in 39 independent nuclei from 26 egg chambers were analyzed.

The correlation between Rhi foci in the nucleus and Vas foci in nuage was quantified as follows: Egg chambers were labeled for Rhi, Vas and the nuclear pore, and images were acquired under identical imaging conditions and analyzed using ImageJ (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://imagej.nih.gov/ij/, 1997-2011.). Background subtraction and signal threshold values were set as the mean image profile value plus three standard deviations. Rhi foci overlapping with the Nuclear Pore Complex were identified and signal was quantified by multiplying the area of the focus by mean intensity. The signal associated with Vasa foci directly opposite Rhi foci were then quantified using the same procedure.

Other procedures

Western blots were quantified using an Odyssey Infrared Imaging System (LI-COR). Rabbit polyclonal anti-UAP56 antibody was used as 1:1000, and the mouse
monoclonal anti-actin antibody (DSHB, JLA20) was used as 1:20. Northern blots were performed as described previously (Klattenhoff et al., 2007).
Chapter 3

UAP56 suppresses piRNA precursor splicing in the Drosophila germline

3.1. Introduction

*Drosophila* piRNAs are essential for germline maintenance and transposon silencing (Siomi et al., 2011). These small silencing RNAs are derived from discrete genomic loci composed of TEs and TE remnants. Mapping piRNAs from germline tissue defined over 100 of these genomic loci, termed piRNA clusters. Based on whether there are piRNAs mapping to both genomic strands or only one strand, the clusters are classified as dual-strand clusters or uni-strand (Brennecke et al., 2007). Previous work from our lab showed that Rhi, an HP1 homolog, is required for dual strand cluster expression and binds to the 42AB dual-strand cluster, suggesting that it mediates epigenetic regulation of cluster expression (Klattenhoff et al., 2009). Transcription of piRNA clusters is largely uncharacterized, although studies in BmN4 cells and mouse testis suggest that clusters are transcribed by RNA polymerase II and produce long single stranded precursors (Kawaoka et al., 2013; Li et al., 2013). A recent study in *Drosophila* showed that Cuff, a protein related to the yeast transcription termination factor Rai1, localizes to the nucleus, interacts with Rhi, and is required for piRNA cluster expression (Pane et al., 2011). After transcription, pre-piRNAs are delivered to nuage, a perinuclear structure at the cytoplasmic face of the nuclear envelope, where the piRNA processing machinery appears to process the precursors into mature piRNAs (Mahowald, 1972). However, there are many missing links in piRNA pathway, and particularly in the nuclear
events of piRNA biogenesis. How the precursors are transcribed? Is there co-
transcriptional processing?

In Chapter 2, I described a novel function for UAP56 in piRNA biogenesis in the
Drosophila germline. UAP56 co-localizes with Rhi at dual strand piRNA clusters in
nucleoplasm, binds to piRNA precursors and facilitates piRNA production from these
loci (Figure 2.1, 2.16 and 2.22). I proposed that UAP56 couples pre-piRNA transcription
and nuclear export to the perinuclear nuage, which facilitates processing into mature
piRNAs (Figure 2.25). Previously, UAP56 was known as a conserved RNA splicing and
export factor in the protein coding gene expression pathway. By comparing piRNA
precursors expressed in uap56 mutant with wild type, and examining UAP56 binding to
transcripts that map to the clusters, I have observed preliminary data suggesting that
uap56 suppresses piRNA precursor splicing. These findings suggest a possible
mechanism for UAP56 mediated piRNA biogenesis.

3.2. Results

UAP56 suppress piRNA precursors splicing

As shown in Chapter 2, UAP56 is required for piRNA production from dual-
strand clusters in germline (Figure 2.18). The E245K substitution in uap56 specifically
disrupts germline piRNA production and leads to TE overexpression, and RNA
immunoprecipitation (RIP) showed that UAP56 interacts with long single stranded
piRNA precursors from the clusters (Figure 2.14, 2.16 and 2.22). However, piRNA
precursors were primarily examined for only a limited number of chromosomal regions
by RT-qPCR (Figure 2.20). To examine genome-wide piRNA precursor expression and
processing, I used non-PolyA selected strand-specific RNA sequencing to profile the total RNA expression from wild type and *uap56* mutant ovaries. An example of the steady state levels of RNA expression patterns in wild type and *uap56* mutants is shown in the Genome Browser screen shot (Figure 3.1). Consistent with the results from whole genome tiling array, I found no difference in mRNA levels (Figure 3.1A). However, I observed an accumulation of precursor transcripts from dual strand clusters (see cluster1/42AB and cluster 3; Figure 3.1A and 3.2A). As described in Chapter 2, I observed a reduction of precursor levels at the 42AB cluster by RT-qPCR, but my RNA-Seq data indicate that those two small regions are not representative of expression across the whole cluster (Figure 3.1A).
Figure 3.1. *uap56* mutations lead to piRNA precursors accumulation at 42AB cluster.

Non-polyA selected strand specific RNA sequencing were used to assay gene and transposon expression in wild type and *uap56^{ts15} / uap56^{28}* ovaries. A. A Genome Brower screen shot showing piRNA reduction and precursor accumulation at 42AB dual strand cluster in *uap56* mutants. The profile of uniquely mapping piRNAs on plus and minus strand in wild type (wt) (dark blue track) and *uap56* mutants (light blue track) shows that piRNA level is decreased in *uap56*. The unique mapping pre-piRNA profile shows higher precursor levels in *uap56* mutants (light purple track) relative to wild type (dark purple track). The approximate positions of qPCR primer pairs used in strand specific RT-qPCR reaction to quantify precursor transcripts are indicated in red bars. B. A Genome Brower screen shot showing the mRNA expression is unchanged in wt and *uap56* mutants. Euchromatic genes, *ppd3* and CG16908 are located on plus and minus strand respectively, and majority of the precursor reads map to exon of the two genes on the corresponding strand. All the tracks are scaled the same for wt and *uap56* mutants.
Figure 3.1

A

- piRNA
- uap
- pre-piRNA
- uap

B

- wt
- pre-mRNA
- uap
Cluster 3 is located in the telomeric region of chromosome 4 and showed a dramatic increase in precursor levels in *uap56* mutants (Figure 3.2A). Cluster 3 contains tandem arrays of repetitive sequences derived from the transposable element TART, which are exclusively oriented on the minus strand. Small RNA sequencing revealed that piRNAs map to both strands of the cluster 3, and *uap56* mutants deplete these mature piRNAs (Figure 3.2A). Compared with wild type, the level of precursor transcripts from both strands were elevated in *uap56* mutant, but plus strand transcripts were much more abundant than transcripts from the opposite strand (Figure 3.2A). And strikingly, the reads distribution in mutant showed clearly defined gaps resembling the pattern produced by splicing of protein-coding gene transcripts (Figure 3.2B). Higher resolution views on the genome browser showed that most reads map to the genome up to consensus splice donor sites, and that homology to the genome resumes precisely at consensus splice acceptor sites. The majority of reads thus span a gap that appears to correspond to an intron, and only a few reads mapping to the intron (Figure 3.2C). In wild type samples, by contrast, very few reads showed gaps corresponding to the putative intron.

To confirm the accumulation of spliced transcripts, I used RT-qPCR with primer sets spanning two gaps in the read mapping pattern in cluster 3 (Figure 3.2B). Consistent with the RNA-Seq results, the level of spliced transcript increased at both locations in *uap56* mutants, with transcripts from the plus strand at much higher levels than the minus strand (Figure 3.2D). These results suggest that wild type UAP56 suppresses processing of precursor transcripts from piRNA clusters.
Figure 3.2. *uap56* mutations disrupt piRNA production and cause precursor accumulation at dual strand cluster 3.

A. A Genome Brower screen shot showing piRNA reduction and precursor accumulation in *uap56* mutations at cluster 3. Dual strand piRNA cluster 3 is located at the telomeric end of chromosome 4 with transposon TART sequences on minus strand (brown bars). piRNAs mapped to both strand in wild type are dramatically decreased in *uap56* mutants. pre-piRNAs mapped to plus strand are highly accumulated, while the minus strand mapping ones show a moderate increase in *uap56* mutants compared with wild type. All the tracks are scaled the same for wt and *uap56* mutants.  

B. A zoom-in view of pre-piRNA distribution in *uap56* mutant over chromosomal region indicated in A. The accumulated piRNA precursors show gaps resembling the pattern produced by splicing of protein-coding gene transcripts. The positions of qPCR primer pairs spanning two introns are indicated in red.  

C. A zoom-in view of a reads mapping gap indicated in B. Majority of reads span the clearly defined gap resemble intron with a few mapping into the intron.  

D. Strand specific RT-qPCR quantification of piRNA precursors spanning the two introns indicated in B. Both show an increase in precursor levels in *uap56* mutants. All signals are normalized to ribosomal RNAs. Error bars represent SE from 3 biological replicates.
Figure 3.2

A

\[ \begin{align*} 
\text{piRNA} & \quad \text{wt} \quad + \\
& \quad - \\
\text{uap} & \quad + \\
& \quad - \\
\text{pre-piRNA} & \quad \text{wt} \\
& \quad + \\
& \quad - \\
\text{uap} & \quad + \\
& \quad - \\
\end{align*} \]

B

C

D

\[ \begin{align*} 
\text{pre-RNA expression} & = \begin{cases} 
3.50E-05 & \text{wt} \\
3.00E-05 & \text{uap-/-} \\
2.50E-05 & \text{cl3-1} \\
2.00E-05 & \text{cl3-2} \\
1.50E-05 & \text{cl3-3} \\
1.00E-05 & \text{cl3-4} \\
5.00E-06 & \text{cl3-5} \\
0.00E+00 & \text{cl3-6} \\
-5.00E-06 & \text{cl3-7} \\
\end{cases} 
\end{align*} \]
The E245K substitution disrupts UAP56 binding to the spliced piRNA precursors

In Chapter 2, the association of UAP56 with piRNA precursors was characterized by RIP followed by RNA sequencing. piRNA precursors were highly enriched in wild type UAP56 IP sample, while E245K substitution led to a modest reduction of precursor binding (Figure 2.22). In light of the finding that precursor splicing increases in uap56 mutants, I have further investigated UAP56 bound piRNA precursors, identified by RIP-Seq as described in Chapter 2. piRNA precursors that were immunoprecipitated with a Flag tagged wild type or mutant UAP56 fusion proteins mapping to a limited region of cluster 3 regions are shown in Figure S3. Flag IP from a wild type Drosophila strain that does not express fusion protein was used as a negative control. Consistent with previous findings (Figure2.22), transcripts from this region of cluster 3 were highly enriched in wild type UAP56 IP, while the uap56sz15 mutant IP showed lower enrichment (Figure 3.3). This accumulation of cluster 3 precursor transcripts in RIP samples was confirmed by strand specific RT-qPCR using primer sets shown in Figure 3.2B (Figure 3.4A). In addition, reads from wild type UAP56 RIP-seq showed the distinct gapped pattern characteristic of splicing (Figure 3.3). To determine if this pattern was due to genomic deletions in the UAP-Venus and sz-Venus strains, genomic DNA was PCR amplified by primers spanning the “intron” in cluster 3, and the products were analyzed by gel electrophoresis. In both of the strains, PCR indicated that the genomic sequences at cluster 3 were intact (Figure 3.4B). To confirm the splicing pattern of the precursor transcripts in RIP sample, pre-piRNAs were amplified by conventional RT-PCR followed by gel electrophoresis. A single RNA fragment consistent with removal of the intron was detected in each IP sample, and Sanger sequencing confirmed that the RT-PCR product
matched to the cluster 3 loci in an interruption corresponding the the intron (Figure 3.4C). Taken together, these results suggest the UAP56 associates with spliced piRNA precursors in *Drosophila* germline, and this association is partially disrupted by E245K substitution.
Figure 3.3. Cluster 3 piRNA precursors bind to UAP56.

The transcripts immunoprecipitated with UAP56-Venus or sz-Venus were analyzed by strand specific RNA-seq. The IP in the genetic background strain (wt) was used as control. A Genome Brower screen shot shows the UAP-Venus input signal in light brown, the UAP-Venus IP signal in orange, the sz-Venus input signal in light green, the sz-Venus IP signal in dark green, wt control input signal in blue, and control IP signal in dark blue. All signals are normalized to ribosomal RNAs. The cluster 3 precursor transcripts were highly enriched over input by wild type UAP56-Venus IP, and sz-Venus IP show a decreased but higher than background enrichment of pre-piRNAs. Transcripts immunoprecipitated with UAP56-Venus or sz-Venus show a splicing pattern.
Figure 3.3

wt
IP
input
UAP56
-V
IP
input
sz-V
IP
Figure 3.4. Spliced piRNA precursor transcripts from cluster 3 bind to UAP56.

A. Strand specific RT-qPCR quantification of piRNA precursors bound to UAP56-Venus and sz-Venus using the primer sets spanning the two “introns”. Both show an increase in precursor levels in UAP56-Venus IP. All signals are normalized to ribosomal RNAs. Error bars represent SE from 3 technical replicates. B. Genomic DNA from UAP56-Venus and sz-Venus fly ovaries was PCR amplified by primers spanning two “introns” in cluster 3, and the products were analyzed by gel electrophoresis. In both of the strains, 186bp or 207bp PCR product amplified by primer sets cl3-1 or cl3-2 was detected, reflecting intact genomic loci at cluster 3. C. The pre-piRNAs bound to UAP56-Venus/sz-Venus were amplified by RT-PCR using primers spanning “introns” followed by gel electrophoresis. A single 125bp or 115bp PCR product was detected by primer sets cl3-1 or cl3-2. It is consistent with removal of the intron from genomic sequences.
Figure 3.4

A

![Graph showing pre-RNA levels for different conditions.](image)

- **RNA levels**
- **Input**
- **IP**

B

<table>
<thead>
<tr>
<th>cl3-1</th>
<th>cl3-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>UAP56-V sz-V</td>
<td>M</td>
</tr>
<tr>
<td>UAP56-V sz-V</td>
<td>M</td>
</tr>
</tbody>
</table>

![Gel electrophoresis images showing bp](image)

C

<table>
<thead>
<tr>
<th>cl3-1+</th>
<th>cl3-1-</th>
</tr>
</thead>
<tbody>
<tr>
<td>w1 UAP-V sz-V</td>
<td>w1 UAP-V sz-V</td>
</tr>
<tr>
<td>Input</td>
<td>Input</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>cl3-2+</th>
<th>cl3-2-</th>
</tr>
</thead>
<tbody>
<tr>
<td>w1 UAP-V sz-V</td>
<td>w1 UAP-V sz-V</td>
</tr>
<tr>
<td>Input</td>
<td>Input</td>
</tr>
</tbody>
</table>

![Additional gel electrophoresis images](image)
3.3. Discussion

In chapter 2, I showed that UAP56 is an essential nuclear protein for piRNA biogenesis that binds to pre-piRNAs during an intermediate step between cluster transcription and precursor processing (Figure 2.1, 2.16 and 2.22). In this chapter, I detected an accumulation of piRNA precursors with striking splicing patterns in uap56 mutants, suggesting UAP56 suppresses cluster transcript processing (Figure 3.2). This observation contrasts with the known function of UAP56 as a spliceosome component that facilitates mRNA splicing (Shen, 2009). Since our mutation does not effect mRNA expression, this splicing suppressor function of UAP56 is specific to the piRNA pathway (Figure 3.1). During protein coding gene expression, UAP56 is recruited to mRNA by U2AF65 and stabilize the interaction of U2 snRNA and pre-mRNA branch point (Fleckner et al., 1997). My RIP-sequencing results showed enrichment for mRNA of about 2 fold, and most of immunoprecipitated coding gene transcripts were spliced mRNAs (Figure 2.22B And 3.1). In the piRNA pathway, UAP56 also binds the spliced form of piRNA precursors, and this interaction is reduced in E245K mutants, suggesting UAP56 is required for the stable interaction with spliced piRNA precursors (Figure 3.3). It is still unclear why UAP56 binds to these spliced piRNA precursors. However, the stability of UAP56 RNP seems to have an inversed relationship with RNA splicing activity, with wild type UAP56/pre-piRNP more stable than sz15'/pre-piRNP. One possible explanation is that the stable association of UAP56 with pre-piRNAs leads to mature piRNA production rather than finishing splicing and producing mRNAs. This mechanism could possibly differentiate piRNA precursor transcripts from protein coding genes transcripts and process them differently. However, this is very speculative, and
Genome-wide analyses and more experimental tests are needed to confirm this hypothesis.

These observations provide an intriguing link between protein coding gene processing and the piRNA pathway. Interestingly, multiple exon-junction components (EJCs) were identified in a screen for germline TE silencing factors (Czech et al., 2013). In addition, females mutants for the EJC genes *mago nashi* and *tunagi/Y14* lay ventralized egg, a phenotype that similar to mutations in *uap56* and other piRNA genes (Micklem et al., 1997; Mohr et al., 2001; Newmark and Boswell, 1994; Newmark et al., 1997). These splicing factors may function with UAP56 to control piRNA precursor processing.

### 3.4. Experimental Procedures

**Total RNA isolation and non-polyA selected strand-specific RNA Sequencing**

Total RNA was extracted from 2-4 day old ovaries from OreR and *uap56^{sz5}/uap56^{28}* females using the RNeasy Kit (Qiagen) and the manufacturer’s instructions. After depletion of ribosomal RNAs using Ribo-Zero (Epicenter Inc), non-polyA selected strand-specific RNA sequencing was performed as described (Zhang et al., 2012).

**Strand-specific Reverse Transcriptase qPCR**

Strand-specific RT-qPCR for cluster transcripts was performed as described previously (Klattenhoff et al., 2009). Primers are listed in Table S1. Signals were normalized to 18s rRNA after subtracting no RT primer background.
### Table 3.1 Primer Sequences

#### RT primers

<table>
<thead>
<tr>
<th>name</th>
<th>sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>cl3-1-plus</td>
<td>CAGGATAGCAGCAATGTC</td>
</tr>
<tr>
<td>cl3-1-minus</td>
<td>GATCACCTCGTCTTCTC</td>
</tr>
<tr>
<td>cl3-2-plus</td>
<td>TACTTTAAACGTTTCTCCTG</td>
</tr>
<tr>
<td>cl3-2-minus</td>
<td>AAAGTGTCACATGGAGTAG</td>
</tr>
</tbody>
</table>

#### qPCR primers

<table>
<thead>
<tr>
<th>name</th>
<th>sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>cl3-1-F</td>
<td>TGCTTTGCACTCAGATTGG</td>
</tr>
<tr>
<td>cl3-1-R</td>
<td>TTTCTGCCAGCTGTAAATG</td>
</tr>
<tr>
<td>cl3-2-F</td>
<td>AGCGAAATGTTTCAGCAAACC</td>
</tr>
<tr>
<td>cl3-2-R</td>
<td>CATCTGCTGCTGGAATGG</td>
</tr>
</tbody>
</table>
Chapter 4

General Discussion

4.1 Introduction

piRNAs and the pathway play essential roles in maintain genome integrity in animal gonad. As a relatively new member of small non-coding RNAs, piRNA biogenesis and function are still under intense investigation. In the past decade, more and more new components were identified and added to the piRNA pathway. However, most of them are found to function in the cytoplasmic processing, and very few progresses have been made to understand the piRNA pathway in nucleus. The genomic origin study showed piRNAs are mostly derived from discrete heterochromatin loci and some of the piRNA pathway components, e.g. Piwi and Rhi, may regulate chromatin status through histone modification. These suggest the importance of the nuclear regulation of piRNA pathway. And the recent findings of de novo piRNA production from the foreign genomic fragments inserted into piRNA clusters further support the chromatic context rather than genomic sequence itself is essential for piRNA production.

The main function of piRNAs is defense against TE over-activation in germline. However, how piRNAs target TEs is still an open question. Although the Ping-Pong amplification cycle proposed a very attempting silencing model based on the sequence complimentary between piRNAs and TEs, there are evidences suggest this may not be the whole case. For example, in most piRNA mutants, suppressed piRNAs and over-activated TEs often do not show a good sequence correlation. This suggests the Ping-Pong cycle may not be the main silencing mechanism. Meanwhile, a good portion of the piRNAs antisense to TEs are associated with Piwi proteins and shuttled back into nucleus
after maturation from precursor transcripts. Although the detailed nuclear TE-targeting by Piwi is still elusive, it is possible that rather than repressing individual TE gene Piwi may suppress the whole TE-rich locus by altering the chromatin modification, which could lead to much broader effect than direct targeting homologous sequences.

During my thesis study, I discovered the novel function of UAP56 as a new nuclear component of piRNA pathway, and my results suggest that this DEAD box protein functions in connecting piRNA precursor transcription to nuclear export and processing in the nuage. These findings contribute to understanding the epigenetic regulation of nuclear piRNA biogenesis, and establish a link between the communications between two different cellular compartments. And since UAP56 has a pre-identified role in protein-coding gene expression, my findings suggest a cross talking between the protein-coding gene pathway and piRNA pathway, and proposed a model for increase specificity and efficiency of RNA processing.

4.2 piRNA cluster transcription

piRNAs are produced from long single stranded precursors from the discrete genomic loci. However, the size of the transcription units and nature of the transcription machinery remain open questions. In Drosophila, our strand specific RT-qPCR studies detects precursors of 300-500bp, and our random primed strand specific RNA sequencing indicates that the primary transcripts are likely to be much longer, but we have not been unable to assemble de novo complete transcripts from any of the major clusters. However, the major piRNA producing regions are now well-defined and it may be possible to computationally define sequence motif linked to cluster transcription, and new
sequencing approaches that generate long reads may allow assembly of transcripts from the repeat region within these loci.

The piRNAs that uniquely map to clusters show a significant anti-sense bias relative to the transposon fragments they are comprised of (Brennecke et al., 2007). Because the transposon fragments that make up dual strand clusters are oriented in both directions, the bias in piRNA accumulation over clusters switches between genomic strands. It was previously unclear if this reflected a bias in precursor accumulation or a bias in processing. My RNA sequencing studies detected piRNA precursor transcripts from both strands of the dual clusters, with relatively little asymmetry with respect to genome strand (Figure 2.21B). This observation suggests that the bias in piRNA accumulation is due to differences in precursor processing, not precursor production. It has been proposed that the observed piRNA bias is generated through interactions between cluster specific piRNAs and sense strand transposon transcripts during the Ping-Pong cycle, consistent with my observations (Brennecke et al., 2007). However, this has not been tested and other post-transcriptional mechanisms are possible.

Recent studies in BmN4 cells, mouse testis and worms identified 5’-Caped and 3’-polyA tailed piRNA precursors, ChIPseq reveals Pol II association with piRNA clusters (Gu et al., 2012; Kawaoka et al., 2013; Li et al., 2013). These findings suggest piRNA pathway utilizes the same transcription machinery as protein coding genes. By contrast, *Drosophila* piRNAs are produced predominantly from clusters and repeated elements that match cluster sequences (Brennecke et al., 2007). Then, how are gene and cluster transcripts distinguished? *Drosophila* piRNA clusters are generally localized to pericentromeric and subtelomeric heterochromatin regions, raising the possibility that
chromatin context plays critical a role. However, many heterochromatic regions, including protein coding loci, do not produce piRNAs.

Rhi is a germline specific HP1 homolog that may epigenetically mark piRNA clusters. Mutations in rhi disrupt piRNA production by germline-specific dual-strand clusters, and ChIP-qPCR indicates that Rhi binds to the major dual strand cluster at 42AB (Klattenhoff et al., 2009). Intriguingly, the RNA binding protein UAP56 colocalizes with Rhi at nuclear foci, and binds to the transcripts from the same genomic regions (Figure 2.1 and 2.21). Rhi is required for UAP56 localization and UAP56 is required to maintain Rhi localization in later stage egg chamber (Figure 2.4 and 2.5). These observations suggest that Rhi may recruit UAP56 to pre-piRNAs, and that UAP56 binding to nascent transcripts may spread Rhi through cluster transcription unit. The DEAD box protein Vasa is a conserved component of nuage, and I have found that Vasa containing nuage foci localize across the nuclear envelop from cluster foci in the nucleus (Figure 2.1). Furthermore, cluster transcripts co-immunoprecipitate with Vasa, and small RNA sequencing and tiling array indicate rhi, uap56 and vasa mutants show similar defects in piRNA production and TE silencing (Figure 2.14, 2.17, 2.18 and 2.22D). By contrast, Vasa is not required for Rhi-UAP56 colocalization (Figure 2.9). I therefore propose that Rhi, UAP56 and Vasa define components of a germline piRNA biogenesis pathway that spans the nuclear envelope. A recent study showed that Cuff, a protein related to the yeast transcription termination factor Rai1, binds to Rhi at dual strand piRNA clusters and facilitate piRNA production (Pane et al., 2011). It is likely that Cuff functions with Rhi and Uap56 in the nuclear compartment of this pathway. It will be very interesting to
further examine the relationship between UAP56 and Cuff, which may function together to facilitate piRNA precursor processing and nuclear export.

Our confocal images show Rhi and UAP56 colocalize at nuclear foci. It is hard to determine whether these signals are from the projection of the signal from the cortex or real internal signal at this level. The three dimensional reconstruction of the fluorescent signal will help to solve the problem. Beside accumulation at nuclear peripheral, Rhi and UAP56 also have profound signals at the internal nuclear foci. The functional significance of this internal localization still remains to be determined.

Piwi is the only nuclear protein that binds to piRNAs. Piwi primarily binds to piRNAs antisense to TEs, and Piwi nuclear localization appears to depend on loading of mature piRNAs (Li et al., 2009; Wang and Elgin, 2011). These finding suggest that Piwi functions at a downstream target recognition and silencing step. However, it has been reported that Piwi is also required for germline piRNA production and knockdown piwi decreases germline piRNA level (Malone et al., 2009; Rozhkov et al., 2013).

### 4.3 Co-transcriptional regulation of piRNA production

UAP56, first identified as a 56-kDa protein associated with U2AF$_{65}$, is recruited by U2AF$_{65}$ to mRNAs and required for stable interaction between U2 snRNP and the pre-mRNA branch point (Fleckner et al., 1997). In the protein coding gene pathway, transcription and splicing are linked, and suggesting that UAP56 is loaded co-transcriptionally (Shen, 2009). Intriguingly, I have found that an E245K mutation that specifically disrupts UAP56 function in piRNA pathway also leads to accumulation of spliced forms of piRNA precursors (Figure S1 and S2). A recent germline RNAi screen
study implicated multiple exon-junction components (EJCs) in piRNA mediated TE silencing (Czech et al., 2013). And chromosomal mutations in the EJC gene mago, nashi and tsunagi/Y14 produce ventralization phenotypes that strikingly similar to mutations in uap56 and other piRNA gene (Micklem et al., 1997; Mohr et al., 2001; Newmark and Boswell, 1994; Newmark et al., 1997). UAP56 and the EJC may therefore have a previously unrecognized role in co-transcription suppression of piRNA cluster transcript splicing. Stalled splicing could differentiate piRNA precursors from protein coding gene transcripts, which are either un-spliced or are efficiently spliced to produce mature mRNAs.

A recent study on influenza RNA production suggests a co-transcription model of UAP56 bind to the transcribing nascent transcripts from both of the genomic strand and keep them from forming double-stranded secondary structure (Wisskirchen et al., 2011). Multiple evidences suggested that the transcription of piRNA precursors from dual-strand clusters is co-dependent on each other. The very same mechanism of UAP56 could be applied here as well.

### 4.4 piRNA transcription and nuclear export

DEAD box proteins function as ATP-dependent RNA clamps and play central roles at several steps in the RNA export pathway (Linder and Jankowsky, 2011). UAP56, for example, is involved in spliceosome assembly and export of matures mRNAs from the nucleus. Studies in multiple systems show that mutation or knockdown of uap56 leads to accumulation of transcripts in the nucleus(Gatfield et al., 2001; Jensen et al., 2001; Luo et al., 2001; MacMorris et al., 2003; Strasser and Hurt, 2001). During
export, UAP56 associates with the THO complex and recruits nuclear export adaptor ALY to form the TREX (transcription-export) complex on spliced mRNAs, thus coupling mRNA splicing and export (Dufu et al., 2010; Katahira, 2012; Strasser et al., 2002). In the piRNA pathway, cluster transcripts associate with UAP56 and Vasa, which localize to opposite sides of the nuclear envelope (Figure 2.1 and 2.22). The related DEAD box protein Dbp5 binds to mRNAs, but interactions with nucleoporins stimulates ATP hydrolysis and transcript release (Montpetit et al., 2011). This suggests a piRNA precursor delivery mechanism in which UAP56 binds cluster RNAs co-transcriptionally, and releases this cargo after interactions with the nuclear pore stimulating ATP hydrolysis. We speculate that the transcripts then funnel through nuclear pore and are bound by Vas as they emerge on the other side. Consistent with this model, RNAi screens in Drosophila have identified proteins in THO and nuclear pore complexes as potential piRNA pathway components (Czech et al., 2013; Handler et al., 2013; Muerdter et al., 2013). Vasa co-immunoprecipitates with the PIWI protein Aub, which is required for ping-pong amplification (Lim and Kai, 2007; Malone et al., 2009). Vasa may therefore deliver piRNA precursor transcripts to the piRNA processing machinery.

4.5 piRNA production and protein coding gene expression

Rhi and most other piRNA pathway genes are germline specific and don’t appear to function in gene expression, but UAP56 is ubiquitously expressed and essential for both piRNA and mRNA biogenesis. How does UAP56 accurately and efficiently function in both pathways?
Despite of the majority of TE mapping piRNAs in the *Drosophila* germline, there is a subset of piRNAs derived from protein coding genes and preferentially from 3’ UTR. And the biogenesis of these genic piRNAs depends on the primary piRNA components, Piwi (Robine et al., 2009). We identified same set of genic piRNA with a size peak at 26nt, which is consistent with Piwi binding piRNAs (Figure 2.19B and D). Intriguingly, *uap56* and *rhi* mutants showed the increase in genic piRNA production, while *vasa* did not (Figure 2.19E-G). This observation suggested Rhi and UAP56 are not only required for piRNA production but also essential for the specificity of the machinery. *rhi* and *uap56* mutants loose the stringency of piRNA production from TE derived transcripts, and more protein-coding gene transcripts might be fed into piRNA pathway to generate more genic piRNAs (Figure 4.1). Detailed analysis of the genic piRNAs in both mutants will be beneficial.

UAP56 is a conserved member of DEAD-box family proteins. X-ray crystallographic structure analysis revealed eIF4AIII and other DEAD-box proteins have conserved surface patches (Caruthers et al., 2000; Caruthers and McKay, 2002; Subramanya et al., 1996). Site-specific mutational study showed certain surface motifs were required for specific functions of the protein suggesting the surface binding sites of regulatory factors (Shibuya et al., 2006). I speculate that UAP56 functions together with a piRNA pathway specific factor. The E245K mutation specifically disrupts UAP56 function in piRNA pathway, and protein structure analysis indicates that this is a surface residue (Figure 2.14 and 2.23). A piRNA pathway specific regulatory factor could associate with UAP56 through the domain defined by this residue, and then perturb the conformation of the protein.
Figure 4.1. Nuclear piRNA pathway

Protein coding mRNAs and pre-piRNAs are transcribed and exported in two parallel pathways in *Drosophila* nurse cells. Rhi and Cuff are associated with dual-strand piRNA cluster and mediate transcription from the locus. UAP56 is recruited to the locus by an unknown regulatory factor and then escorts the nascent cluster transcripts to the nuclear periphery. Nuage component, Vasa, picks up piRNA precursors at the other side of nuclear envelope and feed them into the processing machinery to generate the mature piRNAs. At the same time, protein-coding transcripts are bound by UAP56, and mediate their export to the cytoplasmic translation machinery to generate proteins. By closely localize piRNA machineries to both sides of nuclear envelope, the cell is able to process piRNA cluster transcripts specifically and efficiently over mRNAs. Only small amount of genic piRNAs are produced from the 3’UTR of mRNA transcripts in wild type egg chamber. Mutations of *rhi, cuff* or *uap56* disrupt the nuclear co-localization of the piRNA machineries, and the system looses the specificity and efficiency, increasing the genic piRNA production and reducing piRNAs from the clusters. Some of the cluster transcripts may even be processed by the splicing machinery and show prominent splicing patterns.
Figure 4.1
UAP56 and Rhi colocalize at light level, suggesting that Rhi could be the regulatory factor. However, I was unable to detect an interaction between Rhi and UAP56 by coimmunoprecipitation. Mass spectroscopic analysis of proteins that bind to wild type UAP56 and the sz15 allele should directly identify the hypothesized binding partner. The accumulation of spliced piRNA precursors in uap56 mutants implies an intriguing splicing suppression mechanism for UAP56 in piRNA pathway, which is very different from its function in splicing function and transport of coding gene pathway. Binding to a piRNA pathway component may help UAP56 to co-transcriptionally suppress splicing, and thus distinguish pre-piRNAs from mRNAs.

In wild type nurse cells, UAP56 is dispersed throughout the nucleoplasm and colocalizes with Rhi to distinct foci. Mutations in rhi and the E245K substitution in UAP56 that disrupts piRNA biogenesis block UAP56 localization to these foci, suggesting that they are piRNA precursor transcription sites (Figure 2.4). Vas carrying pre-piRNAs and much of the piRNA processing machinery accumulate in nuage granules across the nuclear envelope from these nuclear foci (Figure 2.1 and Figure 2.22D). Linking the site of precursor transcription to the processing machinery may increase the efficiency of piRNA biogenesis and reduce ectopic piRNA production from mRNAs, which are generated from loci that are dispersed in the nucleus. High resolution imaging of piRNA precursor and mRNA transport across the nuclear membrane could help test this model.

4.6 Open Questions
piRNA mediated TE silencing is a germline mechanism that maintains genome integrity. To achieve high specificity and efficiency, I propose that each step of the piRNA pathway has to be spatially and biochemically connected, and that UAP56 plays an essential role in establishing these connections by linking nuclear expression with cytoplasmic processing piRNA. Unlike Rhi, which is rapidly evolving and germline specific, UAP56 is highly conserved, ubiquitously expressed in both soma and germline, and functions in both general protein expression and piRNA production. Why and how the organism utilizes such a general regulator in tissue and spatially specific processes is not understood, but these links are likely to underlie many cellular tissues and cell specific pathways. Studies on the piRNA related UAP56 functions might therefore provide insight into other developmentally regulated processes.
Reference


Lachner, M., O'Carroll, D., Rea, S., and Mechtler, K. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. 10. 1038/35065132 410, 116-120.


