



universität
wien

DISSERTATION

Titel der Dissertation

The Genetic Makeup of
the *Drosophila* piRNA pathway

Verfasser

Dominik Handler

angestrebter akademischer Grad

Doctor of Philosophy (PhD)

Wien, 2014

Studienkennzahl lt. Studienblatt: A 094 490

Dissertationsgebiet lt. Studienblatt: Molekulare Biologie

Betreuerin / Betreuer: Dr. Julius Brennecke

Table of Contents

1	ABSTRACT	3
2	ZUSAMMENFASSUNG	4
3	PRELUDE	5
4	INTRODUCTION.....	6
4.1	TYPES OF SELFISH GENETIC ELEMENTS	6
4.1.1	<i>Biased gene converters.....</i>	6
4.1.2	<i>Meiotic drivers.....</i>	6
4.1.3	<i>Post segregation distorters.....</i>	7
4.1.4	<i>Cytoplasmic drivers.....</i>	7
4.2	TRANSPOSABLE ELEMENTS	7
4.2.1	<i>Class I transposable elements.....</i>	8
4.2.2	<i>Class II transposable elements</i>	9
4.2.3	<i>The role of TEs in genome evolution.....</i>	10
4.3	RESHAPING GENE EXPRESSION BY RNAi	10
4.4	<i>DROSOPHILA</i> OOGENESIS	12
4.5	THE <i>DROSOPHILA</i> piRNA PATHWAY	13
4.5.1	<i>Somatic piRNA biogenesis and transposon silencing.....</i>	15
5	PUBLICATIONS	19
5.1	A GENETIC SCREEN IDENTIFIES THE GENETIC FRAMEWORK OF THE SOMATIC piRNA PATHWAY IN <i>DROSOPHILA</i>	19
5.2	THE EXON JUNCTION COMPLEX IS REQUIRED FOR DEFINITION AND EXCISION OF NEIGHBORING INTRONS IN <i>DROSOPHILA</i>	59
5.3	OTHER PUBLICATIONS.....	87
6	DISCUSSION	87

6.1	EVALUATION OF THE SOMATIC piRNA PATHWAY SCREEN	87
6.1.1	<i>Decisions underlying the design of the screen</i>	87
6.1.2	<i>Quality measures of the screen</i>	88
6.1.3	<i>Participation of screen hits in somatic and germline piRNA pathways</i>	90
6.2	SCREEN RESULTS.....	91
6.2.1	<i>A genetic link to transcription elongation</i>	91
6.2.2	<i>Screen hits involved in RNA transport</i>	93
6.2.3	<i>A role of the nuclear pore complex in the somatic piRNA pathway</i>	93
6.2.4	<i>Factors potentially involved in transposon silencing</i>	95
6.3	IMPLICATIONS OF THE SCREEN FOR OTHER FIELDS OF BIOLOGY	96
6.4	THE EMERGING ROLE OF MITOCHONDRIA IN THE piRNA PATHWAY	97
6.4.1	<i>Several piRNA biogenesis factors are associated with mitochondria</i>	97
6.4.2	<i>Reasons for mitochondrial sequestering of biogenesis factors</i>	98
6.4.3	<i>The piRNA pathway impacts the localization of mitochondria in the cell</i>	99
7	OUTLOOK	99
8	REFERENCES.....	100
9	APPENDIX.....	108
9.1	A SYSTEMATIC ANALYSIS OF DROSOPHILA TUDOR DOMAIN-CONTAINING PROTEINS IDENTIFIES VRENO AND THE TDRD12 FAMILY AS ESSENTIAL PRIMARY piRNA PATHWAY FACTORS.	108
9.2	ABBREVIATIONS	147
9.3	CURRICULUM VITAE	149
10	ACKNOWLEDGEMENTS.....	153

1 Abstract

The piRNA pathway is a small RNA based defense mechanism that acts in animal gonads. Its main function is the protection of the genome against the detrimental effects of transposable element (TE) activity. Argonaute proteins of the PIWI clade (in *Drosophila*: Aubergine (Aub), Argonaute-3 (AGO3) and Piwi) in conjunction with their bound small RNAs (piRNAs) form the core of the piRNA pathway. Active transposons are sensed by sequence complementarity between transposon transcripts and piRNAs. Depending on the Argonaute protein, active TEs are silenced at the transcriptional level (TGS mediated by Piwi) or at the post-transcriptional level (PTGS) by slicer mediated cleavage of the transposon RNA (Aub and Ago3).

In flies the piRNA pathway is active in the germline and the somatic follicle cell of the ovary. In the somatic cells a linear pathway is active where single-stranded piRNA precursor transcripts are processed into primary piRNAs. These are loaded into Piwi, which mediates TGS of TEs in the nucleus. In the germline, besides this linear pathway a secondary piRNA pathway is active in which Aub and AGO3 reciprocally cleave sense and antisense TE transcripts, leading to the amplification of silencing competent piRNAs.

The mechanisms behind the processes required for a functional piRNA pathway are largely unknown. Therefore, I performed an RNAi screen to identify novel factors involved in the *Drosophila* somatic piRNA pathway. Besides the 9 known factors this led to the identification of nearly ~30 high confidence candidates with a putative role in the piRNA pathway. Many of the identified factors can be grouped into functional categories that suggest important functions for transcription elongation, the exon junction complex, RNA export, the nuclear pore complex and chromatin regulation in the pathway.

From the set of novel piRNA pathway factors I selected two for further follow up studies. The first study focused on CG2183/Gasz, a factor that is anchored into the outer mitochondrial membrane. My studies revealed an important function for Gasz in the recruitment of upstream biogenesis factors to the mitochondrial membrane, presumably to present piRNA precursors to the mitochondria-anchored Zucchini endo-nuclease. In the second follow-up study I could show that the exon junction complex is not directly involved in piRNA biology but is indispensable for the splicing of *piwi* mRNA.

Taken together, the insight obtained by studying Gasz and the exon junction complex indicate that the genetic piRNA pathway screen uncovered a wealth of exciting entry points towards a molecular understanding of the piRNA pathway as well as processes that are intricately linked to it.

2 Zusammenfassung

Der sogenannte piRNA pathway ist ein Abwehrmechanismus, der das Erbgut in der tierischen Keimbahn vor der Aktivität mobiler genetischen Elemente (Transposons) schützt. Der piRNA pathway basiert auf der Funktion kleiner RNA Moleküle, den sogenannten piRNAs, welche von Argonaut Proteinen der PIWI Familie gebunden sind. Die drei PIWI Proteine in *Drosophila* sind Aubergine (Aub), Argonaute-3 (AGO3) und Piwi. Aktive Transposons werden über Sequenzkomplementarität zwischen piRNAs und Transposon mRNAs erkannt. Dies führt entweder dazu, dass die Transkription des Transposons unterbunden wird (involviertes Protein hier ist Piwi) oder dass die Transposon RNA von Aub oder AGO3 durch Schneiden unschädlich gemacht wird.

In *Drosophila* ist der piRNA pathway in der Keimbahn und in den umgebenden somatischen Zellen aktiv. In den somatischen Zellen ist ein vereinfachter und linearer piRNA pathway aktiv. In diesem werden einzelsträngige piRNA Vorläufertranskripte im Zytoplasma in piRNAs prozessiert. Diese werden in Piwi geladen, welches das einzig vorhandene PIWI Protein in diesen Zellen ist. Der finale Piwi-piRNA Komplex wird in den Zellkern transportiert, wo er die Transkription von Transposons unterbindet. In der Keimbahn ist zusätzlich zu diesem linearen Mechanismus noch ein sekundärer piRNA pathway aktiv. In diesem haben Aub und AGO3 eine Schlüsselrolle indem sie Transposontranskripte gegensätzlicher Komplementarität schneiden, was zur Amplifikation sogenannter sekundärer piRNA führt.

Wie die Prozesse des piRNA pathways im Detail funktionieren ist weitgehend unbekannt. Deshalb habe ich alle in den somatischen Zellen expremierten Gene auf etwaige Funktionen im piRNA pathway untersucht. Neben den schon bekannten Genen, habe ich ~50 zusätzliche Gene entdeckt, welche potentielle Funktionen im piRNA pathway haben.

Anschließend habe ich die genaue Funktion für einige dieser Gene im piRNA pathway untersucht. Die erste Studie zeigte, dass das mitochondrial gebundene Protein CG2183/Gasz eine essentielle Komponente in der Rekrutierung von Vorläufer-RNA bindenden Proteinen zu der mitochondrialen Nuklease Zucchini darstellt. Des weiteren, zeigte ich dass der Exon Junction Komplex keine zentrale Rolle im piRNA pathway selber spielt, sondern vielmehr für das korrekte Spleissen der *piwi* mRNA benötigt wird.

Anhand der beiden detaillierten Studien wird deutlich dass der genetische Screen zahlreiche vielversprechende Ansatzpunkte aufgedeckt hat, die fuer das molekulare Verständnis des piRNA pathway instrumental sein werden.

3 Prelude

The aim of my PhD studies was to systematically identify the set of proteins that act in the *Drosophila* somatic piRNA pathway and to characterize several of the identified novel factors in detail. Considering that the main function of the piRNA pathway is the selective silencing of selfish genetic elements in order to maintain genome integrity, I will first give an introduction into the diversity and biology of selfish genomic elements. I will then introduce the animal piRNA pathway with a particular focus on the pathway that acts in the *Drosophila* ovary. The scientific work that underlies this thesis has been published in two first author publications, which constitute the central portion of the thesis. Selected aspects related to the scientific findings but that are not covered within the publications are discussed in the last part of the thesis.

4 Introduction

4.1 Types of selfish genetic elements

Selfish genetic elements are stretches of DNA in the genome that share the ability to increase their own transmission rate over that of the remaining genome¹. Such elements are found in probably every prokaryotic and eukaryotic genome and can comprise a large fraction of it. The impact that selfish elements have on the evolution of their host genomes is many-fold: They can act as important drivers of gene or genome evolution, they can be largely neutral, or they challenge genomic integrity due to their mobile character.

Selfish genetic elements can be classified into five groups (biased gene converters, meiotic drivers, postsegregation drivers, cytoplasmatic drivers and transposable elements (TEs)).

4.1.1 Biased gene converters

Biased gene converters are elements that have the ability to insert themselves from one region of the genome into another region sharing homologous sequence. The most prominent group of biased gene converters are homing endonucleases (HEs), which exist in bacteria, archaea and eukaryotes.

In bacteria HEs can be found very often as self-splicing group I introns, which encode for the homing endonuclease. If expressed, the endonucleases can cleave DNA at a specific recognition sequence. Once the endonuclease cleaves, the double strand break is repaired by the cell's internal DNA repair mechanisms. If the repair is done via homologous recombination, the intron encoding for the endonuclease will serve as the template and this way the HE will be inserted at a previously naïve locus²⁻⁴.

4.1.2 Meiotic drivers

Meiotic drivers are elements that have the ability to modify the segregation of chromosomes during meiosis. Depending on the mechanism these events can be classified as chromosomal or genic drive.

Chromosomal drive occurs mostly during female meiosis where only one out of the four meiotic alleles will participate in the development of a functional oocyte. Usually this process is random. Chromosomal driver elements have the ability to modify the transmission ratio to favor the allele they are located on. This process has also been implicated in the evolution and expansion of centromeres^{2,5}

Genic drive can also occur during spermatogenesis where each meiosis product will develop into a mature sperm cell. Here, genic driver elements gain an advantage because sperm cells with the element have an increased survival rate in comparison to those lacking the element^{2,6}.

4.1.3 Post segregation distorters

Post segregation distorters (PSDs) share the ability to harm or even kill the next generation of an organism if the element was not transmitted.

PSDs are very common in bacteria and on bacterial plasmids. They often function via a toxin-antidote mechanism. The “parental” cell produces a modifier protein (which is harmful for the host) and the corresponding rescue protein in parallel. The rescue protein has a much lower stability than the modifier protein. If the selfish element is not transmitted to the next generation, the transmitted long-lived modifier protein will harm or kill the daughter cell.^{7,8}

4.1.4 Cytoplasmic drivers

Cytoplasmic drivers are heritable organelles and intracellular microbes that have the ability to change the ratio of sex determination.

Mostly, only females have the ability to transmit heritable organelles like mitochondria, chloroplasts or cytoplasmic microbes to the next generation. Therefore it is in the interest of the heritable organelle/microbe to skew the ratio of sex determination towards females⁹.

An example for a microbe changing the sex ratio is the proteobacterium *Wolbachia* that is only transmitted by females and that can be found throughout arthropods. Dependent on the host and the strain of *Wolbachia* the sex ratio is changed by causing sperm-egg incompatibility (only if the same strain of *Wolbachia* is present in the sperm and in the egg the progeny will survive), induction of male feminization (genomic males develop as females), parthenogenesis (no males are needed for reproduction) or male killing¹⁰.

4.2 Transposable elements

Transposable elements (TEs) were first discovered in maize in the mid 1940s by Barbara McClintock¹¹. Transposable elements are DNA sequences that have the ability to move or multiply within the genome. TEs can make up large portions of eukaryotic genomes. Their contribution ranges from ~5% in *Tetraodon nigroviridis* (green spotted puffer), ~15% in *Drosophila melanogaster*, ~38% in mouse, ~45% in human and up to 85% in the maize genome¹²⁻¹⁶.

TEs can be classified into retrotransposons (class I) and DNA transposons (class II) based on their mode of transposition. Whereas class I TEs transpose via an RNA intermediate (copy and paste), class II TEs transpose via a cut and paste mechanism¹⁷.

4.2.1 Class I transposable elements

The group of retrotransposons can be further subdivided based on mechanistic features, organization of their sequence elements and reverse transcriptase phylogeny into LTR retrotransposons, non-LTR retrotransposons, tyrosine-recombinase elements and Penelope-like elements.

The general transposition process for these elements requires the following steps: transcription of the element, translation of transposon proteins, reverse transcription of the RNA, generation of double stranded cDNA and the integration into the genome.

4.2.1.1 LTR retrotransposons

The main feature of LTR retrotransposons is the presence of long terminal repeats (LTRs), which are flanking the ORFs on both sides. Typically this class of TEs encodes genes for structural and enzymatic proteins. The *gag* gene encodes for structural proteins, which are necessary for the assembly of a virus like particle. Inside this particle reverse transcription of the TE transcript into DNA takes place. Subsequently the cDNA copy will be integrated into the host genome. The *pol* gene encodes multiple functional proteins. These are a protease (to cleave the *pol* polyprotein), a reverse transcriptase (to convert the transposon transcript into cDNA) and an integrase (to insert the cDNA into the genome)¹⁷⁻²⁰.

The first step of transposition is the transcription of the transposon by the host cell RNA polymerase II. The RNA is exported to the cytoplasm and the *gag* and *pol* genes are being translated. The structural proteins encoded in the *gag* gene form a viral like particle that contains the proteins translated from the *pol* gene and the transposon RNA. Within the particle the RNA is reverse transcribed into double stranded cDNA, which will be integrated into the host genome^{20,21}.

Typical LTR retrotransposons are bound to replication within the host cell and can only get transmitted from the mother to the child (vertical transfer). Although similar in structure, retroviruses have the additional ability of cell-to-cell transmission that in addition allows them the infection of other individuals (horizontal transfer). This ability relies on the additional *env* gene that encodes for the structural proteins that build the envelope of the viral particle. During evolution some LTR retrotransposons acquired an *env* gene by hijacking the *env* gene of a retrovirus. A prominent example for this ability is the *gypsy* element. *Gypsy* elements can be found in at least 19 species of *Drosophila*. Based on phylogenetic studies it was shown that the phylogeny of *gypsy* does not follow the host phylogeny. This is seen as evidence that the presence of the *env* gene in the *gypsy* element allowed for horizontal transmission of *gypsy* within the group of *Drosophilids*²²⁻²⁵.

4.2.1.2 Non-LTR retrotransposons

The group of non-LTR retrotransposons is distinct from LTR retrotransposons by the lack of inverted or tandem terminal repeats. These elements typically encode multiple open reading frames, which are flanked by UTRs¹⁹.

Non-LTR retrotransposons can be further subdivided into long interspersed elements (LINEs) and short interspersed elements (SINEs).

LINEs contain one or two open reading frames encoding for a reverse transcriptase (RT) and other proteins required for transposition^{19,26}.

In contrast, SINEs do not encode for a reverse transcriptase (RT). As SINEs are still dependent on a reverse transcription function, it is thought that they hijack a functional reverse transcriptase encoded by LINEs. Therefore LINEs are also called autonomous and SINEs are called non-autonomous retrotransposons²⁷⁻²⁹.

4.2.1.3 Tyrosine-Recombinase elements

Tyrosine-recombinase elements have been identified within the DNA and RNA classes of transposable elements. These elements have in common, that integrase and reverse transcriptase are replaced by a tyrosine-recombinase. Although the mode of transposition is not well understood, it is known that these elements transpose via a circular extra-chromosomal DNA intermediate, which is integrated into the genome by the tyrosine-recombinase³⁰⁻³².

4.2.1.4 Penelope-like elements

Penelope-like elements are named after an element identified in *Drosophila virilis*. These elements can exist with or without LTRs in the host genome. Penelope-like elements share only minor sequence homology with other retro-elements. Nevertheless, a cryptic and distant version of a reverse transcriptase enzyme has been identified. In addition many of these elements encode for an endonuclease that is similar to endonucleases of self-splicing group I introns.^{19,33-35}

4.2.2 Class II transposable elements

DNA transposons use a cut-and-paste mechanism to transpose within the genome.

In many species DNA transposons have only a minor contribution to the total transposon content and seem to be mostly inactive. In humans or mouse they comprise ~5%, in *Drosophila* ~20 % of the total genomic transposon content. Some organisms like *C. elegans* however, show a high content of class II transposons (~80%)³⁶.

In most cases, terminal inverted repeats flanking the element can be identified. Cut-and-paste transposons encode for a transposase, which is required for the transposition process.

Transposition can occur via a double stranded (classic cut-and-paste TEs) or a single stranded (rolling-circle replication, Helitrons) DNA intermediate³⁶.

Class II TEs are not able to directly increase their copy number in the genome and therefore rely on the host machinery. One way to increase the copy number is the repair of double strand breaks caused by a jumping transposon by the cellular homologous repair mechanisms.

4.2.3 The role of TEs in genome evolution

As mentioned before, TEs can comprise large portions of host genomes. Initially, they were often considered as genomic junk without any function. Nowadays the transposition activity of TEs is considered as an important driver of genome evolution. New insertions of TEs can cause insertions or deletions in the host genome that can disrupt or change gene function. TEs have also been implicated in shaping regulatory elements for the transcription of host genes³⁷. Finally, there are examples where TEs or TE encoded genes got domesticated by the host and are used for various functions³⁸.

Beside these positive effects of transposon activity, TEs and their activity are often considered to be negative for host fitness. Considering the mutagenic character of TEs, their activity is especially dangerous in the germline. Uncontrolled activity in the germline has immediate effects that will be transmitted to future generations. The fact that in many organisms uncontrolled TE activity results in sterility emphasizes the need for an efficient system to control transposons. To overcome this threat, organisms developed various defense strategies against TEs. As introduced above, TEs are very heterogenic and fast-evolving. Moreover, they are often transmitted horizontally, even between different species. Any defense system against transposons must therefore be highly adaptable on the one side but also highly specific on the other side in order to not target vital cellular processes.

It is interesting to note that in metazoans the most common defense strategy against TEs relies on small RNA silencing pathways that utilize the general principles of RNA interference (RNAi).

4.3 Reshaping gene expression by RNAi

At the center of every RNAi pathway acts an RNA induced silencing complex (RISC). The core of the RISC consists of a small RNA (21-30nt) bound to an Argonaute protein. Whereas the small RNA provides target specificity for the RISC, the Argonaute protein exhibits the functional component of the complex. Different Argonaute proteins complexed with different functional small RNA populations exist in eukaryotic cells. Depending on their characteristics, associated proteins and on their cellular concentration, the various RSCs can

cause nearly complete silencing of gene expression or only modulate gene expression patterns
39-41

Central to every small RNA pathway is the biogenesis of the small RNAs and their loading into the correct Argonaute protein. After RISC assembly, the small RNA acts as a guide to recruit RISC to target RNAs via sequence complementarity. Dependent on the involved Argonaute protein, different complementarity requirements between small RNA and target

Type	Length (nt)	Argonaute protein	Function
Micro RNAs (miRNAs)	21-23	Argonaute 1 (Ago1)	Post-transcriptional regulation of endogenous gene expression via target transcript destabilization and translational repression; ubiquitously expressed
Small interfering RNAs (siRNAs)	21	Argonaute 2 (Ago2)	Post-transcriptional repression of endogenous and exogenous RNA transcripts via target RNA cleavage; ubiquitously expressed
Piwi-interacting RNAs (piRNAs)	23-30	Piwi, Aubergine (Aub) and Argonaute 3 (AGO3)	Post-transcriptional and transcriptional silencing of TEs in gonads

Table 1 *Drosophila* small RNA classes (modified from ⁴²)

are necessary for stable target recruitment. Once RISC is bound to a target it can exhibit its function, which largely depends on the identity of the Argonaute protein^{41,43-45}.

Animals and plants express a multitude of different Argonaute proteins (examples). Accordingly, different RNAi pathways can be distinguished based on RISC characteristics such as identity of the involved Argonaute protein or identity and length of the bound small RNA population. Table 1 lists the three main classes of small RNAs that are expressed in *Drosophila* as well as the Argonaute proteins that they are bound to⁴².

The major role of the miRNA pathway is the regulation of gene expression and modulation of protein translation. This is achieved by affecting RNA stability or by repressing translation.

The major role of the siRNA pathway is the defense against viral infections, which is achieved by direct cleavage of the target RNA. The siRNA pathway has been exploited widely as a method of gene silencing⁴¹.

As mentioned before TE silencing is most important in the germline, which is reflected by the activity of the piRNA pathway. Whereas the miRNA and siRNA pathways are active in all

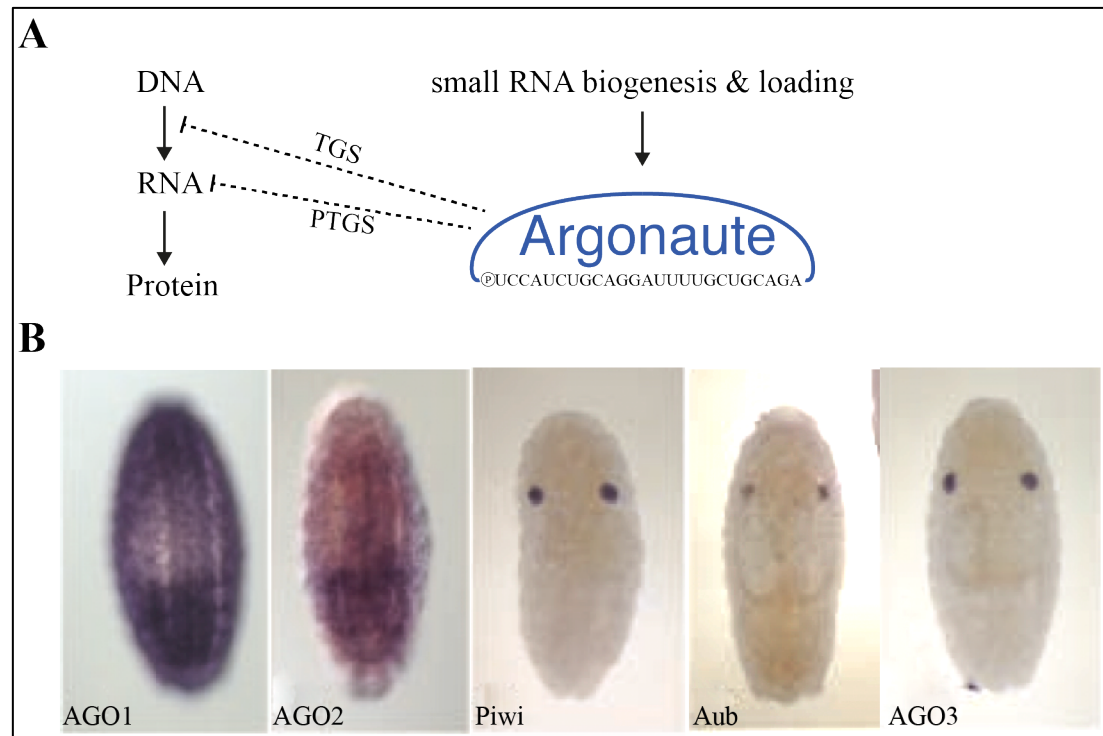


Fig1 small RNA pathways in *Drosophila*. A. Illustration of the general concept of small RNA silencing pathways (TGS=transcriptional gene silencing [Piwi], PTGS=post transcriptional gene silencing [AGO1, AGO2, Aub, AGO3]) B. Shown are in situ hybridization images against the Argonaute proteins in *Drosophila* embryos. (images from ⁴⁶)

tissues, the piRNA pathway is restricted to gonads. This is also reflected by the expression pattern of the involved Argonaute proteins as shown in Fig1 B. Whereas Ago1 and Ago2 are ubiquitously expressed the proteins of the PIWI clade (Piwi, Aub and AGO3) are expressed in gonads only ⁴⁶. Because the piRNA pathway is active in different cell types of the *Drosophila* ovary, I will briefly introduce *Drosophila* oogenesis in the next section.

4.4 *Drosophila* oogenesis

To appreciate the diversity of the *Drosophila* piRNA pathway, it is first necessary to describe the structure of the *Drosophila* ovary. Each ovary consists of approximately 18 ovarioles, which represent individual egg production lines⁴⁸. The structure of an ovariole is illustrated in Fig2. Oogenesis begins in the germarium where two to three stem cells are located. These divide asymmetrically, thereby generating a new stem cell and a daughter cell⁴⁹. The newly derived daughter cell undergoes four mitotic divisions to generate a 16-cell cyst. As these mitotic divisions do not feature full cytokinesis, the cytoplasm of the 16 cells stays interconnected through actin-based ring canals^{50,51}. At the end of the mitotic divisions the germline cyst becomes enveloped by somatic follicle cells, thus forming an egg chamber⁵².

Already early on, one of the 16 cells is defined as the future oocyte. As the oocyte is transcriptionally inactive throughout its maturation, it relies on the supply with cytoplasmic components from the other 15 cells. During egg chamber maturation, these so-called nurse cells grow massively in size and their genome undergoes multiple rounds of endo-replication so that they are highly polyploidy at later developmental stages^{52,53}.

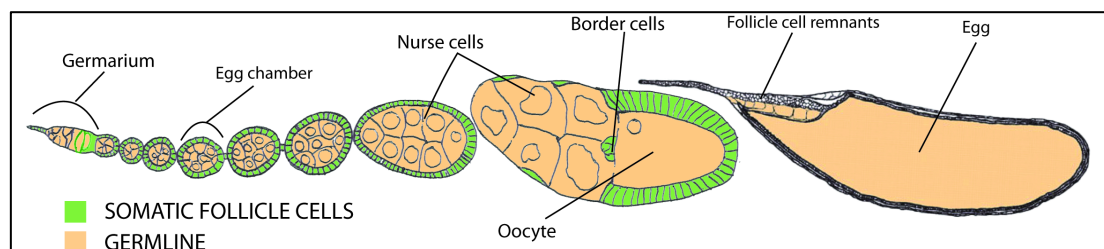


Fig2 Structure of a *Drosophila melanogaster* ovariole with germline cells in beige and somatic follicle cells in green. Illustrated to the left is the germarium where the germline stem cells reside. These stem cells divide and form egg chambers that are encapsulated by somatic follicle cells. Each egg chambers produces one mature egg that is indicated on the right hand side (Figure from ⁴⁷).

During its maturation the oocyte is positioned at the posterior pole of the egg-chamber. At later stages a layer of somatic follicle cells moves posterior and covers the growing oocyte.

At the end of oocyte maturation nurse cells transfer most of their cytoplasmic content into the oocyte and subsequently get degraded. After the follicle cells fully encapsulate the oocyte, they deposit the shell of the egg and undergo apoptosis⁵².

4.5 The *Drosophila* piRNA pathway

In the *Drosophila* ovary the piRNA pathway is active in all germline cells but also in the somatic follicle cells that encapsulate each egg chamber. Although both piRNA pathways control overall TE silencing, they utilize partially different concepts to achieve this goal. This is also reflected in the expression of the PIWI-clade Argonaute proteins: While Piwi is expressed in ovarian somatic and germline cells, Aub and AGO3 are expressed exclusively in the germline. This separation in expression is shown in Fig3A by immunostainings of the PIWI proteins in wildtype egg chambers. In the following, I will give a brief overview of the *Drosophila* piRNA pathway, highlighting the different concepts of the somatic and the germline pathway, which is in addition illustrated in Fig 3B.

In both cell types, precursor transcripts that are processed into mature piRNAs are transcribed from piRNA clusters, which are genomic loci that contain large arrays of transposon remnants⁵⁵. In the soma, piRNA clusters seem to resemble genic transcription units, where a promoter region initiates unidirectional cluster transcription by RNA Polymerase II (RNA Pol II). In contrast, germline piRNA clusters do typically not contain canonical promoter regions. Transcription of these clusters requires the action of the so-called RDC complex (consisting of the proteins Rhino, Deadlock and Cutoff). The RDC might either capture active RNA Pol

II from neighboring transcription units by suppressing their termination or it allows low level transcription initiation within the heterochromatic cluster region⁵⁶. As a general feature, RDC dependent piRNA clusters are transcribed in both directions⁵⁵.

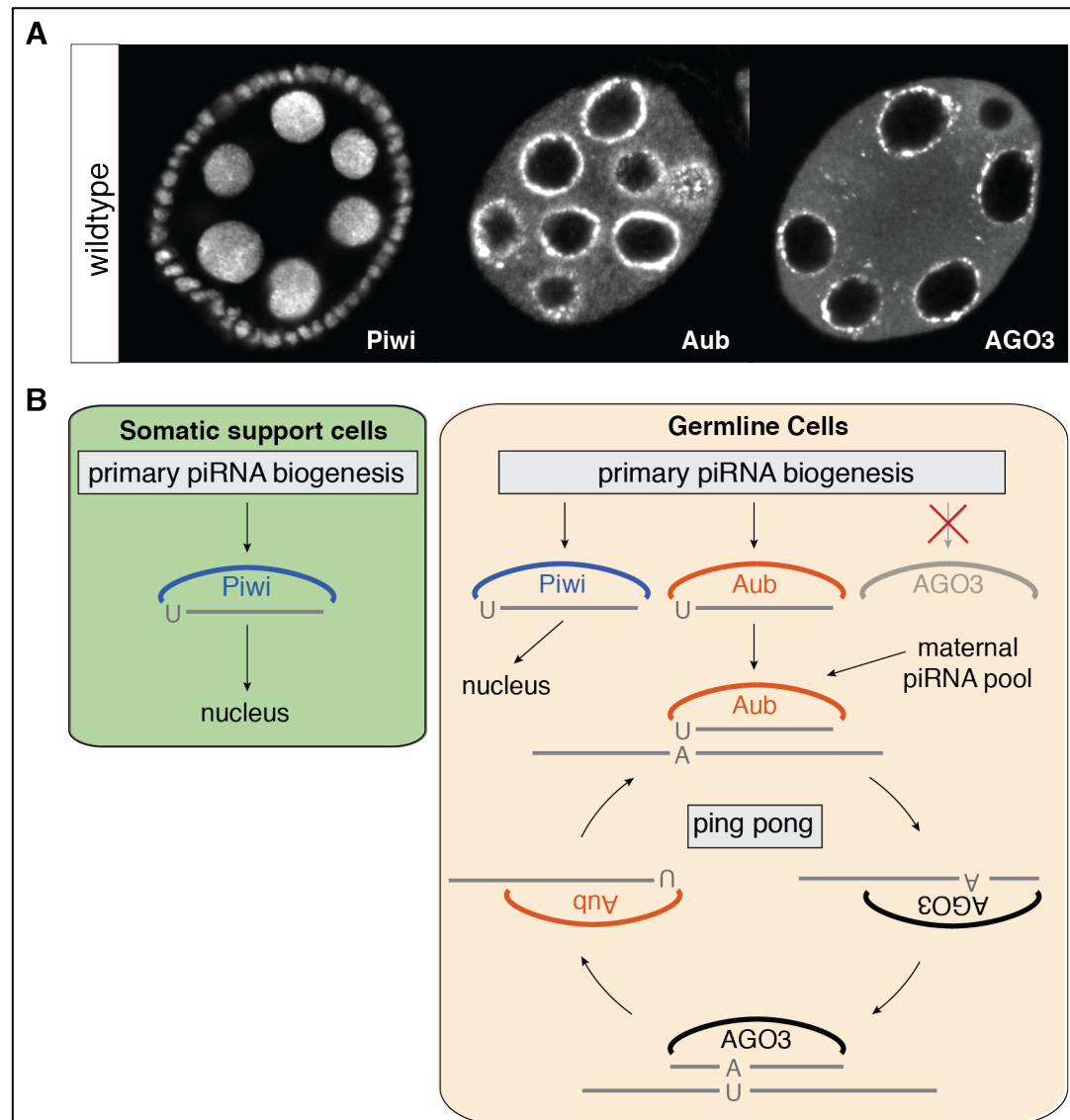


Fig3 Overview of the *Drosophila* piRNA pathway **A** Immuno-fluorescence stainings of the three PIWI-clade proteins in a wildtype *Drosophila* egg chamber (images from ⁵⁴). **B** Schematic representation of the *Drosophila* somatic and the germline piRNA pathway (modified from ⁴⁷).

After transcription, piRNA cluster transcripts must be exported to the cytoplasm as all known piRNA biogenesis factors are located there. Whereas the protein involved in this process are completely unknown for the somatic pathway, it was shown that in the germline the general RNA export factor UAP56 is involved in piRNA cluster export⁵⁷. Once in the cytoplasm, piRNA cluster transcripts are processed into mature piRNAs. Two piRNA biogenesis pathways are generally distinguished. The co-called primary biogenesis pathway leads to the generation of primary piRNAs from single stranded precursors. This pathway is thought to be active in the ovarian soma and germline and fuels Piwi (soma and germline) as well as

Aub (germline). The secondary piRNA biogenesis pathway relies on the endo-nucleolytic cleavage activity of the Aub and AGO3 proteins and is therefore only active in germline cells. The secondary piRNA biogenesis pathway is also referred to as the ping-pong cycle and it is a hallmark of the piRNA pathway in animals ranging from sponges to mammals. The ping-pong cycle gets initiated once a complementary transcript is cleaved by an Aub RISC. The cleavage product (typically a transposon transcript) is then processed into a novel piRNA which is loaded into AGO3. The AGO3 bound sense piRNAs in turn can cleave piRNA cluster transcripts to generate a novel antisense piRNAs, which is identical to the initiator piRNA and which will be loaded into Aub^{55,58}. The ping-pong cycle therefore acts as an amplification loop that enforces the silencing capacity of the piRNA pathway against active transposons.

How mature PIWI RISC silences transposons is largely dependent on the identity of the Argonaute protein. Piwi, the only nuclear Argonaute protein in *Drosophila*, has been shown to silence transposon expression in the nucleus by preventing their transcription (transcriptional gene silencing; TGS)⁵⁹. In contrast, cytoplasmic Aub and AGO3 silence transposons by cleavage of transposon RNAs (post-transcriptional gene silencing; PTGS).

As my PhD project focused exclusively in the somatic ovarian piRNA pathway, I will give a more detailed introduction of this simplified piRNA system in the next chapter.

4.5.1 Somatic piRNA biogenesis and transposon silencing

In the somatic follicle cells of the *Drosophila* ovary a simple, linear piRNA pathway is active in which Piwi is the only participating Argonaute protein. As the involved piRNAs are produced via primary piRNA biogenesis, this pathway is also referred to as a primary piRNA pathway.

Studies on the somatic primary piRNA pathway are greatly aided by the availability of a cell culture system. Initially a cell line called fGS/OSS was established from ovarian stem cells⁶⁰. This cell line was composed of germline stem cells (fGS) and somatic sheet cells (OSS) and germline cells showed expression of the germline marker Vasa. From the original OSS line, a stable sub-clone called Ovarian Somatic Cells (OSCs) was established later on⁶¹. Importantly, it has been shown that both cell lines only express Piwi but not Aub and AGO3 and that a fully active primary piRNA pathway is active in them^{60,61}. The availability of these cell lines featuring only the linear somatic pathway has been of great importance in the understanding of the somatic piRNA pathway. This is because a clean separation of somatic follicle cells either genetically or mechanically from whole ovaries is challenging.

In the somatic primary piRNA pathway two main sources of piRNAs have been identified. On the one side these are TE antisense piRNAs that are derived from piRNA cluster transcripts. On the other side it has been shown that ~17% of somatic piRNAs are derived

from 3' UTRs of mRNAs^{61,62}. Several hundreds of mRNAs have been shown to be piRNA source transcripts, albeit at different levels. The function of these genic sense piRNAs is not known but they might merely be tolerated byproducts of the piRNA biogenesis machinery.

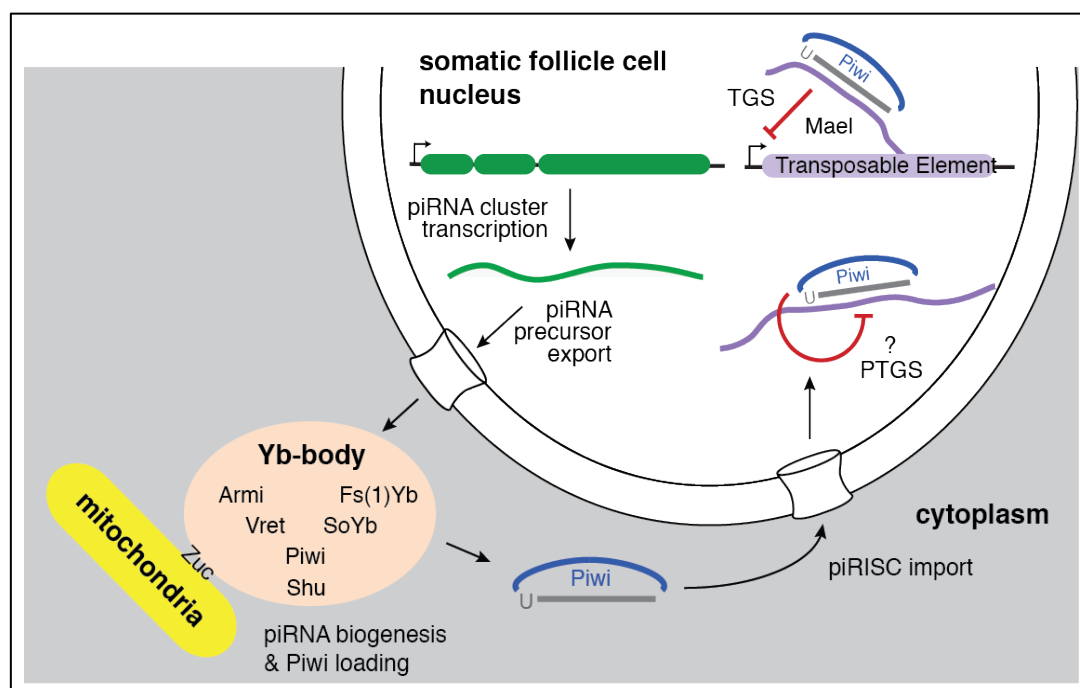


Fig4 Detailed illustration of the *Drosophila* somatic piRNA pathway. Known pathway members are positioned by their published function. (TGS=transcriptional gene silencing; PTGS=post transcriptional gene silencing)

Somatic piRNA clusters are genomic arrays of transposon fragments that are expressed from a single Pol II promoter. As most of the contained TE sequences are inactive TE fragments or remnants⁵⁵, piRNA clusters are considered as a repository that stores sequences of previous TE encounters. piRNA clusters are found on essentially all chromosomes and are often located at the border between euchromatin and peri-centromeric or telomeric heterochromatin⁵⁵.

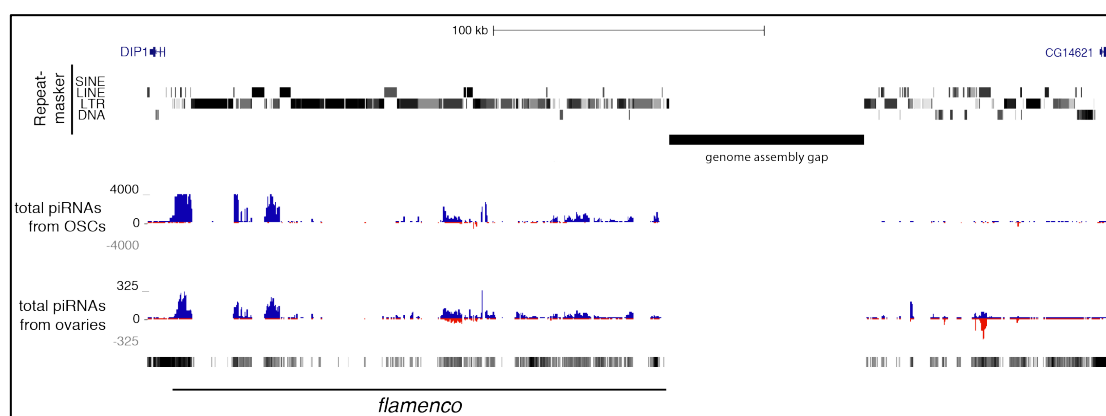


Fig5UCSC Browser screenshot illustrating the distribution of uniquely mapping piRNAs over *flamenco*.

All piRNA clusters in the soma are uni-directional, meaning that they are transcribed in only one direction. The best-studied example of a uni-directional cluster is *flamenco*, which is the

major cluster in somatic follicle cells^{55,63-66}. As can be seen in Fig5, piRNAs are mostly originating from one strand of *flamenco*. Genetic experiments suggest that *flamenco* gives rise to a single long transcript of ~180KB length^{55,65,67,68}. Molecular experiments identified a strong promoter at the beginning of the *flamenco* cluster. It is therefore believed that a single promoter that resembles genic promoters drives all transcription at the *flamenco* locus^{55,65}. Whether transcription of somatic piRNA clusters requires besides canonical transcription steps additional regulatory steps and therefore specialized factors is not known.

Genic piRNAs map nearly exclusively to 3' UTRs of most expressed genes in somatic follicle cells. However, not all cellular mRNAs give rise to piRNAs and there does not seem to be an overall correlation between genic piRNA levels and transcript abundance^{61,62}. How transcripts are selected for piRNA biogenesis and if genic piRNAs have a biological function is unknown.

Prior to their processing into piRNAs, precursor transcripts have to be exported to the cytoplasm. In the case of genic precursors it seems likely that canonical mRNA export pathways are being utilized. Whether piRNA cluster transcripts also utilize canonical mRNA export routes or whether their translocation requires specialized factors is currently unknown. The processing of all piRNA precursor transcripts is believed to be a cytoplasmic process. Based on the subcellular localization of essential piRNA biogenesis factors, it has been suggested that piRNA biogenesis occurs in peri-nuclear structures called Yb-bodies. These RNA-protein accumulations are defined by the presence of the Tudor-domain containing protein Fts(1)Yb (short: Yb). Yb has been shown to be an essential factor for somatic piRNA biogenesis⁶⁹⁻⁷¹. Yb physically interacts with the potential RNA helicase Armitage, which also accumulates in Yb-bodies^{69,70}. How piRNAs are mechanistically processed in a step-wise manner from long cluster transcripts is unknown, but it has been shown that Armitage interacts with 25-70nt long piRNA intermediates⁷⁰. Several proteins have been identified to be important for primary piRNA biogenesis and many of these are enriched in Yb-bodies. Among these are the Tudor-domain protein Vreteno^{54,72}, the Tudor domain containing RNA helicase Sister of Yb⁵⁴ and the Hsp90 co-chaperone Shutdown^{73,74}).

Besides those biogenesis factors localizing to Yb-bodies, several other essential piRNA biogenesis factors localize intriguingly to mitochondria⁷⁵. In fact, these proteins always contain a single helix-spanning transmembrane domain that seems to anchor the respective factor to the outer mitochondrial membrane. The most prominent of these factors is Zucchini^{69,70,76}. In vitro studies suggest that Zucchini is an endo-nuclease that is involved in the 5' end formation of primary piRNAs⁷⁷⁻⁸⁰. However, Zucchini appears not to be required for the generation of Armitage bound piRNA intermediates⁷⁰. Besides Zuc two additional factors associated with the outer mitochondrial were identified. Whereas the function of Minotaur⁸¹

remains elusive, the protein Partner of PIWIs (PAPI)⁸² was implicated in recruiting a trimming exonuclease required for piRNA 3' end maturation to PIWI proteins.

It is believed that after initial nucleolytic processing steps single stranded piRNA precursors are being loaded into Piwi for their subsequent maturation. How loading is regulated mechanistically is unclear but it has been shown to require Shutdown probably in concert with Hsp90^{73,74}. In the current model, the precursors that are loaded into Piwi are subsequently trimmed by an unknown 3' to 5' exonuclease to the mature piRNA length. Such a trimming activity has not been identified in *Drosophila* up to now, but it has been shown to be involved in *Bombyx mori* piRNA biogenesis⁸³. After the final size of the piRNA has been defined, the last step is the 2'O-methylation of the piRNA 3' end by the methyl-transferase Hen1⁸⁴⁻⁸⁶.

While several proteins essential for primary piRNA biogenesis have been identified, their respective molecular functions are not known. In addition it is likely that additional players are involved in piRNA biogenesis. For example, the identity of the trimmer nuclease involved in piRNA 3' end formation is unknown.

When mapping primary piRNAs to their precursor transcripts, a seemingly random pattern of biogenesis emerges. The only hallmark common to most primary piRNAs is their strong enrichment for a uridine residue at their first 5' position. It is, however, unclear whether this bias emerges from a cleavage preference of the involved nuclease (e.g. Zucchini) or from selective stabilization of 1U piRNAs by Piwi. In fact, there is in vitro evidence that Argonaute proteins themselves can display pronounced preferences to bind small RNAs starting with a certain nucleotide^{87,88}.

After mature Piwi RISC is formed, it translocates to the nucleus⁸⁹. There has been a long debate whether nuclear Piwi silences transposons post-transcriptionally or by preventing transcription of its targets via TGS. Although Piwi has been shown in vitro to exhibit slicer activity, it is unlikely that it induces classical PTGS via target cleavage as known for the siRNA pathway. First, Piwi's localization to the nucleus is required for target silencing^{61,90}. Secondly, it has been shown that the activity of the Piwi slicer domain is not required for transposon silencing^{70,90}.

Recently, several studies have shown that Piwi silences TEs primarily or even exclusively at the transcriptional level via TGS⁵⁹. Target silencing is accompanied by formation of an H3 Lysine 9 methylated heterochromatin domain at the target locus⁵⁹. How Piwi orchestrates TGS and which additional factors are involved in this process has been entirely unclear at the start of my PhD project.

5 Publications

During the time of my diploma I was involved in a project that aimed for the optimization of transgenic RNAi in the *Drosophila* germline. The optimized RNAi system is based on the expression of short hairpin RNAs (shRNAs) and is very efficiently depleting target transcripts in germline cells⁹¹.

For my diploma project I utilized this shRNA system to determine the role of Tudor domain containing proteins in the somatic and the germline piRNA pathway. I identified several factors involved in piRNA biology and showed that CG4771/Vreteno, CG31755/BoYb and CG11133/SoYb are required for piRNA biogenesis.

Handler D., Olivieri D., Novatchkova M., Gruber F. S., Meixner K., Mechtler K., Stark A., Sachidanandam R. & Brennecke J. **A systematic analysis of *Drosophila* TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors.** The EMBO Journal (2011). doi:10.1038/emboj.2011.308

Due to the implications of the RNAi strategies presented in this study for the screen, I attached this publication in the Appendix.

In addition another targeted small-scale screen performed in our lab uncovered novel factors involved in primary piRNA biogenesis⁶⁹. Together, these two studies illustrated the potential of RNAi based screens for the identification of piRNA pathway components in the *Drosophila* ovary. Based on this, I performed a genome wide screen tailored towards the identification of piRNA pathway components in the somatic tissue. The results of the screen, in conjunction with the detailed characterization of CG2183/Gasz, a novel piRNA biogenesis factor, constitute the first publication that underlies this thesis. The second topic that is covered in this thesis is a follow-up study on the role of the exon junction complex in the piRNA pathway. This project was performed in close collaboration with Rippei Hayashi.

5.1 A genetic screen identifies the genetic framework of the somatic piRNA pathway in *Drosophila*

At the beginning of my PhD studies, the knowledge about the different molecular steps participating in the piRNA pathway was very limited. My major aim was to fill some of these gaps by systematically identifying the set of protein players that are participating in the primary piRNA pathway. To do so, I conducted a tissue specific RNAi screen in *Drosophila* ovarian follicle cells.

As the readout for the screen I took advantage of a previously published transposon activity reporter, which is under control of the piRNA pathway^{69,92}. This so-called *gypsy-lacZ* reporter proofed to be a highly sensitive yet very robust sensor for piRNA pathway integrity.

Rather than performing a forward genetic screen, I decided to take advantage of the genome wide collection of transgenic RNAi lines available at the Vienna RNAi Center (VDRC⁹³). Wherever possible, we tested two independent RNAi lines for every gene expressed in somatic follicle cells. All positively scoring lines were retests using secondary assays and were also tested for their involvement in the germline piRNA pathway.

The screen identified all but one of the known piRNA pathway genes. Besides those, a total of 53 additional genes scored as strong candidates for participation in the somatic piRNA pathway. 18 out of these showed also clear involvement in the germline piRNA pathway.

From the set of genes required for both, the somatic and the germline piRNA pathway, I selected two genes for further characterization.

The first was *CG9754*, which encodes a polypeptide with no homology to proteins from non-*Drosophilid* species and which does not harbor domains that resemble any domain in databases. I showed, that *CG9754* localizes to the nucleus and that it most likely acts downstream of piRNA biogenesis. This places this factor into the process of Piwi mediated transcriptional target silencing.

The second gene I characterized in more detail was *CG2183/gasz*. My data clearly demonstrated that *Gasz* is an essential factor for primary piRNA biogenesis and that it localizes to mitochondria indistinguishable from the localization of Zucchini. I furthermore provided evidence that *Gasz* connects upstream factors involved in piRNA precursor binding (especially the RNA helicase Armitage) to the mitochondria associated endo-nuclease Zucchini that is implicated in the generation of piRNA 5' ends.

Status: Released Publication

Mol Cell. 2013 Jun 6;50(5):762-77. doi: 10.1016/j.molcel.2013.04.031. Epub 2013 May 9

Authors: Dominik Handler, Katharina Meixner, Manfred Pitzka, Kathrin Lauss, Christopher Schmied, Fanz Sebastian Gruber and Julius Brennecke

Contributions: D. Handler and J. Brennecke planned the experiments; D. Handler, K. Meixner, M. Pitzka, and K. Lauss performed the screen. C. Schmied and F.S. Gruber generated genetic tools; D. Handler performed all experiments. J. Brennecke and D. Handler analyzed the data and wrote the paper.

The Genetic Makeup of the *Drosophila* piRNA Pathway

Dominik Handler,¹ Katharina Meixner,¹ Manfred Pizka,¹ Kathrin Lauss,¹ Christopher Schmied,¹ Franz Sebastian Gruber,¹ and Julius Brennecke^{1,*}

¹Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), Dr. Bohrgasse 3, 1030 Vienna, Austria

*Correspondence: julius.brennecke@imba.oeaw.ac.at

<http://dx.doi.org/10.1016/j.molcel.2013.04.031>

SUMMARY

The piRNA (PIWI-interacting RNA) pathway is a small RNA silencing system that acts in animal gonads and protects the genome against the deleterious influence of transposons. A major bottleneck in the field is the lack of comprehensive knowledge of the factors and molecular processes that constitute this pathway. We conducted an RNAi screen in *Drosophila* and identified ~50 genes that strongly impact the ovarian somatic piRNA pathway. Many identified genes fall into functional categories that indicate essential roles for mitochondrial metabolism, RNA export, the nuclear pore, transcription elongation, and chromatin regulation in the pathway. Follow-up studies on two factors demonstrate that components acting at distinct hierarchical levels of the pathway were identified. Finally, we define CG2183/Gasz as an essential primary piRNA biogenesis factor in somatic and germline cells. Based on the similarities between insect and vertebrate piRNA pathways, our results have far-reaching implications for the understanding of this conserved genome defense system.

INTRODUCTION

The *Drosophila melanogaster* genome contains ~15%–20% of transposable elements (TEs) (Hoskins et al., 2002; Kaminker et al., 2002). Uncontrolled activity of TEs triggers defects in genome integrity due to DNA breaks, insertional mutagenesis, and illegitimate recombination (Levin and Moran, 2011; Slotkin and Martienssen, 2007). This is particularly evident in the germline, where TEs have evolved to be especially active but are suppressed by the piRNA pathway. The piRNA pathway is a small RNA silencing system that is defined via the involvement of PIWI family proteins bound to ~22–31 nt PIWI-interacting RNAs (piRNAs) (Malone and Hannon, 2009; Siomi et al., 2011). Defects in the piRNA pathway allow overall development to occur, but emerging animals are sterile, presumably due to damage of their germ cell genomes (Klattenhoff et al., 2007).

Similar to other small RNA pathways, the central players in the piRNA pathway are Argonaute proteins bound to small RNAs that provide target specificity (Ghildiyal and Zamore, 2009). However, in contrast to the siRNA and the miRNA pathways, whose molecular frameworks are well described, nearly all

aspects of the piRNA pathway are poorly understood. In flies, most piRNAs are derived from piRNA clusters, large loci that contain high proportions of TE sequences (Brennecke et al., 2007). piRNA cluster transcripts are believed to be exported to the cytoplasm where numerous factors participate in the processes of piRNA biogenesis and piRNA loading into PIWI proteins. Most identified piRNA biogenesis factors are enriched in distinct foci that are in proximity to the nuclear envelope (Yb-bodies in ovarian somatic cells and nuage in germline cells) (Siomi et al., 2011). Yb-bodies and nuage are tightly associated with mitochondria, and the essential piRNA biogenesis factor Zucchini is integrated into the outer mitochondrial membrane (Ipsaro et al., 2012; Nishimasu et al., 2012; Saito et al., 2010). After formation of piRNA-induced silencing complexes (pi-RISCs), the identity of the involved PIWI protein defines their respective fates. Piwi-piRNA complexes are transported into the nucleus, where they guide transcriptional silencing (Sienski et al., 2012; Wang and Elgin, 2011; Le Thomas et al., 2013; Rozhkov et al., 2013). The two other family members, Aubergine (Aub) and AGO3, remain in the cytoplasm and are engaged in the ping-pong cycle, a piRNA amplification loop that requires complementary RNAs from TEs and piRNA clusters (Brennecke et al., 2007; Gunawardane et al., 2007).

The *Drosophila* ovary contains two major cell types: germline cells derived from primordial germ cells and somatic support cells of mesodermal origin. Both cell types silence TEs by the piRNA pathway, but the respective pathways differ considerably (Senti and Brennecke, 2010; Siomi et al., 2011). While somatic support cells express only Piwi, germline cells also express Aub and AGO3. Ping-pong amplification and many factors involved in it are thus restricted to germline cells (Li et al., 2009; Malone et al., 2009).

Motivated by the gaps in our understanding of the piRNA pathway, we performed a genetic screen in the *Drosophila* ovary for factors involved in the somatic piRNA pathway. Many identified genes could be grouped into functional categories, and these indicate the identification of key factors acting at all steps of the pathway, ranging from piRNA cluster expression to cluster transcript export to piRNA biogenesis and loading and finally to piRNA-mediated silencing. Our work will therefore serve as a key resource for the genetic and mechanistic dissection of this conserved genome defense pathway.

RESULTS

A Screen for Somatic piRNA Pathway Factors

The *Drosophila* ovary consists of two major cell types (somatic support cells and germline cells) that harbor different piRNA

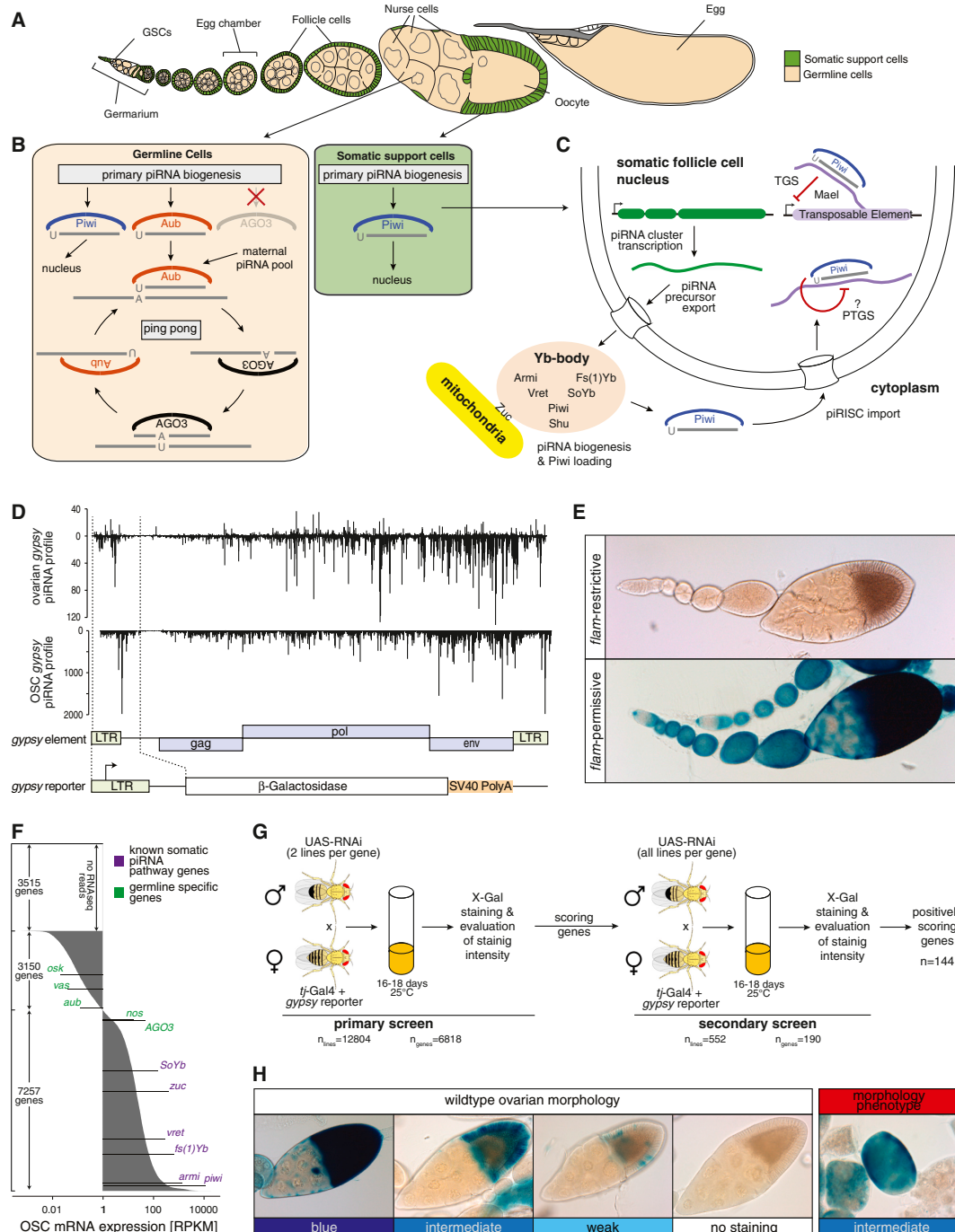


Figure 1. An RNAi Screen for Somatic piRNA Pathway Factors

(A) Cartoon of a *Drosophila* ovary with somatic cells in green and germline cells in beige.

(B) Schematic representations of the *Drosophila* germline and somatic piRNA pathways focusing on the three PIWI family proteins and the biogenesis routes of their bound piRNAs.

(C) Detailed model of the somatic piRNA pathway. Known pathway members are placed at their functional positions based on literature (TGS, transcriptional gene silencing; PTGS, posttranscriptional gene silencing).

(D) Illustration of the gypsy-lacZ reporter. Shown are normalized profiles of ovarian and OSC piRNAs (sense up, antisense down) mapping to the gypsy TE and the gypsy sequence portion present in the reporter.

(legend continued on next page)

pathway versions (Figures 1A–1C). Though strong similarities exist between both pathways, it is unclear how distinct their genetic makeup is. Having an effective transgenic RNAi assay system for somatic piRNA pathway activity at hand (Olivieri et al., 2010), we screened the set of genes expressed in somatic support cells for uncharacterized piRNA pathway factors.

We took advantage of a reporter construct that expresses β -galactosidase (β -gal) under control of the *gypsy* LTR (Figures 1D and 1E; Sarot et al., 2004). The resulting RNA contains ~700 nt of *gypsy* sequence. Many piRNAs isolated from ovaries or from cultured ovarian somatic cells (OSCs) are complementary to this region and guide efficient reporter silencing in wild-type flies (*flamenco* restrictive). Flies harboring a functional piRNA pathway but lacking the ability to produce *gypsy* complementary piRNAs (*flamenco* permissive) exhibit strong reporter desilencing. Similarly, knockdown of somatic piRNA pathway components such as Armitage (Arm), Zucchini (Zuc), or Vreteno (Vret) in a *flamenco* restrictive background results in near loss of *gypsy*-derived piRNAs and reporter desilencing (Handler et al., 2011; Olivieri et al., 2010; Saito et al., 2010; Zamparini et al., 2011). As the *gypsy-lacZ* reporter assays the endpoint of the piRNA pathway (namely piRNA-mediated silencing in *trans*), the screen should capture factors involved at all steps of the piRNA pathway (piRNA cluster biology, piRNA biogenesis, and piRNA-mediated silencing) (Figure 1C).

To filter for genes with possible functions in the somatic piRNA pathway, we determined the gene expression profile of OSCs by RNA-seq. OSCs were derived from a somatic stem cell population of the germarium and express a functional piRNA pathway (Niki et al., 2006; Saito et al., 2009). At a conservative cutoff (1 read per kilobase per 1 million sequenced reads [RPKM = 1]), OSCs express 7,257 genes, and all known somatic piRNA pathway factors were well above this cutoff (Figure 1F).

We used the collection of transgenic RNAi lines from the Vienna RNAi Center (VDRC) (Dietzl et al., 2007) and screened two lines per expressed gene. For the primary screen, this resulted in a total of 12,804 crosses for 6,818 genes (94% of expressed genes). Per cross, five F1 females were dissected and ovaries were stained for β -gal activity. Genes with at least one line exhibiting detectable reporter derepression were retested with all VDRC lines available for this gene (Figure 1G). Dissected ovaries from the secondary screen were mounted on glass slides, and β -gal activity was determined using a classification scheme with six values ranging from strong to weak (representatives shown in Figure 1H). This resulted in 144 genes where at least one knockdown line led to detectable β -gal activity (Table S1). In addition to reporter activity, we classified overall ovarian morphology as “wild-type,” “distorted,” or “rudimentary” (Figure 1H; Table S1).

High Sensitivity, Specificity, and Reproducibility of the Screen

For an overall evaluation of the screen, we considered three questions: (1) What is the false negative rate? (2) What is the false positive rate? (3) How well does the *gypsy-lacZ* reporter mirror derepression of endogenous TEs?

The number of known genes acting in the somatic piRNA pathway is limited ($n = 8$) and not well suited for a false negative estimation. To determine the efficacy of VDRC RNAi lines, we evaluated the ovarian morphology resulting from knockdowns of genes encoding for ribosomal proteins ($n = 83$). Defective ribosome biogenesis results in a “rudimentary ovary” phenotype, and based on this, ~83% of VDRC lines elicit efficient target knockdown (Figure 2A). As we typically tested two independent lines per gene, this resulted in a ~92% positive rate at the gene level. Similar results were obtained when other housekeeping categories were analyzed. Consequently, the set of 663 genes resulting in rudimentary ovaries upon knockdown was highly enriched in basic cellular processes (Figure 2B). We observed a correlation between the “rudimentary ovary” phenotype and gene expression levels, with knockdowns of nearly 40% of the 700 most highly expressed genes preventing ovary development (Figure 2C). We conclude that transgenic RNAi is potent in knocking down genes with very high mRNA levels. As we also tested 358 genes below the RPKM cutoff of 1, we can estimate that off-target effects are not prevalent in this screen, as only 0.8% of lines targeting these “nonexpressed” genes resulted in altered ovarian morphology.

Eight genes have been assigned to the somatic piRNA pathway, and with the exception of the Tdrd12 member SoYb (the available VDRC line is nonfunctional), all of these were identified among the top scoring candidates (Figure 2D) (Haase et al., 2010; Handler et al., 2011; Malone et al., 2009; Olivieri et al., 2010, 2012; Preall et al., 2012; Qi et al., 2011; Saito et al., 2010; Sienski et al., 2012; Zamparini et al., 2011). We conclude that the screen identified most somatic piRNA pathway factors whose knockdown results in analyzable ovaries.

The false positive rate depends on the extent of the *gypsy-lacZ* sensor being influenced only by the piRNA pathway and also on the extent of screen artifacts such as off-targets. Remarkably, only 2% (144 genes) of all tested genes showed any β -gal activity, and only 0.7% (49 genes) showed activity in the range of known pathway members (Figure 2E). Also, for the positive genes, we observed a correlation with expression levels (Figure 2F). This was, however, less pronounced compared to the morphology results (Figure 2C). Importantly, very few of the weakly expressed genes and none of the 358 tested nonexpressed genes derepressed the reporter upon knockdown.

To more directly assess the specificity of the screen, we performed two tests. First, we generated 30 shRNA lines (Ni et al.,

(E) Shown are β -gal stainings of ovarioles as readout for the *gypsy-lacZ* reporter from *flamenco* restrictive (upper panel) or *flamenco* permissive flies (lower panel). (F) Bar chart showing expression levels (average RPKM values (log10 scale) obtained from two independent RNA-seq experiments) of all annotated *Drosophila* genes in OSCs. Several known somatic piRNA pathway factors (violet) and germline-specific control genes (green) are indicated.

(G) Illustration of the screen workflow. Indicated are the numbers of tested RNAi lines and the corresponding number of genes for the primary and secondary screens.

(H) Shown are β -gal stainings of representative egg chambers indicating major staining categories used for the evaluation of the screen crosses (left images, wild-type morphology; right image, distorted morphology).

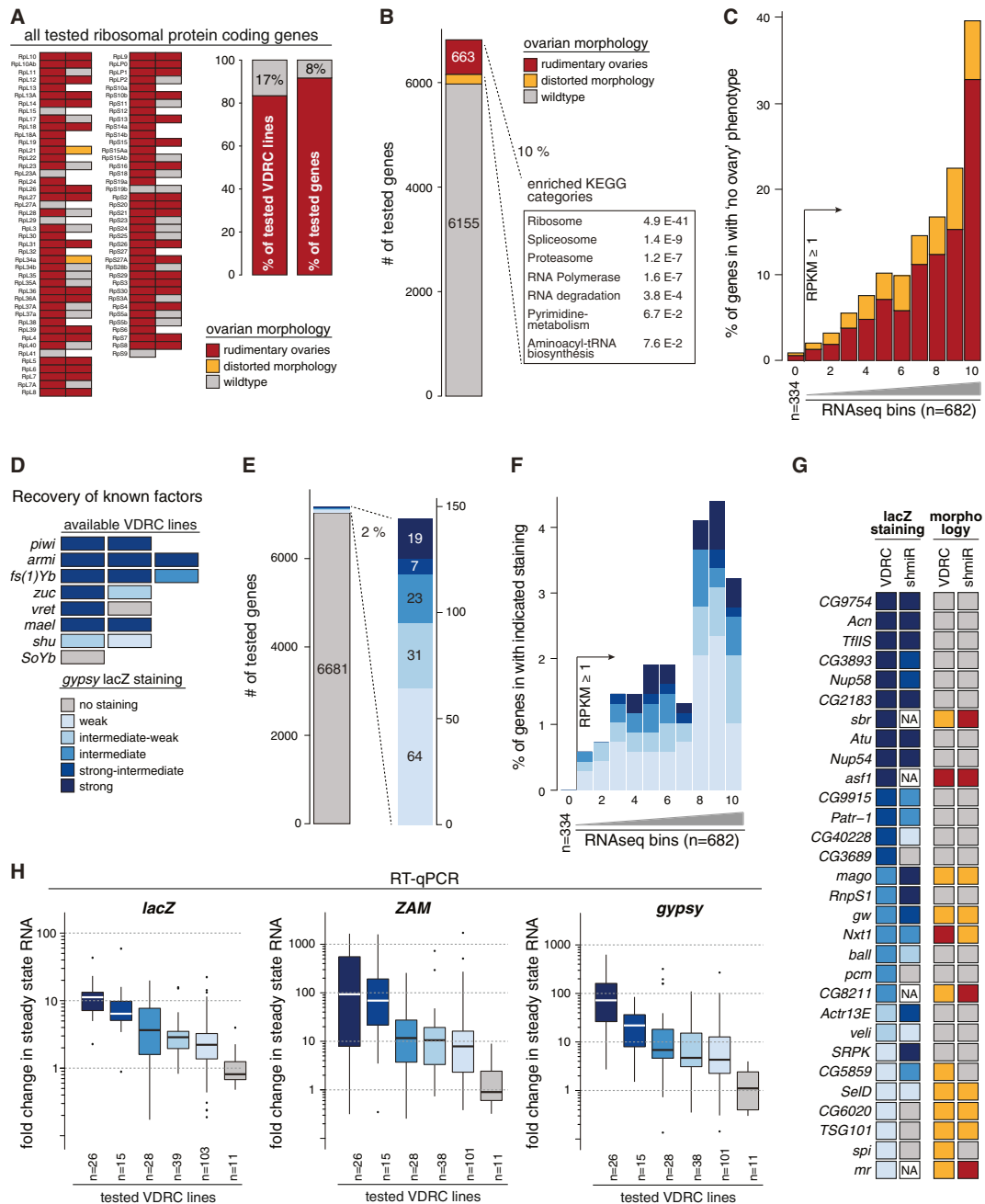


Figure 2. High Sensitivity, Specificity, and Reproducibility of the Screen

(A) Indicated to the left is the impact on ovarian morphology observed upon knockdown of the 83 genes encoding for ribosomal proteins. The two bar charts are based on this analysis and indicate the percentage of effective VDRC lines (left) and the corresponding false negative rate (8%) at the gene level if approximately two lines per gene were tested (right).

(B) Bar chart illustrating the ovarian morphology phenotype observed for all genes tested in the screen. At least one line per gene had to fall into the indicated categories. Also shown are the most enriched gene ontology (GO) categories (p values corrected for multiple testing) for the set of 663 genes that classified for the “no ovary” phenotype.

(C) Indicated are the percentages of genes flagged with the “no ovary” or “distorted morphology” phenotypes when all tested genes were split into ten bins according to their expression level (gray triangle). Bins 1–10 are equally sized bins (n = 682) of all expressed genes (RPKM > 1), while bin 0 contains 334 randomly tested genes expressed below RPKM = 1.

(D) Shown are the gypsy-lacZ staining results from the screen for available VDRC lines targeting the eight known piRNA pathway factors.

(legend continued on next page)

2011) to reassess the screening results with independent knock-downs (Figure 2G). Out of 26 lines that could be analyzed (four lines resulted in “rudimentary ovaries”), 21 (>80%) confirmed the screen result. Considering that not all shRNA lines will be functional, RNAi off-target effects can be considered to be very low in the screen. We then asked whether the β -gal results correlate with endogenous TE derepression. We performed RT-qPCR experiments on ovarian RNA after knocking down each positively scoring gene in somatic follicle cells. We measured steady-state RNA levels of β -gal as well as of the endogenous retro-elements *ZAM* and *gypsy*, which are under control by the somatic piRNA pathway. In all cases, the classification of β -gal reporter activity correlated highly with the measured RNA levels (Figure 2H; Table S2). Taken together, our analyses indicate a very high sensitivity and specificity of the screen.

Classification of Germline-Specific, Soma-Specific, and Common piRNA Factors

The existence of distinct piRNA pathways in the ovarian soma and germline poses the question about their genetic similarities. We therefore tested the identified candidates from the soma screen for their involvement in germline TE repression.

We depleted screen candidates in the germline by transgenic RNAi based on the *nanos* GAL4 driver coupled to UAS-Dcr2 and VDRC lines (Handler et al., 2011; Wang and Elgin, 2011). To monitor piRNA pathway integrity in germline cells, we generated a TE repression sensor. This reporter expresses nuclear β -gal under control of the *nanos* promoter and harbors a target sequence for *Burdock* piRNAs in its 3' UTR (Figure 3A). The retro-element *Burdock* is a prototypic germline TE, as ovarian piRNAs exhibit a pronounced ping-pong signature and are specifically found in germline cells but not in OSCs (Figures 3A and 3B). *Burdock* piRNA levels are sensitive to loss of various germline piRNA pathway factors (Figure 3C), and knockdown of pathway factors as diverse as AGO3, Vreteno, and Armitage strongly derepressed the sensor (Figure 3D). This suggested that the *Burdock* sensor is a valid tool to monitor germline piRNA pathway integrity. Knockdown of any known major germline piRNA pathway factor led to robust sensor activation (Figure 3E). Even knockdown of candidates with moderate impacts on TE silencing (Tudor, Hen1, CG9925) (Handler et al., 2011; Horwich et al., 2007; Nishida et al., 2009) derepressed the sensor, albeit to lesser extents (Figure 3E). Of note, every tested VDRC line targeting a validated germline piRNA pathway factor was highly efficient in this system. We conclude that the germline knockdown system is very robust, allowing meaningful conclusions about the tissue specificity of identified piRNA pathway factors.

Based on the *Burdock*-lacZ sensor, 26 factors scoring at least weak-intermediate in the screen also scored in the germline (Figure 3E). Remarkably, ~30 genes scored specifically in the somatic piRNA pathway only. Among these were some of the strongest scoring screen candidates, and several were validated via independent shRNA lines (Figure 2G). Our results strengthen large similarities between the soma and the germline piRNA pathways but also pinpoint significant differences.

To substantiate the validity of *Burdock*-lacZ as an accurate proxy for germline piRNA pathway integrity, we measured steady-state RNA levels of the two germline-specific TEs, *HeT-A* and *Burdock*. We observed strong correlations with the *Burdock*-lacZ results (Figure 3F; Table S3). Knockdown of candidates strongly scoring in the germline derepressed endogenous TEs to an extent that was highly similar to knockdowns of established germline piRNA pathway factors. On the other hand, no significant changes in endogenous TE levels were seen upon knockdown of factors that scored in the soma but not in the germline. The housekeeping gene *Actin 5C* served as control.

Taken together, our results define the set of common factors acting in the somatic and germline piRNA pathways and provide genetic entry points to dissect the differences between the two pathways. Reaching a similar status for the class of “germline-specific” factors would require a targeted germline screen as presented in the accompanying manuscript by Hannon and colleagues (Czech et al., 2013).

Identification of Key Processes Involved in the Somatic piRNA Pathway

A comparison of all 144 screen hits to the set of 6,818 tested genes in respect to tissue gene expression profiles (FlyAtlas, <http://www.flyatlas.org>; Chintapalli et al., 2007) indicated that the expression of scoring genes was higher in ovaries and to a lesser extent in the larval CNS compared to whole flies (Figure 4A). These results are in agreement with the piRNA pathway acting specifically or preferentially in gonads.

The set of 144 candidate piRNA pathway genes was highly enriched for several functional gene ontology (GO) terms (Figure 4B). Surprisingly, a number of top categories were associated with mitochondrial metabolism. Indeed, 23% of all screen hits (38 genes) could be annotated as mitochondrial proteins (Figure 4C). All of these derepressed the *gypsy*-lacZ sensor only moderately and impacted ovarian morphology to varying degrees. Based on RT-qPCR results, knockdown of identified mitochondrial factors increased the expression level of endogenous TEs moderately but significantly (Figure 4D). As many other factors whose knockdown impaired oogenesis did not result in sensor activation, we conclude that mitochondrial integrity is

(E) Bar chart summarizing the *gypsy*-lacZ staining results ranging from weak to strong for all genes tested in the screen.

(F) Indicated are the percentages of genes scoring with the indicated *gypsy*-lacZ intensities when all tested genes were split into ten bins according to their expression level (gray triangle). Bins 1–10 are equally sized bins ($n = 682$) of all expressed genes (RPKM > 1), while bin 0 contains 334 randomly tested genes expressed below RPKM = 1.

(G) Compared are *gypsy*-lacZ intensities as well as the morphology phenotypes for 30 screen-positive genes tested with shRNA lines or VDRC lines (NA, not analyzable due to a “no ovary” phenotype).

(H) Shown are box plots displaying the fold changes in steady-state RNA levels (based on RT-qPCR) of *lacZ*, *ZAM*, and *gypsy* normalized to control knockdowns. Tested were all RNAi lines (numbers given at the bottom) falling into the five staining categories (color coded) and 11 control lines (gray). Box plots show median (line), 25th–75th percentile (box) \pm 1.5 interquartile range; circles represent outliers.

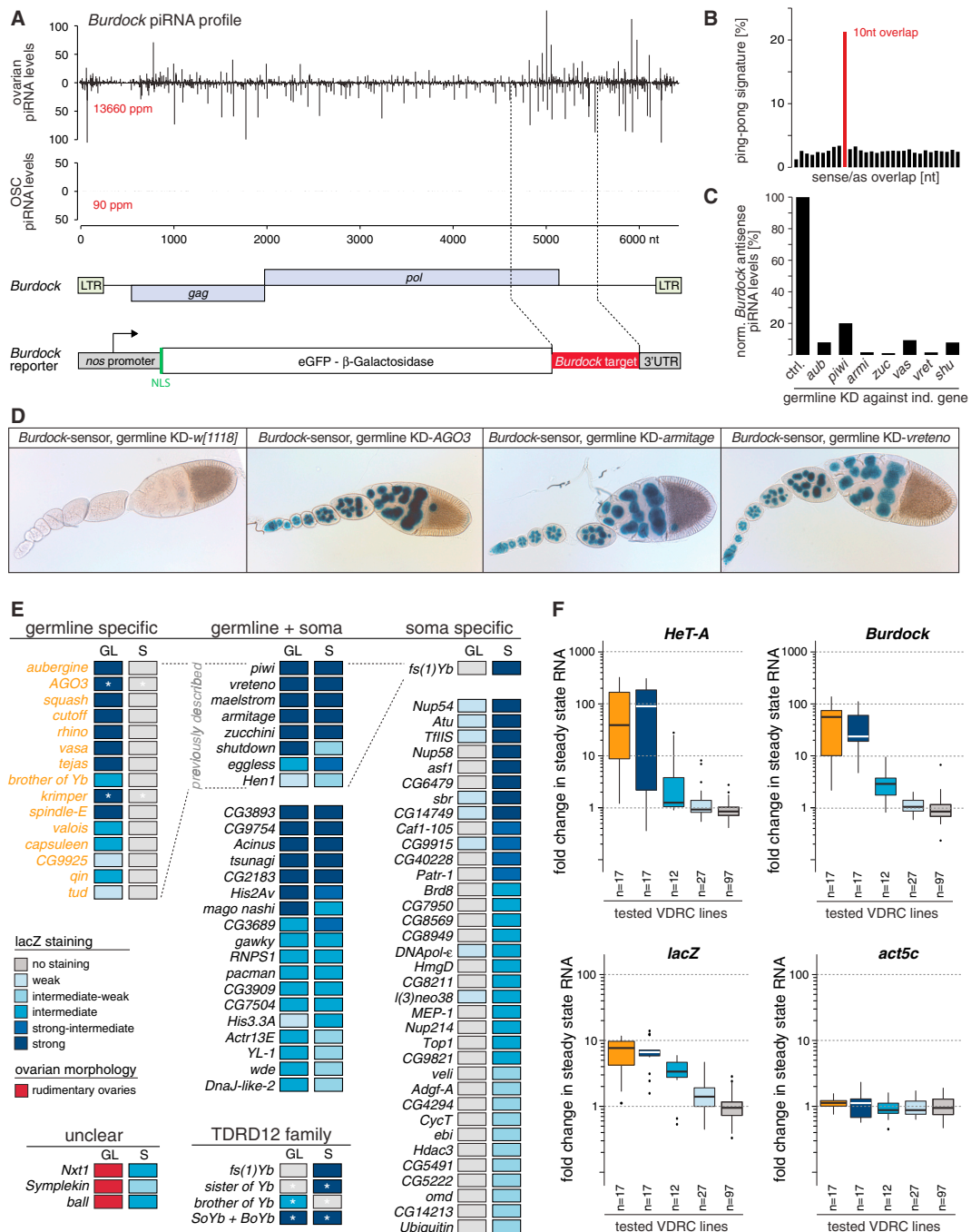


Figure 3. Genetic Classification of Germline-Specific, Soma-Specific, and Common piRNA Factors

(A) Illustration of the Burdock-lacZ reporter. Shown are normalized profiles of ovarian and OSC piRNAs (sense up, antisense down) mapping to the Burdock element and the Burdock sequence portion present in the 3' UTR of the reporter, which expresses β -gal under control of the germline-specific *nanos* promoter. (B) Bar diagram displaying the sense/antisense overlap patterns of the ovarian piRNA population mapping to the Burdock element. The red bar at 10 nt indicates a significant ping-pong signature. (C) Bar plot illustrating normalized piRNA levels (%) antisense to the Burdock TE in ovaries from indicated germline-specific knockdowns (MTD \times shRNA) compared to control levels. (D) Shown are β -gal stainings of egg chambers expressing the Burdock-lacZ reporter and germline knockdowns (KD) for the indicated genes (*w[1118]* serves as negative control).

(legend continued on next page)

important for full piRNA pathway activity. Interestingly, the essential piRNA biogenesis factor Zuc is an integral protein of the outer mitochondrial membrane, and piRNA biogenesis is proposed to occur in close proximity to mitochondria. We note that knockdown of mitochondrial genes in the germline generally did not result in measurable *Burdock-lacZ* sensor derepression, suggesting that the somatic piRNA pathway is more sensitive to mitochondrial malfunction.

We repeated the GO analysis after subtracting annotated mitochondrial proteins. Many of the significantly enriched categories (Figure 4E) were linked to RNA biology, nucleo-cytoplasmic transport, or chromatin biology. In combination with manual annotations, we propose the involvement of several functionally or genetically related gene groups in the piRNA pathway (Figure 4F). In the cytoplasm, several factors required for piRNA biogenesis (Yb, Armi, Vret, Shu) are enriched in Yb-bodies, which are in direct vicinity of mitochondria (Szakmary et al., 2009). Three factors with relationships to processing bodies (Gawky, Pacman, Patr-1) suggest a connection between the piRNA pathway and P-bodies. Notably, several piRNA pathway factors colocalize with P-bodies in so-called piP-bodies in mouse male germ cells (Aravin et al., 2009), and Yb-bodies have been observed in proximity to P-bodies in *Drosophila* follicle cells (Olivieri et al., 2010).

More intriguingly, we identified several factors with a nuclear function or that act at the interface between nucleus and cytoplasm. In each category, at least one factor was among the strongest-scoring candidates, and independent shRNA lines validated several factors in each category.

The two core exon junction complex (EJC) (Tange et al., 2004) factors Mago and Tsunagi/Y14 as well as the EJC accessory factors RnpS1 and Acinus scored strongly in the assay. The third core factor, eIF4AIII, could not be analyzed due to a “rudimentary ovary” phenotype. All identified EJC factors also classified as essential pathway factors in the germline (Figure 3E), strongly implicating the EJC in piRNA pathway biology.

The central RNA export factors small bristles (Nxf1) and Nxt1 strongly derepressed the sensor. These factors are implicated in general mRNA export, and ovarian morphology was distorted in the respective knockdowns. Nevertheless, the extent of sensor and TE derepression upon knockdown of these factors hints at a direct involvement of this protein complex, potentially at the level of piRNA cluster transcript export. Interestingly, Nxf2 and Nxf3 are specifically expressed in ovaries (but not in OSCs), and their knockdown in the germline results in TE silencing defects (unpublished observations), strongly substantiating the central importance of RNA export processes in the piRNA pathway.

Four nuclear pore complex factors (Nup54, Nup58, CG11092, Nup214) and one accessory factor (CG14749) scored in the

screen. The two FG repeat proteins Nup54 and Nup58 were among the strongest-scoring candidates, and given that ovarian morphology was normal in the respective knockdowns, we postulate that an intricate and specific connection exists between the nuclear pore complex and piRNA biology. All of these factors appear to be specifically involved in the somatic piRNA pathway (Figure 3E). Our finding might relate to a recent study in germline cells that found piRNA cluster loci specifically at nuclear pores to allow for specific and/or efficient coupling of piRNA cluster transcription/export to the perinuclear piRNA biogenesis machinery (Zhang et al., 2012).

Several strongly scoring factors are annotated to be involved in aspects of RNA Polymerase II biology. For example, the identified factors Atu, CG3909, CG9915, CG40228, and TflIS are either annotated Paf1 complex members or are linked to TflIS, a key factor required to rescue backtracked RNA Pol II (Sims et al., 2004). These proteins might be involved in Piwi-mediated transcriptional silencing. However, the group of factors with links to Pol II did not score in the germline. As Piwi-mediated TE silencing is unlikely to differ between somatic and germline cells, we favor the hypothesis that factors ensuring efficient Pol II activity are important for aspects of piRNA cluster biology. *flamenco*, for example, is a locus that spans at least 180 kb of repetitive sequence, and all evidence suggests that it is transcribed from a single promoter. Productive transcription through such a large locus might well require factors that reactivate backtracked polymerase. The soma-specific requirement of these factors suggests that transcription of germline piRNA clusters differs from that of *flamenco*. This might relate to their bidirectional transcription and the specific requirement for Rhino and Cutoff for their transcription/processing (Klattenhoff et al., 2009; Pane et al., 2011).

Piwi-mediated silencing of TEs occurs predominantly at the transcriptional level (Siensi et al., 2012; Wang and Elgin, 2011; Le Thomas et al., 2013; Rozhkov et al., 2013). Besides Piwi and Maelstrom, we identified several factors with chromatin-associated functions that might participate in this process. Among these are the histone variant His2Av, the histone chaperone Asf1, the H3K9 methyltransferase Eggless and its cofactor Windei, as well as others. Of note, most of these factors also scored in the germline (Figure 3E).

Uncharacterized Factors in piRNA Biogenesis and Piwi-Mediated Silencing

Many genes identified in the screen do not belong to a functional category. Among these, we chose CG2183 and CG9754, two uncharacterized genes, for follow-up studies. Knockdown of either gene in somatic follicle cells resulted in strong *gypsy-lacZ* sensor derepression (Figure 5A, left) as well as in derepression of the endogenous TEs *gypsy* and *ZAM* (Figure 5A, right).

(E) Listed are all genes known to be specific for the germline piRNA pathway (orange set) and all genes scoring in the somatic screen (intermediate-weak or stronger; no mitochondrial genes). Based on the staining intensities observed with the *Burdock-lacZ* reporter (GL/left columns) or the *gypsy-lacZ* reporter (S/right columns), genes were grouped into germline-specific, common, and soma-specific classes. All genes previously linked to the piRNA pathway are marked as “previously described.” Genes not interpretable in the germline test are shown in the lower left panel. Asterisks indicate genes tested with an shRNA line because no VDRC line was available (*AGO3*, *krimper*) or the VDRC line is not functional (*SoYb*).

(F) Box plots showing fold deregulations of *HeT-A*, *Burdock*, *lacZ*, and *act5c* RNA levels upon knockdown of all screen-scoring VDRC lines. Lines were grouped into three staining categories (blue) based on their effect on the *Burdock-lacZ* reporter. Germline-specific factors (orange) are represented as a separate group.

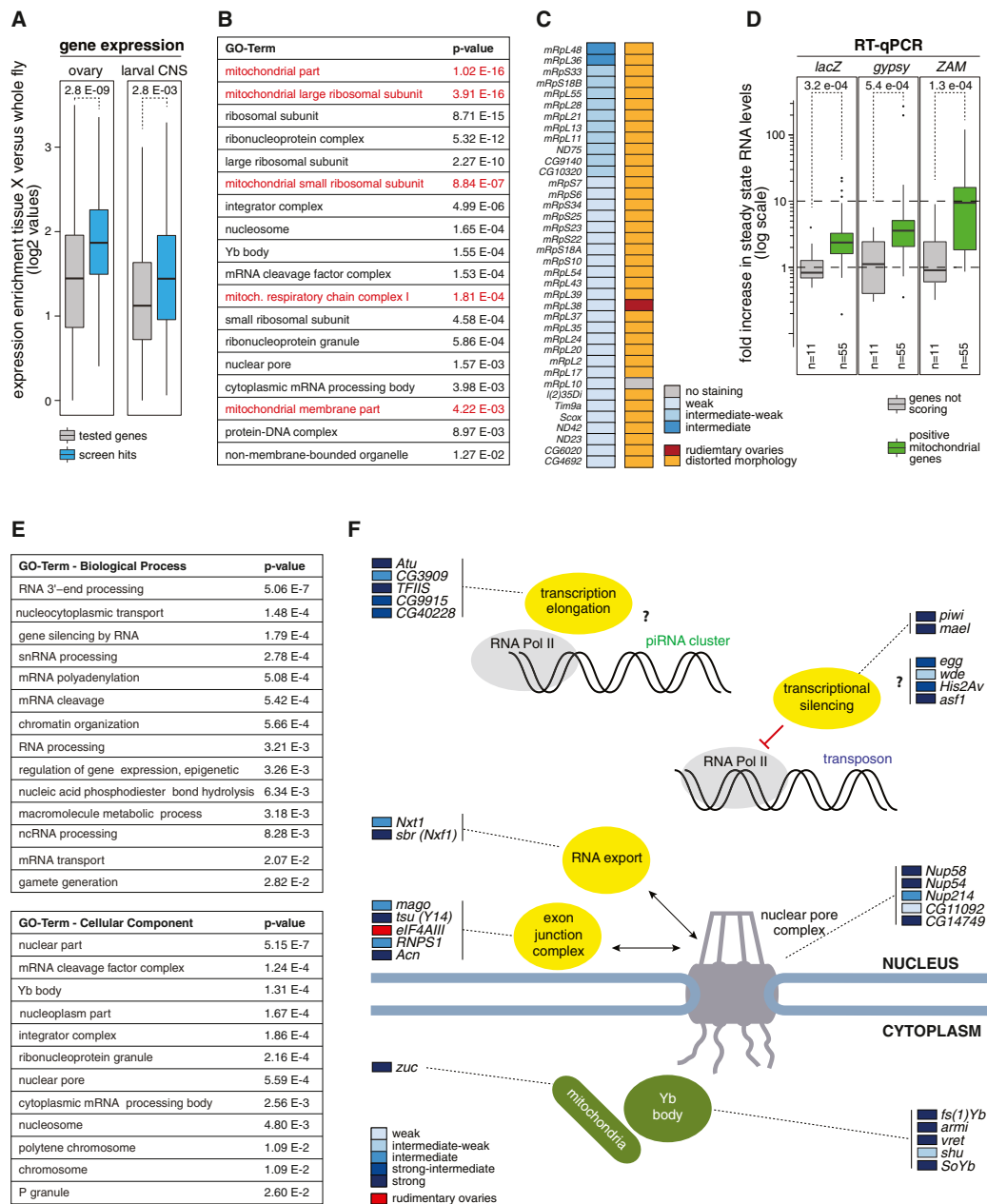


Figure 4. Key Processes and Factors Involved in the Somatic piRNA Pathway

(A) Box plots showing the fold enrichment of gene expression (based on FlyAtlas) in ovaries or larval CNS versus whole flies for all tested genes (gray) and the set of positive screen hits (blue); p values were determined by Wilcoxon signed rank test.

(B) Shown are the most significantly enriched GO terms among the 144 scoring genes in the screen (p values corrected for multiple testing). Mitochondria-related terms are in red.

(C) Listed are all genes with annotated mitochondrial function and their respective *gypsy-lacZ* staining and ovarian morphology phenotypes observed in the somatic screen.

(D) Box plots showing fold increases in *lacZ*, *gypsy*, and *ZAM* steady-state RNA levels for the set of mitochondrial gene knockdowns (green box) compared to nonscoring genes (gray box); p values were determined by Wilcoxon signed rank test. Box plots are defined in Figure 2H.

(E) Significantly enriched GO terms among screen hits without mitochondria associated genes (p values corrected for multiple testing).

(F) Cartoon depicting functional groups of genes identified in the screen. Factors were grouped and placed into nucleus or cytoplasm based on their annotated functions or by identification of orthologous genes with annotated functions. Genes involved in the various processes are indicated together with the *gypsy-lacZ* staining results.

The extent of TE derepression was comparable to knockdowns of *Armi*, an essential piRNA biogenesis factor. We confirmed these findings with independent shRNA transgenes (Figure 2G, not shown). In OSCs, siRNA-mediated knockdown of CG2183 or CG9754 led to strong increases in RNA levels of *gypsy* and *mdg1*, both of which are piRNA targets in this cell line (Figure 5B). Both genes classified also as essential piRNA pathway factors in germline cells. We based this on the *Burdock-lacZ* sensor results and RT-qPCR analyses of the germline-repressed TEs *HeT-A* and *blood* (Figure 5C). Overall, knockdown of these two genes resulted in highly similar defects in TE repression.

However, when analyzing the central effector protein Piwi as well as piRNA levels, clear differences between the two factors emerged. While depletion of CG2183 in follicle cells or germline cells resulted in a near loss of Piwi, depletion of CG9754 had no detectable impact on Piwi levels or localization (Figures 5D and 5E). Multiple piRNA biogenesis pathways feed into the three PIWI family proteins in ovaries, complicating analyses of piRNA biogenesis. We therefore probed for an involvement of CG2183 or CG9754 in piRNA biogenesis in OSCs, which exhibit only primary piRNA biogenesis. Judged by northern analysis, levels of mature piRNAs were reduced to ~25% upon CG2183 depletion but were unchanged upon CG9754 depletion (Figure 5F). We sequenced small RNA populations from CG2183 versus GFP siRNA-treated OSCs. This confirmed a ~4-fold reduction of piRNAs originating from annotated TEs, as well as those originating uniquely from the *flamenco* cluster or from the *traffic jam* 3' UTR (Figure S1). Based on these results, CG2183 is a likely primary piRNA biogenesis factor.

CG9754, on the other hand, is to our knowledge most probably an uncharacterized factor essential for the Piwi-mediated silencing process. Cells depleted for CG9754 exhibited nuclear Piwi loaded with piRNAs, yet the Piwi-RISC was not able to elicit efficient TE silencing. In support for this, GFP-CG9754 localized to the nucleus (where Piwi-mediated silencing occurs) while GFP-CG2183 localized to the cytoplasm (where piRNA biogenesis occurs) both in OSCs as well as in ovaries (Figures 5G and 5H). Both GFP transgenes include the respective endogenous regulatory regions. The GFP fusion proteins were clearly detectable in ovarian somatic and germline cells (GFP-CG2183 was more abundant in germline cells) in support of both proteins being essential piRNA pathway factors in both pathways.

CG2183 Colocalizes with Zucchini on Mitochondria and Is an Essential piRNA Biogenesis Factor

The subcellular localization of GFP-CG2183 was highly reminiscent of mitochondrial patterns and of the reported Zuc localization (Saito et al., 2010). In fact, ectopically expressed Zuc-GFP and YFP-CG2183 in OSCs exhibited an indistinguishable pattern that nearly fully overlapped with mitochondria (Figure 6A). Also, in ovaries, Zuc-GFP and GFP-CG2183 colocalized significantly with mitochondria (Figure S2). When analyzing the Zuc and CG2183 polypeptides with TMHMM (Sonnhammer et al., 1998), a prediction tool for transmembrane (TM) helices, a significant TM helix was predicted for the Zuc N terminus (Saito et al., 2010) and the CG2183 C terminus (Figure S2). Deletion of the respective TM helices resulted in a uniform distribution of both

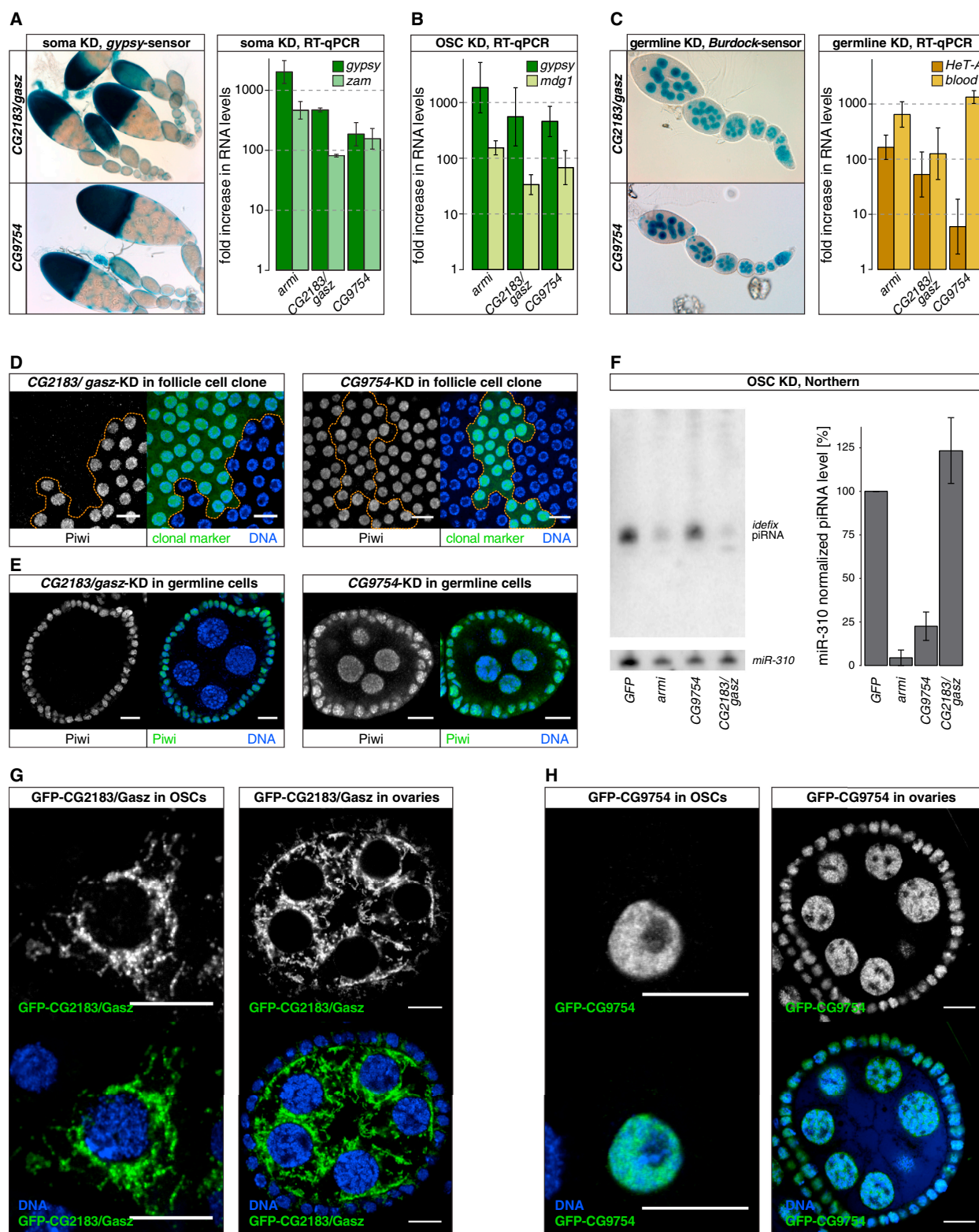
proteins in OSCs (Figure 6B). This indicated that CG2183 is a transmembrane protein like Zuc, most probably of the outer mitochondrial membrane.

Our results suggested a link between CG2183 and Zuc. Indeed, knockdown of CG2183 in the germline resulted in a loss of Piwi, no change in Aub localization, and a mild delocalization of AGO3 from nuage, a pattern indistinguishable from the effects reported upon Zuc knockdown (Figure 6C) (Olivieri et al., 2010). To extend this analysis, we sequenced small RNAs from ovaries with CG2183 germline knockdown and compared piRNA populations to control as well as to Zuc or *Armi* knockdowns (Figure 6E). If compared to the respective control libraries (normalized to miRNA levels), piRNA populations (23–30 nt small RNAs) were reduced to 25%, 20%, or 23% upon knockdown of CG2183, Zuc, or *Armi*, respectively. piRNA populations mapping to the major germline piRNA cluster 42AB collapsed nearly entirely in all three knockdowns (Figure 6E), while piRNAs derived from the soma-specific *flamenco* cluster were essentially unchanged, in agreement with the knockdown being germline specific. The similarities in piRNA populations from CG2183, Zuc, or *Armi* knockdowns extended to TE piRNA profiles. Highly similar patterns of piRNA losses were seen in all three cases (Figure 6F). Zuc and *Armi* are required for primary piRNA biogenesis but not for ping-pong-mediated piRNA biogenesis per se and have been classified as type I piRNA biogenesis factors (Olivieri et al., 2012). Loss of either factor has a highly stereotypical impact on piRNA populations. While piRNAs mapping to most TEs are severely reduced, a handful of TEs exhibit nearly unchanged piRNA levels, and these appear to be almost exclusively produced via ping-pong based on a strong increase in the 10 nt overlap of sense/antisense piRNAs. A detailed ping-pong analysis indicated that CG2183 classifies as a third type I biogenesis factor. Adenosine levels at position 10 of sense piRNAs (an indicator of ping-pong piRNA biogenesis) were increased to almost identical levels in all three knockdowns (Figure 6G). Similarly, ping-pong signatures were elevated for the same subset of TEs in CG2183 or Zuc knockdowns (Figure 6H). The two prototypical TEs *F*-element (increased ping-pong) and *I*-element (collapsed piRNA profile) exemplify the differential impact of loss of type I biogenesis factors on piRNA populations (Figure S2).

Taken together, CG2183 classifies as a type I piRNA biogenesis factor like Zuc and *Armi*, and its subcellular localization on mitochondria hints at an intricate connection to the endonuclease Zuc.

CG2183 Encodes *Drosophila* Gasz and Recruits Armitage to Mitochondria

Protein BLAST searches identified orthologs of CG2183 in mammals, and intriguingly, the mouse ortholog GASZ (germ cell protein with ankyrin repeats, sterile alpha motif, and leucine zipper) has been linked to piRNA biogenesis (Ma et al., 2009). Based on the overall sequence identity of 26%, a similar domain organization (five ankyrin repeats and a sterile alpha motif), a predicted C-terminal transmembrane domain in both proteins, and a demonstrated colocalization of GASZ with mitochondria, we termed CG2183 *Drosophila* Gasz.



(legend on next page)

Ankyrin repeats and sterile alpha motifs often mediate protein-protein interactions. Given the localization of Gasz and the endonuclease Zuc on mitochondria, we speculated that Gasz might serve as an adaptor to recruit other biogenesis factors to mitochondria. We first verified that Gasz and Zuc localize to mitochondria independent of each other, as would be predicted for two proteins harboring TM domains (Figure S3). We then tested whether any of the known piRNA biogenesis factors showed altered subcellular localization upon loss of Gasz. In follicle cells as well as in cultured OSCs, the biogenesis factors Yb, Armi, Vret, and Shu are cytoplasmic and are enriched in perinuclear Yb-bodies that depend on the RNA helicase Yb. Yb-bodies are in close association with mitochondria and are hypothesized to be the sites of primary piRNA biogenesis. We generated clones of cells expressing either *gasz* or *zuc* dsRNA hairpins in the follicular epithelium and stained for known biogenesis factors (Figures 7A and 7B). As reported previously, cells with Zuc knockdowns exhibited significantly enlarged Yb-bodies (evidenced by the accumulation of Armi and Vret). This likely results from a clustering of mitochondria into one or two perinuclear spheres per cell (Figure 7B) (Olivieri et al., 2010, 2012). Cells depleted for Gasz showed no defects in Yb-body localization of Yb, Armi, or Vret. Vret levels were elevated in Yb-bodies, but Armi levels were not, indicating no general enlargement of Yb-bodies and presumably also no clustering of mitochondria (Figure 7A). The enlarged Yb-bodies in Zuc mutant cells also stain positive for Piwi, indicating that presumably unloaded Piwi is recruited to Yb-bodies/mitochondria and cannot be released in the absence of Zuc (Olivieri et al., 2010; Saito et al., 2010). Interestingly, loss of Gasz also resulted in Piwi accumulation in Yb-bodies, albeit to a lesser degree, probably because of no apparent mitochondrial clustering (Figures 7C and 7D; Figure S4). These results once more highlighted the similarities between Zuc and Gasz, yet they did not provide any clues toward the molecular function of Gasz.

Drosophila somatic ovarian cells are special, as here piRNA biogenesis depends on the soma-specific RNA Helicase Yb,

the defining factor of Yb-bodies (Szakmary et al., 2009). We hypothesized that localization of biogenesis factors to Yb-bodies might not be the relevant assay to test whether Gasz serves as an adaptor to recruit biogenesis factor(s) to the mitochondrial surface. We therefore analyzed the localization of piRNA biogenesis factors in germline cells (which lack Yb-bodies but do perform primary piRNA biogenesis) upon depletion of Zuc or Gasz. Striking differences were observed for the RNA Helicase Armi. In wild-type nurse cells, Armi localized to the cytoplasm and was enriched in diffuse “clouds,” which colocalized with mitochondria (Figure 7E). Remarkably, Armi accumulated even more in areas of mitochondrial clusters upon knockdown of Zuc but was evenly dispersed and lacked mitochondrial colocalization in cells depleted for Gasz (Figures 7F and 7G). Given that Piwi accumulated around mitochondrial clusters in Zuc-depleted but not in Gasz-depleted cells (Figure S5), we propose that Gasz serves as the adaptor protein that recruits an Armi-Piwi complex to mitochondria, where the endonucleolytic activity of Zuc generates the 5' end of pre-piRNAs that will be loaded into Piwi.

Our results define a second mitochondrial transmembrane protein in primary piRNA biogenesis and provide evidence that Gasz couples biochemical reactions occurring at the mitochondrial surface (e.g., Zuc-mediated cleavage) with the upstream sorting of piRNA precursors.

DISCUSSION

The mechanistic understanding of the piRNA pathway is very limited, and systematic studies are hampered due to its gonad-restricted activity. Therefore, our *in vivo* RNAi screen provides an invaluable entry point into the dissection of this silencing system. In the following, we critically assess two key aspects that are relevant to this work.

Specificity and Validity of the Screen Results

Transgenic RNAi screens based on the expression of long dsRNA hairpins must be rigorously evaluated in terms of their

Figure 5. CG2183/Gasz and CG9754 Are Essential for the Somatic piRNA Pathway

(A) Left panels show β -gal stainings of ovarioles as readout for *gypsy-lacZ* silencing upon soma knockdown of *CG2183/gasz* or *CG9754*. The bar diagram depicts fold increases in RNA levels of indicated TEs in ovaries with soma knockdown of *armi*, *CG2183/gasz*, or *CG9754* (averages of three biological replicates; error bars, SD; normalized to control knockdown).

(B) Displayed are fold increases in RNA levels of indicated TEs in OSCs upon siRNA-mediated knockdown of *armi*, *CG2183/gasz*, or *CG9754* siRNA in OSCs (averages of three biological replicates; error bars, SD; normalized to control knockdown).

(C) Left panels show β -gal stainings of ovarioles as readout for *Burdock-lacZ* silencing upon germline knockdown of *CG2183/gasz* or *CG9754*. The bar diagram depicts fold increases in RNA levels of indicated TEs in ovaries with germline knockdown of *armi*, *CG2183/gasz*, or *CG9754* (averages of three biological replicates; error bars, SD; normalized to control knockdown).

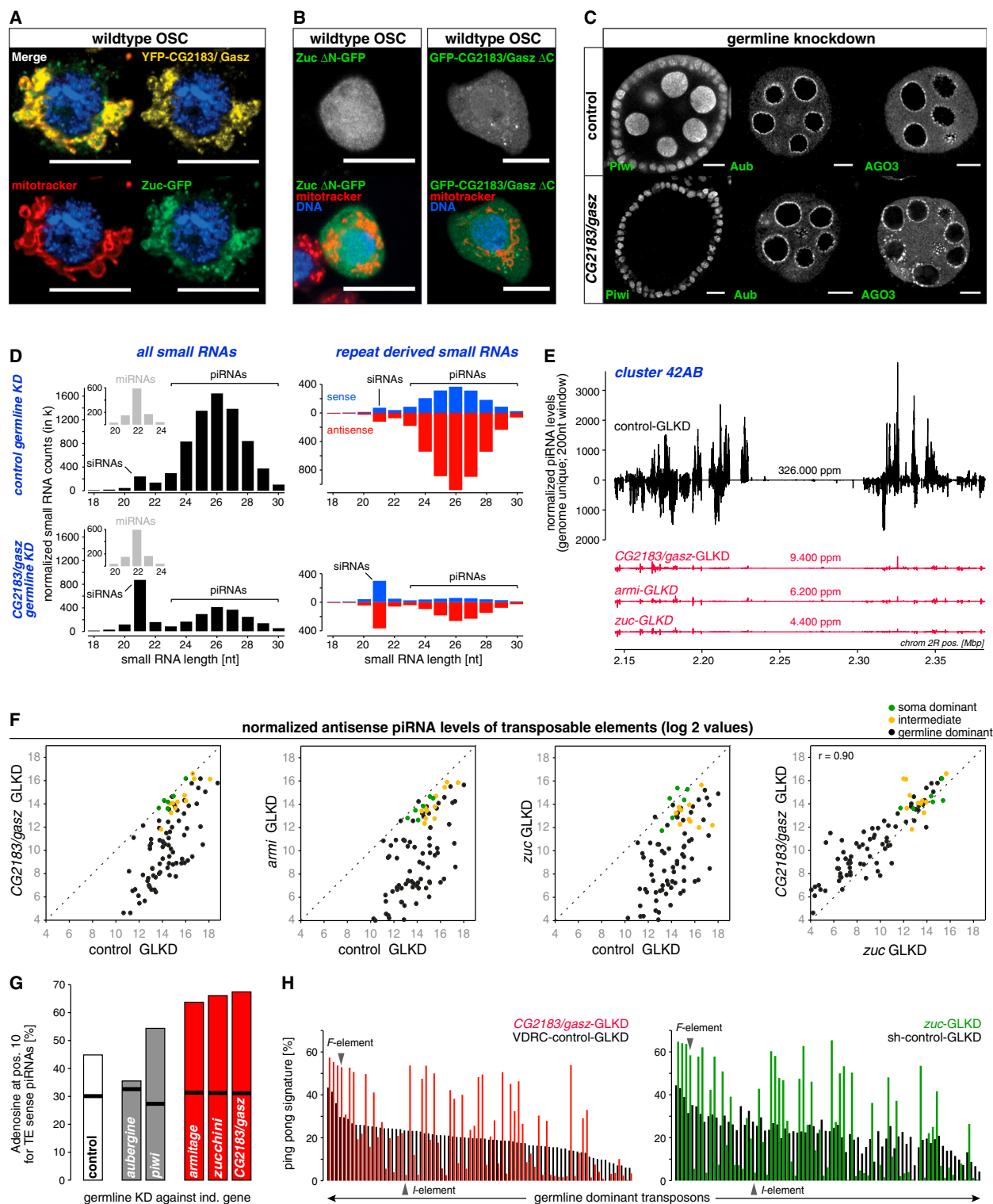
(D) Confocal sections (scale bars, 10 μ m) through the follicular epithelium of egg chambers stained for Piwi (monochrome panel), DNA (blue), and the clonal marker (green). Cells within the clone (dashed lines mark clone boundaries in the monochrome panels) express dsRNAs against *CG2183/gasz* (left panel) or *CG9754* (right panel).

(E) Confocal sections (scale bars, 10 μ m) through egg chambers stained for Piwi (monochrome panel and green) and DNA (blue). Knockdown of *CG2183/gasz* (left) and *CG9754* (right) was specifically activated in germline cells.

(F) Northern blot analysis of piRNA levels in OSC total RNA upon siRNA-mediated knockdowns of *GFP* (control), *armi*, *CG9754*, or *CG2183/gasz*. One representative blot probed for *idexif* piRNA (top) and then reprobated for *miR-310* (bottom) is shown. The bar diagram indicates quantified results (normalized to *miR-310*) based on three independent experiments (error bars, SD).

(G) Confocal sections (scale bars, 10 μ m) of OSCs (left panels) or egg chambers (right panels) expressing GFP-*CG2183/Gasz* stained for DNA (blue). Monochrome panels show the GFP signal separately.

(H) Confocal sections (scale bars, 10 μ m) of OSCs (left panels) or egg chambers (right panels) expressing GFP-*CG9754* stained for DNA (blue). Monochrome panels show the GFP signal separately. See also Figure S1.



(legend on next page)

false negative and false positive rates. The false negative rate is influenced by the availability of VDRC lines (94% for this screen), by the number of genes that cannot be evaluated due to overall defects in ovarian development (10% for this screen), and by the knockdown efficiency of the VDRC lines toward their respective target genes (estimated to be ~90% if two lines per gene were tested). Based on this, we conservatively estimate that our screen captured up to 80% of the key factors involved in the somatic piRNA pathway. We point out that the *gypsy-lacZ* sensor faithfully recapitulates defects at all levels of the piRNA pathway, such as *flamenco* mutations, piRNA biogenesis defects, or defects in piRNA-mediated silencing.

Our screen is relatively blind toward genes that are essential for cellular viability and that therefore could not be evaluated. It is clear that at some point piRNA-specific processes will connect to more general cellular biology. For example, the key heterochromatin protein Su(var)205 (HP1) is likely to be important for piRNA-mediated silencing (Wang and Elgin, 2011), but it could not be evaluated as its knockdown in follicle cells prevented ovarian development.

We identified 144 genes with potential involvement in the somatic piRNA pathway. This set includes (1) true positives, (2) factors with indirect impact on the pathway, and (3) RNAi artifacts. Considering that the *gypsy-lacZ* sensor picks up even moderate defects in TE silencing, the number of factors with indirect impacts will be significant. For example, at least 38 factors can be clearly assigned to aspects of mitochondrial metabolism. As expected for such “indirect” factors, these scored weakly in the screen. We note that the identification of processes or factors that impact the piRNA pathway indirectly might still be valuable. For example, mitochondrial integrity and/or number likely influence TE silencing, as at least two piRNA biogenesis factors are mitochondrial transmembrane proteins. Mitochondrial function is diminished upon aging. As aging has been linked to increased mobility of TEs, it is tempting to speculate that aspects of this might be linked to modulations of piRNA pathway activity.

False positives due to screening artifacts are common for transgenic RNAi screens. They stem from off-target effects of generated siRNAs but can also arise from gain-of-function con-

ditions of genes located in the vicinity of the VDRC transgene insertion. Both aspects, however, are expected to impact our screen only marginally. First, the set of core factors acting in the pathway is unlikely to exceed 30–40 genes, making off-target effects less likely. Second, the screen was based on manual evaluation of a transgenic reporter assay, a highly specific phenotype. And third, only very few genes (if any) will impair TE silencing if overexpressed. These arguments are strongly supported by an experimental validation of nearly all of the 30 tested candidates by an independent RNAi line.

Taken together, our screen represents a highly enriched set of somatic piRNA pathway genes. We speculate that ~30 factors are either core factors of the somatic piRNA pathway or have direct impact on its function. This testifies to the remarkable complexity of the piRNA pathway compared to the siRNA or miRNA pathways.

Differences and Commonalities between Somatic and Germline piRNA Pathways

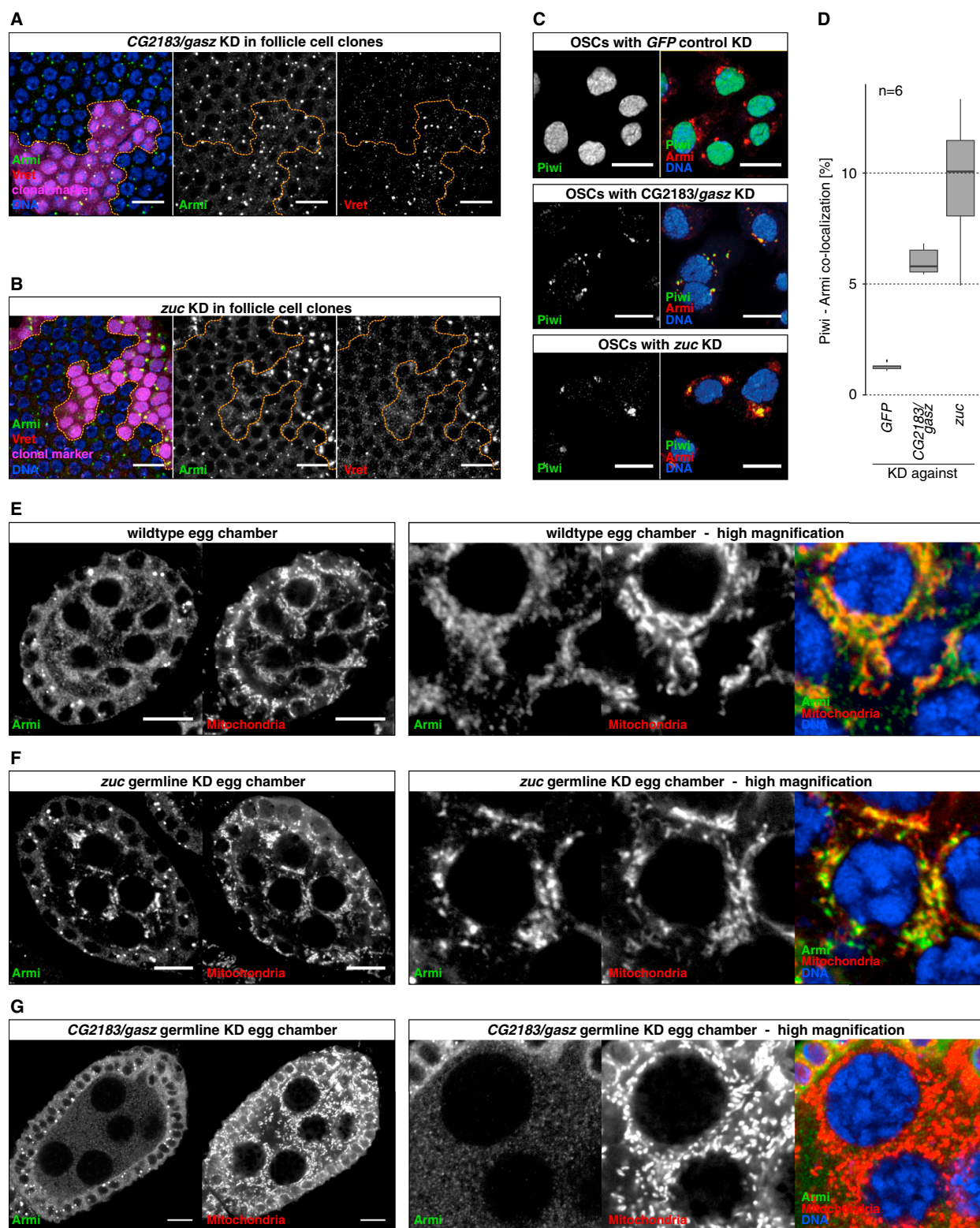
At least two key differences distinguish the piRNA pathways in somatic or germline cells of the *Drosophila* ovary. First, expression of Aub and AGO3 as well as the process of secondary piRNA biogenesis is germline specific (Li et al., 2009; Malone et al., 2009). Second, only germline piRNA clusters are transcribed in both directions. Most of the germline-specific piRNA pathway factors listed in Figure 3E are involved in either of these two aspects. These are Aub/AGO3 and ping-pong factors such as Spindle-E or Vasa on the one side and Rhino and Cutoff on the other side (Klattenhoff et al., 2009; Li et al., 2009; Malone et al., 2009; Pane et al., 2011).

Primary piRNA biogenesis and Piwi-mediated transcriptional silencing are much more likely to be similar between the two cell types, consistent with the literature (Handler et al., 2011; Olivieri et al., 2012; Preall et al., 2012; Sienski et al., 2012; Wang and Elgin, 2011; Zamparini et al., 2011). Several of the newly identified “common factors” acting in both cell types are likely to classify into one of these two categories, including Gasz (primary biogenesis) or CG9754 (silencing).

Considering this, it was surprising to find numerous genes with strong involvement in the somatic pathway that were seemingly dispensable for TE silencing in the germline. Many of these

Figure 6. CG2183/Gasz Is an Uncharacterized piRNA Biogenesis Factor

- (A) Confocal sections (scale bars, 10 μ m) of OSCs transfected with YFP-CG2183/Gasz (yellow) and Zuc-GFP (green) expression constructs and stained for mitochondria (MitoTracker; red) and DNA (blue).
- (B) Confocal sections (scale bars, 10 μ m) of OSCs transfected with Zuc-GFP and GFP-CG2183/Gasz expression constructs (green) lacking the respective transmembrane domains and stained for mitochondria (MitoTracker; red) and DNA (blue).
- (C) Confocal sections (scale bars, 10 μ m) through egg chambers stained for Piwi, Aub, or AGO3. Control knockdown or CG2183/gasz knockdown was specifically activated in germline cells.
- (D) Left panels show length profiles of normalized small RNA populations from ovaries with control (upper) or CG2183/gasz (lower) germline knockdown split into miRNAs (small insets) and remaining RNAs (siRNA and piRNA populations indicated). Right panels show respective length profiles of repeat-derived small RNAs only (red, antisense; blue, sense).
- (E) Normalized piRNA profiles (genome unique; sense up, antisense down) from ovaries with indicated germline knockdowns mapping to cluster 42AB.
- (F) Scatter plots showing normalized antisense piRNA levels (log2 values) mapping to soma-dominant (green), intermediate (yellow), or germline-dominant (black) TEs from ovaries with indicated germline knockdowns (Pearson correlation [r] based on all TEs).
- (G) Bar chart displaying the adenosine content at position 10 for sense piRNAs mapping to TEs isolated from ovaries with indicated germline knockdowns. Black lines indicate the expected level based on the average 10A content at positions 2–9 and 11–23.
- (H) Shown are ping-pong signatures of germline-dominant TEs based on piRNAs from ovaries with indicated germline knockdowns. TEs were ordered according to their ping-pong signature in the VDRC control library. See also Figure S2.



(legend on next page)

“soma-specific” factors can be grouped into three major processes: nuclear RNA export, the nuclear pore complex, and transcriptional elongation. We believe that these processes are either carried out by related factors in germline cells (e.g., Nxf2 and Nxf3 are germline-specific RNA export factors) or that they are indeed required specifically in somatic cells.

All in all, the identified somatic piRNA pathway factors and the established genetic and cell-biological tools will advance investigations on the *Drosophila* piRNA pathway. Given that several of the key processes and involved factors are conserved in vertebrates, our data will also influence studies in the model systems zebrafish and mouse.

EXPERIMENTAL PROCEDURES

Drosophila Stocks

Fly stocks are listed in the [Supplemental Information](#).

RT-qPCR Analysis

Primer sequences and details are given in the [Supplemental Information](#).

shRNA Transgenes

shRNA lines were cloned into the Valium-20 vector (Ni et al., 2011) modified with a white selection marker and integrated into the attP2 landing site reported in Groth et al., 2004. Hairpin sequences are listed in the [Supplemental Information](#).

Small RNA Cloning

Small RNA cloning and sequencing was performed as in Jayaprakash et al., 2011. Details are given in the [Supplemental Information](#).

Antibodies

α -Piwi, α -Aub, and α -AGO3 (rabbit) were described in Brennecke et al., 2007; α -Armi, α -Vret, and α -Yb (rabbit) were described in Handler et al., 2011; α -Piwi and α -Armi (mouse) were described in Saito et al., 2010.

Immunocytochemistry

IF staining of ovaries and OSCs was done as in Olivieri et al., 2010. All primary antibodies were used at 1:500.

Northern Blot

Total RNA was isolated from respective knockdowns and separated on a 15% Urea-PAA gel. After transfer onto a membrane, radioactively labeled probes were hybridized over night. Probe sequences and details are given in the [Supplemental Information](#).

Cell Culture

OSCs were cultured as described as in Niki et al., 2006 and transfected with Cell Line Nucleofector Kit V (Amaxa Biosystems; program T-029). Details are given in the [Supplemental Information](#).

ACCESSION NUMBERS

All Illumina data sets (Table S4) were deposited at GEO (GSE45894).

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures, nine tables, and Supplemental Experimental Procedures and can be found with this article online at <http://dx.doi.org/10.1016/j.molcel.2013.04.031>.

ACKNOWLEDGMENTS

We thank Brennecke lab members for support and discussions. We thank the VDRC for transgenic RNAi lines; the CSF NGS unit for Illumina sequencing; the IMP/IMBA bio-optics unit for support; P. Ducheck and S. Lopez for transgenic lines; R. Sachidanandam, D. Jurczak, A. Stark, D. Gerlach, and R. Maier for bioinformatics support; and M. Siomi for OSCs, α -Armi, and α -Piwi. The Brennecke laboratory is supported by the Austrian Academy of Sciences, the European Community's 7th Framework Programme (FP7/2007-2013; ERC grant # 260711EU), and the Austrian Science Fund (FWF; Y 510-B12). D.H. and J.B. planned the experiments; D.H., K.M., M.P., and K.L. performed the screen. C.S. and F.S.G. generated genetic tools; D.H. performed all experiments. J.B. and D.H. analyzed the data and wrote the paper.

Received: February 5, 2013

Revised: April 5, 2013

Accepted: April 5, 2013

Published: May 9, 2013

REFERENCES

- Aravin, A.A., van der Heijden, G.W., Castañeda, J., Vagin, V.V., Hannon, G.J., and Bortvin, A. (2009). Cytoplasmic compartmentalization of the fetal piRNA pathway in mice. *PLoS Genet.* 5, e1000764.
- Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128, 1089–1103.
- Chintapalli, V.R., Wang, J., and Dow, J.A. (2007). Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nat. Genet.* 39, 715–720.
- Czech, B., Preall, J.B., McGinn, J., and Hannon, G.J. (2013). A Transcriptome-wide RNAi Screen in the *Drosophila* Ovary Reveals Factors of the Germline piRNA Pathway. *Mol. Cell* 50, Published online June 6, 2013. <http://dx.doi.org/10.1016/j.molcel.2013.04.031>.
- Dietzl, G., Chen, D., Schnorrer, F., Su, K.C., Barinova, Y., Fellner, M., Gasser, B., Kinsey, K., Oppel, S., Scheiblaue, S., et al. (2007). A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature* 448, 151–156.
- Ghildiyal, M., and Zamore, P.D. (2009). Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.* 10, 94–108.
- Groth, A.C., Fish, M., Nusse, R., and Calos, M.P. (2004). Construction of transgenic *Drosophila* by using the site-specific integrase from phage ϕ C31. *Genetics* 166, 1775–1782.
- Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* 315, 1587–1590.
- Haase, A.D., Fenoglio, S., Muerdter, F., Guzzardo, P.M., Czech, B., Pappin, D.J., Chen, C., Gordon, A., and Hannon, G.J. (2010). Probing the initiation

Figure 7. CG2183/Gasz Recruits Armitage to Mitochondria

(A and B) Confocal sections (scale bars, 10 μ m) through the follicular epithelium of egg chambers stained for Armi (green), Vret (red), and DNA (blue). Knockdown of CG2183/gasz (A) or *zuc* (B) was clonally induced (clonal marker in magenta; clone borders marked by dashed line). (C) Confocal sections (scale bars, 10 μ m) of OSCs with indicated siRNA-mediated knockdowns stained for Piwi (green), Armi (red), and DNA (blue). (D) Quantification of Piwi-Armi colocalization based on (C). The fraction of Piwi-positive pixels colocalizing with Armi-positive pixels is indicated. Box plots are based on six quantified images per knockdown with ~30 cells each. (E–G) Confocal sections (scale bars, 10 μ m) through egg chambers with indicated genotype stained for Armi (green), mitochondria (red), and DNA (blue). Overview panels are shown to the left and high magnification images to the right. Colocalization of Armi and mitochondria in the merge panels appears yellow. See also Figures S3–S5.

- p>and effector phases of the somatic piRNA pathway in
- Drosophila*
- .
- Genes Dev.*
- 24, 2499–2504.
- Handler, D., Olivieri, D., Novatchkova, M., Gruber, F.S., Meixner, K., Mechtler, K., Stark, A., Sachidanandam, R., and Brennecke, J. (2011). A systematic analysis of *Drosophila* TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. *EMBO J.* 30, 3977–3993.
- Horwich, M.D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., and Zamore, P.D. (2007). The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr. Biol.* 17, 1265–1272.
- Hoskins, R.A., Smith, C.D., Carlson, J.W., Carvalho, A.B., Halpern, A., Kaminker, J.S., Kennedy, C., Mungall, C.J., Sullivan, B.A., Sutton, G.G., et al. (2002). Heterochromatic sequences in a *Drosophila* whole-genome shotgun assembly. *Genome Biol.* 3, H0085.
- Ipsaro, J.J., Haase, A.D., Knott, S.R., Joshua-Tor, L., and Hannon, G.J. (2012). The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. *Nature* 491, 279–283.
- Jayaprakash, A.D., Jhabdo, O., Brown, B.D., and Sachidanandam, R. (2011). Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Res.* 39, e141.
- Kaminker, J.S., Bergman, C.M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D.A., Lewis, S.E., Rubin, G.M., et al. (2002). The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biol.* 3, H0084.
- Klattenhoff, C., Bratu, D.P., McGinnis-Schultz, N., Koppetsch, B.S., Cook, H.A., and Theurkauf, W.E. (2007). *Drosophila* rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. *Dev. Cell* 12, 45–55.
- Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Khurana, J.S., Zhang, F., Schultz, N., Koppetsch, B.S., Nowosielska, A., et al. (2009). The *Drosophila* HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell* 138, 1137–1149.
- Le Thomas, A., Rogers, A.K., Webster, A., Marinov, G.K., Liao, S.E., Perkins, E.M., Hur, J.K., Aravin, A.A., and Tóth, K.F. (2013). Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev.* 27, 390–399.
- Levin, H.L., and Moran, J.V. (2011). Dynamic interactions between transposable elements and their hosts. *Nat. Rev. Genet.* 12, 615–627.
- Li, C., Vagin, V.V., Lee, S., Xu, J., Ma, S., Xi, H., Seitz, H., Horwich, M.D., Syrzycka, M., Honda, B.M., et al. (2009). Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* 137, 509–521.
- Ma, L., Buchold, G.M., Greenbaum, M.P., Roy, A., Burns, K.H., Zhu, H., Han, D.Y., Harris, R.A., Coarfa, C., Gunaratne, P.H., et al. (2009). GASZ is essential for male meiosis and suppression of retrotransposon expression in the male germline. *PLoS Genet.* 5, e1000635.
- Malone, C.D., and Hannon, G.J. (2009). Small RNAs as guardians of the genome. *Cell* 136, 656–668.
- Malone, C.D., Brennecke, J., Dus, M., Stark, A., McCombie, W.R., Sachidanandam, R., and Hannon, G.J. (2009). Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell* 137, 522–535.
- Ni, J.Q., Zhou, R., Czech, B., Liu, L.P., Holderbaum, L., Yang-Zhou, D., Shim, H.S., Tao, R., Handler, D., Karpowicz, P., et al. (2011). A genome-scale shRNA resource for transgenic RNAi in *Drosophila*. *Nat. Methods* 8, 405–407.
- Niki, Y., Yamaguchi, T., and Mahowald, A.P. (2006). Establishment of stable cell lines of *Drosophila* germ-line stem cells. *Proc. Natl. Acad. Sci. USA* 103, 16325–16330.
- Nishida, K.M., Okada, T.N., Kawamura, T., Mituyama, T., Kawamura, Y., Inagaki, S., Huang, H., Chen, D., Kodama, T., Siomi, H., and Siomi, M.C. (2009). Functional involvement of Tudor and dPRMT5 in the piRNA processing pathway in *Drosophila* germlines. *EMBO J.* 28, 3820–3831.
- Nishimasu, H., Ishizu, H., Saito, K., Fukuhara, S., Kamatani, M.K., Bonnefond, L., Matsumoto, N., Nishizawa, T., Nakanaga, K., Aoki, J., et al. (2012). Structure and function of Zucchini endoribonuclease in piRNA biogenesis. *Nature* 491, 284–287.
- Olivieri, D., Sykora, M.M., Sachidanandam, R., Mechtler, K., and Brennecke, J. (2010). An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in *Drosophila*. *EMBO J.* 29, 3301–3317.
- Olivieri, D., Senti, K.A., Subramanian, S., Sachidanandam, R., and Brennecke, J. (2012). The cochaperone shutdown defines a group of biogenesis factors essential for all piRNA populations in *Drosophila*. *Mol. Cell* 47, 954–969.
- Pane, A., Jiang, P., Zhao, D.Y., Singh, M., and Schüpbach, T. (2011). The Cutoff protein regulates piRNA cluster expression and piRNA production in the *Drosophila* germline. *EMBO J.* 30, 4601–4615.
- Preall, J.B., Czech, B., Guzzardo, P.M., Muerdter, F., and Hannon, G.J. (2012). shutdown is a component of the *Drosophila* piRNA biogenesis machinery. *RNA* 18, 1446–1457.
- Qi, H., Watanabe, T., Ku, H.Y., Liu, N., Zhong, M., and Lin, H. (2011). The Yb body, a major site for Piwi-associated RNA biogenesis and a gateway for Piwi expression and transport to the nucleus in somatic cells. *J. Biol. Chem.* 286, 3789–3797.
- Rozhkov, N.V., Hammell, M., and Hannon, G.J. (2013). Multiple roles for Piwi in silencing *Drosophila* transposons. *Genes Dev.* 27, 400–412.
- Saito, K., Inagaki, S., Mituyama, T., Kawamura, Y., Ono, Y., Sakota, E., Kotani, H., Asai, K., Siomi, H., and Siomi, M.C. (2009). A regulatory circuit for piwi by the large Maf gene traffic jam in *Drosophila*. *Nature* 461, 1296–1299.
- Saito, K., Ishizu, H., Komai, M., Kotani, H., Kawamura, Y., Nishida, K.M., Siomi, H., and Siomi, M.C. (2010). Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in *Drosophila*. *Genes Dev.* 24, 2493–2498.
- Sarot, E., Payen-Groschêne, G., Bucheton, A., and Pélisson, A. (2004). Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the *Drosophila melanogaster* flamenco gene. *Genetics* 166, 1313–1321.
- Senti, K.A., and Brennecke, J. (2010). The piRNA pathway: a fly's perspective on the guardian of the genome. *Trends Genet.* 26, 499–509.
- Sienski, G., Dönertas, D., and Brennecke, J. (2012). Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* 151, 964–980.
- Sims, R.J., 3rd, Belotserkovskaya, R., and Reinberg, D. (2004). Elongation by RNA polymerase II: the short and long of it. *Genes Dev.* 18, 2437–2468.
- Siomi, M.C., Sato, K., Pezic, D., and Aravin, A.A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. *Nat. Rev. Mol. Cell Biol.* 12, 246–258.
- Slotkin, R.K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* 8, 272–285.
- Sonnhammer, E.L., von Heijne, G., and Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 6, 175–182.
- Szakmary, A., Reedy, M., Qi, H., and Lin, H. (2009). The Yb protein defines a novel organelle and regulates male germline stem cell self-renewal in *Drosophila melanogaster*. *J. Cell Biol.* 185, 613–627.
- Tange, T.O., Nott, A., and Moore, M.J. (2004). The ever-increasing complexities of the exon junction complex. *Curr. Opin. Cell Biol.* 16, 279–284.
- Wang, S.H., and Elgin, S.C. (2011). *Drosophila* Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. *Proc. Natl. Acad. Sci. USA* 108, 21164–21169.
- Zamparini, A.L., Davis, M.Y., Malone, C.D., Vieira, E., Zavadil, J., Sachidanandam, R., Hannon, G.J., and Lehmann, R. (2011). Vreteno, a gonad-specific protein, is essential for germline development and primary piRNA biogenesis in *Drosophila*. *Development* 138, 4039–4050.
- Zhang, F., Wang, J., Xu, J., Zhang, Z., Koppetsch, B.S., Schultz, N., Vreken, T., Meignin, C., Davis, I., Zamore, P.D., et al. (2012). UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. *Cell* 151, 871–884.

Molecular Cell, Volume 50

Supplemental Information

The Genetic Makeup of the *Drosophila* piRNA Pathway

Dominik Handler, Katharina Meixner, Manfred Pizka, Kathrin Lauss, Christopher
Schmied, Franz Sebastian Gruber, and Julius Brennecke

Figure S1.

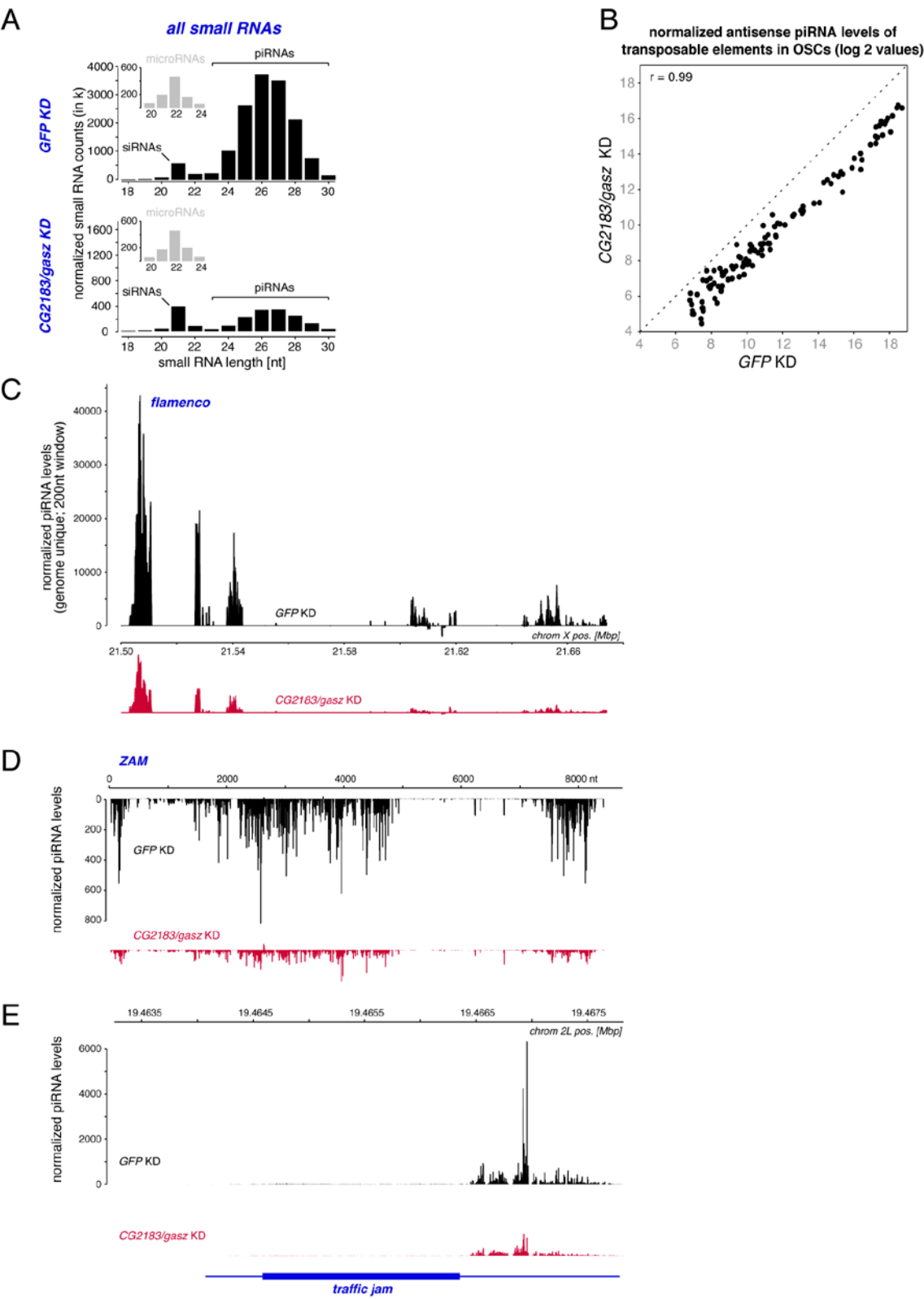


Figure S1. (relates to main Figure 5)

(A) Shown are length profiles of small RNAs (normalized to total miRNA counts) from OSCs transfected with siRNAs against *GFP* (upper panel) or *CG2183/gasz* (lower panel). Small RNAs were split into miRNAs (small insets) and remaining RNAs (siRNAs and piRNAs).

(B) Scatter plot showing log₂ values of normalized antisense piRNA levels isolated from OSCs transfected with *GFP* or *CG2183/gasz* siRNAs mapping to annotated TEs in OSCs.

(C) Shown are normalized profiles of genome unique piRNAs (sense up; antisense down) from *GFP* or *CG2183/gasz* siRNA knockdowns in OSCs mapping to the *flamenco* cluster.

(D, E) Shown are normalized profiles of piRNAs (sense up; antisense down) from *GFP* or *CG2183/gasz* siRNA knockdowns in OSCs mapping to the retro-element *ZAM* **(D)** or the traffic jam (tj) locus **(E)**.

Figure S2.

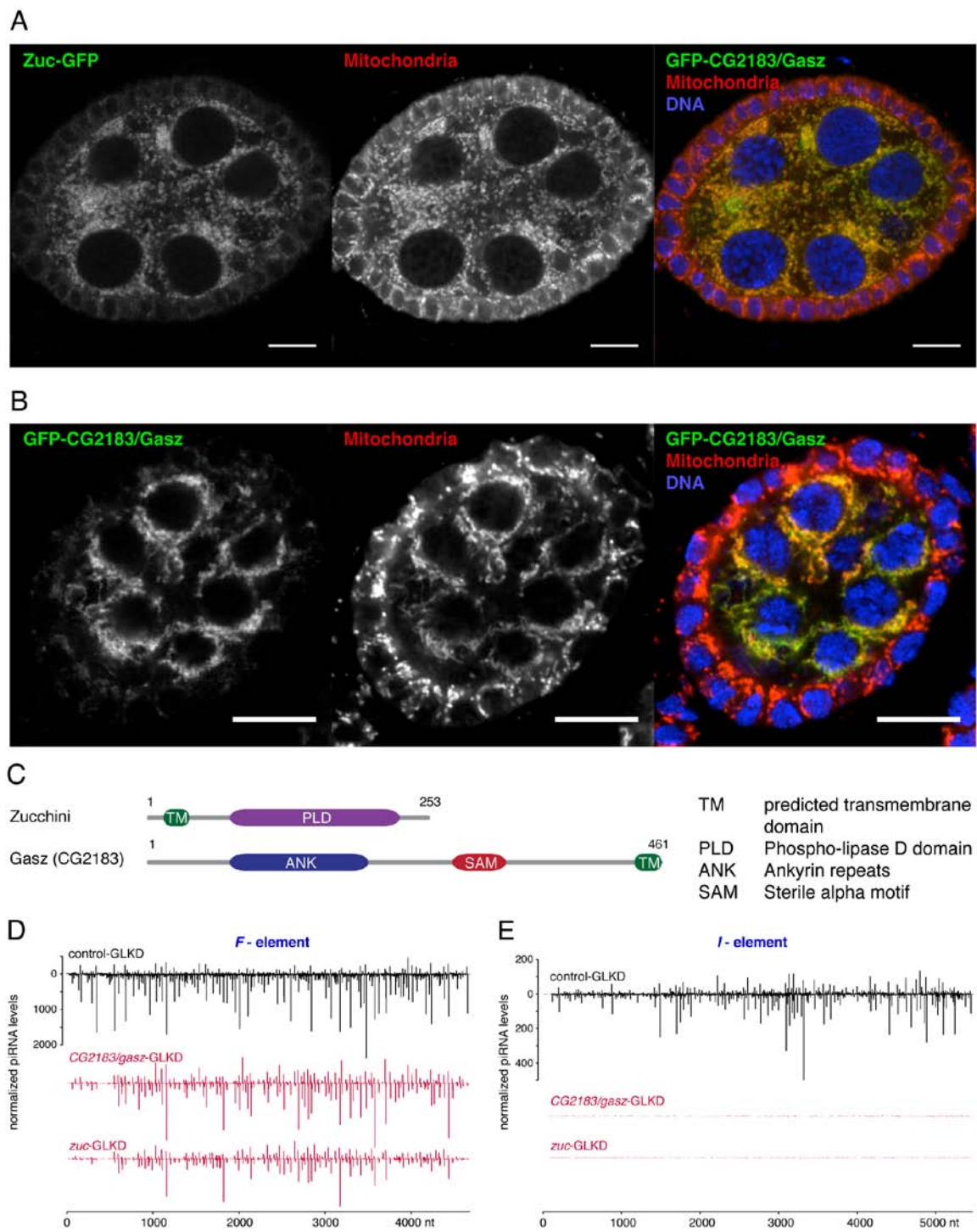


Figure S2. (relates to main Figure 6)

(A, B) Confocal sections of egg chambers expressing GFP tagged Zucchini **(A)** or GFP tagged CG2183/Gasz **(B)** stained for mitochondria (red) and DNA (blue).

(C) Cartoon depicting the predicted domain architecture of Zucchini and CG2183/Gasz. Shown are the full-length protein sequences annotated with their protein domains. Transmembrane helices were predicted with TMHMM.

(D) Cartoon depicting the predicted domain architecture of Zucchini and CG2183/Gasz. Shown are the full-length protein sequences annotated with their protein domains. Transmembrane helices were predicted with TMHMM.

(E-F) Shown are normalized piRNA profiles (sense up; antisense down) from ovaries with indicated germline knockdowns mapping to the *F*-element **(E)** or to the *I*-element **(F)**.

Figure S3.

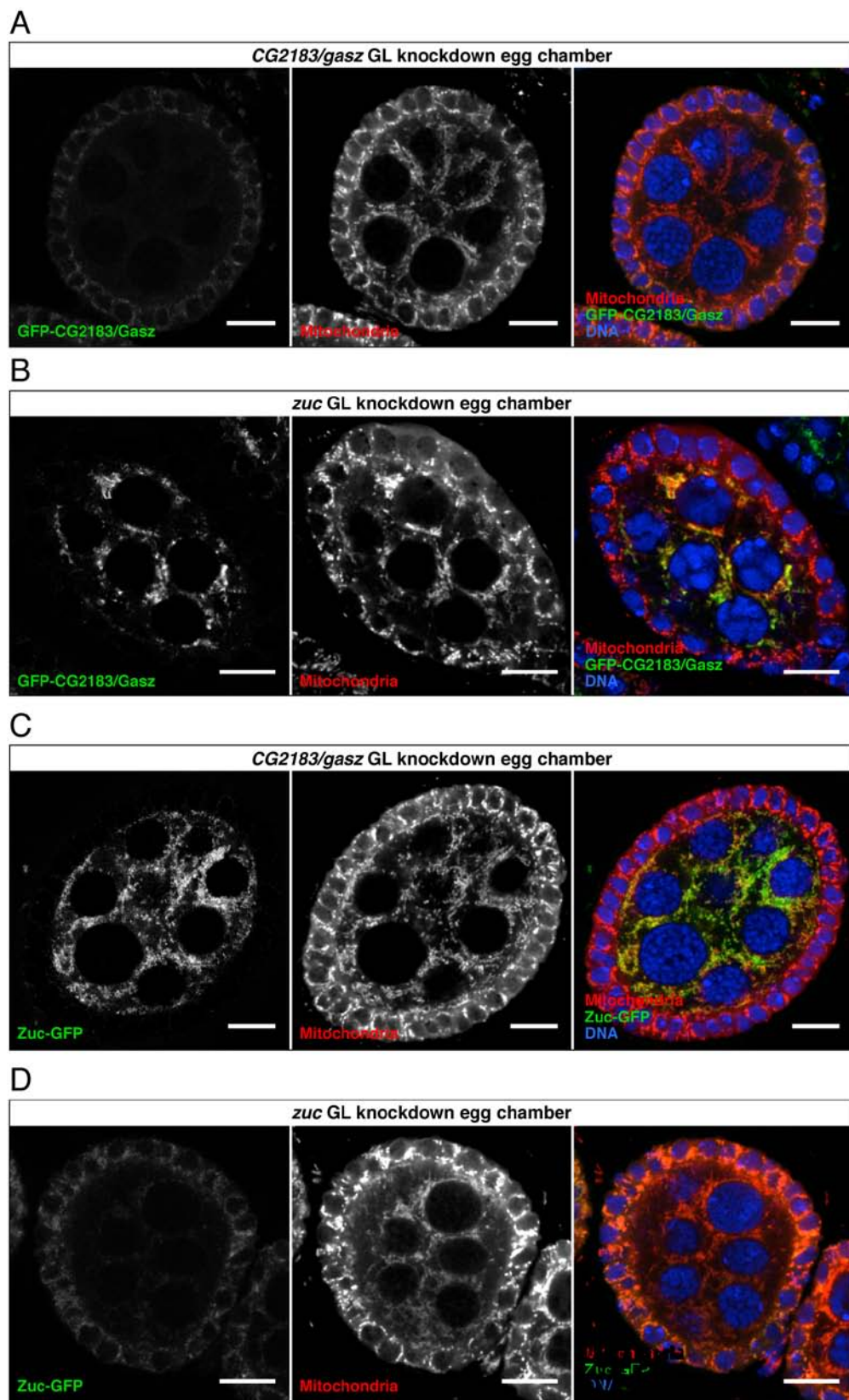


Figure S3. (relates to main Figure 7)

(A, B) Confocal sections of egg chambers expressing GFP-CG2183/Gasz stained for mitochondria (red) and DNA (blue). Shown are germline specific knockdowns for *CG2183/gasz* **(A)** or *zuc* **(B)**.

(C, D) Confocal sections of egg chambers expressing Zuc-GFP stained for mitochondria (red) and DNA (blue). Shown are germline specific knockdowns for *CG2183/gasz* **(C)** or *zuc* **(D)**.

Figure S4.

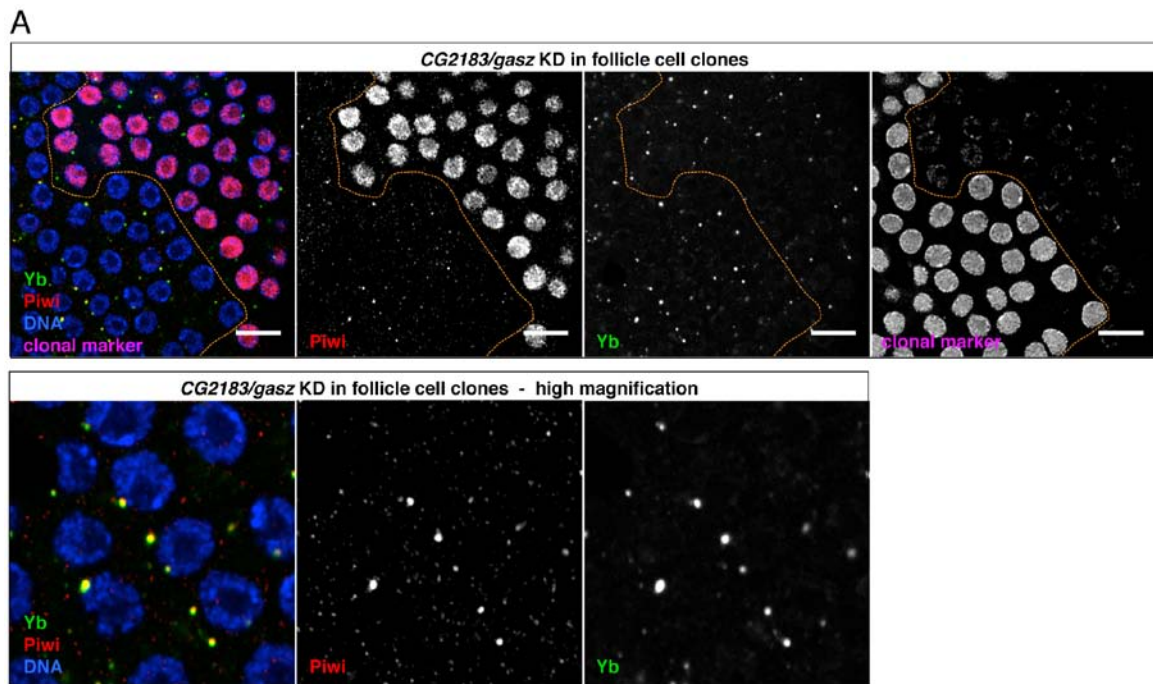


Figure S4. (relates to main Figure 7)

(A) The upper panels show a confocal section through the follicular epithelium of an egg chamber stained for Yb (green), Piwi (red) and DNA (blue). Knockdown of *CG2183/gasz* has been clonally induced (clonal marker in magenta; outlined by dashed line). The lower panels represent high magnification images confirming co-localization of Piwi and Yb.

Figure S5.

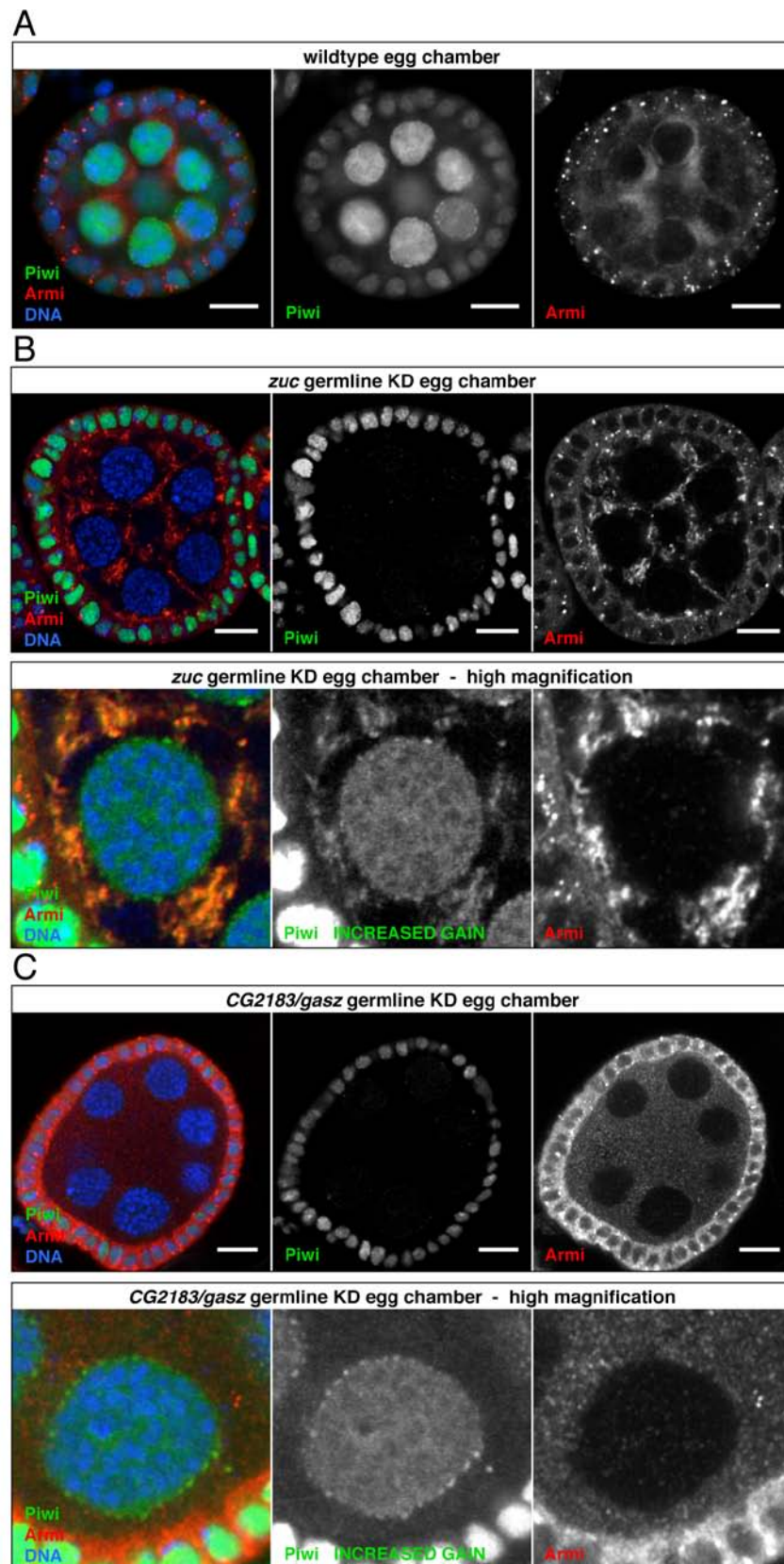


Figure S5. (relates to main Figure 7)

(A-C) Confocal sections of egg chambers expressing GFP-Piwi and stained for Armi (red) and DNA (blue). Shown are egg chambers from wildtype flies **(A)**, germline specific *zuc* knockdown **(B)** or germline specific *CG2183/gasz* knockdown **(C)**. For panels **(B)** and **(C)** high magnification images with increased gain for the Piwi channel are shown in addition.

Table S4. Illumina sequencing data sets used in this study

Origin	Method	Genotype	GEO
Flies	small RNA-seq	<i>UAS-Dcr-2; NGT; nosGAL4, Burdock-lacZ</i> x <i>w[1118]</i>	GSE45894
		<i>UAS-Dcr-2; NGT; nosGAL4, Burdock-lacZ</i> x VDRC-13762	GSE45894
		MTD x shRNA- <i>armi</i>	GSE38728
		MTD x shRNA- <i>zuc</i>	GSE38728
		MTD x <i>w[1118]</i>	GSE38728
OSC	RNA-seq	Wildtype (replicate #1)	GSE45894
		Wildtype (replicate #2)	GSE45894
	small RNA-seq	si-GFP KD	GSE45894
		si- <i>CG2183/gasz</i> KD	GSE45894

Table S5. Used fly stocks

MTD-GAL4 (Ni et al., 2011)
<i>y,w,hsFlp122;;act<CD2>GAL4, UAS-GFP</i> (Olivieri et al., 2010)
Short hairpin RNA (shRNA) lines were cloned into the Valium-20 or the Valium-22 vector (Ni et al, 2011) modified with a white selection marker and integrated into the <i>attp2</i> landing site (Markstein et al., 2008). Hairpin sequences, see Table S5
<i>armi</i> shRNA, <i>SoYb</i> shRNA, <i>BoYb</i> shRNA, (Handler et al., 2011)
<i>zuc</i> shRNA (TRiP # GL00111);
eGFP_CG2183/Gasz and eGFP_CG9754 were cloned by inserting N-terminal eGFP via bacterial recombineering into genomic rescue constructs and integrated into the <i>attp2</i> landing site.
eGFP_Zuc was cloned by inserting C-terminal eGFP via bacterial recombineering into a genomic rescue construct that was integrated into the <i>attp2</i> landing site.
<i>flam^(R);;gypsy-lacZ</i> (Sarot et al., 2004)
<i>flam^(P);;gypsy-lacZ</i> (Sarot et al., 2004)
<i>tj-GAL4, gypsy-lacZ/Cyo</i> ;
<i>UAS-Dcr-2; NGT; nosGAL4, Burdock-lacZ/Tm3,Ser</i>
All flies have been aged 5-7 days at 25°C before analysis

Table S6. Primer for QPCR analysis

Gene	Sequence
<i>rp49</i>	fw: CCGCTTCAAGGGACAGTATCTG rv: ATCTCGCCGCAGTAAACGC
<i>lacZ</i>	fw: AATGTTGATGAAAGCTGGCTAC rv: GCTCAGGTCAAATTCAGACG
<i>zam</i>	fw: ACTTGACCTGGATACACTCACAAC rv: GAGTATTACGGCGACTAGGGATAC
<i>gypsy</i>	fw: CAACAATCTGAACCCACCAATCT rv: TATGAACATCATGAGGGTGAACG
<i>HeT-A</i>	fw: CGCGCGGAACCCATCTTCAGA rv: CGCCGCAGTCGTTTGGTGAGT
<i>Burdock</i>	fw: CGGTAAAATCGCTTCATGGT rv: ACGTTGCATTTCCCTGTTTC
<i>Act5c</i>	fw: AAGTTGCTGCTCTGGTTGTCG rv: GCCACACGCAGCTCATTGTAG
<i>mdg1</i>	fw: AAACCTCCAACCTCCCAATCcG rv: AGTGGTCCCTCGCAGTCGTT
<i>blood</i>	fw: AACCAATAGAAAGAAGCCACCGAAC rv: AGTCATGGACTATTGAGGGTGTTG

Table S7. Oligos used for cloning of shRNA constructs

Gene	Sequence
	Cloned into W20
<i>CG9754</i>	fw: ctagcagtAACGGCTACGCTGTACAAGAAatggtatattcaagcataTTCTTGACAGCGTAGCCGTTgcg rv: aattcgcAACGGCTACGCTGTACAAGAAatggttgaatataactaTTCTTGACAGCGTAGCCGTTactg
<i>Acn</i>	fw: ctagcagtACCGCCTATCAGACTACTAGAtagttatattcaagcataTCTAGTAGTCTGATAGGCGGTgcg rv: aattcgcACCGCCTATCAGACTACTAGAtatgcttgaatataactaTCTAGTAGTCTGATAGGCGGTactg
<i>Tj11S</i>	fw: ctagcagtCACGGATATGAAGTACAAGAAatggtatattcaagcataTTCTTGACTTCATATCCGTGgcg rv: aattcgcCACGGATATGAAGTACAAGAAatgcttgaatataactaTTCTTGACTTCATATCCGTGactg
<i>CG3893</i>	fw: ctagcagtGCCGTGTGATCTACAAAGACatagttatattcaagcataTGCTTTGTAGATCACACGGCGcg rv: aattcgcGCCGTGTGATCTACAAAGACatgcttgaatataactaTGCTTTGTAGATCACACGGCactg
<i>Nup58</i>	fw: ctagcagtACCACAAGGAGCACTACCTAatagttatattcaagcataTTAGGTAGTGCTCCTTGTGGTgcg rv: aattcgcACCACAAGGAGCACTACCTAatgcttgaatataactaTTAGGTAGTGCTCCTTGTGGTactg
<i>CG2183</i>	fw: ctagcagtTCCCTTGTTTCATTGACTTCAAtagttatattcaagcataTTGAAGTCAATGAACAAGGGAagcg rv: aattcgcTCCCTTGTTTCATTGACTTCAAtatgcttgaatataactaTTGAAGTCAATGAACAAGGGAactg
<i>sbr</i>	fw: ctagcagtAAGCGATGCTCTCCATATCAatagttatattcaagcataTTGATATGGAGAGCATCGCTTgcg rv: aattcgcAAGCGATGCTCTCCATATCAAtatgcttgaatataactaTTGATATGGAGAGCATCGCTTactg
<i>Atu</i>	fw: ctagcagtCCGCTCGTAGTAAGAGATCAatagttatattcaagcataTTGATCTCTTACTACGAGCGGgcg rv: aattcgcCCGCTCGTAGTAAGAGATCAAtatgcttgaatataactaTTGATCTCTTACTACGAGCGGactg
<i>Nup54</i>	fw: ctagcagtAGCGAAGATACTTGAACATAAatggtatattcaagcataTTATGTTCAAGTATCTTCGCTgcg rv: aattcgcAGCGAAGATACTTGAACATAAatgcttgaatataactaTTATGTTCAAGTATCTTCGCTactg
<i>asf1</i>	fw: ctagcagtCCGCGTGGGCTACTATGTGAatagttatattcaagcataTTCACATAGTAGCCACGCGGgcg rv: aattcgcCCGCGTGGGCTACTATGTGAAtatgcttgaatataactaTTCACATAGTAGCCACGCGGactg
<i>CG9915</i>	fw: ctagcagtGTGGTTCTAGACGTTCTAGAAatggtatattcaagcataTTCTAGAACGCTCTAGAACCAcgcg rv: aattcgcGTGGTTCTAGACGTTCTAGAAatgcttgaatataactaTTCTAGAACGCTCTAGAACCAactg
<i>Patr-1</i>	fw: ctagcagtCACGCCGTATATCGACGACTAtagttatattcaagcataTAGTCGTCGATATACGCGTGgcg rv: aattcgcCACGCCGTATATCGACGACTAtatgcttgaatataactaTAGTCGTCGATATACGCGTGactg
<i>CG40228</i>	fw: ctagcagtCACCTCCGAAGAGAAAGAATAtagttatattcaagcataTATTCTTTCTCTTCGGAGGTGgcg rv: aattcgcCACCTCCGAAGAGAAAGAATAtatgcttgaatataactaTATTCTTTCTCTTCGGAGGTGactg
<i>CG3689</i>	fw: ctagcagtACCAACGATAGCGCTACCTAatagttatattcaagcataTTAGGTAGCGCTATCGTTGGTgcg rv: aattcgcACCAACGATAGCGCTACCTAatgcttgaatataactaTTAGGTAGCGCTATCGTTGGTactg
<i>mago</i>	fw: ctagcagtCACCACCTCAAGCATAGCCTAtagttatattcaagcataTAGGCTATGCTTGAGGTGGTGgcg rv: aattcgcCACCACCTCAAGCATAGCCTAtatgcttgaatataactaTAGGCTATGCTTGAGGTGGTGactg
<i>RnpS1</i>	fw: ctagcagtACCCAGTTCAATAGGTTCAatagttatattcaagcataTTGAACCTATTGAAGTGGGGTgcg rv: aattcgcACCCAGTTCAATAGGTTCAAtatgcttgaatataactaTTGAACCTATTGAAGTGGGGTactg
<i>gw</i>	fw: ctagcagtTAGCGCTTCTAACTTAACTAatagttatattcaagcataTTAGTTAAGTTAGAAGCGCTAgcg rv: aattcgcTAGCGCTTCTAACTTAACTAAtatgcttgaatataactaTTAGTTAAGTTAGAAGCGCTAactg
<i>Nxt1</i>	fw: ctagcagtCTCCGTCAAGTTTCGCAGATCAatagttatattcaagcataTGATCTGCGAAGTTGACGGAGgcg rv: aattcgcCTCCGTCAAGTTTCGCAGATCAatgcttgaatataactaTGATCTGCGAAGTTGACGGAGactg
<i>ball</i>	fw: ctagcagtACGGCGGTTGATTACGTACAtagttatattcaagcataTGACGTGAATCAACCGCCGTgcg rv: aattcgcACGGCGGTTGATTACGTACAtatgcttgaatataactaTGACGTGAATCAACCGCCGTactg
<i>pcm</i>	fw: ctagcagtACCGACGAACGAGACTATAGAtagttatattcaagcataTCTATAGTCTCGTTCGTCGGTgcg rv: aattcgcACCGACGAACGAGACTATAGAtatgcttgaatataactaTCTATAGTCTCGTTCGTCGGTactg
<i>CG8211</i>	fw: ctagcagtCTGGTGCTCAACTATTCGATAtagttatattcaagcataTATCGAATAGTTGAGCACCAGgcg rv: aattcgcCTGGTGCTCAACTATTCGATAtatgcttgaatataactaTATCGAATAGTTGAGCACCAGactg
<i>Actr13E</i>	fw: ctagcagtGCGACGCGAAGTCACAGTTGAtagttatattcaagcataTCAACTGTGACTTCGCGTCGCGcg rv: aattcgcGCGACGCGAAGTCACAGTTGAtatgcttgaatataactaTCAACTGTGACTTCGCGTCGCGactg
<i>veli</i>	fw: ctagcagtCACGAGAAGGCCGTAGAGCTAtagttatattcaagcataTAGCTCTACGGCCTTCTCGTGcg rv: aattcgcCACGAGAAGGCCGTAGAGCTAtatgcttgaatataactaTAGCTCTACGGCCTTCTCGTGactg
<i>SRPK</i>	fw: ctagcagtCACCTTGAAAGGACAATCAAatagttatattcaagcataTTTGATTGTCCTTTCAAGGTGcg rv: aattcgcCACCTTGAAAGGACAATCAAatgcttgaatataactaTTTGATTGTCCTTTCAAGGTGactg
<i>CG5859</i>	fw: ctagcagtCCGCAATGCGTTGCTATGTAAatagttatattcaagcataTTACATAGCAACGCATTGCGGcg rv: aattcgcCCGCAATGCGTTGCTATGTAAatgcttgaatataactaTTACATAGCAACGCATTGCGGactg
<i>SelD</i>	fw: ctagcagtCAGTACCAAGATGACAGAGAAatagttatattcaagcataTTCTCTGTCATCTTGGTACTGcg rv: aattcgcCAGTACCAAGATGACAGAGAAatgcttgaatataactaTTCTCTGTCATCTTGGTACTGactg
<i>CG6020</i>	fw: ctagcagtCAGGGACTTCGAGACCAAGAAatagttatattcaagcataTTCTTGGTCTCGAAGTCCCTGcg rv: aattcgcCAGGGACTTCGAGACCAAGAAatgcttgaatataactaTTCTTGGTCTCGAAGTCCCTGactg
<i>TSG101</i>	fw: ctagcagtCCGAGTTCAGGAGAAAGTTAatagttatattcaagcataTTAACTTTCTCCTGAACCTCGGcg rv: aattcgcCCGAGTTCAGGAGAAAGTTAatgcttgaatataactaTTAACTTTCTCCTGAACCTCGGactg

<i>spi</i>	fw: ctacgagcCAAGGAGATCGACAATACTTAtagttatattcaagcataTAAGTATTGTCGATCTCCTTGgcg rv: aattcgcCAAGGAGATCGACAATACTTAtatgcttgaatataactaTAAGTATTGTCGATCTCCTTGactg
<i>mr</i>	fw: ctacgagcATCGATGGTTGTAGACATATAtagttatattcaagcataTATATGTCTACAACCATCGATgcg rv: aattcgcATCGATGGTTGTAGACATATAtatgcttgaatataactaTATATGTCTACAACCATCGATactg
	Cloned into W22
<i>armi</i>	fw: ctacgagcTTCGCTGTCAAGCTAAGTTTAtagttatattcaagcataTAACTTAGCTTGACAGCGAAgcg rv: aattcgcTTCGCTGTCAAGCTAAGTTTAtatgcttgaatataactaTAACTTAGCTTGACAGCGAAactg
<i>krimp</i>	fw: ctacgagcCAGATTGGGAGACTACGAATAtagttatattcaagcataTATTCGTAGTCTCCCAATCTGgcg rv: aattcgcCAGATTGGGAGACTACGAATAtatgcttgaatataactaTATTCGTAGTCTCCCAATCTGactg
<i>AGO3</i>	fw: ctacgagcTTGGTTGGAAATATTGTCTTAtagttatattcaagcataTAAGACAATATTTCCAACCAAgcg rv: aattcgcTTGGTTGGAAATATTGTCTTAtatgcttgaatataactaTAAGACAATATTTCCAACCAAactg
<i>SoYb</i>	fw: ctacgagcAAGGTTCAAAGTATCAGCGAAtagttatattcaagcataTTCGCTGA TACTTTGAACCTTgcg rv: aattcgcAAGGTTCAAAGTATCAGCGAAatgcttgaatataactaTTCGCTGATACTTTGAACCTTactg
<i>BoYb</i>	fw: aattcgcCAGCATCAGTTGTGCGATCAAtagttatattcaagcataTTGATCGCACAACTGATGCTGactg rv: ctacgagcCAGCATCAGTTGTGCGATCAAtagttatattcaagcataTTGATCGCACAACTGATGCTGgcg
	Other
<i>zuc</i>	TRiP.GL0111 / Bloomington stock: 35227

Table S8. Probes used for Northern Blots

Name	Sequence
<i>Idefix</i>	AAACTACTGGCAATCGTTTGGGAA
<i>miR-310</i>	AAAGGCCGGGAAGTGTGCAATA

Table S9. siRNAs used for RNAi in OSCs

Gene	sequence
<i>GFP</i>	Guide: ACUUCAGGGUCAGCUUGCCTT Passenger: GGCAAGCUGACCCUGAAGUTT
<i>armi</i>	Guide: UAAACUUAGCUUGACAGCGTT Passenger: CGCUGUCAAGCUAAGUUUATT
<i>CG9754</i>	Guide: UUCUUGUACAGCGUAGCCGUU Passenger: CGGCUACGCUGUACAAGAAUU
<i>CG2183/gasz</i>	Guide: UUGAAGUCAAUAGAACAAGGGAUU Passenger: CCUUGUUCAUUGACUUCAACGUU
<i>zuc</i>	Guide: UUGUUGUGCAUCAAGUUCGTT Passenger: CGAACUUGAUGCACAACAATT

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

X-Gal staining

Ovaries from 5-7 day old flies were dissected into ice cold PBS (max 30 min), fixed in 0.5% Glutaraldehyde/PBS (RT, 15 min), and washed with PBS. The staining reaction was performed with staining solution (10mM PBS, 1mM MgCl₂, 150 mM NaCl, 3 mM potassium ferricyanide, 3 mM potassium ferrocyanide, 0.1% Triton, 0.1% X-Gal) at room temperature over night (*gypsy* reporter) or for 2 hours (*Burdock* reporter).

Cell culture

Act>GFP-CG9754, Act>GFP-CG2183/Gasz and Act>Zuc-GFP have been constructed by Gateway cloning using full cDNA amplicons of the genes.

Act>Zuc Δ N-GFP and Act>GFP-CG2183/Gasz Δ C have been constructed by gateway cloning. Sequences used for cloning were *zuc* cDNA (bp 116-759) and *CG2183/gasz* cDNA (bp 4-1305)

Transposon QPCR analysis

cDNA was prepared via random priming of 1 μ g total RNA isolated from ovaries of 5-7 day old flies. Quantitative PCR was performed using a homemade QPCR master mix (20 mM Tris pH 8.3, 100 mM KCl, 5 mM MgCl₂, 0.5 mM each dNTP, 1x Evagreen, 40 μ L/ml TAQ).

Each experiment was performed in biological triplicates with technical duplicates. Relative RNA levels were calculated by the 2- Δ CT method (Livak and Schmittgen, 2001) and normalized to rp49 levels. Fold enrichments were calculated in comparison to respective RNA levels obtained from heterozygous flies, from flies not harboring a

knockdown hairpin or from control siRNA transfections into OSCs.

Northern Blot analysis

Total RNA was isolated from respective knockdowns and separated on a 15% polyacrylamide urea gel. RNA was transferred to Amersham Hybond-NX (RPN303T) membrane and crosslinked by EDC (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide) for 1 hour (Pall and Hamilton, 2008). The membrane was pre-hybridized in Church Buffer and hybridized to probes overnight at 37°C. The membrane was washed 3 times 10 minutes with 2xSSC, 0.1% SDS and exposed.

small RNA cloning

Small RNA cloning and sequencing was performed as described (Jayaprakash et al., 2011). In brief, 20 µg of total RNA was isolated from ovaries or OSCs by TRIzol and Phenol/Chloroform extraction, was resolved on a denaturing polyacrylamide gel and RNAs corresponding to 18-28 nt were isolated and subjected to ligations of 3'-, and 5'-adaptors followed by reverse transcription and PCR amplification; libraries were sequenced on GAII or HiSeq2000 platforms (Illumina).

The RNA cloning strategy introduces 4 random nucleotides at 3' end of the 5' linker and 5' end of the 3' linker, which reduces ligation biases (Jayaprakash et al., 2011). Reads were first stripped of the 3' adaptor and then the introduced 4 random nucleotides at each end of the read were removed.

Reads were mapped to the genome (100% match; release 5). For piRNA cluster mapping we considered genome-unique mappers, for TE mappings (Repbase; (Jurka et al., 2005) all mappers (up to 3 MM) have been considered. Libraries were normalized to 1 Mio

miRNA reads. Small RNAs mapping to rRNAs, tRNAs and snoRNAs were excluded. The calculation of TE piRNA levels was based on antisense piRNAs. Ping-pong signatures were calculated as previously described (Malone et al., 2009).

RNA sequencing (RNA-seq)

mRNA from wildtype OSCs was selected with Dynabeads Oligo(dT) (Invitrogen) from total RNA, fragmented and reverse transcribed with random hexamers. Strand-specific libraries were prepared using the UDG-digestion-based strategy, cloned with NEBNext ChIP-Seq Library Prep Reagent Set for Illumina (NEB) and sequenced on a Genome Analyzer II (Illumina).

This yielded ~6–20 million genome- and transcriptome-mappable reads. For the computational analyses, we first extracted high quality bases from every read (6–36 nt) and mapped these to the *Drosophila* genome as well as to the FlyBase transcriptome. Uniquely aligned reads were used for quantification of gene expression levels according to coordinates in the Flybase gene annotation (r5.38) by calculating RPKM values.

Gene expression enrichment analysis

Expression data was obtained from FlyAtlas (Chintapalli et al., 2007). For genes tested multiple times an average value was calculated. Ratios between expression values of individual tissues and the value for “whole fly” have been calculated for all genes. Tissue enrichment analysis was performed using Wilcoxon Rank Sum and Signed Rank Test implementation in R. The lists used for statistical analysis contained all tested genes (background set) or all genes scoring positively in the screen.

KEGG tem analysis

KEGG analysis was done with DAVID (Huang et al., 2009b; 2009a) online tool. All tested genes have been used as background list, whereas genes showing a ‘no ovary’ phenotype were used for analysis. Presented P-values have been corrected for multiple testing using the Benjamini-Hochberg method.

GO term analysis

GO analysis was performed using GOrilla (Eden et al., 2007; 2009) online tool using two ranked lists of genes. As target set all positively scoring or all positively scoring genes depleted for mitochondrial genes have been used. All genes tested in the screen have been used as background list. Presented P-values have been corrected for multiple testing using the Benjamini-Hochberg method.

Statistical analyses

We used statistical packages implemented in R 2.15.0 for all calculations and plots in this study. For data visualization in box plot format we used standard features of ggplot2 boxplot function: horizontal bar represents median, the box depicts 25th and 75th percentile (lower and upper quartile respectively), whiskers represent sample minimum (lower) and maximum (upper); outliers are shown as circles. Statistical significances in Fig. 4D and Fig. 7D were calculated with Wilcoxon Rank Sum and the Signed Rank Test implementation in R.

Image quantification

The images were copied to a 50% downsampled map where the initial segmentation was performed. The Armitage and DAPI channels were added together to get a good representation of the whole cell and a multiresolution segmentation was used to define object borders. By evaluating intensity and standard deviation those objects belonging to cells were defined and classified. Objects of the cell class were merged, the borders refined by growing using surface tension and then the objects were synchronized back onto the main map. From within the cell class a quantile of pixel intensities of 95% from the Armi channel was calculated and used as a threshold to segment the Yb-bodies. The sum of intensities from the Piwi channel within Yb-bodies was calculated and divided by the sum of intensities of the surrounding cytoplasm. This gives a relative measure of how much of the Piwi signal colocalized with Yb-bodies.

SUPPLEMENTAL REFERENCES

- Chintapalli, V.R., Wang, J., and Dow, J.A.T. (2007). Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nat Genet* 39, 715–720.
- Eden, E., Lipson, D., Yogev, S., and Yakhini, Z. (2007). Discovering motifs in ranked lists of DNA sequences. *PLoS Comput. Biol.* 3, e39.
- Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10, 48.
- Handler, D., Olivieri, D., Novatchkova, M., Gruber, F.S., Meixner, K., Mechtler, K., Stark, A., Sachidanandam, R., and Brennecke, J. (2011). A systematic analysis of *Drosophila* TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. *The EMBO Journal*.
- Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37, 1–13.
- Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4, 44–57.
- Jayaprakash, A.D., Jabado, O., Brown, B.D., and Sachidanandam, R. (2011). Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Res* 39, e141.
- Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 110, 462–467.
- Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2^{(-Delta Delta C(T))} Method. *Methods* 25, 402–408.
- Malone, C.D., Brennecke, J., Dus, M., Stark, A., McCombie, W.R., Sachidanandam, R., and Hannon, G.J. (2009). Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell* 137, 522–535.
- Markstein, M., Pitsouli, C., Villalta, C., Celniker, S.E., and Perrimon, N. (2008). Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. *Nat Genet* 40, 476–483.
- Olivieri, D., Sykora, M.M., Sachidanandam, R., Mechtler, K., and Brennecke, J. (2010). An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in *Drosophila*. *The EMBO Journal* 29, 3301–3317.
- Pall, G.S., and Hamilton, A.J. (2008). Improved northern blot method for enhanced detection of small RNA. *Nat Protoc* 3, 1077–1084.

Sarot, E., Payen-Groschêne, G., Bucheton, A., and Péliisson, A. (2004). Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the *Drosophila melanogaster* flamenco gene. *Genetics* 166, 1313–1321.

5.2 The exon junction complex is required for definition and excision of neighboring introns in *Drosophila*

In collaboration with Rippei Hayashi I investigated the role of the exon junction complex (EJC) in the *Drosophila* piRNA pathway. We selected the EJC for further characterization because multiple EJC components scored strongly in the screen. Also, EJC components have been shown to be important for the expression of heterochromatic genes and we therefore hypothesized that the EJC might play an important role in piRNA cluster transcription or nuclear export of cluster transcripts. Surprisingly, we instead found that the function of the EJC in the piRNA pathway is to ensure proper splicing of *piwi* mRNA. This allowed us to dissect the role that the EJC plays in splicing.

In the following, I will first give a short introduction into the biology of the exon junction complex.

The EJC is a multi-protein complex that is deposited upstream of exon-exon boundaries by the splicing machinery⁹⁴. Three proteins, eIF4AIII, Tsunagi and Mago, build the core of the complex. The complex is directly bound to the RNA ~26nt upstream of exon junctions^{95,96}. RNA binding is sequence independent and is mediated by the DEAD-box domain of eIF4AIII. After export of the fully spliced and poly-adenylated mRNA, the fourth core component, Barentz (Btz), binds to the EJC⁹⁷.

Deposition of the core EJC at exon-exon boundaries serves as a binding platform for an array of accessory factors⁹⁸⁻¹⁰⁰. Several functions are assigned to the EJC and these seem to vary with the different associated factors. Among the functions of the EJC are nonsense mediated mRNA decay (co-factor: Upf3¹⁰¹), nuclear mRNA export (co-factors: UAP56¹⁰², Aly¹⁰³ and NXF1¹⁰³) and mRNA quality control (co-factors: Acinus (Acn)¹⁰⁴, RnpS1^{105,106} and SAP18¹⁰⁷).

Although most actions of the EJC occur post splicing, the EJC has been shown to also interact with proteins involved in splicing^{108,109}. This caused the question, if the EJC is actually involved directly in splicing. Indeed it was shown that the nuclear core (eIF4AIII, Mago and Tsu) of the EJC and the auxiliary factor RnpS1 are necessary for the faithful splicing of several mRNAs in *Drosophila*^{110,111}. Loss of any of these factors leads to the exclusion of exons in mRNAs displaying atypical large introns (so-called exon-skipping). Affected genes are often located in heterochromatin and contain several large introns. The most prominent gene showing defects is *rolled* which encodes for the signaling kinase ERK^{110,111}.

The EJC core components Mago, Nashi (Mago) and Tsunagi (Tsu) as well as the accessory factors RnpS1 and Acinus (Acn) scored as strong hits in the screen. Follow-up experiments showed that eIF4AIII is also required for a functional piRNA pathway. Depletion of this factor in all follicle cells resulted in a non-interpretable phenotype during the screen as

ovarian development was blocked. However, eIF4AIII depletion in clones of cells confirmed its involvement in the piRNA pathway.

Based on several lines of evidence we could show that the EJC is critically involved in splicing of the *piwi* pre-mRNA. In contrast to the literature where only exon skipping phenotypes were reported for EJC deficient cells^{110,111}, we observed intron-inclusion of the fourth intron of *piwi*. Our work also demonstrated for the first time a clear involvement of the accessory factor Acinus in EJC mediated splicing processes.

To dissect the mechanism behind the splicing defects we generated transgenic flies where we modified intron 4 of *piwi*. We were able to show, that the poor poly-pyrimidine tract of *Drosophila piwi* intron 4 in combination with the extension of intron 4 length by transposon insertions underlie the requirement of the EJC for successful splicing. Furthermore we showed that deposition of the EJC onto neighboring exon-exon boundaries facilitates splicing of *piwi* intron 4. We also analyzed the impact of EJC loss genome wide and could show that ~286 introns of *Drosophila melanogaster* show an intron retention phenotype. Taken together, our work illustrates that the EJC is critically involved in the splicing of poor introns and that this is facilitated by its prior deposition to a flanking exon-exon junction.

Status: Released Publication

Genes Dev. 2014 Aug 15;28(16):1772-85. doi: 10.1101/gad.245738.114. Epub 2014 Jul 31.

Authors: Rippei Hayashi*, Dominik Handler*, David Ish-Horowicz and Julius Brennecke (*equal contribution)

Contributions: R.Hayashi, D. Handler, D. Ish-Horowicz and J. Brennecke planned the experiments. R. Hayashi and D. Handler performed all experiments. J. Brennecke, R. Hayashi and D. Handler analyzed the data. R.Hayashi, D. Handler, D. Ish-Horowicz and J. Brennecke wrote the paper.

In close collaboration with Rippei Hayashi I was involved in: Fig1 C-F; Fig2 C-D; Fig3 B,D,E; Fig4 A; Fig6 A; Fig7 B; FigS2 A,C; FigS5

The exon junction complex is required for definition and excision of neighboring introns in *Drosophila*

Rippe Hayashi,^{1,3} Dominik Handler,^{1,3} David Ish-Horowicz,² and Julius Brennecke¹

¹Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), 1030 Vienna, Austria; ²MRC Laboratory for Molecular Cell Biology, University College London, London WC1E 6BT, United Kingdom

Splicing of pre-mRNAs results in the deposition of the exon junction complex (EJC) upstream of exon–exon boundaries. The EJC plays crucial post-splicing roles in export, translation, localization, and nonsense-mediated decay of mRNAs. It also aids faithful splicing of pre-mRNAs containing large introns, albeit via an unknown mechanism. Here, we show that the core EJC plus the accessory factors RnpS1 and Acinus aid in definition and efficient splicing of neighboring introns. This requires prior deposition of the EJC in close proximity to either an upstream or downstream splicing event. If present in isolation, EJC-dependent introns are splicing-defective also in wild-type cells. Interestingly, the most affected intron belongs to the *piwi* locus, which explains the reported transposon desilencing in EJC-depleted *Drosophila* ovaries. Based on a transcriptome-wide analysis, we propose that the dependency of splicing on the EJC is exploited as a means to control the temporal order of splicing events.

[**Keywords:** Piwi–piRNA pathway; SR proteins Acinus and RnpS1; exon junction complex; intron definition; splicing; transposon silencing]

Supplemental material is available for this article.

Received May 19, 2014; revised version accepted July 11, 2014.

Expression of protein-coding genes involves a series of interlinked and interdependent molecular events, including nuclear steps such as pre-mRNA transcription, intron removal via splicing, 5' and 3' modifications, and export, followed by cytoplasmic events such as mRNA translation and degradation. A central factor connecting nuclear pre-mRNA maturation to mRNA fate is the exon junction complex (EJC), a multisubunit protein complex that is deposited ~24 nucleotides (nt) upstream of individual exon–exon boundaries after splicing (Le Hir et al. 2000a).

Assembly of the EJC on an RNA depends on tight RNA binding of a trimeric nuclear complex of the sequence-independent DEAD-box RNA clamp eIF4AIII and a heterodimer of Mago and Y14/Tsunagi (Tsu) (Shibuya et al. 2004; Gehring et al. 2009). Upon nuclear mRNA export to the cytoplasm, these three factors recruit a fourth EJC core component, Barentsz (Btz)/MLN51 (Degot et al. 2004).

The EJC is a binding platform for several peripheral factors, which provide mechanistic links to various processes in mRNA fate control (Tange et al. 2004; Bono and

Gehring 2011). Among these are the nuclear mRNA export factors UAP56, REF/Aly, and TAP/NFX1:p15 (Kataoka et al. 2000; Luo et al. 2001); the nonsense-mediated mRNA decay factor Upf3 (Gehring et al. 2003); and the ASAP complex (Acinus [Acn], RnpS1, and SAP18), whose subunits have been individually implicated in mRNA quality control, pre-mRNA splicing, and transcriptional regulation (Zhang et al. 1997; Mayeda et al. 1999; Schwert et al. 2003; Sakashita et al. 2004).

Most molecular processes that depend on the EJC are post-splicing events. However, the EJC core also interacts with proteins implicated in splicing (e.g., the splicing coactivators/alternative splicing factors SRm160 and Pinin) (Le Hir et al. 2000b; Li et al. 2003), suggesting that the EJC plays a direct role in splicing itself. Indeed, the nuclear EJC core factors eIF4AIII, Mago, and Tsu, together with RnpS1, are required for the faithful splicing of several *Drosophila* pre-mRNAs harboring large introns, often encoded at heterochromatic gene loci (Ashton-Beaucage et al. 2010; Roignant and Treisman 2010). Loss of the EJC results in frequent exon skipping and reduced mRNA levels, the most prominent example being the ERK-encoding *rolled* transcript.

³These authors contributed equally to this work.

Corresponding author: julius.brennecke@imba.oeaw.ac.at

Article published online ahead of print. Article and publication date are online at <http://www.genesdev.org/cgi/doi/10.1101/gad.245738.114>. Freely available online through the *Genes & Development* Open Access option.

© 2014 Hayashi et al. This article, published in *Genes & Development*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0>.

Hayashi et al.

It is unclear how the EJC facilitates splicing mechanistically—in particular, whether the EJC aids splicing of neighboring introns after it has been deposited at flanking spliced junctions or whether individual EJC subunits directly facilitate splicing independently of their assembly into an EJC.

Here, we demonstrate that the nuclear EJC core factors Mago, Tsu, and eIF4AIII and the peripheral EJC components RnpS1 and Acn are essential for the accurate transcript splicing of *piwi*, which encodes an essential component of the ovarian defense pathway for transposable element (TE) silencing in *Drosophila*. We show that the nuclear EJC is required for splicing of *intron4* of *piwi* and that expression of *piwi* transcripts lacking this intron is EJC-independent. In the absence of the EJC recruited to a flanking exon–exon junction, *intron4* is inefficiently spliced, even in wild-type cells. A transcriptome-wide analysis reveals that several other mRNAs also exhibit an intron retention phenotype upon EJC depletion, indicating that the EJC facilitates the removal of adjacent/nearby introns with suboptimal splicing characteristics.

Results

The nuclear EJC is required for transposon silencing in ovaries

TE silencing in *Drosophila* ovarian tissues in both the germline and associated soma is mediated by the piRNA pathway, a small RNA silencing system. Three Argonaute proteins of the PIWI clade (nuclear Piwi and cytoplasmic Aubergine and AGO3) are at its core. Each is bound to a population of 22- to 30-nt single-stranded piRNAs that act as sequence-specific guides to identify complementary targets for their transcriptional and post-transcriptional silencing (Senti and Brennecke 2010; Siomi et al. 2011).

Recent reverse genetic screens have begun to identify the genetic framework of the *Drosophila* piRNA pathway (Czech et al. 2013; Handler et al. 2013; Muerdter et al. 2013). Besides factors involved in piRNA biogenesis or piRNA-mediated silencing, several unexpected players were identified. Among these are components of the EJC, whose depletion results in desilencing of piRNA pathway repressed TEs.

We retested 15 annotated core and peripheral EJC factors for a role in ovarian TE silencing. Using tissue-specific RNAi (Dietzl et al. 2007; Ni et al. 2011), we depleted individual factors in either germline tissue (which comprises accessory nurse cells interconnected with each other and with the oocyte) or the overlying somatic follicular epithelium (Fig. 1A,B; Materials and Methods). piRNA pathway integrity was monitored with *lacZ* reporter transgenes that are repressed by the pathway in wild-type cells (Fig. 1C,D) and via quantitative RT–PCR (qRT–PCR) to measure steady-state RNA levels of marker TEs (Fig. 1E,F; Materials and Methods). For both assays, depletion of the essential piRNA biogenesis factor Armitage (Armi) served as positive control.

Only the EJC core factors Mago and Tsu and the peripheral factors RnpS1 and Acn give strong, consistent

derepression of the reporters and endogenous TEs in both the ovarian germline (*blood* and *HeT-A*) (Fig. 1C,E) and somatic cells (*gypsy* and *mdg1*) (Fig. 1D,F). Depletion of the core EJC factor eIF4AIII prevents ovarian development entirely and could not be evaluated. Importantly, depletion of the cytoplasmic core EJC factor Btz does not result in any observable TE derepression.

Similar results were obtained in cultured ovarian somatic stem cells (OSCs), which are also subject to piRNA-mediated silencing of TEs (Niki et al. 2006; Saito et al. 2009). RNAi-mediated depletion of Mago, Tsu, RnpS1, and Acn, but not of Btz, causes derepression of *gypsy* (Supplemental Fig. S1). Together, our results suggest an involvement of the nuclear EJC plus the factors RnpS1 and Acn in the ovarian piRNA pathway.

The EJC is required for a post-transcriptional step in piwi expression

To place the EJC into the hierarchy of the piRNA pathway (illustrated in Fig. 2A), we took advantage of the observation that defects in piRNA biogenesis (e.g., loss of Armi) cause severe reductions in Piwi protein levels, presumably because unloaded Piwi is unstable (Fig. 2C; Olivieri et al. 2010; Saito et al. 2010). Interestingly, depletion of Mago, Tsu, RnpS1, or Acn, but not Btz, also causes a strong reduction of Piwi levels in soma and the germline (Fig. 2C,D [see also B for a summary of knockdown strategies]). The clonal knockdown approach in somatic cells allowed us to show that the other nuclear EJC core factor, eIF4AIII, is also required for Piwi accumulation (Fig. 2C).

Because the EJC is known to control RNA fate, we hypothesized that loss of the EJC impairs expression or specification of piRNA precursor transcripts, which are transcribed from discrete, often heterochromatic loci termed piRNA clusters (Senti and Brennecke 2010). However, depletion of the EJC in OSCs results in only moderate and insignificant changes in steady-state RNA levels of *flamenco*, the major somatic piRNA cluster (Brennecke et al. 2007). Also, integrity of the Yb body, the cytoplasmic piRNA processing center in somatic ovarian cells (Szakmary et al. 2009; Olivieri et al. 2010; Saito et al. 2010), does not depend on the EJC (Supplemental Fig. S2).

We therefore analyzed *piwi* expression itself using fluorescent in situ hybridization (FISH). In wild-type egg chambers, an *exon8* antisense probe detected abundant *piwi* transcript in the cytoplasm of nurse cells (Fig. 2E). In addition, signal within nurse cell nuclei was observed in strong foci that presumably reflect nascent transcripts at the sites of transcription. Both signals are absent in *piwi*[06843] mutant egg chambers (germline clones), where a P element inserted into the first intron perturbs *piwi* transcription. Consistent with the loss of Armi resulting in post-translational destabilization of Piwi protein (Olivieri et al. 2010; Saito et al. 2010), *piwi* transcript patterns in the cytoplasm and nucleus are normal in Armi-depleted egg chambers (Fig. 2E). In contrast, depletion of Acn or RnpS1 results in the severe reduction of cytoplasmic *piwi* transcript levels, while nuclear foci are unaffected, indicating that primary *piwi* transcription is normal (cf. the signals

Intron definition by the exon junction complex

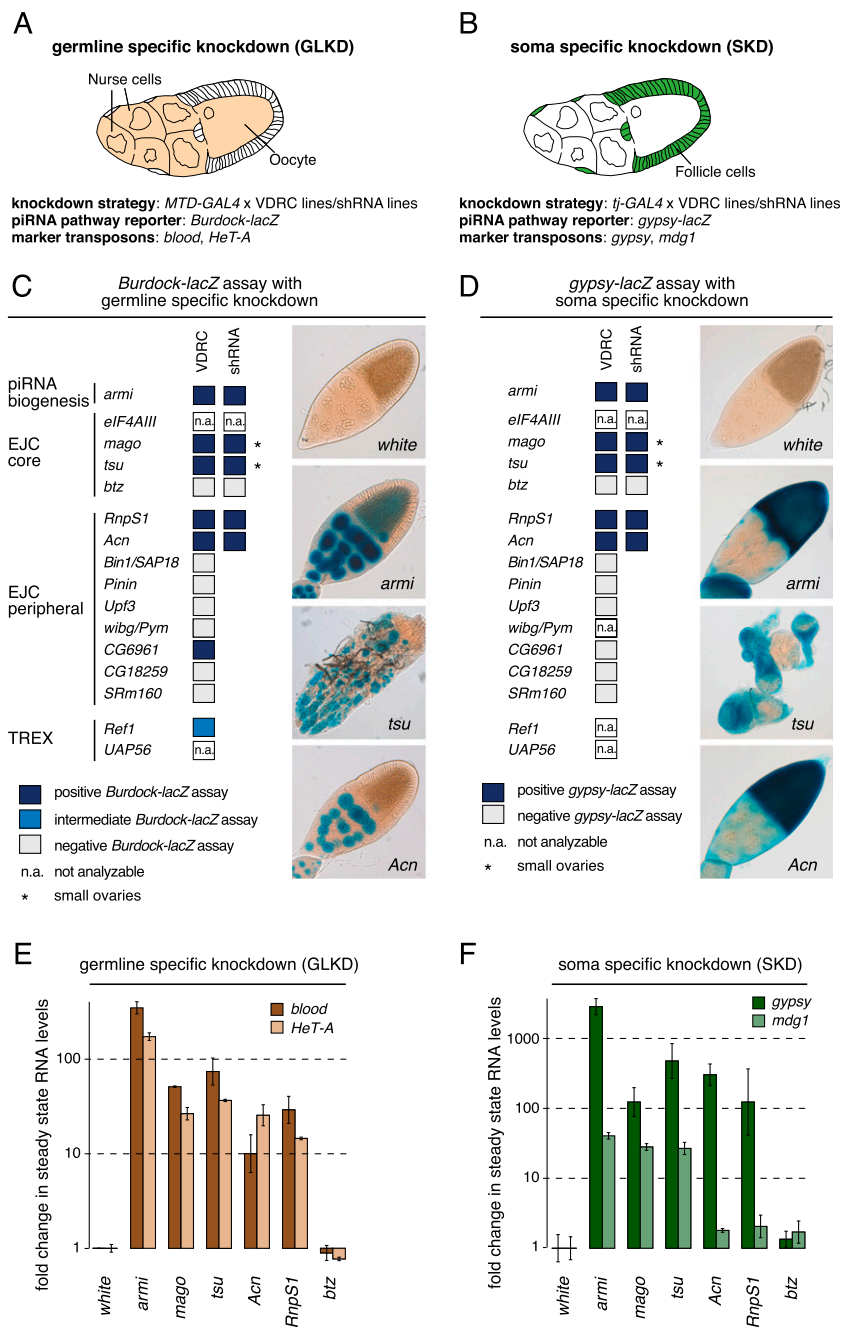


Figure 1. The EJC is required for transposon silencing in ovarian soma and the germline. (A,B) Cartoons of *Drosophila* egg chambers to illustrate the tissue-specific gene depletions in germline (A) or somatic (B) cells. The affected cell types (colored), knockdown strategies, TE reporters, and marker TEs expressed in the respective tissues are indicated. (C,D) The results of the TE reporter assays upon germline (C) or soma (D) knockdown of EJC factors. Defective TE silencing is reflected by *lacZ* expression (blue). The depletion of EJC core factors *mago* and *tsu*, but not *btz*, as well as the depletion of peripheral factors *RnpS1* or *Acn* consistently leads to *lacZ* expression in both soma and the germline. Stained egg chambers of representative phenotypes are shown: *white* for no staining, *tsu* for strong staining but abnormal morphology, and *Acn* for strong staining and no morphological defects. Knockdown of piRNA biogenesis factor *armi* served as control. (E,F) Fold changes in steady-state RNA levels of endogenous TEs in ovaries upon depletion of the indicated genes in the germline (E) or soma (F). RNA levels are normalized to *rp49* levels, and values indicate averages of three biological replicates relative to control knockdowns; error bars indicate standard deviation. The depletion of EJC core factors (*mago* and *tsu*) and peripheral factors (*RnpS1* and *Acn*) causes strong TE derepression almost comparable with *armi* depletions.

in nurse cell nuclei in Fig. 2E). Thus, depleting EJC activity impairs the processing, export, or stability of *piwi* transcripts.

EJC factors are required for *piwi* intron4 splicing

Because the nuclear EJC and *RnpS1* are required for faithful splicing of numerous mRNAs (Ashton-Beaucage et al. 2010; Roignant and Treisman 2010), we analyzed *piwi* splicing patterns by RT-PCR using total RNA from OSCs depleted for individual EJC factors by RNAi. A primer pair mapping to the first and the last *piwi* exon exclusively amplifies the expected spliced wild-type *piwi*

cDNA from control and *Armi*-depleted cells (Fig. 3A,B, amplicon A).

Depletion of *Mago*, *Tsu*, *RnpS1*, or *Acn*, but not *Btz*, results in additional, longer products, suggestive of intron inclusion, and shorter products that are indicative of exon skipping (*Mago* and *Tsu* only) (Fig. 3B). We focused on the major longer fragment, considering that it is observed in all four knockdown samples. Sequencing showed that it comprises the expected *piwi* cDNA plus the entire *intron4* ($n = 10$ of 10); three clones contained *intron6* in addition.

Intron retention in cells with reduced EJC activity was unexpected because loss of the EJC has previously been shown to be associated with skipping of exons, in partic-

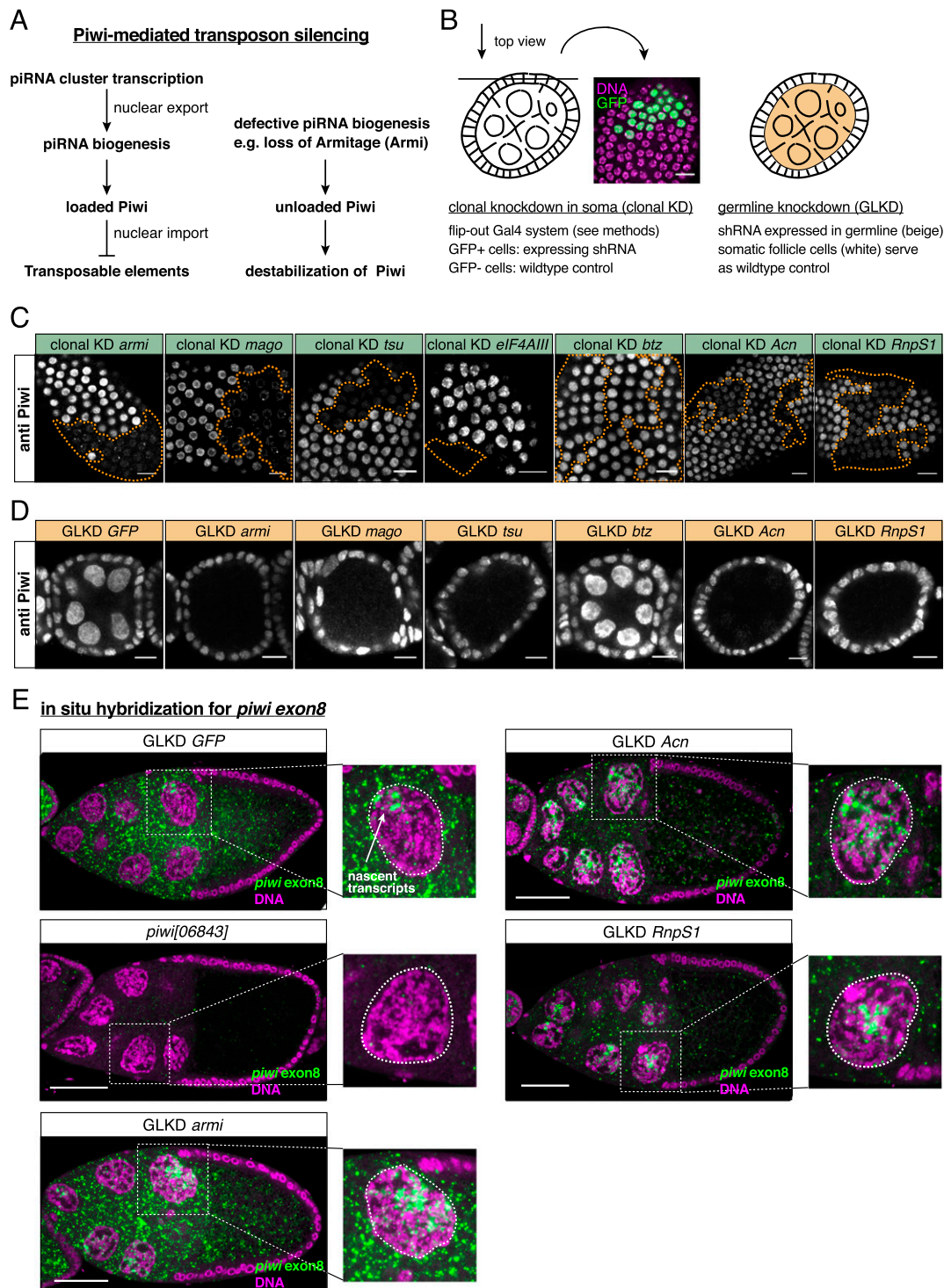


Figure 2. The EJC is required for *piwi* mRNA and protein expression. (A) Illustration of Piwi-mediated transposon silencing. piRNA precursor transcripts are expressed from heterochromatic piRNA clusters (e.g., *flamenco*) and exported to cytoplasmic piRNA biogenesis sites. Upon loading with a piRNA, Piwi enters the nucleus, where it guides transcriptional silencing of TEs. Defective piRNA biogenesis (e.g., caused by loss of Armi) prevents Piwi loading, which renders it unstable. (B) Illustration of applied RNAi knockdown strategies. Indicated factors were depleted in either clones of somatic cells (left) or all germline cells (right). Confocal images in C and D represent surface sections and cross-sections through stage 3–6 egg chambers, respectively. (C,D) Confocal sections of egg chambers with clones of somatic follicle cells (C) or germline cells (D) expressing shRNAs against EJC factors stained for Piwi. Bars, 10 μ m. Piwi levels are greatly reduced in cells depleted for EJC factors except Btz. shRNAs against *GFP* and *armi* served as controls. (E) Confocal sections through stage 9–10 egg chambers expressing shRNAs against the indicated genes in the germline stained for *piwi* exon8 (green) and DNA (magenta). Bars, 50 μ m. Egg chambers from *piwi*[06843] germline clones served as negative control. Cytoplasmic *piwi* staining is lost upon *Acn* or *RnpS1* knockdown but not upon *armi* knockdown. Nuclear staining (individual nurse cell nuclei enlarged) of *piwi* reflects nascent transcripts and thus active transcription.

Intron definition by the exon junction complex

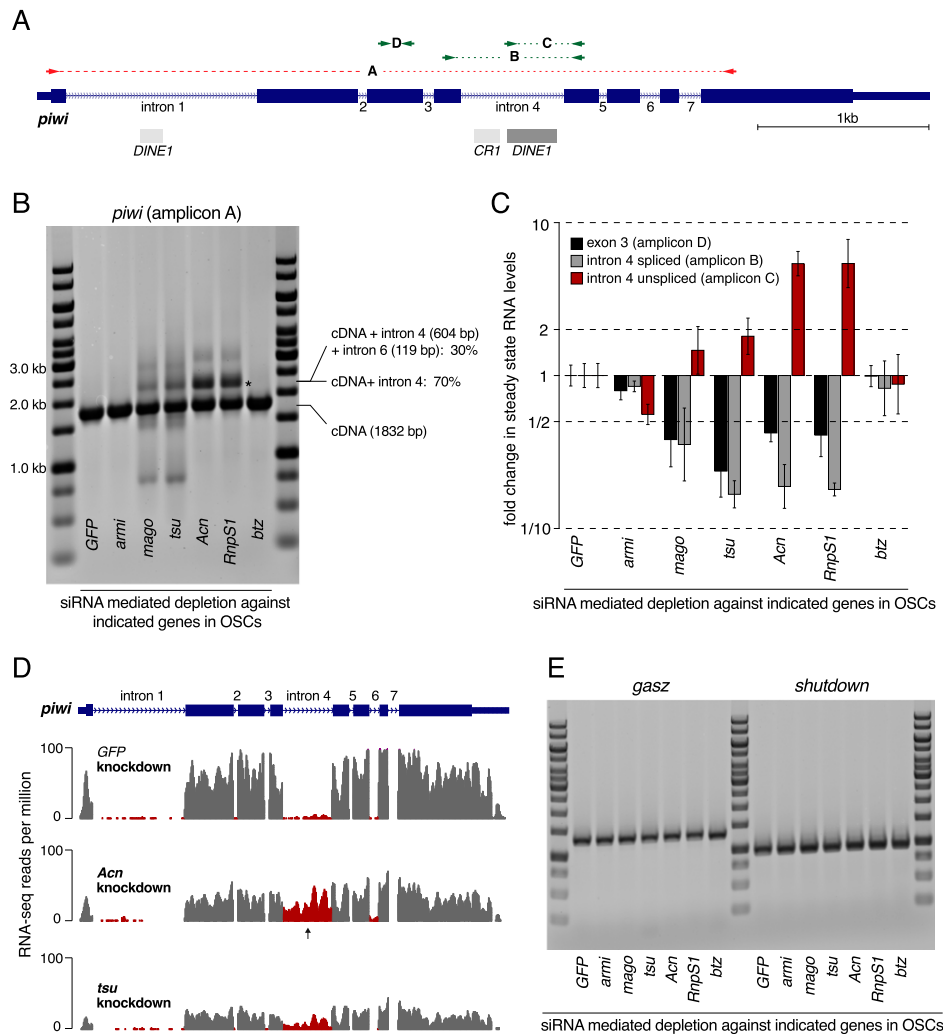


Figure 3. EJC factors are required for *piwi* *int4* splicing. (A) Cartoon of the genomic *piwi* locus showing exons as blue boxes, introns as lines, TE fragments as gray boxes, and the locations of RT-PCR primers used in B–D. (B) Agarose/EtBr gel showing *piwi* RT-PCR products (primers in *exon1* and *exon8*; amplicon A in A; size markers at the borders) amplified from total RNA isolated from OSCs depleted for the indicated genes. *piwi* cDNAs containing *int4* or both *int4* and *int6* (indicated by an asterisk) are amplified along with wild-type *piwi* cDNA in *mago*, *tsu*, *Acn*, or *RnpS1* knockdowns. (C) Shown are the fold changes in steady-state RNA levels of all *piwi* transcripts (*exon3*; black bars) and those *piwi* transcripts with spliced (gray bars) or unspliced (red bars) *int4* amplified from total RNA isolated from OSCs depleted for the indicated genes. Values are normalized to *rp49* levels, and averages of three biological replicates relative to control knockdowns are shown, with error bars indicating standard deviation. The levels of *piwi* transcripts with spliced and unspliced *int4* are decreased and increased by twofold to fourfold in *mago*, *tsu*, *Acn*, or *RnpS1* knockdowns, respectively. (D) RNA-seq profiles at the *piwi* locus obtained from polyA-selected RNA from OSCs depleted for the indicated genes (exonic reads in gray; intronic reads in red; arrows mark *int4*), showing the retention of *int4* in *Acn* and *tsu* knockdowns. (E) Agarose/EtBr gel showing *gasz* and *shutdown* RT-PCR products (primers in the respective first and last exons) amplified from total RNA isolated from OSCs depleted for the indicated genes. Unlike *piwi*, the splicing of these transcripts is not affected by the depletion of EJC factors.

ular those adjacent to very long introns (Ashton-Beaucage et al. 2010; Roignant and Treisman 2010). We therefore confirmed the retention of *piwi* *intron4* (*int4*) by qPCR. Compared with control cells, we observed an approximately fourfold decrease of the spliced species (amplicon B) and an approximately fourfold increase of the unspliced species (amplicon C) in *Acn*- or *RnpS1*-depleted cells (Fig. 3C). The accumulation of the unspliced species is evident but less pronounced in *Mago*- or *Tsu*-depleted cells, perhaps because depletion of these core factors has a broader impact on *piwi* splicing patterns or causes a general reduction in

cell viability. We observed a similar degree of *int4* retention also in ovaries depleted for EJC factors specifically in germline cells (Supplemental Fig. S3).

A highly similar intron retention pattern was found by sequencing polyA-selected RNAs: *int4* is substantially retained, and *int6* is slightly overrepresented in *Acn*- or *Tsu*-depleted OSCs (Fig. 3D). These data also confirm a general reduction in *piwi* steady-state RNA levels, especially in *Tsu*-depleted cells, presumably a consequence of nonsense-mediated decay (NMD) because inclusion of *int4* adds a premature stop codon.

Intron retention is not a general phenotype in EJC-depleted cells; splicing of several other transcripts (e.g., of the piRNA pathway factors *gasz* or *shutdown*) is normal in the various OSC knockdowns tested (Fig. 3E). Thus, the reduced Piwi activity is probably due to a selective failure of *piwi* RNA splicing.

piRNA pathway defects in EJC-depleted ovaries are largely due to retention of piwi intron4

To check whether *int4* retention is indeed the major block to Piwi expression in EJC-depleted cells, we tested whether removing the intron renders *piwi* EJC-independent. We generated transgenic flies that express an N-terminal GFP-Piwi fusion in the context of a fully functional genomic construct containing the entire *piwi* locus (*GFP-piwi*) (Sienski et al. 2012) and flies expressing an equivalent gene that lacks *int4* (*GFP-piwi[Δint4]*).

As expected, GFP-Piwi levels from the former construct are sensitive to the depletion of EJC factors in ovarian somatic and germline cells (Fig. 4A; Supplemental Fig. S4A). In contrast, GFP-Piwi from *GFP-piwi[Δint4]* is insensitive to depletion of Tsu or Acn (Figs. 4A; Supplemental Fig. S4A) while remaining sensitive to reduced piRNA synthesis by Armi depletion. Retention of *int4* is therefore causal to the reduced Piwi levels in EJC-depleted cells.

We further tested whether the EJC-independent *GFP-piwi[Δint4]* transgene rescues the TE derepression phenotype in EJC-depleted flies. We depleted Acn specifically in ovarian germline or somatic cells of flies carrying either the wild-type *GFP-piwi* or the *GFP-piwi[Δint4]* construct. TE derepression in the germline caused by the depletion of Acn, but not of Armi, is greatly rescued by the *GFP-piwi[Δint4]* transgene and not by the wild-type *GFP-piwi* transgene (Fig. 4B). The transgene also rescues TE derepression in the soma: In Acn-depleted somatic cells, expression of the *gypsy-lacZ* reporter is largely repressed in the presence of *GFP-piwi[Δint4]* (Fig. 4C).

The *GFP-piwi[Δint4]* construct fails to rescue germline TE derepression due to Tsu depletion (data not shown), perhaps because these flies also show reduced germline expression of AGO3 (Supplemental Fig. S4B), a second central Argonaute protein involved in the piRNA pathway. AGO3 pre-mRNA has very large introns and thus resembles transcripts known to require the EJC for efficient splicing. Indeed, AGO3 mRNA level is reduced upon Tsu but not Acn depletion. We did not detect misspliced AGO3 mRNA species, possibly due to efficient NMD (Supplemental Fig. S4C,D). We note that the *GFP-piwi[Δint4]* construct does reduce desilencing of *mdg1* in Tsu-depleted follicle cells in which AGO3 is not expressed (Supplemental Fig. S4E). Taken together, our results indicate that missplicing of *piwi int4* is the major cause of TE derepression in ovaries depleted for EJC components.

Splicing of piwi intron4 requires deposition of the EJC at a nearby splice junction

To understand why splicing of *piwi int4* is sensitive to EJC levels, we compared the *int4* sequence in *Drosophila*

melanogaster with that in other *Drosophilids* (Fig. 5A). In the *melanogaster* subgroup, *int4* harbors one or two TE remnants that extend the intron to ~200–700 nt, whereas the intron is devoid of TE fragments and very short (~60 nt) in the more distantly related species. Also, the 3' end of *piwi int4* lacks a robust polypyrimidine tract (pY; defined as at least seven consecutive pyrimidines) in nearly all of these species. Assembly of spliceosome components at 3' splice junctions involves recognition of the pY tract by the U2 snRNP factor U2AF65 (Zamore et al. 1992). Introns with a poor pY tract can be spliced in vitro, but only if they are <90 nt (Guo et al. 1993).

This analysis raised the possibility that *int4* is a sub-optimal intron in the *melanogaster* subgroup. To test this idea, we generated transgenic flies harboring the aforementioned genomic *GFP-piwi* construct but lacking either all introns or all except *int4*. The intron-less construct drives readily detectable GFP-Piwi expression in ovarian somatic and germline cells (Fig. 5C). However, addition of *int4* abolishes expression of GFP-Piwi almost completely (Fig. 5C), indicating that even wild-type cells cannot splice *piwi int4* if it is present in isolation.

Replacing the 3'-most 50 nt of *int4* with the equivalent part of *piwi int7* that contains a clear pY tract (the *pY-fixed* construct) (Fig. 5B) largely rescues *int4* splicing, as does shortening of *int4* to 60 nt (Fig. 5D,E). A length reduction to 106 nt does not restore splicing, consistent with the aforementioned length cutoff of ~90 nt for introns with poor pY tracts (Guo et al. 1993). These results suggest that *piwi int4* is a very poorly defined intron due to the lack of an optimal pY tract coupled with its large size.

Because *piwi int4* is efficiently spliced in the context of the wild-type locus but poorly spliced when isolated, perhaps prior splicing of flanking introns is required for *int4* definition. We tested this idea directly by adding either *int3* or *int5* at their natural positions to the transgene that contains only *int4*. Addition of *int3* has no measurable impact, but the presence of *int5* results in significant GFP-Piwi expression to about half the level observed in the *pY-fixed* construct (Fig. 5F). Most importantly and as shown for the wild-type construct, GFP-Piwi expression from the *int4/5* construct is EJC-dependent, being approximately threefold reduced in germline cells upon germline-specific depletion of Acn (Fig. 5F).

These results strongly favor a model in which splicing of *int5* with subsequent EJC deposition facilitates *int4* splicing. To understand why addition of *int5* but not of *int3* rescues *int4* splicing, we analyzed the impact of intron identity, absolute intron distance to *int4*, and relative intron positioning (upstream versus downstream) on GFP-Piwi expression. Intron identity plays no role because an intron derived from the *armi* locus rescues expression as efficiently as *piwi int5*, while *piwi int5* placed at the *int3* position lacks rescuing activity (Fig. 5G). Increasing the distance between *int4* and *int5* from 200 nt (wild type) to 500 or 900 nt impairs rescue, while shortening the distance has no impact (Fig. 5H), suggesting that proximity to the splice junction is important. To test whether the upstream introns fail to rescue because they are too distant from the *int4* 3' end, we examined whether

Intron definition by the exon junction complex

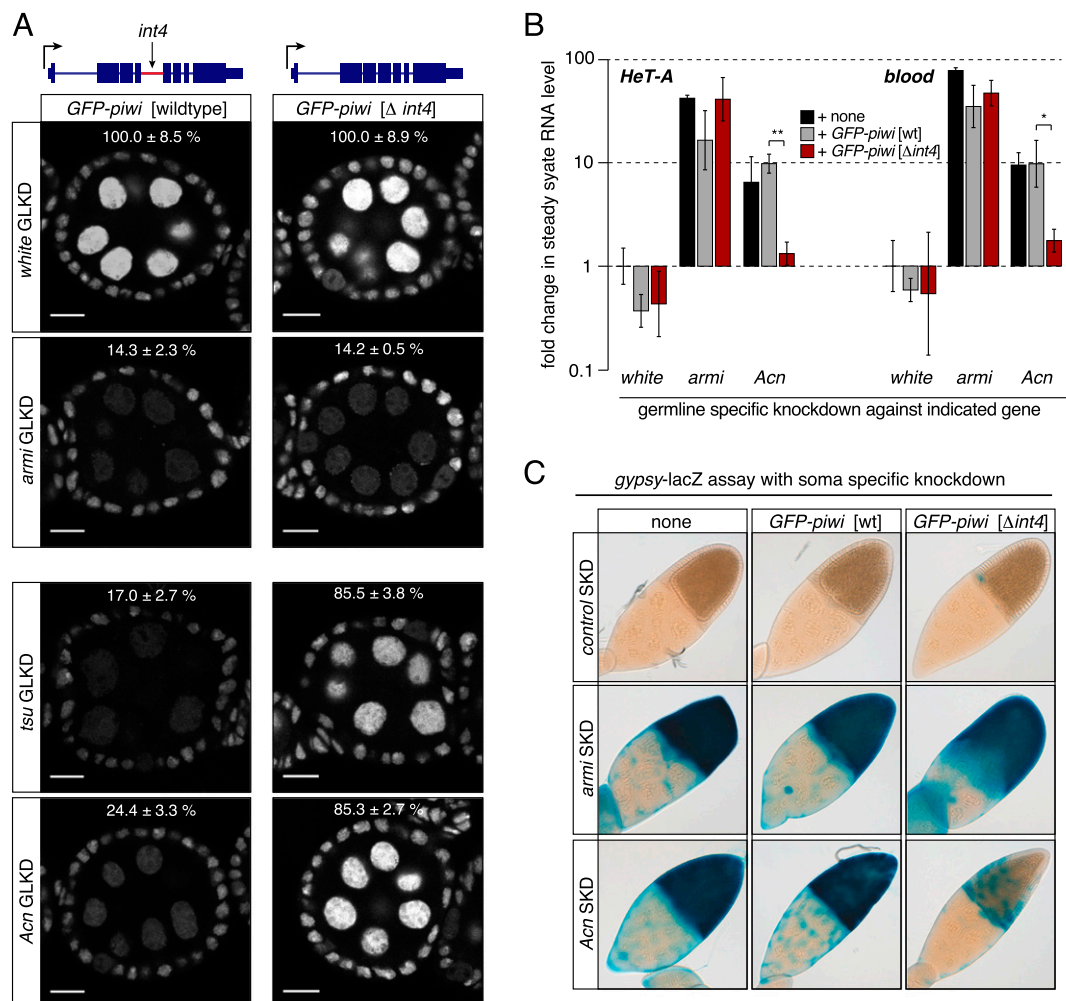


Figure 4. A *piwi* transgene lacking *int4* rescues transposon derepression in EJC-depleted cells. (A) Expression of wild-type *GFP-piwi*, but not of *GFP-piwi* $\Delta int4$, is dependent on EJC factors. Shown are confocal sections through egg chambers expressing *GFP-piwi* or *GFP-piwi* $\Delta int4$ depleted for the indicated genes in germline cells. Bars, 10 μ m. Relative GFP levels (\pm standard deviation) from the indicated *GFP-piwi* transgenes in comparison with the *white* control knockdown are indicated (GFP fluorescence levels in nurse cell nuclei normalized to those in follicle cell nuclei from three egg chambers). (B) Displayed are fold changes in steady-state RNA levels of *HeT-A* and *blood* in ovaries depleted for the indicated genes in the germline and expressing the indicated *GFP-piwi* transgenes. RNA levels are normalized to *rp49* levels, and averages of three biological replicates relative to control knockdowns are shown. Error bars indicate standard deviation. (*) $P < 0.01$; (**) $P < 0.001$. TE derepression caused by *Acn* knockdown is rescued by *GFP-piwi* $\Delta int4$ but not by wild-type *GFP-piwi*. (C) Shown are the X-gal stainings of egg chambers depleted for the indicated genes in somatic follicle cells and expressing the *gypsy-lacZ* reporter and the indicated *GFP-piwi* transgenes. *GFP-piwi* lacking *intron4*, but not wild-type *GFP-piwi*, greatly reduces the *gypsy-lacZ* expression induced by the depletion of *Acn*.

an upstream intron can rescue the splicing of a shorter *int4* from *Drosophila yakuba* (235 nt), which, when isolated, is also splice-defective in wild-type cells (Fig. 5I). This shorter *int4* is indeed rescued by the upstream presence of the *melanogaster int3* (Fig. 5, cf. I and F), indicating that both upstream and downstream introns can rescue *int4* splicing if placed nearby, presumably due to prior deposition of the EJC to a flanking exon–exon junction.

Splicing of first introns tends to depend on the EJC if followed by a large second intron

To investigate whether *piwi int4* is indicative of a more general role of the EJC in efficient intron removal, we

sequenced polyA-selected RNAs from OSCs depleted for *Tsu*, *Acn*, or *GFP* (Materials and Methods). We considered ~ 1500 introns for which a significant quantitative analysis was possible (Materials and Methods).

An analysis of intronic RNA sequencing (RNA-seq) reads indicates that splicing of most introns is EJC-independent (median changes in intron retention index upon *tsu* and *Acn* knockdown are both 1.00). Nevertheless, 286 introns exhibit pronounced retention characteristics (twofold to 10-fold higher than in control cells) in cells depleted for either *Tsu* or *Acn* (Fig. 6A). We verified the RNA-seq results for several EJC-dependent introns by performing RT experiments or by qPCR using amplicons that discriminate spliced and nonspliced transcripts (shown

Hayashi et al.

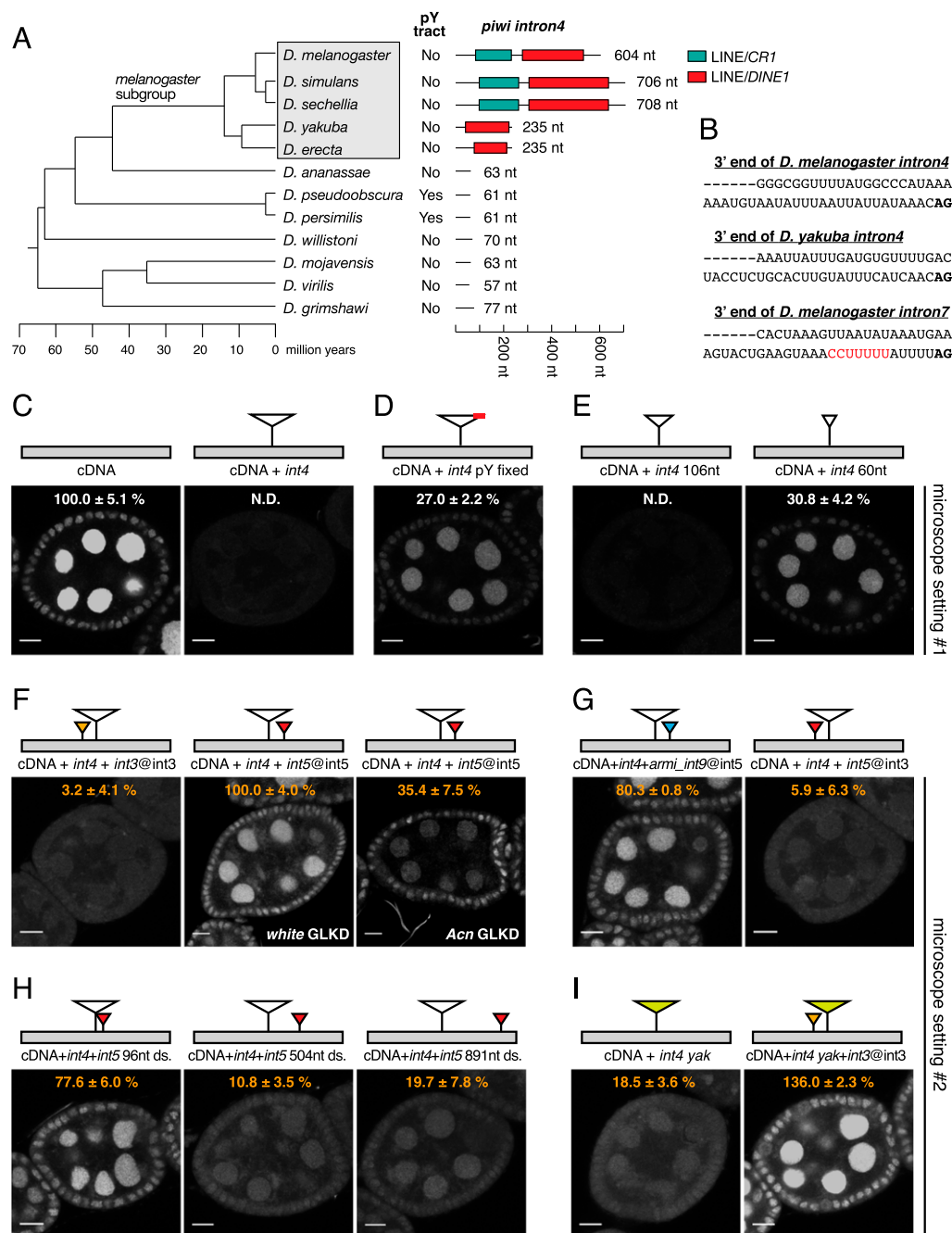


Figure 5. The EJC facilitates *piwi int4* splicing from a nearby splice junction. (A) Architecture of *piwi int4* from 12 *Drosophila* species (phylogenetic tree is at the left). Indicated are TE fragments (colored), intron length, and the presence of a pY tract. *piwi int4* is enlarged in the *melanogaster* subgroup (boxed) due to insertions of TE fragments. (B) Depicted are the last 50 nt of the indicated introns, with pY tract in red. *piwi int4* from *D. melanogaster* and *D. yakuba* have suboptimal pY tracts. (C–I) *piwi int4* is not efficiently spliced when it is placed in isolation. The features of *int4* that contribute to the inefficient splicing and the effects of additional flanking introns are examined. Shown are confocal sections through egg chambers expressing the indicated *GFP-piwi* transgenes (depicted as cartoons above each image). Bars, 10 μ m. In F, shRNAs against *white* or *Acn* are expressed in germline cells (GLKD). The transgene cartoons represent the *piwi* cDNA (gray), *melanogaster int4* (white large triangles), *yakuba int4* (green large triangles), and additional indicated introns (small colored triangles). The indicated GFP levels (\pm standard deviation) were calculated by measuring GFP fluorescence in nurse cell nuclei of three egg chambers. Microscope settings were identical in C–E and F–I, respectively. GFP intensity was set to 100% for the left panel in C and the middle panel in F. For a global comparison, the intensity of *GFP-piwi* [cDNA + *int4* + *int5@int5*] is 13.0 \pm 2.0% of the intensity of *GFP-piwi* [cDNA].

Intron definition by the exon junction complex

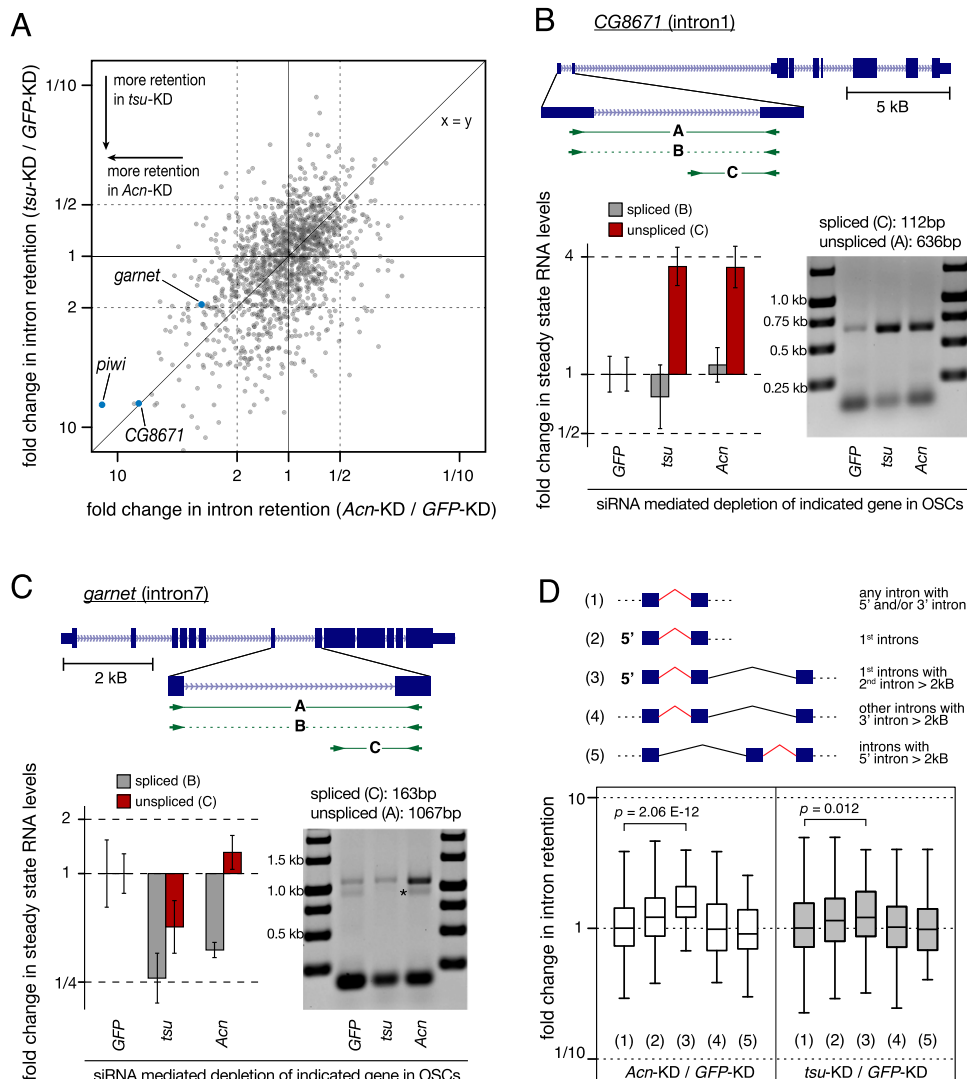


Figure 6. Transcriptome-wide analysis of EJC-dependent introns. (A) Scatter plot showing the fold changes in the retention of individual introns of OSC transcripts upon *Acn* or *tsu* knockdown relative to the control knockdown. Validated hits are indicated (blue dots). *piwi int4* is among the most prominent introns that are retained both in *Acn* and *tsu* knockdowns. (B,C) Experimental validations of enhanced retention of *CG8671 int1* and *garnet int7* in *tsu*- or *Acn*-depleted OSCs using polyA-selected RNAs. Shown at the top is a cartoon of the respective locus (exons as boxes and introns as lines) with the location of the primers in RT-PCR. The bar charts indicate fold changes in steady-state levels of the indicated *CG8671* (B) or *garnet* (C) transcripts. (Gray) Spliced; (red) unspliced. Amplicon identity is given above. The Agarose/EtBr gel images show RT-PCR products obtained with primers mapping to the flanking exons. The asterisk in C marks a nonspecific product. (D, top) The cartoon depicts the analyzed groups of introns, considering their location in the transcript and the identity of flanking introns. The box plot shows fold changes in intron retention for the groups defined above upon *Acn* or *tsu* depletion. First introns followed by a large second intron (group 3) are significantly more retained in EJC-depleted cells (*P*-values are calculated by the Wilcoxon rank-sum test).

for *CG8671* and *garnet* in Fig. 6B,C). Of note, *piwi int4* is among the most affected intron transcriptome-wide.

As a poor pY tract is an apparent characteristic of *piwi int4*, we inspected whether EJC-dependent introns generally exhibit a weak pY tract. However, the group of introns harboring a poor pY tract ($n = 272$) (Materials and Methods) is not significantly more dependent on the EJC than the average intron (median changes in intron retention index upon *tsu* or *Acn* knockdown are 1.03 and 1.02, respectively; *P*-values > 0.5). Also, the suspicious TE insertions in *piwi int4* do not seem to suggest a general

pattern: TE remnants are only rarely found in *tsu*- or *Acn*-dependent introns (only four out of the 30 most affected introns harbor TE remnants).

Loss of the nuclear EJC has been shown to result in exon skipping for transcripts containing large introns (Ashton-Beaucage et al. 2010; Roignant and Treisman 2010). This appears not to be the case for intron retention because the intron retention index is not significantly increased for either introns that are themselves >2 kb ($n = 162$; median changes in intron retention index upon *tsu* or *Acn* knockdown are 0.93 and 0.92, respectively) or introns within

Hayashi et al.

transcripts containing at least one intron >10 kb ($n = 353$; median changes in intron retention index upon *tsu* or *Acn* knockdown are 1.00 and 0.94, respectively).

Manual inspection of many gene loci containing an EJC-dependent intron indicated that affected introns are often first introns and that they are frequently followed by a large second intron (e.g., *CG8671* in Fig. 6B; see also Supplemental Table S1). The population of first introns followed by a large second intron (>2 kb; $n = 77$) is significantly more likely to be inefficiently spliced in *tsu*-depleted (median = 1.21; $P = 0.012$) or *Acn*-depleted (median = 1.46; $P = 2.06 \times 10^{-12}$) cells compared with control cells (Fig. 6D). In comparison, the group of all first introns ($n = 320$) or internal introns immediately followed ($n = 64$) or preceded ($n = 99$) by a large intron shows no significantly different behavior in EJC-depleted cells compared with control cells (Fig. 6D). Together, our results demonstrate an unanticipated dependency of first introns that are followed by a large second intron on the EJC for their splicing.

Exon skipping and intron retention are genetically separable events

Both exon skipping and intron retention are defects in EJC-depleted cells, arguing that the two processes are mechanistically related. However, we noticed one clear difference between the two phenotypes: Loss of *Acn* strongly impacts intron retention yet was not previously reported to have an impact on exon skipping (Ashton-Beaucage et al. 2010; Roignant and Treisman 2010).

We therefore analyzed splicing patterns of the large intron-containing *rolled* and *light* transcripts in OSCs depleted for EJC core factors or RnpS1 or *Acn*. Consistent with previous reports, depletion of Mago or *Tsu* leads to severe exon skipping (Fig. 7A; Supplemental Fig. S5; Ashton-Beaucage et al. 2010; Roignant and Treisman 2010). The same is true, although to a lesser extent, for depletion of RnpS1 but not for *Acn*, whose depletion results in very mild exon skipping (Fig. 7A; Supplemental Fig. S5). Similarly, protein levels of the Rolled MAP kinase ERK are strongly reduced in vivo in cell clones depleted for EJC core factors (Mago, *Tsu*, and eIF4AIII), only mildly reduced upon RnpS1 depletion, and almost wild type in *Acn*-depleted cells (Fig. 7B). In contrast, loss of any involved EJC factor affects Piwi protein levels equally under comparable experimental conditions (Fig. 2C,D).

As exon skipping results in reduced levels of affected transcripts (presumably due to NMD), depletion of the core EJC affects preferentially levels of transcripts harboring long introns (Ashton-Beaucage et al. 2010; Roignant and Treisman 2010). We confirmed that transcript levels are increasingly more affected as a function of increasing intron length upon *Tsu* depletion in OSCs (Fig. 7C). Again, no such trend is detectable upon *Acn* depletion, confirming its limited impact on exon skipping.

Taken together, our observations indicate that exon skipping and intron retention are related yet genetically separable processes. While depletion of nuclear

core EJC factors results in both phenotypes, depletion of *Acn* only has a clear effect on the intron retention phenotype.

Discussion

Genetic screens have uncovered a surprising link between the EJC and the piRNA pathway in *Drosophila*. We dissected the underlying molecular cause and uncovered several principles that connect the EJC to intron definition and splicing. The major outcome of our study is as follows:

- (1) Loss of the nuclear EJC or the accessory factors RnpS1 or *Acn* impairs the *D. melanogaster* piRNA pathway due to impaired splicing of the *piwi* transcript.
- (2) Deposition of the EJC to flanking introns, but not direct recruitment by the spliceosome, underlies the EJC's impact on splicing.
- (3) Many introns—preferentially first introns followed by large second introns—require the EJC for their splicing.
- (4) Exon skipping and intron retention depend on related but distinct aspects of EJC function.

Roles of the EJC in the piRNA pathway

Defective splicing of *piwi int4* is a major molecular defect underlying the involvement of the EJC in the *D. melanogaster* piRNA pathway (Figs. 3, 4). During evolution, one or two LINE fragments appear to have inserted into *int4*, which, together with its poor pY tract, renders it too long to be spliced efficiently. Based on these criteria, an involvement of the EJC in *piwi* expression appears restricted to the *melanogaster* subgroup (which branched from the *ananassae* subgroup ~45 million years ago) (Fig. 5; Tamura et al. 2004). It is unclear whether a suboptimal *int4* is simply tolerated given its rescue via the EJC or whether the dependence of *int4* on the EJC provides a selective advantage.

In the case of *Acn* depletion, TE repression is nearly entirely restored by expression of an EJC-independent *piwi* transgene. Loss of *Tsu* or Mago, on the other hand, results in additional defects that impair TE silencing. It is currently unclear whether splicing of other piRNA pathway factor transcripts requires the core EJC. A strong candidate is AGO3, which is expressed as an ~150-kb pre-mRNA with several large introns; indeed, AGO3 transcript levels are affected upon depletion of core EJC but not peripheral EJC factors (Supplemental Fig. S4).

EJC assembly to flanking splice junctions occurs prior to its contribution to splicing

Because deposition of the minimal EJC core (pre-EJC: eIF4AIII, Mago, and *Tsu*) at the upstream exon occurs shortly before or during exon ligation and therefore depends on a successful splicing event (Gehring et al. 2009), most studies on the EJC have focused on its role in controlling mRNA fate downstream from splicing. It came as a surprise that the EJC also impacts the process

Intron definition by the exon junction complex

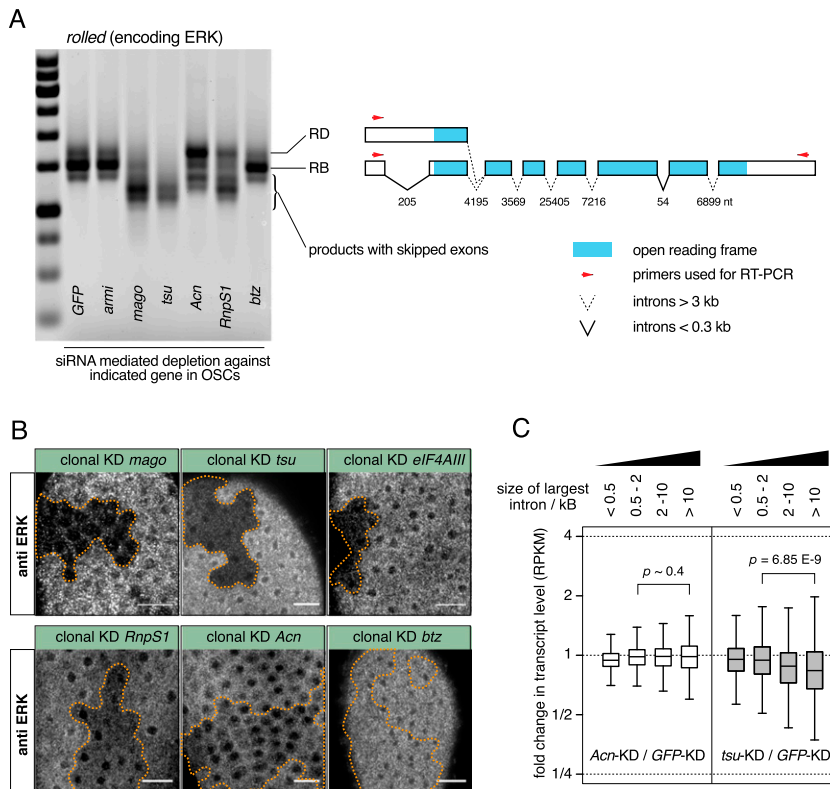


Figure 7. Depletion of core EJC factors but not *Acn* results in exon skipping. (A) Agarose/EtBr gel image showing RT-PCR products of *rolled* transcripts amplified from total RNA from OSCs depleted for the indicated genes. Illustrated on the right is the *rolled* locus showing exons as boxes, with the ORF depicted in blue, introns as solid (<0.3 kb) and dashed (>3 kb) lines, and arrows as primers used in RT-PCR (modified from Ashton-Beaucage et al. 2010, © 2010, with permission from Elsevier). Each intron is labeled with their size (in nucleotides). *rolled*-RB is the major product in wild-type cells, and *rolled*-RD is an annotated *rolled* transcript with *intron1* retained that still contains the wild-type ORF. Knockdown of *mago*, *tsu*, and, to a lesser degree, *RnpS1* causes exon skipping of *rolled*, whereas the splicing is largely unaffected by the knockdown of *Acn*. (B) Confocal sections through the follicular epithelium of stage 3–6 egg chambers stained for ERK (encoded by *rolled*), showing that the depletion of *mago*, *tsu*, or *eIF4AIII* leads to a severe reduction in ERK protein level, which is moderately or only very mildly affected by the depletion of *RnpS1* or *Acn*, respectively. Bars, 10 μ m. Cells within the clone (boundaries marked by dashed line) express shRNAs against the indicated genes. (C) Box plot showing fold changes in the

steady-state RNA level of transcript groups containing increasing intron length upon *Acn* (left) or *tsu* (right) depletion in OSCs. Grouping of transcripts was according to the largest intron (x in kilobases): from left, $0 < x < 0.5$; $0.5 < x < 2$; $2 < x < 10$; $10.0 < x$. The level of transcripts with large introns is decreased upon depletion of *tsu* but not of *Acn* (P -values based on Wilcoxon rank-sum test).

of splicing itself (Ashton-Beaucage et al. 2010; Roignant and Treisman 2010). However, analyzing the molecular basis for this contribution to splicing has been difficult due to the complex architecture of previously described target transcription units (long genes with large introns harboring numerous TE insertions, often situated within the heterochromatin). In particular, it was unclear whether the role of the EJC in the splicing of transcripts containing large introns reflects a noncanonical role of EJC components in splicing or whether canonically deposited EJCs at flanking splice junctions aid the definition of flanking introns (discussed in Ashton-Beaucage and Therrien 2011).

The finding that the EJC is also required for splicing of *int4* at the compact and euchromatic *piwi* locus provides a powerful system to address this important question. Our collective data strongly indicate that conventional assembly of the EJC at splice junctions can facilitate the subsequent splicing of neighboring introns that pose a challenge for the spliceosome, such as ones that are atypically large (Ashton-Beaucage et al. 2010; Roignant and Treisman 2010) or lack strong splicing signals (this study). In the case of *piwi int4*, a neighboring EJC (upstream or downstream) can assist the splicing if it is bound within ~ 250 nt of the 3' splice site (Fig. 5F,H). Interestingly, this distance requirement resembles that of the minimum size of an intron in the intron definition model of splicing (Fox-Walsh et al. 2005) and indicates a direct

role of the EJC in aiding splicing rather than a licensing model in which EJC deposition anywhere on the message promotes splicing of suboptimal introns. The EJC has been shown to interact with a number of SR proteins (Singh et al. 2012), suggesting that the EJC acts as an RNA-bound scaffold to recruit splicing machineries via SR proteins.

A temporal order in splicing instructed by the EJC?

Previous observations point to a preferential role of the EJC in the splicing of complex transcripts; namely those that contain unusually large introns, often harboring TE insertions. Why, however, would first introns followed by large second introns be disproportionately dependent on the EJC (Fig. 6D)? These first introns themselves are not unusually large and exceptional, yet their dependency on a flanking EJC suggests that they are intrinsically difficult to splice. An attractive, albeit speculative, model is that evolution has selected this dependency in order to splice first introns only after successful splicing of the large downstream intron. Large introns are likely to contain cryptic polyadenylation/cleavage sites, and their usage is suppressed by deposition of U1 onto the nascent transcript (Kaida et al. 2010). Delaying intron splicing to a time point after transcription and splicing of the large downstream intron would therefore increase gene expression fidelity. Interestingly, first introns have been reported to be

Hayashi et al.

less efficiently spliced compared with internal introns in *Drosophila* (Khodor et al. 2011).

A related hypothesis is that the EJC licenses splice events during alternative splicing. Alternatively spliced introns often have weak splice recognition sites (Clark and Thanaraj 2002; Garg and Green 2007), and it has been shown that splicing of neighboring introns influences the selection of alternative splice sites (Fededa et al. 2005; Glauser et al. 2011). It would be intriguing if the order of intron splicing impacts alternative splicing in an EJC-dependent manner, thereby increasing the transcript repertoire in higher organisms.

Exon skipping versus intron retention and the role of Acinus

Intron retention and exon skipping have two different molecular causes: defective intron and exon definition, respectively (De Conti et al. 2013). Nevertheless, depletion of core EJC factors (eIF4AIII, Mago, and Tsu) or of RnpS1 can give rise to both phenotypes, suggesting that they are linked (Ashton-Beaucage et al. 2010; Roignant and Treisman 2010; this study). Indeed, we hypothesize that both phenotypes manifest after a first successful round of splicing. In the case of exon skipping, we propose that the reinforcement of the exon definition of the newly formed exon pair is compromised (see also Ashton-Beaucage and Therrien 2011). In the case of intron retention, we propose that EJC loss results in a failure to license a weak nearby splice site.

Despite these similarities, it appears that these two processes are at least to some extent distinct, as loss of Acn leads more or less exclusively to intron retention and only marginally to exon skipping (Fig. 7). Our genome-wide analysis shows a global correlation between intron retention levels in Acn- and Tsu-depleted cells (Fig. 6A). Also, Acn enhances the splicing of *piwi intron4* in an EJC-dependent manner (Fig. 5F). Assembly of the EJC core is therefore probably a prerequisite for Acn to help defining neighboring splice sites. Acn forms a stable complex with the SR protein RnpS1 to interact with the EJC (Tange et al. 2005) and has recently been implicated in the regulation of alternative splicing (Michelle et al. 2012). Further genetic and biochemical analyses on Acn (and RnpS1) should be the key to understanding their distinctive roles in intron definition and exon definition.

Materials and methods

Fly husbandry and strains

All flies were kept at 25°C. *MTD-GAL4* and *piwi[06843]* strains were obtained from the Bloomington *Drosophila* Stock Center, *tj-GAL4* strain was obtained from the *Drosophila* Genomics Resource Center. The *lacZ* sensor flies for *Burdock* and *gypsy* were described in Handler et al. (2013) and Sarot et al. (2004), respectively. shRNA lines were generated by cloning short hairpins (for sequences, see Supplemental Table S3) into the *Valium-20* vector modified with a *white* selection marker (Ni et al. 2011) and inserting these into the *attP2* landing site on chromosome 3 (Groth et al. 2004). *GFP-Piwi* flies were described in Sienski et al. (2012). Knockdown crosses with Vienna

Drosophila Resource Center lines (Dietzl et al. 2007) were with a *UAS-Dcr2*; *NGT Gal4*; *nosGal4* driver line (Handler et al. 2013). shRNA lines were used for the *Act-Gal4* flip-out clones (Olivieri et al. 2010) except for *fs(1)Yb*, for which a Vienna *Drosophila* Resource Center line (GD25437) was used.

For all experiments, flies were aged for 4–7 d. Knockdown experiments were always accompanied by a negative control of the same genetic background (carrying an shRNA against either *GFP* or *white*). During the last 2 d, flies were kept on apple juice plates supplemented with yeast paste at medium density to promote egg laying and uniform ovary morphology.

Cell culture

OSCs were cultured as described (Niki et al. 2006; Saito et al. 2009). In each experiment, 3×10^6 to approximately 4×10^6 cells were transfected twice with 200 pmol of preannealed siRNAs using Cell Line Nucleofector kit V (Amaxa Biosystems, program T-029) and cultured for 4–5 d (2 d plus 2–3 d) before harvesting (see Supplemental Table S4 for siRNA sequences).

Pacman[GFP-piwi] constructs

The construction of *GFP-piwi* was described previously (Sienski et al. 2012). All other *GFP-piwi* *Pacman* plasmids were constructed via Gibson assembly (Gibson et al. 2009). *D. yakuba* genomic DNA was obtained from the *Drosophila* Species Stock Center. The 3' end of *D. melanogaster piwi intron4* was replaced with the sequence from *piwi intron7* using gBlock DNA fragments (Integrated DNA Technologies). The resulting sequences of shortened intron4 are as follows: *intron4_60nt*, GTAAGACTT TAAACTATATTTAAAT-deletion-CCATAAAAAATGTAATAT TTAATTATTATAAACAG; and *intron4_106nt*, GTAAGACTT TAAACTATATTTAAATTAACAAGCTCTTGTGTCGCAAC-deletion-GCAGTTTAGGGCGGTTTATGGCCATAAAAAA TGAATATTTAATTATTATAAACAG.

Immunohistochemistry, RNA FISH, and image analysis

Rabbit α -Ago3 was described in Brennecke et al. (2007). Mouse monoclonal α -Piwi was raised against the first 150 N-terminal amino acids.

In situ hybridization was performed as described in Lecuyer et al. (2007). Digoxigenin (DIG)-labeled probes were in vitro transcribed from *piwi exon8* cloned into pBluescriptII using the following primers: 5'-ACTGACagcttGCCTCAATGGATCTA CAGCAAA-3' and 5'-ACTGACgtaccCGAACTTGTTCGA GACCAGATA-3'. In situ signal was developed using fluorescein tyramide conjugate (Perkin Elmer).

All fluorescent images were taken on an LSM 780 (Zeiss). GFP intensity was quantified by ImageJ version 1.48f (Figs. 4, 5; Supplemental Fig. S3).

RT-PCR analysis

Total RNA was isolated from OSCs or ovaries using Trizol (Ambion). cDNA was prepared via random priming of total RNA unless otherwise stated. qPCR was performed using standard techniques. All experiments were performed in biological triplicate with technical duplicates. Relative RNA levels were normalized to *rp49* levels. Fold enrichments were calculated in comparison with respective RNA levels obtained from flies expressing a control hairpin or from control siRNA transfections into OSCs (Schmittgen and Livak 2008). Primer sequences for qRT-PCR and RT-PCR analyses are listed in Supplemental Tables S5 and S6.

RNA-seq

PolyA-plus or rRNA-depleted RNA from biological duplicates of OSCs treated with siRNA against *GFP*, *tsu*, or *Acn* was selected with Dynabeads Oligo(dT) (Invitrogen) from total RNA, fragmented, and reverse-transcribed with random hexamers. Strand-specific libraries were prepared using the UDG digestion-based strategy (Zhang et al. 2012) and cloned with NEBNext ChIP-seq library preparation reagent set for Illumina (New England Biolabs), and 50-bp paired-end reads were sequenced on a HiSeq2000 (Illumina) instrument. This yielded 10 million–20 million genome-mappable reads per sample. For the computational analyses, we first extracted high-quality bases from every read (6–50 nt) and mapped these to the *Drosophila* genome as well as to the transcript and intron annotations in FlyBase release 5.56, allowing up to three mismatches. Uniquely aligned reads were used for quantification of gene expression levels and intron levels. Transcripts that scored >10 RPKM (reads per kilobase per million) in *GFP* knockdown ($n = 4203$) and introns whose RPKM values were >10% of their transcript RPKM value in any of biological duplicates ($n = 1478$) were analyzed. The intron retention index was calculated as the intron RPKM level divided by the transcript RPKM level in the respective samples.

Statistical analysis

We used statistical packages implemented in R 2.15.3 for all calculations and plots. For data visualization in box plot format, we used standard features (horizontal bars represent median, the box depicts 25th and 75th percentile, and whiskers represent the 1.5 interquartile range). Statistical significances in Figures 6D and 7D were computed with Mann-Whitney *U*-test; other *P*-values were calculated using Student's *t*-test.

Acknowledgments

We thank members of the Brennecke laboratory for support and comments on the manuscript; J. Cokceza for fly injections; D. Jurczak for bioinformatics support; the Campus Science Support Facilities GmbH (CSF) Next-Generation Sequencing unit for Illumina sequencing; the Institute of Molecular Pathology (IMP)/Institute of Molecular Biotechnology (IMBA) Biooptics Unit for support; and the Vienna *Drosophila* Resource Center, the *Drosophila* Genomics Resource Center, and the Bloomington Stock Center for flies. We thank FlyBase for the continuous genome annotation efforts (version r5.56). All deep sequencing data sets are summarized in Supplemental Table S2. This work was supported by the Austrian Academy of Sciences, the European Community's 7th Framework Programme (FP7/2007-2013; ERC grant no. 260711EU), and the Austrian Science Fund (FWF; Y 510-B12). R.H. is supported by an SFB grant from the FWF (no. F4303). D.I.H. is supported by the University College London and Cancer Research U.K.

References

Ashton-Beaucage D, Therrien M. 2011. The exon junction complex: a splicing factor for long intron containing transcripts? *Fly (Austin)* **5**: 224–233.

Ashton-Beaucage D, Udell CM, Lavoie H, Baril C, Lefrançois M, Chagnon P, Gendron P, Caron-Lizotte O, Bonnell E, Thibault P, et al. 2010. The exon junction complex controls the splicing of mapk and other long intron-containing transcripts in *Drosophila*. *Cell* **143**: 251–262.

Bono F, Gehring NH. 2011. Assembly, disassembly and recycling: the dynamics of exon junction complexes. *RNA Biol* **8**: 24–29.

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

Clark F, Thanaraj TA. 2002. Categorization and characterization of transcript-confirmed constitutively and alternatively spliced introns and exons from human. *Hum Mol Genet* **11**: 451–464.

Czech B, Preall JB, McGinn J, Hannon GJ. 2013. A transcriptome-wide RNAi screen in the *Drosophila* ovary reveals factors of the germline piRNA pathway. *Mol Cell* **50**: 749–761.

De Conti L, Baralle M, Buratti E. 2013. Exon and intron definition in pre-mRNA splicing. *Wiley Interdiscip Rev RNA* **4**: 49–60.

Degot S, Le Hir H, Alpy F, Kedinger V, Stoll I, Wendling C, Seraphin B, Rio MC, Tomasetto C. 2004. Association of the breast cancer protein MLN51 with the exon junction complex via its speckle localizer and RNA binding module. *J Biol Chem* **279**: 33702–33715.

Dietzl G, Chen D, Schnorrer F, Su KC, Barinova Y, Fellner M, Gasser B, Kinsey K, Oppel S, Scheiblaue S, et al. 2007. A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature* **448**: 151–156.

Fededa JP, Petrillo E, Gelfand MS, Neverov AD, Kadener S, Nogues G, Pelisch F, Baralle FE, Muro AF, Kornblihtt AR. 2005. A polar mechanism coordinates different regions of alternative splicing within a single gene. *Mol Cell* **19**: 393–404.

Fox-Walsh KL, Dou Y, Lam BJ, Hung SP, Baldi PF, Hertel KJ. 2005. The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proc Natl Acad Sci* **102**: 16176–16181.

Garg K, Green P. 2007. Differing patterns of selection in alternative and constitutive splice sites. *Genome Res* **17**: 1015–1022.

Gehring NH, Neu-Yilik G, Schell T, Hentze MW, Kulozik AE. 2003. Y14 and hUpf3b form an NMD-activating complex. *Mol Cell* **11**: 939–949.

Gehring NH, Lamprinak S, Hentze MW, Kulozik AE. 2009. The hierarchy of exon-junction complex assembly by the spliceosome explains key features of mammalian nonsense-mediated mRNA decay. *PLoS Biol* **7**: e1000120.

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA 3rd, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**: 343–345.

Glauser DA, Johnson BE, Aldrich RW, Goodman MB. 2011. Intragenic alternative splicing coordination is essential for *Caenorhabditis elegans* slo-1 gene function. *Proc Natl Acad Sci* **108**: 20790–20795.

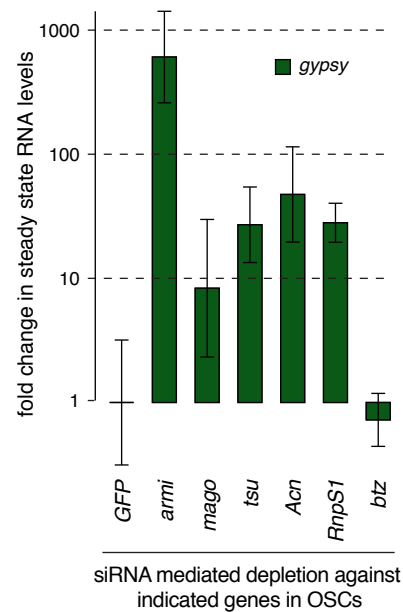
Groth AC, Fish M, Nusse R, Calos MP. 2004. Construction of transgenic *Drosophila* by using the site-specific integrase from phage ϕ C31. *Genetics* **166**: 1775–1782.

Guo M, Lo PC, Mount SM. 1993. Species-specific signals for the splicing of a short *Drosophila* intron in vitro. *Mol Cell Biol* **13**: 1104–1118.

Handler D, Meixner K, Pizka M, Lauss K, Schmied C, Gruber FS, Brennecke J. 2013. The genetic makeup of the *Drosophila* piRNA pathway. *Mol Cell* **50**: 762–777.

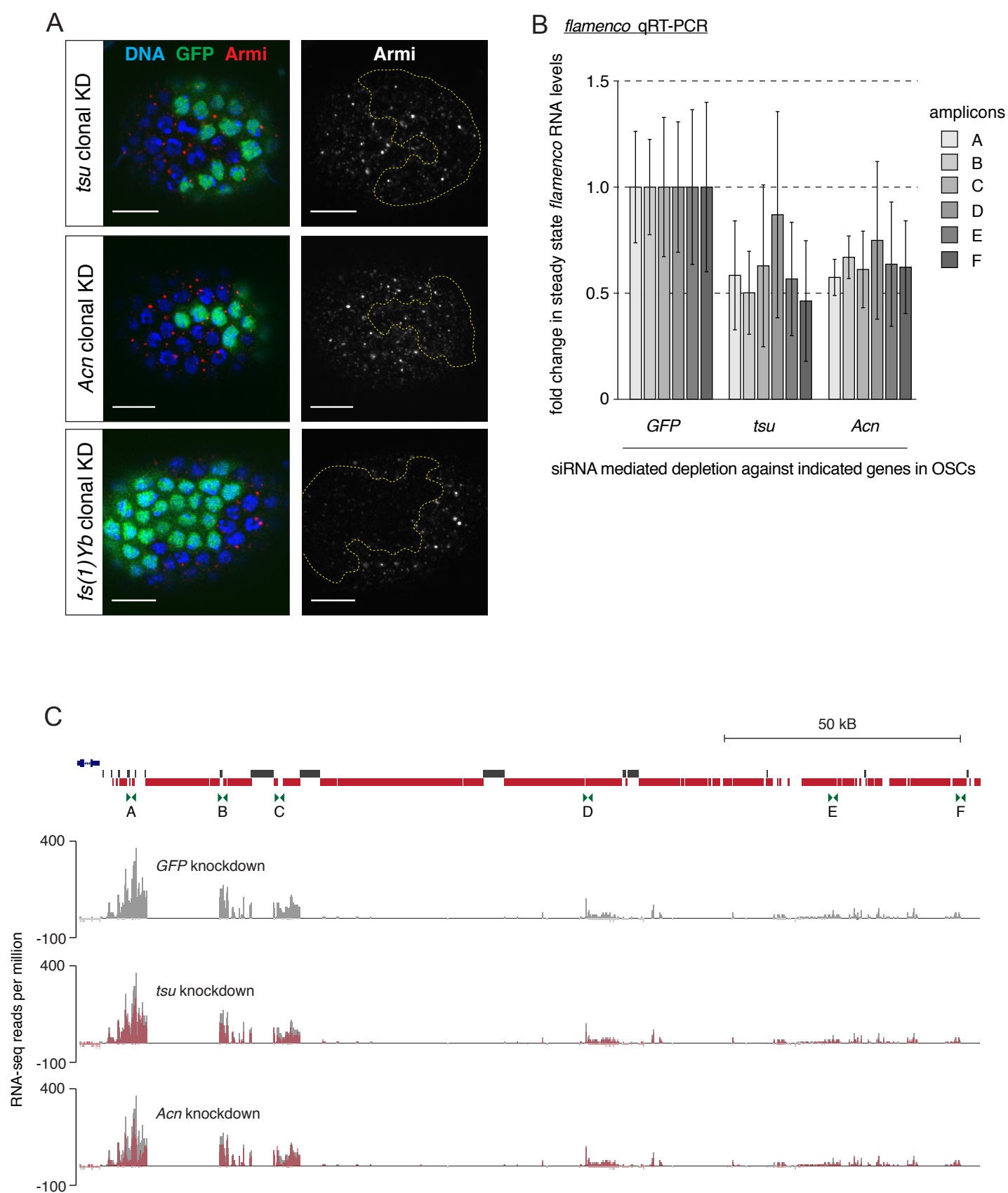
Kaida D, Berg MG, Younis I, Kasim M, Singh LN, Wan L, Dreyfuss G. 2010. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* **468**: 664–668.

Kataoka N, Yong J, Kim VN, Velazquez F, Perkinson RA, Wang F, Dreyfuss G. 2000. Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm. *Mol Cell* **6**: 673–682.



Supplementary Figure S1. The EJC is required for transposon silencing in OSCs.

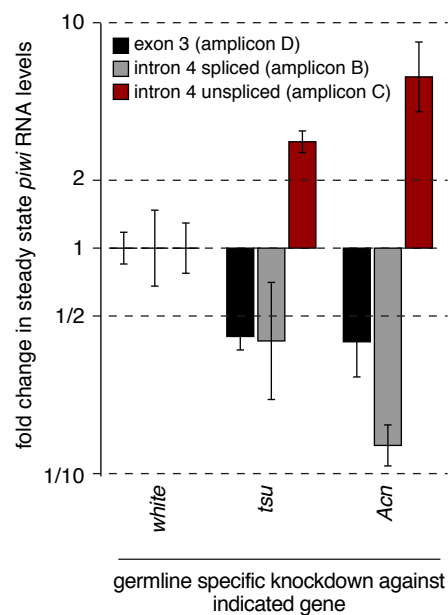
Fold changes in steady-state RNA levels of *gypsy* in OSCs treated with siRNAs against indicated genes, showing that the depletion of EJC factors except for *btz* leads to an enhanced expression of *gypsy*. RNA levels are normalized to *rp49* levels, and values indicate averages of three biological replicates relative to control knockdowns (error bars: stdev.).



Supplementary Figure S2. Early steps in piRNA biogenesis are not affected in EJC-depleted somatic cells.

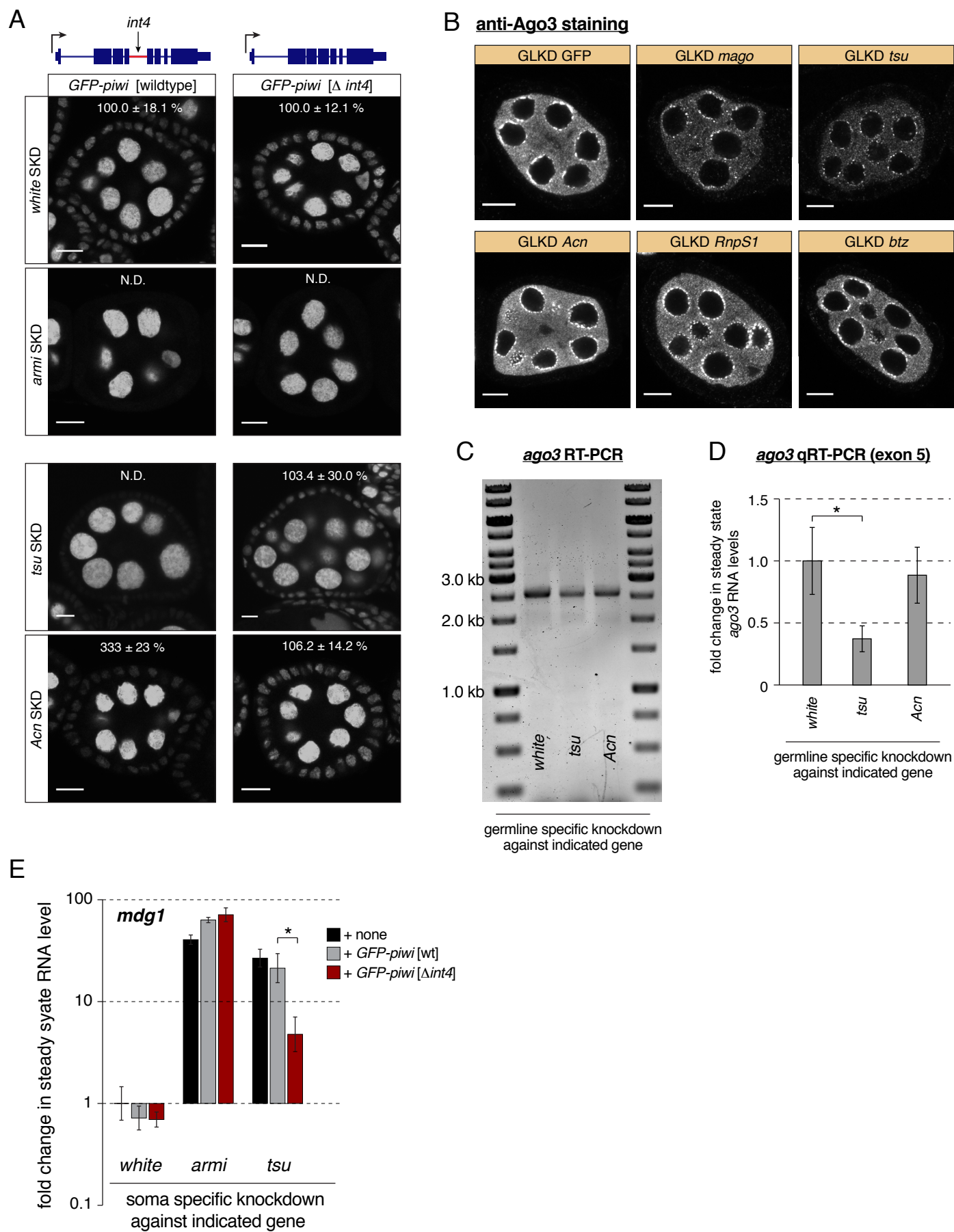
(A) Confocal sections (scale bars = 10µm) of egg-chambers with clones of somatic follicle cells (marked by GFP expression) expressing shRNAs against EJC factors stained for Armi (red) and DNA (blue). Yb bodies (piRNA processing sites marked by cytoplasmic Armi foci) are abolished in *Fs(1)Yb*-depleted cells (control), but remain intact upon knockdowns of *Tsu* or *Acn*.

(B and C) Changes in steady-state levels of flamenco in OSCs treated with siRNAs against indicated genes, showing that the depletion of *Tsu* or *Acn* lead only to mild reductions of flamenco levels. RNA levels were measured by qRT-PCR (B; normalized to *rp49* levels; values indicate averages of five biological replicates relative to control knockdowns; error bars: stdev.) or by RNA-seq (C; normalized to one million genome-mappable reads). RNA-seq profiles at the *flamenco* locus were obtained from rRNA-depleted total RNA from OSCs treated with siRNAs against indicated genes. LINE transposons (red) and LTR transposons (black) are indicated. The RNA-seq tracks of *Tsu*- and *Acn*-depleted samples (red) are overlayed to the tracks of the GFP-depleted control sample (gray).



Supplementary Figure S3. The EJC is required for splicing of *piwi int4* in ovaries.

The fold changes in steady-state RNA levels of all *piwi* transcripts (*exon3*, black bars) and those *piwi* transcripts with spliced (grey bars) or unspliced (red bars) *int4*, showing the decrease of spliced *int4* and the increase of unspliced *int4* in *Tsu*- or *Acn*-depleted ovaries. Total RNA isolated from ovaries depleted for indicated genes in the germline was used for qRT-PCR. Values were normalized to *rp49* levels and averages of three biological replicates relative to control knock-downs are shown with error bars indicating Stdev.



Supplementary Figure S4. A *piwi* transgene lacking *int4* rescues TE de-repression in *Tsu*-depleted somatic follicle cells.

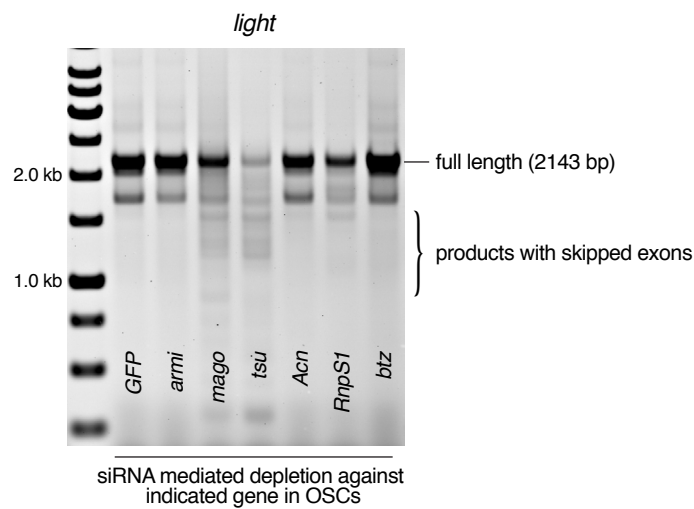
(A) The depletion of *tsu* or *Acn* affects the expression of wildtype *GFP-piwi*, but not of *GFP-piwi* Δ *int4* in follicle cells. Shown are confocal sections (scale bars = 10 μ m) through egg chambers expressing wildtype *GFP-piwi* or *GFP-piwi* that lacks *int4* (Δ *int4*) depleted for indicated genes in somatic follicle cells. All images report relative GFP levels to white SKD from each *GFP-piwi* transgene. Indicated fluorescence levels (\pm Stdev.) were calculated from GFP intensities in nurse cell nuclei normalized to those in follicle cell nuclei from three egg chambers. N.D.: not determined.

(B) Confocal sections (scale bars = 10 μ m) through stage 3-5 egg-chambers expressing shRNAs against EJC factors stained for AGO3 proteins, showing that AGO3 protein levels are decreased upon knockdown of *mago* or *tsu*.

(C) Agarose/EtBr gel showing AGO3 RT-PCR products amplified from poly-A selected RNA isolated from ovaries depleted for indicated genes. Primers mapping to the first and the last exon were used. No mis-spliced species were detected in any knockdown.

(D) Shown are the fold changes in steady-state RNA levels of AGO3 transcripts (*exon5*) amplified from poly-A selected RNA isolated from ovaries depleted for indicated genes. Values were normalized to *rp49* levels and averages of four biological replicates relative to control knockdowns are shown with error bars indicating Stdev. AGO3 transcript levels are decreased in *Tsu*-, but not in *Acn*-depleted cells (*: $p < 0.01$)

(E) Shown are fold changes in steady state RNA levels of *mdg1* in ovaries expressing indicated *GFP-piwi* transgenes and depleted for *white*, *armi* or *tsu* in the soma. The *GFP-piwi* construct lacking *int4* rescues the de-repression of *mdg1* upon depletion of *tsu* (*: $p < 0.01$). RNA levels were normalized to *rp49* levels and averages of three biological replicates relative to control knockdowns are shown (error bars indicate Stdev.).



Supplementary Figure S5. Depletion of core EJC factors but not *Acn* results in exon-skipping of *light* transcripts.

Agarose/EtBr gel image showing RT-PCR products of *light* transcripts amplified from total RNA from OSCs depleted for indicated genes using the primers mapping to the first and the last exons. The depletion of *mago*, *tsu* and to a lesser extent *RnpS1* leads to exon-skipping of *light* transcripts whereas the depletion of *Acn* does not affect the splicing.

Supplementary Table S2. (Illumina sequencing datasets used in this study)

Origin	Method	Genotype	GEO
OSC	RNA-seq (Poly-A+)	si- <i>GFP</i> KD (replicate #1)	GSE58830
	RNA-seq (Poly-A+)	si- <i>GFP</i> KD (replicate #2)	GSE58830
	RNA-seq (Poly-A+)	si- <i>tsunagi</i> KD (replicate #1)	GSE58830
	RNA-seq (Poly-A+)	si- <i>tsunagi</i> KD (replicate #2)	GSE58830
	RNA-seq (Poly-A+)	si- <i>Acinus</i> KD (replicate #1)	GSE58830
	RNA-seq (Poly-A+)	si- <i>Acinus</i> KD (replicate #2)	GSE58830
	RNA-seq (Δ rRNA)	si- <i>GFP</i> KD	GSE58830
	RNA-seq (Δ rRNA)	si- <i>tsunagi</i> KD	GSE58830
	RNA-seq (Δ rRNA)	si- <i>Acinus</i> KD	GSE58830

Supplementary Table S3. (Oligos used for cloning of shRNA constructs)

Gene	Sequence
<i>tsunagi</i>	fw: ctagcagtCTGCGATTACGGAGAAATCAAtagttatattcaagcataTTGATTTCTCCGTAATCGCAGgcg rv: aattcgcCTGCGATTACGGAGAAATCAAtatgcttgaatataactaTTGATTTCTCCGTAATCGCAGactg
<i>eIF4AIII</i>	fw: ctagcagtAAGGGTTTCAAGGAACAGATAtagttatattcaagcataTATCTGTTCCCTTGAAACCCTTgcg rv: aattcgcAAGGGTTTCAAGGAACAGATAtatgcttgaatataactaTATCTGTTCCCTTGAAACCCTTActg
<i>barentsz</i>	fw: ctagcagtTGGCATGGAATTTAAGAAGAAtagttatattcaagcataTTCTTCTTAAATTCCATGCCAgcg rv: aattcgcTGGCATGGAATTTAAGAAGAAatgcttgaatataactaTTCTTCTTAAATTCCATGCCAactg
	Others
<i>GFP</i>	(Handler et al. 2013)
<i>white</i>	(Handler et al. 2013)
<i>armitage</i>	(Handler et al. 2013)
<i>mago nashi</i>	(Handler et al. 2013)
<i>Acinus</i>	(Handler et al. 2013)
<i>RnpS1</i>	(Handler et al. 2013)

Supplementary Table S4. (siRNAs used for RNAi in OSCs)

Gene	sequence
<i>mago nashi</i>	Guide: UCCUCCUGCAUGAUCUCCGUU Passenger: CGGAGAUCAUGCAGGAGGAUU
<i>tsunagi</i>	Guide: UUGAUUUCUCCGUAUUCGCUU Passenger: GCGAUUACGGAGAAAUCAAUU
<i>Acinus</i>	Guide: UUGUUAUCGCGGUCUUUGCUU Passenger: GCAAAGACCGCGAUAAACAAUU
<i>RnpSl</i>	Guide: UUAACAUAUUGCUAUCGGUU Passenger: CCGAUAGCAAUAAUGUAAAUU
<i>barentsz</i>	Guide: UUCUUCUUAUUUCCAUUGCUU Passenger: GCAUGGAAUUUAAGAAGAAUU
	Others
<i>GFP</i>	(Handler et al. 2013)
<i>armitage</i>	(Handler et al. 2013)

Supplementary Table S5. (Primers for QPCR analysis)

Gene	Sequence
<i>flamenco</i> (amplicon A)	fw: GAAGTCTTGGGACACTCATAGGT rv: CCAGAAAATTAAGCGGAAGC
<i>flamenco</i> (amplicon B)	fw: TGCAATTCCCCAAATCTCGATCC rv: GGACACATGGAAGCTTCGAAGAA
<i>flamenco</i> (amplicon C)	fw: TGTAGTCTTTGCAGTGTGTCAGTGT rv: CCTGCATAAACGGATCGGTGATA
<i>flamenco</i> (amplicon D)	fw: TTGCCTCAGTGAAACGCCTAAAA rv: TTCCCTTATTGAACATCACCGCC
<i>flamenco</i> (amplicon E)	fw: GTCAAGTGTCTTTTGCTGTGCTG rv: CCCACCCTTGCATAGTCTCTCTA
<i>flamenco</i> (amplicon F)	fw: AATGTTTCGGGTTCTAGGGTAGC rv: TCAGCAACGAACATGATTACCTAGT
<i>piwi intron4_spliced</i>	fw: AATTCCTGAGCTCTGCCGAGTG rv: TAACTGCTCATGGCACGCATAA
<i>piwi intron4_unspliced</i>	fw: AAACGAAGAAAAATCTGAGCACT rv: TAACTGCTCATGGCACGCATAA
<i>piwi exon3</i>	fw: TTGGGCACCGAAATAACTCA rv: TTACCCGTACTTCGTCCTGATG
<i>ago3 exon5</i>	fw: CCAACAAGATAAGTGGGCGTCA rv: TTGGTCACAGCATGTTTGGAGA
<i>cg8671 intron1_spliced</i>	fw: CAAACGAAAGCAAAGGCAGTTG rv: ATATAGCGGAGACACATCCGAA
<i>cg8671 intron1_unspliced</i>	fw: TGTTCGGTGCTGTTGTGTTTGT rv: ATATAGCGGAGACACATCCGAA

<i>garnet intron7_spliced</i>	fw: CCTTAGAACCCCGGCTAGGAAA rv: CAGAGCTGAATGGAGGCACTGT
<i>garnet intron7_unspliced</i>	fw: CGAACCAAAAACCAAACCCAAC rv: CAGAGCTGAATGGAGGCACTGT
	Others
<i>rp49</i>	(Handler et al. 2013)
<i>HeT-A</i>	(Handler et al. 2013)
<i>blood</i>	(Handler et al. 2013)
<i>mdgl</i>	(Handler et al. 2013)
<i>gypsy</i>	(Muerdter et al. 2013)

Supplementary Table S6. (Primers for RT-PCR analysis)

Gene	Sequence
<i>piwi</i>	fw: CAGGCGTCCACTTAACGAAG rv: GTGTGCTCTTCGCAATGTCA
<i>gasz</i>	fw: AGCAACCTTTGCAAATACGG rv: CTTTTTGCGTTTTTCGCTCTG
<i>shutdown</i>	fw: TGGAAGAAACTTCGAACCGTA rv: CCAAGTTGTACTCGCCCAGA
<i>ago3</i>	fw: TTAGTTTTCTGTTTTTGTGGTA rv: CATCTCGTTGAATACTTTGTCC
<i>rolled</i>	(Ashton-Beaucage et al. 2010)
<i>light</i>	(Ashton-Beaucage et al. 2010)

REFERENCES

- Ashton-Beaucage D, Udell CM, Lavoie H, Baril C, Lefrancois M, Chagnon P, Gendron P, Caron-Lizotte O, Bonneil E, Thibault P et al. 2010. The exon junction complex controls the splicing of MAPK and other long intron-containing transcripts in *Drosophila*. *Cell* **143**: 251-262.
- Handler D, Meixner K, Pizka M, Lauss K, Schmied C, Gruber FS, Brennecke J. 2013. The Genetic Makeup of the *Drosophila* piRNA Pathway. *Mol Cell* **50**: 762-777.
- Muerdter F, Guzzardo PM, Gillis J, Luo Y, Yu Y, Chen C, Fekete R, Hannon GJ. 2013. A Genome-wide RNAi Screen Draws a Genetic Framework for Transposon Control and Primary piRNA Biogenesis in *Drosophila*. *Mol Cell* **50**: 736-748.

5.3 Other publications

During my doctoral time, I was also involved in two additional projects:

1. Mohn F., Sienski G., Handler D. & Brennecke J. **The Rhino-Deadlock-Cutoff Complex Licenses Noncanonical Transcription of Dual-Strand piRNA Clusters in *Drosophila*.** *Cell* **157**, 1364–1379 (2014).

I performed the experiments in: Fig1 A-B; Fig4 A; Fig5 E-F; Fig7 D; FigS1 A-F; FigS6 C

2. Mohn F*, Handler D*. & Brennecke J. **piRNA-guided slicing specifies transcripts for Zucchini dependent, phased piRNA biogenesis.** *Science* [currently under review]

*These authors contributed equally to the work

In close collaboration with Fabio Mohn, I performed the entire experimental and computational work that underlies this study.

6 Discussion

6.1 Evaluation of the somatic piRNA pathway screen

6.1.1 Decisions underlying the design of the screen

Prior to the initiation of the RNAi screen several key decisions had to be taken. The three central ones concerned in which cell type we should screen, which RNAi system we should use and which readout for piRNA pathway integrity we should rely on.

The decision to perform the screen in the somatic tissue of *Drosophila* ovaries was based on biological and technical reasons. One of the main reasons was the simplicity of the somatic piRNA pathway. The sole PIWI protein expressed in the soma is Piwi and only the primary but not the secondary piRNA pathway is active. This eliminated the risk that screen hits are covered by a potential redundancy of the primary and the secondary piRNA pathways.

Another important reason for the somatic tissue was the availability of a functional genome wide RNAi library. At the time of screen initiation, RNAi by long double stranded hairpins was widely used in somatic tissues of *Drosophila* and the near genome wide collection of RNAi lines available at the Vienna *Drosophila* RNAi Center has been the basis for several genome wide screens^{93,112-115}. In contrast, efficient gene knockdown by RNAi in germline cells was only possible by the expression of short hairpin RNAs for which no library existed⁹¹. Only later was it shown that the level of Dicer-2 in the germline limits the efficiency of RNAi by long hairpin RNAs and that overexpression of Dicer-2 together with

the VDRC RNAi lines overcomes that problem¹¹⁶. Based on these observations the primary screen was performed in the somatic tissue utilizing RNAi lines from the VDRC library. After the primary screen, validated genes were tested for their functional contribution to the germline piRNA pathway utilizing VDRC lines in the germline.

We decided to screen only genes expressed in OSCs, which harbor a fully functional primary piRNA pathway. Towards this end, we determined the expression level of genes in wildtype OSCs by RNA sequencing. The cutoff for calling a gene express was set at 1 sequenced read per kilobase per 1 million sequenced reads (1 RPKM). This rather conservative cutoff resulted in 7257 genes selected for testing. Due to limitations of the VDRC library only 6818 of these were tested. In the screen, no strong screen hit was observed for lowly expressed genes (<RPKM10) and out of 334 tested genes with RPKM<1 not a single one showed any reporter staining. In addition nearly no morphology distortions were observed upon depletion of genes expressed at RPKM levels below one. Taken together, these results are very reassuring that the gene expression cutoff for the screen was set at a valid level.

A key factor for the success of the screen was the selection of a suitable readout for piRNA pathway integrity. I took advantage of a reporter that has been shown previously to efficiently detect perturbations in the piRNA pathway^{69,92}.

6.1.2 Quality measures of the screen

Like with every screen the results needed to be carefully evaluated. Key numbers reflecting the quality of a screen are the false-negative and false-positive rates. Both values are largely influenced by effects related to the RNAi system, by secondary consequences of gene knockdowns on tissue development and by the sensitivity and specificity of the reporter.

A clear advantage of RNAi screens over chemical mutagenesis screens is the specific targeting of genes instead of random gene mutation. Extensive mapping of mutations to identify the affected gene is therefore not required. On the other hand, RNAi screens suffer from a certain proportion of non-functional hairpins and from off-target effects that result in increased false negative and false positive rates.

The fraction of non-functional VDRC lines must be evaluated separately for each screen and can be calculated from the number of hairpins that do not result in the examined phenotype. As the number of known piRNA pathway factors was limited, RNAi efficiency could not be calculated reliably based on positive reporter results. Therefore I calculated the RNAi efficiency based on whether depletion of central house keeping genes resulted in a clear ovarian morphology phenotype. Many genes and even gene-groups are known to be essential for the cell. Therefore these gene-groups are well suited for the RNAi efficiency calculation. One of the largest gene-groups known to be essential for cells are ribosomal protein encoding genes. Around 80% of the VDRC lines targeting these genes caused a detectable morphology

phenotype when expressed in follicle cells. Considering that we tested typically two independent VDRC lines per gene, ~93% of the genes encoding ribosomal proteins show a morphology phenotype with at least one of the tested VDRC lines. Similar rates were observed if other essential gene categories were analyzed. Together with the observation that only one gene known to act in the somatic piRNA pathway was missed in the screen, I conclude that the false negative rate of the screen is in the range of 10-20%⁵⁴.

An estimation of the off-target rate (a major cause for false positive hits) is not directly possible. Off-target effects mostly arise when RNAi hairpins target a non-wanted gene non-specifically. The best approach is to test screen hits with independent RNAi lines. We therefore constructed for 30 screen hits shRNA lines and tested these in the same assay. Expression of four of the shRNA lines resulted in a block of oogenesis and therefore these genes could not be evaluated. Of the remaining 26 shRNA lines 80% confirmed the result from the initial screen.

Another off-target effect is the unspecific activation of gene expression by driving expression of the RNAi hairpin construct in case this is inserted close to a gene locus. This effect only applies for RNAi hairpins that are randomly integrated into the genome, which was the case for 78% of the VDRC hairpins used the screen. I argue, however, that artificial activation of gene expression is of minor concern for this screen. This is as it is rather difficult to imagine that overexpression of a gene would lead to piRNA pathway disruption.

As mentioned before depletion of several genes in follicle cells resulted in morphological distortions of the ovary. A key decision for the screen was therefore to manually score the reporter stainings. This way, reductions in tissue size did not affect the evaluation and only complete loss of ovaries prevented the scoring of the phenotype. In the screen ~10% of the genes could not be evaluated due to severe tissue loss. KEGG term analysis revealed that many of these genes are involved in basic cellular processes. It cannot be excluded that reporter phenotypes were masked by severe morphological distortions. Indeed, with eIF4AIII one of the genes resulting in no ovaries has been reported to be involved in the piRNA pathway. eIF4AIII is one of the core proteins of the exon junction complex that are important for the piRNA pathway¹¹⁷.

The sensitivity and the specificity of the reporter assay are crucial for the success of a screen. The reporter I used was shown to be a valid tool for the detection of piRNA pathway perturbations^{69,92}. The combination of direct silencing of the reporter-gene by the piRNA pathway with the manual evaluation of the staining intensity ensured a high sensitivity throughout the screen. Indeed, RNAi mediated depletion of all known piRNA pathway components (with the exception of CG31755 for which no functional VDRC line exists⁵⁴) resulted in robust reporter expression^{54,59,69,70,72,74,118}.

With the high sensitivity of the reporter the rate of false positives must to be evaluated carefully. In total 144 genes scored in the screen. Besides the known pathway members, only ~30 additional genes showed strong reporter de-silencing. The remaining screen hits showed weak or intermediate reporter staining at best. RT-qPCR demonstrated a high correlation of reporter staining intensities with the increase in steady state transposon RNA levels. In addition also the steady state RNA level of the reporter mRNA was directly connected to staining intensities. Therefore the staining intensity in the screen was a good indicator for the strength of the distortion in the piRNA pathway.

Based on all the data presented, the reporter used for the screen shows high sensitivity and specificity for the identification of genes involved in the piRNA pathway. I would estimate that the screen recovered at least 80% of the genes involved in the somatic piRNA pathway.

6.1.3 Participation of screen hits in somatic and germline piRNA pathways

The main focus of the screen has been the identification of novel factors acting in the somatic piRNA pathway. Due to the similarities in the primary pathway in ovarian soma and germline, it was of interest if the factors identified in the somatic screen are also important for the germline pathway. Therefore, I retested all screen hits in the germline with a reporter designed to detect even minor changes in the germline piRNA pathway. The germline reporter (Burdock-lacZ) was developed in our lab and was used in combination with VDRC line driven germline knockdowns. We showed that this combination provides very good sensitivity as RNAi-mediated depletion of different factors involved in the primary and secondary piRNA pathway resulted in reporter deregulation. All 144 hits from the somatic screen were retested with the germline reporter. Of these, 29 scored at least with medium staining intensity. Measurements of steady state RNA levels of the reporter and germline specific transposons showed high correlation with the germline reporter staining intensities.

Interestingly 36 genes scoring highly in the somatic tissue did not show any importance for the germline piRNA pathway although a similar linear piRNA pathway is active in both tissues. This argues for a diversification of the piRNA pathway between the soma and the germline. Indeed, it was shown that the structure and transcription of piRNA clusters differs substantially between soma and germline. Therefore, cluster-related processes are likely to be controlled differently, too. In support of this, it has been shown that for germline cluster transcription the action of a germline specific protein complex consisting of the factors Rhino, Deadlock and Cutoff is required^{56,119,120}.

6.2 Screen results

A bulk analysis of gene functions by gene ontology (GO) enrichments resulted in a remarkable enrichment for GO-terms involved in mitochondrial metabolism. In total 38 genes that were identified in the screen were associated with mitochondrial GO-terms. Nearly all of these genes scored only mildly and specifically in the soma but not in the germline. In addition their depletion resulted in morphological distortions of the ovary. Based on RT-qPCR results depletion of these factors caused a weak but significant de-repression of piRNA pathway repressed transposons. Therefore these hits are most likely real hits that only mildly affect piRNA pathway function in the soma. It seems however rather unlikely that general mitochondrial metabolism is directly involved in the piRNA pathway. A more likely explanation is that upon depletion of these genes mitochondrial number or vitality is reduced and therefore mitochondrial-associated processes of the piRNA pathway are perturbed. Due to the high likelihood that these genes are only indirectly involved in the piRNA pathway I excluded them from further analyses.

When repeating the GO-term analysis after excluding the set of mitochondrial metabolism genes the most enriched cellular component GO-terms were nucleus, the Yb-body and the P-body. The Yb-body term is explained by the efficient identification of the known biogenesis factors in the screen. Interestingly, it has previously been shown that the Yb-body is in close proximity or even in connection with the P-body in the *Drosophila* soma⁶⁹. Also in mouse several proteins involved in piRNA biogenesis colocalize with P-body components in so-called piP-bodies in the male germline¹²¹. Considering that depletion of P-body components typically leads only to mild TE de-repression, the functional implications of P-body components in the piRNA pathway are currently unknown.

Besides the GO enrichment analysis, I also grouped the screen hits manually into functional groups based on gene function studies and the literature. Of note, in every functional group at least one gene was among the highest scoring genes in the screen. In the following paragraphs I will introduce the most prominent functional gene groups and discuss their potential role in the piRNA pathway.

6.2.1 A genetic link to transcription elongation

Based on the literature, several screen hits have a strong link to transcription, in particular to transcription elongation. These genes can be grouped into components of the Paf1 complex and factors implicated in TFIIIS biology.

The Paf1 complex (Paf1c) is a multi-protein complex that was first characterized in yeast¹²²⁻¹²⁵. The yeast core complex consists of the five proteins Paf1, Leo1, Rtf1, Cdc73 and Ctr9, but additional factors have been identified in the human Paf1c (hSki8)¹²⁶. The Paf1c is

implicated in many processes. Among these are regulation of transcription initiation and elongation, modification of histones, phosphorylation of the Pol II C-terminal domain and participation in RNA processing and export events (reviewed in¹²⁷).

The Paf1c member that scored strongly in the screen is the *Drosophila* ortholog of Leo1/Atu. The core Paf1c factors, Paf1/Atms, Ctr9/CG2469 (identified by BLAST search) and Cdc73/Hyx could not be evaluated in the screen as their depletion led to ovary loss. Rtf1 on the other hand showed no reporter staining. Of note, Rtf1 was shown to be not a permanent member of the *Drosophila* Paf1c, possibly explaining this negative result¹²⁸. Besides Atu, the gene CG3909 scored strongly in the screen. CG3909 shows homology to the human and yeast Ski8 genes. Interestingly, CG3909 was shown to interact with the Paf1 core members Atms and Hyx in *Drosophila*¹²⁹. It therefore seems plausible that CG3909 as a bona fide Paf1c member in *Drosophila*. The *Drosophila* Paf1 complex has been implicated in modulation of chromatin structure at actively transcribed genes¹²⁸. Although disruption of *Drosophila* Paf1c does not change RNA Pol II occupancy at active gene loci over the gene-body, transcript levels of heat shock genes are reduced two fold¹²⁸.

Besides the Paf1c, another clear link to transcription biology is based on TfiIS scoring as one of the strongest hits in the screen. TfiIS has been implicated in the regulation of RNA Pol II fidelity. The main function of TfiIS is the rescue of arrested RNA Pol II by multiple mechanisms¹³⁰. Which of the molecular functions that TFIIS exerts on Pol II is required for the piRNA pathway is unclear. Interestingly, TfiIS interacts with the Paf1c^{122,131}. It therefore seems plausible that the Paf1c in conjunction with TFIIS is required for a critical step in piRNA pathway biology.

The two most obvious steps where the Paf1c and TFIIS might impact the pathway are at the level of piRNA cluster transcription or at the level of transcriptional TE silencing. In my opinion, the former is more likely due to several reasons: First, both the Paf1c and TfiIS, have been shown to impact active transcription processes and have not been linked to transcriptional silencing. In addition, all of the screen hits linked to these two complexes score only in the somatic pathway and it seems rather unlikely that the process of piRNA guided transcriptional silencing differs between soma and germline. In contrast, piRNA cluster biology is clearly different in soma and germline. For example, germline clusters require the RDC complex for their transcription, while somatic clusters do not⁵⁶. Maybe the large transcription unit of the *flamenco* cluster- the most prominent piRNA cluster in somatic follicle cells- poses a significant challenge for RNA Pol II that can only efficiently be overcome with the help of dedicated transcription elongation factors.

6.2.2 Screen hits involved in RNA transport

A second clear scheme that emerged from the screen is the process of nuclear RNA export. Two of the strongest hits in the screen were Sbr/Nxf1 and Nxt1. In vivo, Nxf1 and Nxt1 form a stable heterodimer, which is required for mRNA export from the nucleus¹³². The general importance of these genes in RNA export is reflected by their depletion resulting in severe disruptions in oogenesis and ovarian morphology. Still, the severe de-repression of endogenous TEs upon depletion of either factor argues for a direct role of these genes in the somatic piRNA pathway. Potentially these proteins -in conjunction with special nucleoporin components (discussed below)- are critically involved in piRNA cluster transcript export from the site of their transcription to the biogenesis sites in the cytoplasm.

I also tested the involvement of these two RNA export factors in the germline piRNA pathway. While Nxf1 scored weakly in the germline assay, a role for Nxt1 in the germline pathway could not be assessed as its depletion led to rudimentary ovaries. Interestingly, the *Drosophila* genome encodes for multiple *nxf* genes but only one *nxt1* gene. While *nxf1* is expressed ubiquitously, *nxf2* and *3* are preferentially expressed in ovaries. Depletion of either *Nxf2* or *3* via germline specific RNAi resulted in severe transposon de-repression (unpublished observations). This is in line with *Nxf2* being identified as piRNA pathway factor in a screen focusing on the germline piRNA pathway¹³³. These combined findings strongly support a critical role for RNA export processes in the somatic and germline piRNA pathways.

6.2.3 A role of the nuclear pore complex in the somatic piRNA pathway

One remarkable outcome of the screen was the identification of several proteins directly implicated in nuclear pore complex biology. Specifically, these are Nup54, Nup58, Nup214 and CG11092/Nup93-1. The nuclear pore is a large macromolecular complex that spans the two nuclear membranes. A schematic illustration can be seen in Fig6. It allows the regulated transport of molecules from the nucleus to the cytoplasm and vice versa. The pore consists of a structural part that anchors the pore within the nuclear membrane. On this structural part a set of linker proteins attach, which build the structure of the central pore. On the nuclear face of the pore a set of proteins build a basket like structure, while the cytoplasmic face is decorated with several cytoplasmic filaments. The pore itself is filled with a mesh of FG-repeat peptides from proteins that are anchored in the pore complex. Together these various structures ensure the tightly regulated, but efficient transport of molecules across the nuclear membranes (reviewed in ¹³⁴).

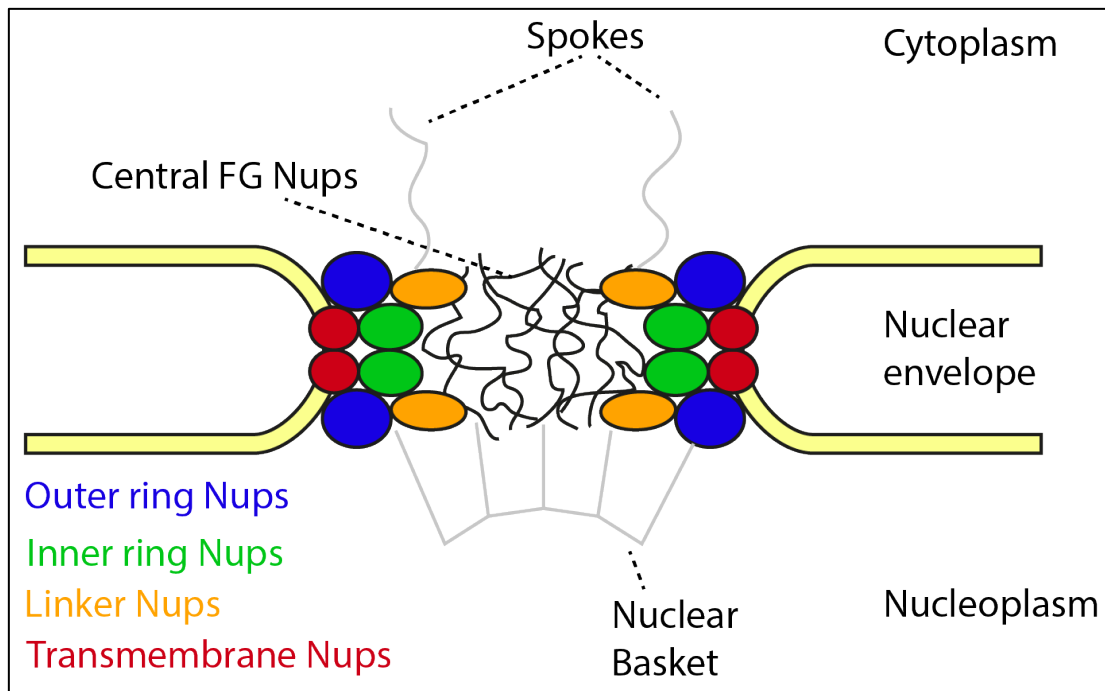


Fig6 Schematic representation of the nuclear pore complex. (modified from ¹³⁴)

The nucleoporins identified in the screen belong to the group of central FG-repeat containing proteins (Nup54 & Nup58), to the cytoplasmic FG-repeat containing proteins (Nup214) and to the linker Nups (Nup93-1). Interestingly the proteins have a different level of impact on the piRNA pathway. Whereas the central pore-filling Nups Nup54 & Nup58 scored with a very strong phenotype, Nup214 scored intermediate and the linker protein Nup93-1 scored only weakly.

Depletion of Nup54 and Nup58 also resulted in strong de-repression of endogenous piRNA pathway controlled TEs. Generally, their importance for the pathway is close to core pathway factors. Although both factors belong to the central FG-repeat containing proteins, other proteins belonging to this group in *Drosophila* (Nup98, Nup62) did not score in the screen. Only depletion of Nup98 resulted in a distorted ovarian morphology and is therefore probably important for general nuclear pore function. Considering that ovarian morphology is wildtype upon Nup54/58 depletion, these results argue for a highly specific role of Nup54 and Nup58 in the piRNA pathway.

Based on the phenotypic strength I would attribute Nup93-1 only an indirect role in the piRNA pathway. Of note, Nup93-1 is one of the linker proteins within the nuclear pore that interacts with the central FG-repeat containing proteins (including Nup54 & Nup58). Therefore loss of Nup93-1 could result in a partial perturbation of Nup54 and Nup58 resulting in a mild piRNA pathway phenotype. Interestingly, in *Drosophila* a paralog to Nup93-1 exists (CG7262/Nup93-2)¹³⁵. CG7262 scored with an intermediate phenotype in the primary screen, but the phenotype was not confirmed by re-tests. One possibility is that the two Nup93

paralogs in *Drosophila* act in a redundant manner and a double knockdown would result in more severe piRNA pathway defects. From the positioning within the nuclear pore, I still propose that the main role for Nup93 is the positioning of the FG-repeat containing nucleoporins, which would imply an indirect role of Nup93 in the piRNA pathway.

The last nucleoporin identified in the screen is Nup214. It is also an FG-repeat containing protein, but it is located on the cytoplasmic side of the nuclear pore. The general function of Nup214 is unclear, making speculations about its function in the piRNA pathway difficult.

At the moment it is unclear what the specific role of the nucleoporins is in the somatic piRNA pathway. One possibility is that they are intricately involved in the efficient export of piRNA cluster transcripts. It has been shown, that the Yb-body –the cytoplasmic center for piRNA biogenesis- is located close to the nuclear membrane next to an accumulation of piRNA precursor transcripts^{69,70,136}. Based on unpublished observations in our laboratory, it appears as if the Yb-body is always juxtaposed to the nuclear location of the *flamenco* DNA locus. It is therefore intriguing to propose a model, wherein nuclear transcription of piRNA clusters is directly coupled to the cytoplasmic processing machinery via specialized nucleoporins and the RNA export factors Nxf1 and Nxt1.

6.2.4 Factors potentially involved in transposon silencing

TE silencing by Piwi has been shown to mainly occur at the transcriptional level⁵⁹ more REFS. The molecular events that underlie TGS, however, are currently unknown. It has been observed that target silencing is accompanied by the local formation of a histone H3 Lysin 9 trimethylation (H3K9me3) domain, which is one of the major repressive histone marks⁵⁹. In fact, Piwi induced TE silencing is responsible for most H3K9me3 islands in euchromatin. Three of the screen hits appear to be directly involved in Piwi mediated TGS and importantly, their disruption does not impact heterochromatin biology at non-Piwi target loci. These are the factors Maelstrom, CG9754 and CG3893. As Maelstrom and CG3893/Gtsf1 were studied in detail by two colleagues in the laboratory^{59,137}, I performed some initial follow-up experiments on CG9754, which were part of the screen publication⁴⁷. Though the combined efforts on these three factors clearly support their integral role in Piwi mediated TGS, clues about their exact molecular functions are still missing. It seems likely, however, that target recognition by Piwi leads to a conformational switch in the Argonaute structure, which initiates the assembly of a chromatin modifying and silencing complex.

As none of the identified piRNA pathway specific silencing factors appears to exhibit any enzymatic activity that would explain the silencing phenotype, it seems likely that these factors at some point link to cellular activities that have evolved for heterochromatin formation and transcriptional silencing in general. Of note, the screen identified multiple other proteins with functions related to chromatin biology. One interesting candidate is the

Lysine methyl-transferase Eggless (Egg/Setdb1). Egg has been shown to be responsible for most methylation marks on H3K9 in euchromatin^{138,139}. For its function Egg requires the presence of its cofactor Windei (Wde)¹⁴⁰, which also scored in the screen.

Another intriguing hit in the screen is the Histone variant H2Av, which is a variant of the main histone H2A. H2Av has multiple functions and is associated with active and silent chromatin. As one of its major functions, H2Av is involved in the formation of H3K9me3 marked repressive chromatin. It has been reported that one of the most upstream steps of H3K9 methylation is the exchange of H2A with H2Av, which initiates a cascade of events resulting in H3K9 methylation^{141,142}.

Based on this it is intriguing to hypothesize that Piwi binding to a nascent target RNA initiates the formation of silent chromatin by utilizing general cell biological machineries that are potent inducers of heterochromatin. Detailed genetic and possibly in vitro experiments will be required to dissect the exact cascade of molecular events that underlie Piwi mediated TGS.

6.3 Implications of the screen for other fields of biology

The piRNA pathway requires specific processes carried out by specific proteins at several key steps. However, some processes in the piRNA pathway require more general cellular processes. Therefore, it is no surprise that the screen uncovered many proteins involved in general cell biological processes. Examples for these processes are factors involved in transcription, RNA transport/export and chromatin remodeling factors. Although these factors are not specific to the piRNA pathway, targeted studies of their role in the piRNA pathway will be of great importance. Such studies very likely will not only provide insight into the functions of these factors in the piRNA pathway, but also will result in a better understanding of the factors itself. In addition, the piRNA pathway can be seen as an isolated environment to study the function of these factors based on a specific phenotype.

One very good example of such a “spin-off” are studies on the EJC we and others^{117,143} performed. Initially, we hypothesized that the EJC has an important function in piRNA cluster biology. In the end we were studying splicing defects on *piwi* mRNA upon EJC depletions. Although this study offered limited insight into piRNA pathway biology in general, it provided key insights into the role of the EJC in splicing. For example, both studies demonstrated a central role of the EJC in the definition and proper splicing of poorly defined introns. Furthermore, they demonstrated a clear genetic link to the co-factor Acinus. Most significantly, however, they showed that the EJC impacts splicing only after its deposition to a splice junction that is close-by to the problematic intron.

Previously the EJC has only been implicated in splicing of large heterochromatic genes that are difficult to study. The fact that *piwi* resembles a pretty average *Drosophila* gene, opened the possibility to study the splicing defects in more detail and to uncover intron features that underlie the dependency of splicing on the EJC. Based on this, we think that our study on the EJC is interesting for many fields of research and provides tools that can lead to a better understanding of EJC biology.

I predict that several other screen hits will similarly offer a unique window to study general cellular processes. For example, the study of the export route of piRNA cluster transcripts could result in a better understanding of RNA export through the nuclear pore in general.

6.4 The emerging role of mitochondria in the piRNA pathway

6.4.1 Several piRNA biogenesis factors are associated with mitochondria

In a follow up of the screen, I showed that Gasz is an essential factor for piRNA biogenesis. Gasz is anchored in the outer mitochondrial membrane by a C-terminal transmembrane domain. The role of Gasz in the pathway is likely to recruit upstream biogenesis factors such as the RNA helicase Armitage to mitochondria⁴⁷. This is consistent with the cellular localization patterns observed for Armi in the germline in wildtype versus Gasz depleted cells. Consistent with this is the fact that Gasz harbors Ankyrin repeats and a SAM domain, both of which are protein domains implicated in specific protein-protein interactions.

Interestingly, mitochondrial localization is a recurrent theme in the pathway. Besides Gasz several other piRNA biogenesis factors were shown to have a close association with mitochondria. Thus, the role of mitochondria is an emerging topic in piRNA biology. In fact, the outer mitochondrial membrane was predicted to serve as platform for piRNA biogenesis factors in most organisms studied ranging from *Drosophila melanogaster* and *Bombyx mori* to *Mus musculus*^{70,73,79,80,144,145}.

The first piRNA biogenesis factor shown to be associated with mitochondria was Zucchini. It harbors an N-terminal transmembrane domain, which is required for its mitochondrial localization^{47,69,70,73,76-78}. Zucchini is a 25kDa protein and contains a single Phospho-lipase D domain. It therefore was initially categorized as member of the phospholipase D family, which is implicated in the generation of phosphatidic acid (PA)¹⁴⁶. Subsequent structural studies revealed that Zuc is an endonuclease generating piRNA 5' ends^{77,78}, which is in line with initial predictions based on Zucchini's similarity to the bacterial endonuclease Nuc⁷⁶. Based on the crystal structures Zuc contains a positively charged groove with the active site in the center. Sterically, this groove can accommodate single stranded but not double stranded nucleic acids, which is in line with the observation that piRNA precursors are single stranded RNAs. Zucchini is probably a sequence independent nuclease, consistent with a modeling

attempt that predicted interactions of the groove in Zucchini with the RNA backbone. In fact, in vitro studies support the proposed sequence independence of the nuclease activity.

Another piRNA biogenesis factor shown to localize to mitochondria in soma and germline is Minotaur (Mino)⁸¹. Remarkably, Mino is a member of the highly conserved GPAT protein family that is involved in the PA pathway. However it turned out that the enzymatic function of catalytic domain in Mino is not required for piRNA biogenesis. The exact role of Mino in piRNA biogenesis is currently unknown.

A fourth piRNA biogenesis factor that localizes to mitochondria is Partner of PIWIs (PAPI)⁸². Describe the protein domains first. Studies in *Bombyx mori* indicated that BmPAPI is anchored to the outer mitochondrial membrane by an N-terminal transmembrane domain¹⁴⁵. Depletion of BmPAPI did not lead to a loss of piRNAs but rather to an increase in their length by 0.5-1nt on average. This predicted that Papi might be involved in the 3' end maturation of piRNAs, potentially by linking PIWI proteins to the trimmer exo-nuclease. Whether the *Drosophila* protein plays a similar function is currently unclear. Its depletion leads to only very mild defects in the ovarian piRNA pathway and might therefore indicate that also in *Drosophila*, the role of Papi is more in the fine-trimming of piRNA 3' ends.

In contrast to *Drosophila* and *Bombyx mori*, the mouse Papi ortholog TDRKH is indispensable for piRNA biogenesis. In TDRKH mutants piRNA populations are still loaded into PIWI proteins but they exhibit a massive increase in length¹⁴⁴. As the 5' end of the extended piRNAs do overlap with mature piRNAs, TDRKH was suggested to be involved in piRNA 3' end maturation via recruiting an unknown 3' to 5' exonuclease. Based on these observations the Tudor-domain containing PAPI and its homologs might be the factor connecting PIWI proteins and the trimming exonuclease generating piRNA 3'ends.

6.4.2 Reasons for mitochondrial sequestering of biogenesis factors

It is currently unclear, why the piRNA pathway has such intricate connections to the outer mitochondrial membrane. One possibility is the necessity to sequester potentially dangerous nucleases from the cytoplasm. It could be very hazardous for the cell to have an unspecific nuclease like Zuc free floating in the cytoplasm. Therefore anchoring of Zuc to the mitochondrial membrane could be a way to spatially restrict Zuc's enzymatic activity.

Alternatively, the anchoring of piRNA pathway associated nucleases could be important in providing substrate specificity by compartmentalization of different steps in piRNA biogenesis. In the *Drosophila* soma the Yb-body is the dominant structure in piRNA biogenesis besides the mitochondrial membrane. piRNA cluster transcripts have been shown to localize in close proximity to the Yb-body^{136,147} or even colocalize with it (my unpublished observations). Therefore, the Yb-body might serve as a platform for piRNA precursor definition, whereas the mitochondrial membrane serves as the platform for piRNA

processing. There is strong evidence, that Armi and Gasz mediate the connection between these compartments. It was shown, that Armi associates with empty Piwi and piRNA precursors in the Yb-body^{69,70}. In addition, my work showed that Gasz is required for the recruitment of Armi and Piwi to mitochondria, consistent with the hypothesis that the direct interaction of Armi and Gasz allows for the sequential order of piRNA biogenesis steps. No structure resembling the Yb-body exists in the germline. It is therefore unknown where the sorting of piRNA precursors and their association with Armi occurs. However, also in the germline Armi is recruited to mitochondria in a Gasz dependent manner. Therefore, it is likely that also in the germline Gasz provides the connection between the upstream precursor sorting and the maturation of piRNAs at mitochondria.

6.4.3 The piRNA pathway impacts the localization of mitochondria in the cell

The influence of the piRNA pathway on the cellular localization of mitochondria is another interesting aspect of the connection between the piRNA pathway and mitochondria. It has been shown that depletion of Zuc results in pronounced clustering of mitochondria around the nucleus in flies and mice^{69,73,79,80,148}. The reasons for this phenomenon are not known. Potentially mitochondria are glued together by long piRNA precursor RNAs that are recruited to multiple mitochondria simultaneously. As the nuclease Zuc is missing, this inevitably leads to mitochondrial clustering. Indeed it has been shown that the Zuc dependent clustering of mitochondria can be prevented by cleaving the most prominent piRNA precursor transcript (flamenco) using a mix of several siRNAs¹⁴⁷. If correct, this would predict that multiple Armi molecules can bind one piRNA precursor at the same time and that these are recruited to multiple mitochondria simultaneously. Furthermore, these data strongly suggest that no other nuclease is able to process piRNA precursor into piRNAs. This is consistent with the inability of identifying additional nucleases in a series of genome-wide screens aimed at finding novel components of the piRNA pathway^{47,133,149}.

Interestingly, I observed that mitochondria show a clear clustering around the nucleus in wildtype germline cells but that this clustering is absent in cells depleted for Gasz. This suggests that the piRNA pathway impacts mitochondrial positioning in wildtype cells, probably due to limited gluing of mitochondria by piRNA precursor transcripts that are in the process of being cleaved.

7 Outlook

The set of recently published genetic screens for piRNA pathway factors in the soma and germline of the *Drosophila* ovary uncovered many novel players that act in this fascinating

genome defense pathway^{47,133,149}. These screens will therefore serve as important entry points for the molecular and mechanistic dissection of the various steps that constitute the pathway. The next big challenge is to study the organization of the individual key-processes and their cell biological and molecular details.

To me, one of the most fascinating areas is related to the question how piRNA clusters are transcribed and how the cluster transcript are transported to the cytoplasm. In the germline the action of a special protein complex (the RDC) is required for piRNA cluster transcription⁵⁶. At which step the RDC impacts cluster transcription, however, is unclear. The only factor shown to participate in the export of germline piRNA cluster transcripts is the general RNA export factor UAP56⁵⁷. Under steady state conditions, UAP56 is a nuclear protein that co-localizes with the RDC at piRNA cluster loci. Interestingly, RDC foci and therefore also UAP56 foci are in close proximity to the nuclear membrane and typically juxtaposed of cytoplasmic accumulations of piRNA biogenesis factors such as the RNA helicase Vasa^{57,118,150-152}. The germline screen¹³³ offers several interesting entry points towards studying these processes. For example, several unknown factors appears to be specifically involved in the germline piRNA pathway and it seems likely that some of these are involved in cluster biology, similar to the RDC. In addition, the RNA export factors Nxf2 and 3 as well as members of the THO complex scored positively. Potentially, evolution has devised special export routes for piRNA cluster transcripts in order to efficiently funnel them into the piRNA biogenesis machinery.

In the soma our knowledge about piRNA cluster biology is even sparser. It is unclear which factors are necessary for the transcription of the >180kb long piRNA cluster *flamenco* that resides at the border between euchromatin and heterochromatin. Whether such a long transcript requires special factors for efficient transcription or special factors for efficient export is unclear, but several screen hits offer exciting entry points towards those ends.

I would like to conclude this thesis by emphasizing how important genetic screens are for the understanding of biological processes. While they are of course important for the discovery of core pathway members, I find it particularly interesting how screens uncover frequently the unexpected links. The piRNA pathway screens are certainly no exception here.

8 References

1. Werren, J. H., Nur, U. & Wu, C. I. Selfish genetic elements. *Trends Ecol Evol (Amst)* **3**, 297–302 (1988).
2. Werren, J. H. Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proc Natl Acad Sci USA* **108 Suppl 2**, 10863–10870 (2011).
3. Edgell, D. R. Selfish DNA: homing endonucleases find a home. *Current biology* :

- CB* **19**, R115–7 (2009).
4. Stoddard, B. L. Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. *Structure* **19**, 7–15 (2011).
5. Malik, H. S. & Henikoff, S. Major evolutionary transitions in centromere complexity. *Cell* **138**, 1067–1082 (2009).
6. Lyttle, T. W. Segregation distorters. *Annu Rev Genet* **25**, 511–557 (1991).
7. Kobayashi, I. Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res* **29**, 3742–3756 (2001).
8. Frank, A. C. & Wolfe, K. H. Evolutionary Capture of Viral and Plasmid DNA by Yeast Nuclear Chromosomes. *Eukaryotic Cell* **8**, 1521–1531 (2009).
9. Eberhard, W. G. Evolutionary consequences of intracellular organelle competition. *Q Rev Biol* **55**, 231–249 (1980).
10. Werren, J. H., Baldo, L. & Clark, M. E. Wolbachia: master manipulators of invertebrate biology. *Nat. Rev. Microbiol.* **6**, 741–751 (2008).
11. McCLINTOCK, B. The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci USA* **36**, 344–355 (1950).
12. Guo, B., Zou, M., Gan, X. & He, S. Genome size evolution in pufferfish: an insight from BAC clone-based *Diodon holocanthus* genome sequencing. *BMC Genomics* **11**, 396 (2010).
13. Deloger, M. *et al.* Identification of expressed transposable element insertions in the sequenced genome of *Drosophila melanogaster*. *Gene* **439**, 55–62 (2009).
14. Mouse Genome Sequencing Consortium *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
15. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
16. Schnable, P. S. *et al.* The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**, 1112–1115 (2009).
17. Finnegan, D. J. Transposable elements. *Curr Opin Genet Dev* **2**, 861–867 (1992).
18. Kazazian, H. H. Mobile elements: drivers of genome evolution. *Science* **303**, 1626–1632 (2004).
19. Eickbush, T. H. & Jamburuthugoda, V. K. The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Res.* **134**, 221–234 (2008).
20. Havecker, E. R., Gao, X. & Voytas, D. F. The diversity of LTR retrotransposons. *Genome Biol* (2004).
21. Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M. Mobile dna ii. (2002).
22. Malik, H. S., Henikoff, S. & Eickbush, T. H. Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. *Genome Res* **10**, 1307–1318 (2000).
23. Herédia, F., Loreto, E. L. S. & Valente, V. L. S. Complex evolution of gypsy in *Drosophilid* species. *Mol. Biol. Evol.* **21**, 1831–1842 (2004).
24. Pelisson, A., Mejlumian, L., Robert, V., Terzian, C. & Bucheton, A. *Drosophila* germline invasion by the endogenous retrovirus gypsy: involvement of the viral env gene. *Insect Biochem Mol Biol* **32**, 1249–1256 (2002).
25. Kim, A. *et al.* Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. *Proc Natl Acad Sci USA* **91**, 1285–1289 (1994).
26. Malik, H. S., Burke, W. D. & Eickbush, T. H. The age and evolution of non-LTR retrotransposable elements. *Mol. Biol. Evol.* **16**, 793–805 (1999).
27. Heidmann, T., Esnault, C. & Maestre, J. Human LINE retrotransposons generate processed pseudogenes. *Nat Genet* **24**, 363–367 (2000).
28. Ostertag, E. M. & Kazazian, H. H. Biology of mammalian L1 retrotransposons. *Annu Rev Genet* **35**, 501–538 (2001).
29. Boeke, J. D. & Stoye, J. P. *Retrotransposons, Endogenous Retroviruses, and the*

- Evolution of Retroelements*. (Cold Spring Harbor Laboratory Press, 1997).
30. Goodwin, T. J. D. & Poulter, R. T. M. A new group of tyrosine recombinase-encoding retrotransposons. *Mol. Biol. Evol.* **21**, 746–759 (2004).
 31. Poulter, R. T. M. & Goodwin, T. J. D. DIRS-1 and the other tyrosine recombinase retrotransposons. *Cytogenet Genome Res* **110**, 575–588 (2005).
 32. Szitenberg, A., Koutsovoulos, G., Blaxter, M. L. & Lunt, D. H. The evolution of tyrosine-recombinase elements in nematoda. *PLoS ONE* **9**, e106630 (2014).
 33. Arkhipova, I. R. Distribution and phylogeny of Penelope-like elements in eukaryotes. *Syst. Biol.* **55**, 875–885 (2006).
 34. Evgen'ev, M. B. & Arkhipova, I. R. Penelope-like elements--a new class of retroelements: distribution, function and possible evolutionary significance. *Cytogenet Genome Res* **110**, 510–521 (2005).
 35. Evgen'ev, M. B. *et al.* Penelope, a new family of transposable elements and its possible role in hybrid dysgenesis in *Drosophila virilis*. *Proc Natl Acad Sci USA* **94**, 196–201 (1997).
 36. Feschotte, C. & Pritham, E. J. DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet* **41**, 331–368 (2007).
 37. Testori, A. *et al.* The role of Transposable Elements in shaping the combinatorial interaction of Transcription Factors. *BMC Genomics* **13**, 400 (2012).
 38. Volff, J.-N. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. *Bioessays* **28**, 913–922 (2006).
 39. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
 40. Ambros, V. The functions of animal microRNAs. *Nature* **431**, 350–355 (2004).
 41. Ghildiyal, M. & Zamore, P. D. Small silencing RNAs: an expanding universe. *Nat Rev Genet* **10**, 94–108 (2009).
 42. Tolia, N. H. & Joshua-Tor, L. Slicer and the argonautes. *Nat Chem Biol* **3**, 36–43 (2007).
 43. Tomari, Y. & Zamore, P. D. Perspective: machines for RNAi. *Genes Dev* **19**, 517–529 (2005).
 44. Okamura, K. & Lai, E. C. Endogenous small interfering RNAs in animals. *Nat Rev Mol Cell Biol* **9**, 673–678 (2008).
 45. Tuschl, T., Zamore, P. D., Lehmann, R., Bartel, D. P. & Sharp, P. A. Targeted mRNA degradation by double-stranded RNA in vitro. *Genes Dev* **13**, 3191–3197 (1999).
 46. Williams, R. W. & Rubin, G. M. ARGONAUTE1 is required for efficient RNA interference in *Drosophila* embryos. *Proc Natl Acad Sci USA* **99**, 6889–6894 (2002).
 47. Handler, D. *et al.* The Genetic Makeup of the *Drosophila* piRNA Pathway. *Mol Cell* (2013). doi:10.1016/j.molcel.2013.04.031
 48. Bate, M. & Martínez Arias, A. The Development of *Drosophila melanogaster* Volume I. *Cold Spring Harbor Laboratory Press* **1**, 2–70 (1993).
 49. Huynh, J.-R. & St Johnston, D. The origin of asymmetry: early polarisation of the *Drosophila* germline cyst and oocyte. *Current biology : CB* **14**, R438–49 (2004).
 50. Mahowald, A. P. & Strassheim, J. M. Intercellular migration of centrioles in the germarium of *Drosophila melanogaster*. An electron microscopic study. *J Cell Biol* **45**, 306–320 (1970).
 51. Warn, R. M., Gutzeit, H. O., Smith, L. & Warn, A. F-actin rings are associated with the ring canals of the *Drosophila* egg chamber. *Exp. Cell Res.* **157**, 355–363 (1985).
 52. Bastock, R. & St Johnston, D. *Drosophila* oogenesis. *Current biology : CB* **18**, R1082–7 (2008).
 53. Edgar, B. A. & Orr-Weaver, T. L. Endoreplication cell cycles: more for less. *Cell* **105**, 297–306 (2001).
 54. Handler, D. *et al.* A systematic analysis of *Drosophila* TUDOR domain-containing

- proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. *The EMBO Journal* (2011). doi:10.1038/emboj.2011.308
55. Brennecke, J. *et al.* Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in *Drosophila*. *Cell* **128**, 1089–1103 (2007).
 56. Mohn, F., Sienski, G., Handler, D. & Brennecke, J. The Rhino-Deadlock-Cutoff Complex Licenses Noncanonical Transcription of Dual-Strand piRNA Clusters in *Drosophila*. *Cell* **157**, 1364–1379 (2014).
 57. Zhang, F. *et al.* UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. *Cell* **151**, 871–884 (2012).
 58. Gunawardane, L. S. *et al.* A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* **315**, 1587–1590 (2007).
 59. Sienski, G., Dönertas, D. & Brennecke, J. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* **151**, 964–980 (2012).
 60. Niki, Y., Yamaguchi, T. & Mahowald, A. P. Establishment of stable cell lines of *Drosophila* germ-line stem cells. *Proc Natl Acad Sci USA* **103**, 16325–16330 (2006).
 61. Saito, K. *et al.* A regulatory circuit for piwi by the large Maf gene traffic jam in *Drosophila*. *Nature* **461**, 1296–1299 (2009).
 62. Robine, N. *et al.* A broadly conserved pathway generates 3'UTR-directed primary piRNAs. *Current biology : CB* **19**, 2066–2076 (2009).
 63. Pelisson, A. *et al.* Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the *Drosophila* flamenco gene. *The EMBO Journal* **13**, 4401–4411 (1994).
 64. Bucheton, A. The relationship between the flamenco gene and gypsy in *Drosophila*: how to tame a retrovirus. *Trends Genet* **11**, 349–353 (1995).
 65. Robert, V., Prud'homme, N., Kim, A., Bucheton, A. & Pelisson, A. Characterization of the flamenco region of the *Drosophila melanogaster* genome. *Genetics* **158**, 701–713 (2001).
 66. Prud'homme, N., Gans, M., Masson, M., Terzian, C. & Bucheton, A. Flamenco, a gene controlling the gypsy retrovirus of *Drosophila melanogaster*. *Genetics* **139**, 697–711 (1995).
 67. Mével-Ninio, M., Péliesson, A., Kinder, J., Campos, A. R. & Bucheton, A. The flamenco locus controls the gypsy and ZAM retroviruses and is required for *Drosophila* oogenesis. *Genetics* **175**, 1615–1624 (2007).
 68. Malone, C. D. *et al.* Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell* **137**, 522–535 (2009).
 69. Olivieri, D., Sykora, M. M., Sachidanandam, R., Mechtler, K. & Brennecke, J. An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in *Drosophila*. *The EMBO Journal* **29**, 3301–3317 (2010).
 70. Saito, K. *et al.* Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in *Drosophila*. *Genes Dev* **24**, 2493–2498 (2010).
 71. Szakmary, A., Reedy, M., Qi, H. & Lin, H. The Yb protein defines a novel organelle and regulates male germline stem cell self-renewal in *Drosophila melanogaster*. *J Cell Biol* **185**, 613–627 (2009).
 72. Zamparini, A. L. *et al.* Vreteno, a gonad-specific protein, is essential for germline development and primary piRNA biogenesis in *Drosophila*. *Development* (2011). doi:10.1242/dev.069187
 73. Olivieri, D., Senti, K.-A., Subramanian, S., Sachidanandam, R. & Brennecke, J. The cochaperone shutdown defines a group of biogenesis factors essential for all piRNA populations in *Drosophila*. *Mol Cell* **47**, 954–969 (2012).
 74. Preall, J. B., Czech, B., Guzzardo, P. M., Muerdter, F. & Hannon, G. J. shutdown is a component of the *Drosophila* piRNA biogenesis machinery. *RNA* **18**, 1446–1457 (2012).

75. Aravin, A. A. & Chan, D. C. piRNAs meet mitochondria. *Dev Cell* **20**, 287–288 (2011).
76. Pane, A., Wehr, K. & Schüpbach, T. zucchini and squash encode two putative nucleases required for rasiRNA production in the *Drosophila* germline. *Dev Cell* **12**, 851–862 (2007).
77. Nishimasu, H. *et al.* Structure and function of Zucchini endoribonuclease in piRNA biogenesis. *Nature* **491**, 284–287 (2012).
78. Ipsaro, J. J., Haase, A. D., Knott, S. R., Joshua-Tor, L. & Hannon, G. J. The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. *Nature* **491**, 279–283 (2012).
79. Watanabe, T. *et al.* MITOPLD is a mitochondrial protein essential for nuage formation and piRNA biogenesis in the mouse germline. *Dev Cell* **20**, 364–375 (2011).
80. Huang, H. *et al.* piRNA-associated germline nuage formation and spermatogenesis require MitoPLD profusogenic mitochondrial-surface lipid signaling. *Dev Cell* **20**, 376–387 (2011).
81. Vagin, V. V. *et al.* Minotaur is critical for primary piRNA biogenesis. *RNA* **19**, 1064–1077 (2013).
82. Liu, L., Qi, H., Wang, J. & Lin, H. PAPI, a novel TUDOR-domain protein, complexes with AGO3, ME31B and TRAL in the nuage to silence transposition. *Development* **138**, 1863–1873 (2011).
83. Kawaoka, S., Izumi, N., Katsuma, S. & Tomari, Y. 3' end formation of PIWI-interacting RNAs in vitro. *Mol Cell* **43**, 1015–1022 (2011).
84. Vagin, V. V. *et al.* A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**, 320–324 (2006).
85. Horwich, M. D. *et al.* The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Current biology : CB* **17**, 1265–1272 (2007).
86. Saito, K. *et al.* Pimet, the *Drosophila* homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. *Genes Dev* **21**, 1603–1608 (2007).
87. Ma, J.-B. *et al.* Structural basis for 5'-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. *Nature* **434**, 666–670 (2005).
88. Frank, F., Sonenberg, N. & Nagar, B. Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* **465**, 818–822 (2010).
89. Cox, D. N., Chao, A. & Lin, H. piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development* **127**, 503–514 (2000).
90. Klenov, M. S. *et al.* Separation of stem cell maintenance and transposon silencing functions of Piwi protein. *Proc Natl Acad Sci USA* **108**, 18760–18765 (2011).
91. Ni, J.-Q. *et al.* A genome-scale shRNA resource for transgenic RNAi in *Drosophila*. *Nat Methods* **8**, 405–407 (2011).
92. Sarot, E., Payen-Groschêne, G., Bucheton, A. & Péliçon, A. Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the *Drosophila melanogaster* flamenco gene. *Genetics* **166**, 1313–1321 (2004).
93. Dietzl, G. *et al.* A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature* **448**, 151–156 (2007).
94. Le Hir, H., Izaurralde, E., Maquat, L. E. & Moore, M. J. The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon-exon junctions. *The EMBO Journal* **19**, 6860–6869 (2000).
95. Shibuya, T., Tange, T. Ø., Sonenberg, N. & Moore, M. J. eIF4AIII binds spliced mRNA in the exon junction complex and is essential for nonsense-mediated

- decay. *Nat Struct Mol Biol* **11**, 346–351 (2004).
96. Gehring, N. H., Lamprinaki, S., Hentze, M. W. & Kulozik, A. E. The hierarchy of exon-junction complex assembly by the spliceosome explains key features of mammalian nonsense-mediated mRNA decay. *PLoS Biol.* **7**, e1000120 (2009).
97. Degot, S. *et al.* Association of the breast cancer protein MLN51 with the exon junction complex via its speckle localizer and RNA binding module. *J Biol Chem* **279**, 33702–33715 (2004).
98. Tange, T. Ø., Nott, A. & Moore, M. J. The ever-increasing complexities of the exon junction complex. *Curr. Opin. Cell Biol.* **16**, 279–284 (2004).
99. Singh, G. *et al.* The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. *Cell* **151**, 750–764 (2012).
100. Bono, F. & Gehring, N. H. Assembly, disassembly and recycling: the dynamics of exon junction complexes. *RNA Biol* **8**, 24–29 (2011).
101. Gehring, N. H., Neu-Yilik, G., Schell, T., Hentze, M. W. & Kulozik, A. E. Y14 and hUpf3b form an NMD-activating complex. *Mol Cell* **11**, 939–949 (2003).
102. Kataoka, N. *et al.* Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm. *Mol Cell* **6**, 673–682 (2000).
103. Luo, M. L. *et al.* Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly. *Nature* **413**, 644–647 (2001).
104. Schwerk, C. *et al.* ASAP, a novel protein complex involved in RNA processing and apoptosis. *Molecular and Cellular Biology* **23**, 2981–2990 (2003).
105. Mayeda, A. *et al.* Purification and characterization of human RNPS1: a general activator of pre-mRNA splicing. *The EMBO Journal* **18**, 4560–4570 (1999).
106. Sakashita, E., Tatsumi, S., Werner, D., Endo, H. & Mayeda, A. Human RNPS1 and its associated factors: a versatile alternative pre-mRNA splicing regulator in vivo. *Molecular and Cellular Biology* **24**, 1174–1187 (2004).
107. Zhang, Y., Iratni, R., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. Histone deacetylases and SAP18, a novel polypeptide, are components of a human Sin3 complex. *Cell* **89**, 357–364 (1997).
108. Le Hir, H., Moore, M. J. & Maquat, L. E. Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions. *Genes Dev* **14**, 1098–1108 (2000).
109. Li, C., Lin, R.-I., Lai, M.-C., Ouyang, P. & Tarn, W.-Y. Nuclear Pnn/DRS protein binds to spliced mRNPs and participates in mRNA processing and export via interaction with RNPS1. *Molecular and Cellular Biology* **23**, 7363–7376 (2003).
110. Roignant, J.-Y. & Treisman, J. E. Exon junction complex subunits are required to splice *Drosophila* MAP kinase, a large heterochromatic gene. *Cell* **143**, 238–250 (2010).
111. Ashton-Beaucage, D., Udell, C., Lavoie, H. & Baril, C. ScienceDirect - Cell : The Exon Junction Complex Controls the Splicing of mapk and Other Long Intron-Containing Transcripts in *Drosophila*. *Cell* (2010).
112. Neumüller, R. A. *et al.* Genome-wide analysis of self-renewal in *Drosophila* neural stem cells by transgenic RNAi. *Cell Stem Cell* **8**, 580–593 (2011).
113. Mummery-Widmer, J. L. *et al.* Genome-wide analysis of Notch signalling in *Drosophila* by transgenic RNAi. *Nature* **458**, 987–992 (2009).
114. Neely, G. G. *et al.* A global in vivo *Drosophila* RNAi screen identifies NOT3 as a conserved regulator of heart function. *Cell* **141**, 142–153 (2010).
115. Pospisilik, J. A. *et al.* *Drosophila* genome-wide obesity screen reveals hedgehog as a determinant of brown versus white adipose cell fate. *Cell* **140**, 148–160 (2010).
116. Wang, S. H. & Elgin, S. C. R. *Drosophila* Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. *Proc Natl Acad Sci USA* **108**, 21164–21169 (2011).
117. Hayashi, R., Handler, D., Ish-Horowicz, D. & Brennecke, J. The exon junction

- complex is required for definition and excision of neighboring introns in *Drosophila*. *Genes Dev* (2014). doi:10.1101/gad.245738.114
118. Malone, C. D. & Hannon, G. J. Molecular evolution of piRNA and transposon control pathways in *Drosophila*. *Cold Spring Harb Symp Quant Biol* **74**, 225–234 (2009).
 119. Le Thomas, A. *et al.* Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. *Genes Dev* **28**, 1667–1680 (2014).
 120. Zhang, Z. *et al.* The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. *Cell* **157**, 1353–1363 (2014).
 121. Aravin, A. A. *et al.* Cytoplasmic compartmentalization of the fetal piRNA pathway in mice. *PLoS Genetics* **5**, e1000764 (2009).
 122. Wade, P. A. *et al.* A novel collection of accessory factors associated with yeast RNA polymerase II. *Protein Expr. Purif.* **8**, 85–90 (1996).
 123. Mueller, C. L. & Jaehning, J. A. Ctr9, Rtf1, and Leo1 are components of the Paf1/RNA polymerase II complex. *Molecular and Cellular Biology* **22**, 1971–1980 (2002).
 124. Squazzo, S. L. *et al.* The Paf1 complex physically and functionally associates with transcription elongation factors in vivo. *The EMBO Journal* **21**, 1764–1774 (2002).
 125. Krogan, N. J. *et al.* RNA polymerase II elongation factors of *Saccharomyces cerevisiae*: a targeted proteomics approach. *Molecular and Cellular Biology* **22**, 6979–6992 (2002).
 126. Zhu, B. *et al.* The human PAF complex coordinates transcription with events downstream of RNA synthesis. *Genes Dev* **19**, 1668–1673 (2005).
 127. Jaehning, J. A. The Paf1 complex: platform or player in RNA polymerase II transcription? *Biochim Biophys Acta* **1799**, 379–388 (2010).
 128. Adelman, K. *et al.* *Drosophila* Paf1 modulates chromatin structure at actively transcribed genes. *Molecular and Cellular Biology* **26**, 250–260 (2006).
 129. Guruharsha, K. G. *et al.* A protein complex network of *Drosophila melanogaster*. *Cell* **147**, 690–703 (2011).
 130. Schweikhard, V. *et al.* Transcription factors TFIIF and TFIIS promote transcript elongation by RNA polymerase II by synergistic and independent mechanisms. *Proc Natl Acad Sci USA* **111**, 6642–6647 (2014).
 131. Kim, J., Guermah, M. & Roeder, R. G. The human PAF1 complex acts in chromatin transcription elongation both independently and cooperatively with SII/TFIIS. *Cell* **140**, 491–503 (2010).
 132. Izaurrealde, E. A novel family of nuclear transport receptors mediates the export of messenger RNA to the cytoplasm. *Eur. J. Cell Biol.* **81**, 577–584 (2002).
 133. Czech, B., Preall, J. B., McGinn, J. & Hannon, G. J. A Transcriptome-wide RNAi Screen in the *Drosophila* Ovary Reveals Factors of the Germline piRNA Pathway. *Mol Cell* 1–13 (2013). doi:10.1016/j.molcel.2013.04.007
 134. Strambio-De-Castillia, C. C., Niepel, M. M. & Rout, M. P. M. The nuclear pore complex: bridging nuclear transport and gene regulation. *Nat Rev Mol Cell Biol* **11**, 490–501 (2010).
 135. FlyBase. FlyBase-actioned changes to gene nomenclature. (2009).
 136. Dennis, C. *et al.* "Dot COM", a Nuclear Transit Center for the Primary piRNA Pathway in *Drosophila*. *PLoS ONE* **8**, (2013).
 137. Dönertas, D., Sienski, G. & Brennecke, J. *Drosophila* Gtsf1 is an essential component of the Piwi-mediated transcriptional silencing complex. *Genes Dev* **27**, 1693–1705 (2013).
 138. Clough, E., Moon, W., Wang, S., Smith, K. & Hazelrigg, T. Histone methylation is required for oogenesis in *Drosophila*. **134**, 157–165 (2007).
 139. Seum, C., Bontron, S., Reo, E., Delattre, M. & Spierer, P. *Drosophila* G9a is a

- nonessential gene. *Genetics* **177**, 1955–1957 (2007).
140. Koch, C. M., Honemann-Capito, M., Egger-Adam, D. & Wodarz, A. Windei, the *Drosophila* homolog of mAM/MCAF1, is an essential cofactor of the H3K9 methyl transferase dSETDB1/Eggless in germ line development. *PLoS Genetics* **5**, e1000644 (2009).
 141. Swaminathan, J., Baxter, E. M. & Corces, V. G. The role of histone H2Av variant replacement and histone H4 acetylation in the establishment of *Drosophila* heterochromatin. *Genes Dev* **19**, 65–76 (2005).
 142. Baldi, S. & Becker, P. B. The variant histone H2A.V of *Drosophila*--three roles, two guises. *Chromosoma* **122**, 245–258 (2013).
 143. Malone, C. D. *et al.* The exon junction complex controls transposable element activity by ensuring faithful splicing of the piwi transcript. *Genes Dev* **28**, 1786–1799 (2014).
 144. Saxe, J. P., Chen, M., Zhao, H. & Lin, H. Tdrkh is essential for spermatogenesis and participates in primary piRNA biogenesis in the germline. *The EMBO Journal* (2013). doi:10.1038/emboj.2013.121
 145. Honda, S. *et al.* Mitochondrial protein BmPAPI modulates the length of mature piRNAs. *RNA* **19**, 1405–1418 (2013).
 146. Choi, S.-Y. *et al.* A common lipid links Mfn-mediated mitochondrial fusion and SNARE-regulated exocytosis. *Nature Cell Biology* **8**, 1255–1262 (2006).
 147. Murota, Y. *et al.* Yb Integrates piRNA Intermediates and Processing Factors into Perinuclear Bodies to Enhance piRISC Assembly. *Cell Rep* (2014). doi:10.1016/j.celrep.2014.05.043
 148. Saito, K. & Siomi, M. C. Small RNA-mediated quiescence of transposable elements in animals. *Dev Cell* **19**, 687–697 (2010).
 149. Muerdter, F. *et al.* A Genome-wide RNAi Screen Draws a Genetic Framework for Transposon Control and Primary piRNA Biogenesis in *Drosophila*. *Mol Cell* **1–13** (2013). doi:10.1016/j.molcel.2013.04.006
 150. Hay, B., Jan, L. Y. & Jan, Y. N. Localization of vasa, a component of *Drosophila* polar granules, in maternal-effect mutants that alter embryonic anteroposterior polarity. *Development* **109**, 425–433 (1990).
 151. Lasko, P. F. & Ashburner, M. Posterior localization of vasa protein correlates with, but is not sufficient for, pole cell development. *Genes Dev* **4**, 905–921 (1990).
 152. Liang, L., Diehl-Jones, W. & Lasko, P. Localization of vasa protein to the *Drosophila* pole plasm is independent of its RNA-binding and helicase activities. *Development* **120**, 1201–1211 (1994).
 153. Nishida, K. M. *et al.* Functional involvement of Tudor and dPRMT5 in the piRNA processing pathway in *Drosophila* germlines. *The EMBO Journal* **28**, 3820–3831 (2009).
 154. Liu, H. *et al.* Structural basis for methylarginine-dependent recognition of Aubergine by Tudor. *Genes Dev* (2010). doi:10.1101/gad.1956010
 155. Kirino, Y. *et al.* Arginine methylation of Aubergine mediates Tudor binding and germ plasm localization. *RNA* **16**, 70–78 (2010).
 156. Kirino, Y. *et al.* Arginine methylation of Piwi proteins catalysed by dPRMT5 is required for Ago3 and Aub stability. *Nature Cell Biology* **11**, 652–658 (2009).
 157. Lim, A. K. & Kai, T. Unique germ-line organelle, nuage, functions to repress selfish genetic elements in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* **104**, 6714–6719 (2007).
 158. Patil, V. S. & Kai, T. Repression of Retroelements in *Drosophila* Germline via piRNA Pathway by the Tudor Domain Protein Tejas. *Current biology : CB* (2010). doi:10.1016/j.cub.2010.02.046
 159. Ni, J.-Q. *et al.* A *Drosophila* resource of transgenic RNAi lines for neurogenetics. *Genetics* **182**, 1089–1100 (2009).

9 Appendix

9.1 A systematic analysis of *Drosophila* TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors.

It has been shown that several proteins involved in the piRNA pathway, foremost Aub, AGO3 and Vasa are symmetrically di-methylated on Arginine residues¹⁵³⁻¹⁵⁶. This modification is a known binding motif for Tudor-domain containing proteins and several of these have been associated with functions in the piRNA pathway (Spindle-E, Krimper, Tejas and Tudor)^{68,84,153,155-158}.

Therefore I conducted a systematic analysis and tested more than 30 Tudor domain containing proteins for their function in the germline and somatic piRNA pathway. For this mini-screen I utilized a novel transgenic RNAi system that is based on the expression of short hairpin RNAs¹⁵⁹. In this study I could show that four previously not reported Tudor-domain containing proteins are implicated in the piRNA pathway. I characterized one of these factors, CG4771/Vreteno, in more detail and showed that it is required for primary piRNA biogenesis.

A systematic analysis of *Drosophila* TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors

Dominik Handler¹, Daniel Olivieri¹,
Maria Novatchkova^{1,2},
Franz Sebastian Gruber¹, Katharina
Meixner¹, Karl Mechtler^{1,2}, Alexander
Stark², Ravi Sachidanandam³
and Julius Brennecke^{1,*}

¹Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), Vienna, Austria, ²Institute of Molecular Pathology (IMP), Vienna, Austria and ³Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, NY, USA

PIWI proteins and their bound PIWI-interacting RNAs (piRNAs) form the core of a gonad-specific small RNA silencing pathway that protects the animal genome against the deleterious activity of transposable elements. Recent studies linked the piRNA pathway to TUDOR biology as TUDOR domains of various proteins bind symmetrically methylated Arginine residues in PIWI proteins. We systematically analysed the *Drosophila* TUDOR protein family and identified four previously not characterized TUDOR domain-containing proteins (CG4771, CG14303, CG11133 and CG31755) as essential piRNA pathway factors. We characterized CG4771 (Vreteno) in detail and demonstrate a critical role for this protein in primary piRNA biogenesis. Vreteno physically and/or genetically interacts with the primary pathway components Piwi, Armitage, Yb and Zucchini. Vreteno also interacts with the Tdrd12 orthologues CG11133 (Brother of Yb) and CG31755 (Sister of Yb), which are essential for the primary piRNA pathway in the germline and probably replace the function of the related but soma-specific factor Yb.

The EMBO Journal advance online publication, 23 August 2011; doi:10.1038/emboj.2011.308

Subject Categories: RNA

Keywords: *Drosophila*; piRNAs; Piwi; transposons; Tudor

Introduction

The PIWI-interacting RNA (piRNA) pathway is an animal-specific small RNA pathway that silences selfish genetic elements such as transposons in gonads (Malone and Hannon, 2009; Khurana and Theurkauf, 2010; Senti and Brennecke, 2010). At the core of this pathway act Argonaute

proteins from the PIWI clade and their bound small RNAs, generally referred to as piRNAs. Mutations in PIWI proteins or in factors involved in piRNA biogenesis or piRNA-mediated silencing lead to de-silencing of transposons, to widespread DNA damage and ultimately result in sterility.

The analyses of piRNA populations from vertebrates and invertebrates have provided genuine insight into the genomic origin of piRNAs (Aravin *et al*, 2006, 2007, 2008; Girard *et al*, 2006; Lau *et al*, 2006; Vagin *et al*, 2006; Brennecke *et al*, 2007; Li *et al*, 2009; Malone *et al*, 2009; Robine *et al*, 2009; Saito *et al*, 2009). The three major piRNA sources are long RNAs originating from discrete genomic loci typically enriched in transposon sequences (piRNA clusters), transcripts from active transposons and finally mRNAs from numerous endogenous genes.

The genetic and mechanistic principles of piRNA biogenesis are only poorly understood but sequence analyses of piRNA populations indicated that two modes of piRNA biogenesis exist (reviewed in Senti and Brennecke, 2010). On the one hand, during primary piRNA biogenesis presumably single-stranded precursor transcripts are processed in a seemingly random manner into 23–30 nt primary piRNAs (Lau *et al*, 2009; Li *et al*, 2009; Malone *et al*, 2009; Saito *et al*, 2009). On the other hand, transposon sense transcripts (typically from active elements) and antisense transcripts (typically from piRNA clusters) participate in the process of secondary piRNA biogenesis: Here, piRNA-mediated cleavage of the target transcript triggers the production of a novel piRNA with the reciprocal polarity (Brennecke *et al*, 2007; Gunawardane *et al*, 2007). Hallmarks of this so-called ping-pong amplification of piRNAs are conserved from sponges to mammals (Aravin *et al*, 2007; Grimson *et al*, 2008).

The existence of two distinct piRNA biogenesis branches is particularly evident in the *Drosophila* ovary. Within ovarian germ cells, the three PIWI proteins Piwi, Aubergine and Argonaute 3 (Ago3) are co-expressed and piRNAs are generated via the primary and secondary pathways. The two major players of the secondary ping-pong pathway are Aubergine and Ago3 with Aubergine binding primarily cluster derived antisense piRNAs, while Ago3 is primarily complexed with transposon mRNA-derived sense piRNAs (Brennecke *et al*, 2007; Gunawardane *et al*, 2007; Li *et al*, 2009; Malone *et al*, 2009). In contrast, the surrounding follicle cells (somatic origin) express exclusively Piwi and piRNAs are produced only via the primary pathway (Lau *et al*, 2009; Li *et al*, 2009; Malone *et al*, 2009; Saito *et al*, 2009).

As all three PIWI proteins are expressed in germline cells, accurate systems must be in place to guarantee controlled piRNA biogenesis and PIWI loading. Several recent studies indicate that modular interactions between PIWI proteins and TUDOR domain-containing proteins are part of this control system (Chen *et al*, 2009; Kirino *et al*, 2009, 2010; Nishida

*Corresponding author. Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), Dr Bohrgasse 3, Vienna 1030, Austria. Tel.: +43 179 044 4508; Fax: +43 179 044 110; E-mail: julius.brennecke@imba.oeaw.ac.at

Received: 24 June 2011; accepted: 3 August 2011

et al, 2009; Reuter *et al*, 2009; Vagin *et al*, 2009). The TUDOR domain is a member of the TUDOR 'royal family', which among others also contains Chromo, plant Agenet, MBT and PWWP domains (Maurer-Stroh *et al*, 2003). The core TUDOR domain spans ~60 amino acids and folds into a strongly bent anti-parallel β -sheet with five strands forming a barrel-like fold (Sprangers *et al*, 2003; Chen *et al*, 2009; Friberg *et al*, 2009; Liu *et al*, 2010a, b). A key function of this domain is to facilitate protein-protein interactions, which often depend on the post-translational methylation of Lysine or Arginine residues in target proteins. Indeed, several methylated Arginine residues have been identified in PIWI-family proteins and at least in some cases specific interactions between PIWI and TUDOR proteins require the symmetric di-methylation of Arginine residues (sDMAs) in PIWI proteins (Kirino *et al*, 2009, 2010; Nishida *et al*, 2009; Reuter *et al*, 2009; Vagin *et al*, 2009; Huang *et al*, 2011b). Based on the observed specificity of PIWI-TUDOR interactions, it is possible that an intricate sDMA code allows the controlled recruitment of selected TUDOR domain-containing proteins at specific points of the life cycle of PIWI-piRNA complexes.

In *Drosophila*, six (Tudor, Spindle-E, Krimper, Tejas, Yb and Papi) out of the roughly 20 proteins implicated in the piRNA pathway contain TUDOR domains (Boswell and Mahowald, 1985; Gillespie and Berg, 1995; Lim and Kai, 2007; Malone *et al*, 2009; Nishida *et al*, 2009; Olivieri *et al*, 2010; Patil and Kai, 2010; Qi *et al*, 2010; Saito *et al*, 2010; Liu *et al*, 2011). We therefore decided to systematically analyse all fly TUDOR domain-containing proteins for an involvement in the piRNA pathway. This led to the identification of four novel TUDOR proteins as essential piRNA pathway factors. We characterized in detail the role of CG4771 (Vreteno), a tandem TUDOR domain-containing protein. Vreteno localizes to Yb bodies in follicle cells and to nuage in germline cells and is required for primary piRNA biogenesis in both cell types. Vreteno interacts with the three fly Tdrd12 proteins Yb, CG11133 (Brother of Yb) and CG31755 (Sister of Yb), which have partially overlapping functions in the somatic and germline piRNA pathways.

Results

Identification and classification of TUDOR domain-containing proteins in *Drosophila*

We mined the *Drosophila melanogaster* proteome for TUDOR-clan domains (Pfam CL0049) using sensitive sequence-profile (HMMer) and profile-profile comparison methods (Soding *et al*, 2005). Supplementary Table S1 lists all identi-

fied proteins and specifies the individual subclasses (see also Figure 1A). For further analysis we focused on the TUDOR-clan domains TUDOR and SMN, which both have been reported to bind sDMA residues (Selenko *et al*, 2001; Sprangers *et al*, 2003; Cote and Richard, 2005; Liu *et al*, 2010a, b). This resulted in 22 proteins containing at least one TUDOR/SMN domain.

An alignment of all TUDOR/SMN domains contained in this set indicates three subgroups (Supplementary Figure S1). Groups A (Smn, CG13472 and CG17454) and B (Otu, CG3251) show similarity only to the ~60 amino-acid TUDOR core. All other sequences cluster together in group C and share significant similarity also N- and C-terminal to the TUDOR core. Characteristic for group C are two 100% conserved amino acids, an Arginine in β 4 and an Aspartate in the loop linking β 5 and β 6 of the extended TUDOR structure (Supplementary Figure S1; marked in green in Figure 1B; Liu *et al*, 2010a). Based on structural studies, group C sequences represent extended TUDOR domains, which are characterized by a core TUDOR domain tightly interacting with an OB-fold that consists of the N-terminal and C-terminal extensions (Liu *et al*, 2010a, b). So far, every TUDOR domain-containing protein that has been linked to the piRNA pathway belongs to the extended TUDOR group.

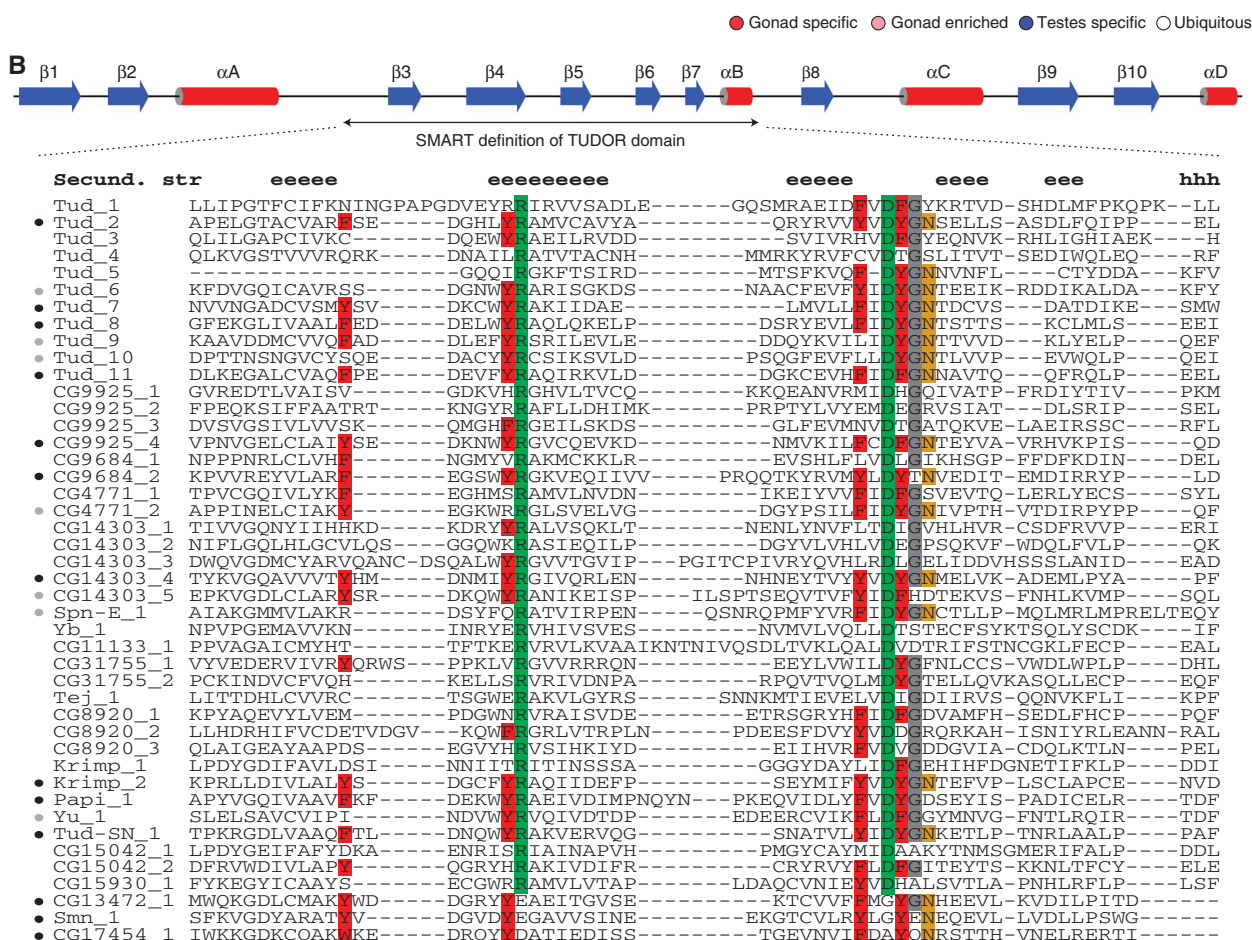
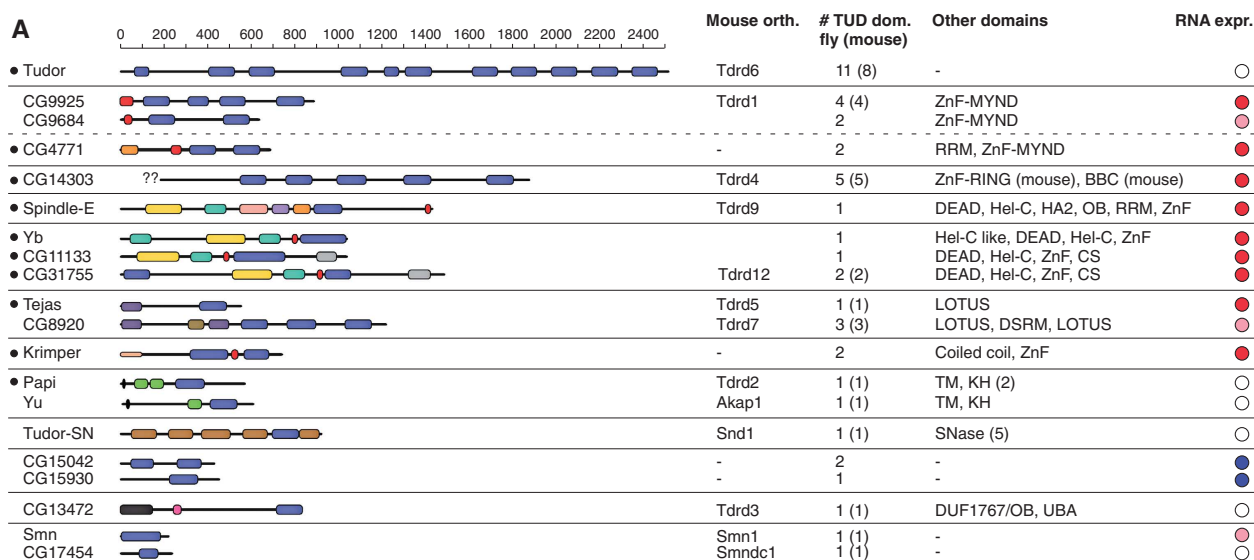
To further characterize the set of proteins harbouring extended TUDOR domains, we annotated all additionally contained protein domains and searched for the corresponding mouse orthologues (Figure 1A; Supplementary Figure S2). Most of the fly proteins exhibit strong similarity to their mouse counterparts and the listed pairs in Supplementary Figure S2 are supported by multiple independent orthology assignment methods. CG14303 was linked to Rnf17 based on automated orthology identification (Inparanoid, Compara, OMA) and similarities in the TUDOR domains and an N-terminal B-Box C-terminal domain. We note that the N-terminus of CG14303 is not annotated in FlyBase (lack of EST data), indicating that the similarities between CG14303 and Rnf17 might also include the RING-type zinc finger found in Rnf17. Notably, an N-terminal RING finger could be identified in the *Apis* and *Bombyx* CG14303/Rnf17 orthologues. Murine Tdrd12 was assigned to CG11133 and CG31755 based on OrthoMCL (v2:OG2_82474 and v4:OG4_21213). Since both of these proteins share a similar domain composition with the piRNA pathway protein Yb, it appears that Tdrd12 has radiated in *Drosophila* into three proteins. Indeed, all three fly proteins are more related to each other than to the single mouse or human Tdrd12 proteins. Less obvious was the assignment of Tdrd1, a 4 \times TUDOR domain protein with an

Figure 1 Characterization of the *Drosophila* TUDOR proteins. (A) Cartoon showing all *Drosophila melanogaster* proteins containing TUDOR/SMN domains (blue boxes). All other significant protein domains identified via HHpred searches are indicated with coloured boxes and their identity is given to the right from N to C (ZnF: zinc finger; RRM: RNA recognition motif; BBC: B-Box C-terminal domain; DEAD: DEAD-Box RNA Helicase; Hel-C: Helicase C-terminal; HA2: Helicase associated domain; OB: oligo-nucleotide binding; CS: HSP20-like domain; DSRM: double-stranded RNA binding; TM: trans-membrane domain; KH: K homology; SNase: Staphylococcus nuclease; DUF: domain of unknown function; UBA: ubiquitin-associated domain). TUDOR proteins implicated in the piRNA pathway (including the ones from this study) marked with a black dot (left). The scale indicates amino-acid positions. The identified mouse orthologues (see Supplementary Figure S1), the number of identified TUDOR domains in fly (mouse) and the expression bias towards gonads in adult flies are shown to the right. Proteins with similar domain composition are grouped together. For CG14303, the '?' indicate the non-annotated N-terminus. (B) The secondary structure cartoon (blue indicates β -strands, red α -helices) denotes the extended TUDOR domain and is based on Liu *et al* (2010a) (see also Supplementary Figure S1). The core TUDOR domain (SMART definition) is shown as an alignment for all identified TUDOR domains ('e' and 'h' above the alignment indicate β -strands and α -helices, respectively). The conserved Arginine and Aspartate residues present in all extended TUDOR domains are highlighted in green, aromatic cage residues in red, the Asparagine involved in sDMA binding in orange and a strongly conserved glycine in grey. To the left, the predicted likelihood of a domain to bind sDMA residues (based on the aromatic cage residues) is indicated with black (likely binder) and grey (potential binder) circles.

N-terminal MYND-type zinc finger. Based on domain composition, Tdrd1 might be the single mammalian counterpart of fly CG9925, CG9864 and CG4771, all of which encode besides multiple TUDOR domains also a MYND zinc finger (CG4771 contains in addition an RRM domain). Fly proteins with no assignable mouse counterparts are Krimer as well as the two

testes specific proteins CG15042 and CG15930 (the two TUDOR domains of Krimper and CG15042 are highly similar, potentially suggesting a common ancestor). Finally, mouse Tdrd8 seems to lack a detectable fly orthologue.

TUDOR/SMN domains often bind peptides with sDMA residues in target proteins. The sDMA-binding pocket resides



within the TUDOR core. It consists of four aromatic residues (Figure 1B, marked in red), whose aromatic rings form a cuboid cage and complex the di-methylated guanidine group (Selenko *et al*, 2001; Sprangers *et al*, 2003; Cote and Richard, 2005; Liu *et al*, 2010a, b). An additional conserved Asparagine (Figure 1B, marked in orange) interacts with the sDMA residue via a hydrogen bond (Liu *et al*, 2010a, b). In sDMA-binding TUDOR domains, the aromatic cage residues are highly conserved and are critical for sDMA binding. We inspected the *Drosophila* extended TUDOR domains for aromatic cage residues. The alignment in Figure 1B indicates that while a set of TUDOR domains harbours all of these important residues at the exact same position, numerous TUDOR domains seemingly lost the ability to bind sDMA residues due to multiple amino-acid exchanges at critical positions. Nevertheless, many of the identified proteins contain at least one TUDOR domain with an intact aromatic cage and therefore likely interact with sDMA residues.

We finally analysed the RNA expression pattern of all TUDOR/SMN genes via the adult *Drosophila* Fly Atlas (Chintapalli *et al*, 2007). This showed a strong bias for genes with extended TUDOR domains to be expressed in ovaries and/or testes, further suggesting a link to piRNA biology (Figure 1A; Supplementary Figure S3).

Defining the set of TUDOR proteins with critical roles in the ovarian piRNA pathway

The implication of several TUDOR proteins in piRNA biology and their often gonad-specific expression prompted us to genetically test all proteins with extended TUDOR domains for their involvement in the piRNA pathway. Defects in the piRNA pathway lead to sterility and to a substantial accumulation of transposon transcripts in ovaries. We therefore assayed these phenotypes in females where individual TUDOR domain-containing proteins were knocked down via RNAi specifically in the ovarian soma (marked in green in Figure 2A and B) or in the germline (marked in beige in Figure 2A and B).

RNAi in the follicular epithelium (soma), where only the primary piRNA pathway is active, was based on *tj*-GAL4 driven dsRNA-hairpin constructs (hp-lines) from the VDRC (Vienna *Drosophila* RNAi Centre) library (Dietzl *et al*, 2007; Olivieri *et al*, 2010). For the germline we expressed short hairpin constructs (sh-lines) with the germline-specific MTD-GAL4 driver, which allows robust knockdowns (Haley *et al*, 2008; Ni *et al*, 2011). In addition, we took advantage of the observation that VDRC hp-lines induce potent RNAi in the germline if expressed in conjunction with Dicer-2 (Sidney Wang and Sarah Elgin, personal communication). Figure 2B illustrates specificity and efficacy of the soma and germline-specific knockdowns using the piRNA biogenesis factor Armitage as an example.

Integrity of the somatic piRNA pathway was monitored via a *gypsy-lacZ* construct that accurately reports piRNA-mediated silencing in follicle cells (Figure 2C; Sarot *et al*, 2004; Olivieri *et al*, 2010). Integrity of the germline piRNA pathway was monitored via the steady-state RNA levels of the two transposons *HeT-A* and *blood* (Figure 2D). In addition, we determined female fertility rates (percentage of hatched eggs) for all knockdowns (Figure 2E). The two germline knockdown approaches yielded in nearly all cases identical results. We attribute the three exceptions (CG9925-hp, *yu*-sh,

CG14303-sh; Figure 2D and E) to off-target effects or non-functional RNAi lines.

Seven TUDOR proteins scored as putative piRNA pathway components (CG4771, CG11133, Tejas, CG14303, Spindle-E, Krimper, Yb). All four factors that had previously been shown to be essential pathway members (Spindle-E, Krimper, Tejas, Yb) were identified. In agreement with the literature, Spindle-E, Krimper and Tejas scored only in the germline knockdowns while Yb scored only in the soma assay (Lim and Kai, 2007; Malone *et al*, 2009; Szakmary *et al*, 2009; Olivieri *et al*, 2010; Patil and Kai, 2010). Papi and Tudor—though previously implicated in the pathway—did not result in transposon desilencing or sterility. This is in agreement with the literature as both proteins are dispensable for fertility and corresponding mutant ovaries contain no or only slightly elevated transposon RNA levels (Nishida *et al*, 2009; Liu *et al*, 2011). We note that the grandchild-less phenotype for Tudor (Boswell and Mahowald, 1985) is recapitulated in the Tudor germline knockdowns.

In addition to the known factors, germline knockdowns of three uncharacterized proteins (CG14303, CG4771, CG11133; Figure 2D and E) resulted in sterility and transposon silencing defects. Out of these, CG4771 was also identified as an essential component for the somatic piRNA pathway (Figure 2C) and we therefore decided to characterize this factor in more detail.

Vreteno (CG4771) is an essential piRNA pathway factor

CG4771 is localized on the third chromosome (Figure 3A) and encodes a protein with two extended TUDOR domains (Figure 3B). The C-terminal TUDOR domain might possess sDMA-binding activity (Figure 1B) and the relevant aromatic cage residues are conserved in distantly related *Drosophila* species (Figure 3B). In addition, CG4771 harbours an N-terminal RRM domain and a highly conserved zinc finger belonging to the MYND family (C2C4HC).

To verify that CG4771 is a piRNA pathway factor, we obtained genetic alleles of this gene. Females homozygous for the P-insertion *HP36220* (Bloomington), which is inserted into the 5'UTR of *CG4771* (Figure 3A) were sterile and laid eggs that exhibited defects in dorso-ventral patterning as evidenced by a high percentage of fused dorsal appendages. This is a common phenotype of piRNA pathway mutants and stems from the activation of the Chk2 DNA damage pathway, presumably caused by widespread DNA damage originating from uncontrolled transposon activity (Chen *et al*, 2007; Klattenhoff *et al*, 2007). However, in *HP36220* mutants only the *blood* element was de-repressed, although germline-specific and soma-specific knockdowns of *CG4771* clearly de-repressed also *HeT-A* and *ZAM*, respectively (Figure 3C). This suggested that *HP36220* is a hypomorphic allele. Indeed, nuclear Piwi localization in the mutant was impaired, yet to a lesser degree than in *armitage* or *Yb* mutants (Figure 3D; Olivieri *et al*, 2010).

We therefore generated an additional allele by mobilizing the *HP36220* element. Out of 280 analysed excision events, one line (CG4771[Δ1]) exhibited a more pronounced phenotype as homozygous females failed to lay eggs. The ovarian morphology of CG4771[Δ1] mutants strongly resembled those of *armitage* or *zucchini* null ovaries (Pane *et al*, 2007; Olivieri *et al*, 2010). In some egg chambers, we observed besides the oocyte nucleus a single giant nurse cell nucleus, indicating

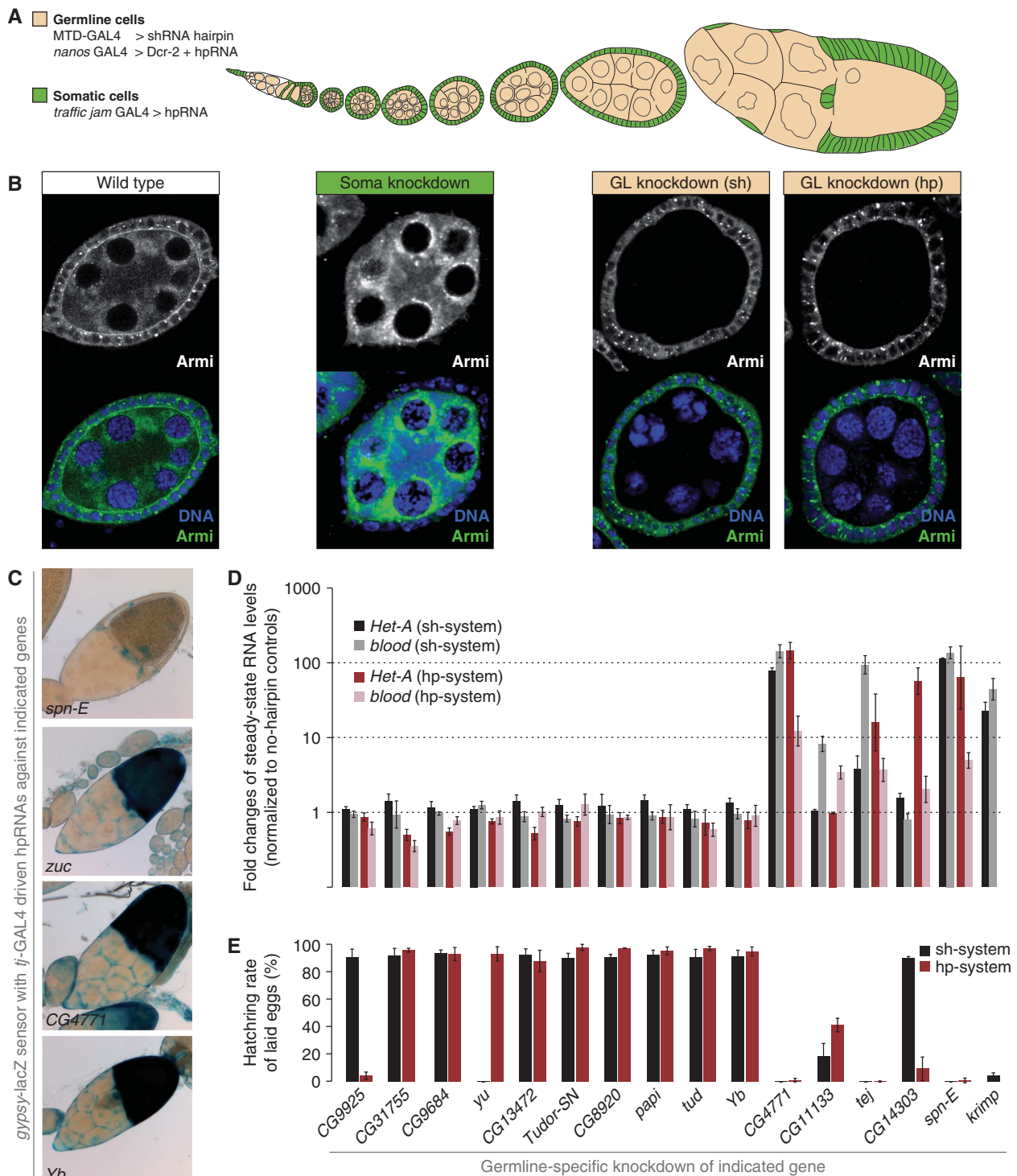


Figure 2 The set of TUDOR proteins involved in the *Drosophila* piRNA pathway. (A) Cartoon of a *Drosophila* ovariole (somatic cells are in green, germline cells are in beige). The RNAi systems used for the two cell types are listed. (B) Immunostaining of Armitage (green) and DNA (blue) in egg chambers expressing RNAi constructs in a tissue-specific manner (left: wild type; middle: soma knockdown via *tj*-GAL4 > hpRNA; right: germline knockdown via MTD-GAL4 > shRNA or NGT-GAL4 > Dcr-2 + hpRNA). Monochrome panels show only the anti-Armitage channel. (C) Bright field images of ovaries stained for β -GAL activity. The individual genotypes represent soma-specific knockdowns of the indicated genes in the background of the *gypsy-lacZ* sensor described in Sarot et al (2004). *zucchini* knockdown serves as a positive control and *spindle-E* as negative control. Of all TUDOR knockdowns, only those against CG4771 or *Yb* resulted in sensor de-repression. (D) Changes in steady-state levels of *Het-A* and *blood* transposon transcripts upon knockdown of individual TUDOR proteins in the germline with the shRNA (black/gray) or the hpRNA (red/rose) knockdown systems (normalized to no-hairpin controls via *rp49*; log scale; $n = 3$; error bars indicate s.d.). Identity of knocked down genes identical to the legend in (E). (E) Fertility rates of females with germline-specific knockdown of indicated TUDOR proteins using the shRNA (black) and the hpRNA (red) systems (~ 200 eggs per experiment; $n = 3$; error bars indicate s.d.).

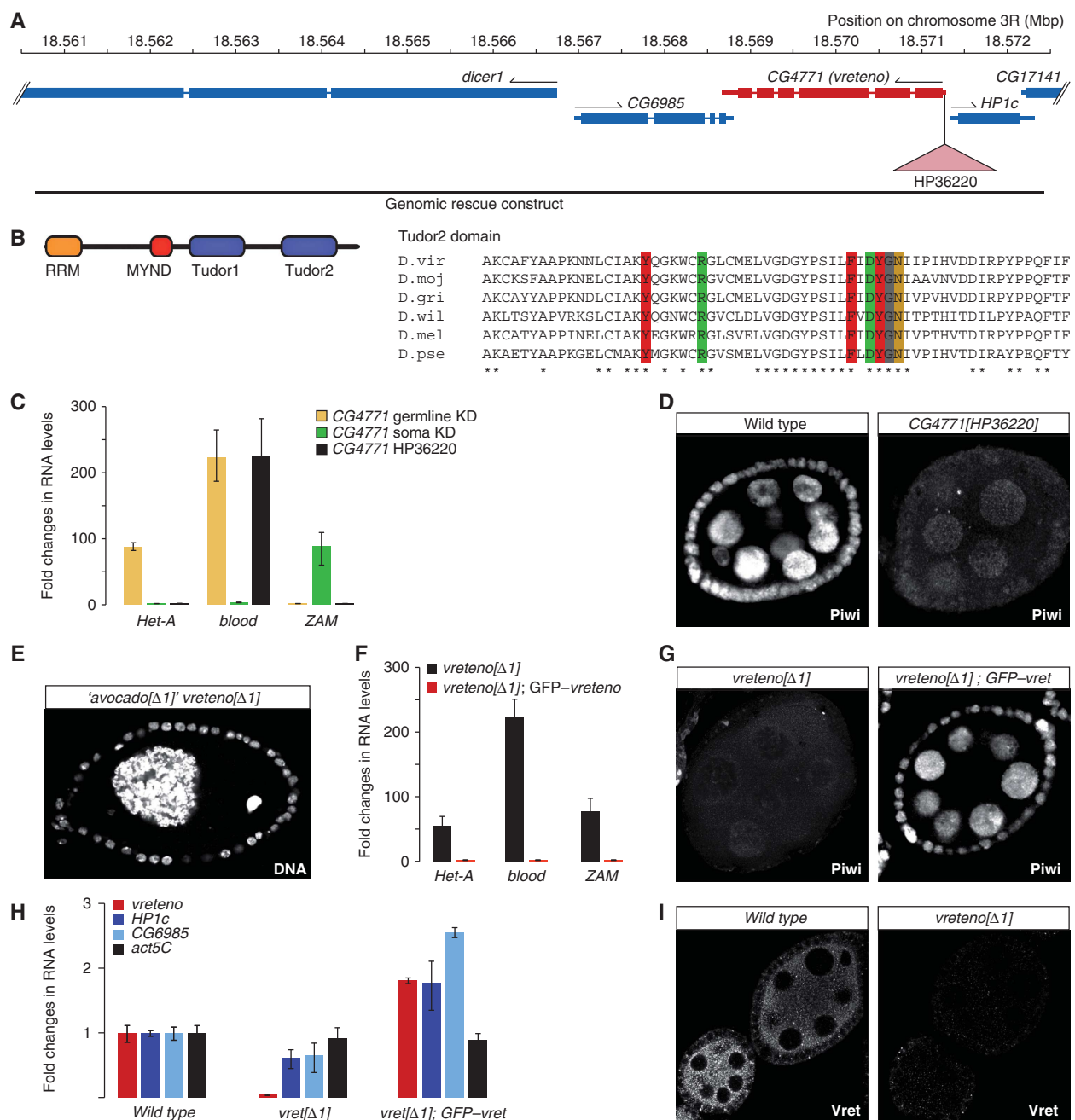


Figure 3 Vreteno is a novel piRNA pathway member. **(A)** Overview of the *CG4771 (vreteno)* genomic locus indicating flanking genes (blue), the HP36220-insertion site (pink triangle) and the extent of the genomic rescue construct. **(B)** Cartoon of the *CG4771* protein domain structure and sequence alignment of the C-terminal TUDOR domain in distantly related *Drosophilids* (*virilis*, *mojavensis*, *grimshawi*, *willistoni*, *melanogaster*, *pseudobscura*). Aromatic cage residues and the conserved Arg/Asp residues colour coded as in Figure 1B. **(C)** Changes in steady-state transposon levels ($n = 3$; s.d.) upon *CG4771* knockdown (normalized to no-hairpin controls) in soma (green) or germline (beige) in comparison to those in *CG4771[HP36220]* mutants (black; normalized to heterozygotes). **(D)** Immunostaining of Piwi in wild-type and *CG4771[HP36220]* mutant egg chambers. **(E)** The occasionally observed egg chamber morphology of *CG4771[Δ1]* (*vreteno*) mutants, which originally led us to name the gene 'avocado' (DNA stained with DAPI). **(F)** RNA levels of *CG4771*, of the flanking genes *HP1c* and *CG6985* and of *actin-5C* in *vreteno[Δ1]* mutant ovaries compared with *vreteno[Δ1]; GFP-vreteno* rescued ovaries (values normalized to *w[1118]* controls). **(G)** Immunostaining of Piwi in *vreteno[Δ1]* mutant egg chambers and in *vreteno[Δ1]* mutant egg chambers expressing a GFP-*vreteno* rescue construct. **(H)** Steady-state RNA levels of the *HeT-A*, *blood* and *ZAM* transposons in *vreteno[Δ1]* mutant ovaries compared with *vreteno[Δ1]; EGFP-vreteno* rescued ovaries (values normalized to heterozygous siblings; $n = 3$; error bars indicate s.d.). **(I)** Immunostaining of Vreteno in wild-type and *vreteno[Δ1]* mutant egg chambers at identical microscope settings.

severe defects in cytokinesis. Based on the morphology of these egg chambers (Figure 3E), we initially named *CG4771* 'avocado'. While this work was under review, *CG4771* was named 'vreteno' in FlyBase by the Lehmann

group ('vreteno' means 'spindle' in Bulgarian, referring to the spindle class phenotype of eggs laid by *CG4771* mutants) and we therefore adopted this name for consistency reasons.

In *vreteno*[$\Delta 1$] homozygous ovaries, germline- and soma-specific transposons were severely de-repressed, indicating the stronger nature of this allele (Figure 3F). This was paralleled by a more pronounced defect in nuclear Piwi accumulation (compare Figure 3G and D). To verify that the *vreteno*[$\Delta 1$] phenotype is due to defects in the *CG4771* locus, we restored fertility (not shown), transposon silencing (Figure 3F) and nuclear Piwi localization (Figure 3G) to wild-type levels by introducing a genomic rescue construct that expresses GFP-tagged Vreteno under its endogenous regulatory regions (Figure 3A). As this rescue construct also contained the complete loci for *HP1c* and *CG6985*, we measured steady-state RNA levels of all three genes in ovaries of *vreteno*[$\Delta 1$] mutants and of the GFP-*vreteno* rescued animals (Figure 3H). This confirmed the specificity of the $\Delta 1$ allele for the *vreteno* locus. Immunofluorescence analysis with an antibody recognizing the Vreteno N-terminus further indicated that also the protein is essentially not detectable in ovaries from *vreteno*[$\Delta 1$] mutants (Figure 3I).

We also analysed the requirement of *vreteno* for the piRNA pathway in males. Towards this end, we measured steady-state levels of the transposons *mdg1* and *copia* as well as of the repetitive *Stellate* locus that is under control of the piRNA pathway in testes. This indicated a requirement of *vreteno* for *copia* and *Stellate* silencing, supported by the observation that silencing was fully restored in males expressing a GFP-*vreteno* rescue construct (Supplementary Figure S4). Taken together, *vreteno* encodes a novel piRNA pathway factor that is essential for the ovarian and testes piRNA pathways.

Vreteno is required for primary piRNA biogenesis in soma and germline

Defects in primary piRNA biogenesis (e.g. in *armitage* or *zucchini* mutants) result in a collapse of piRNA populations in follicle cells, in a severe reduction of most germline piRNA species and in defects in Piwi's nuclear accumulation accompanied by a significant loss of Piwi protein (Pane et al, 2007; Malone et al, 2009; Haase et al, 2010; Olivieri et al, 2010; Saito et al, 2010). Delocalization and decreased levels of Piwi were also observed in the *vreteno*[$\Delta 1$] mutant (Figure 3D and G). We therefore analysed the effects of loss of Vreteno on PIWI-family proteins and on piRNA populations in the ovarian soma and germline.

Similar to an RNAi-mediated Armitage knockdown, knockdown of Vreteno in follicle cells led to an almost complete loss of Piwi protein in these cells (Figure 4A), indicating defects in primary piRNA biogenesis. Indeed, analysis of small RNA populations obtained from *vreteno*[$\Delta 1$] mutant ovaries showed that piRNAs originating from the *flamenco* cluster, which gives rise to primary piRNAs in follicle cells (Lau et al, 2009; Li et al, 2009; Malone et al, 2009), were almost entirely lost in *vreteno*[$\Delta 1$] mutants (Figure 4B). Highly similar profiles were obtained from ovaries mutant for *armitage*, *zucchini* or *Yb*, the only known factors involved in primary piRNA biogenesis. In contrast, piRNA populations from *spindle-E* mutant ovaries showed no impact, in agreement with *spindle-E* functioning exclusively in the germline pathway (Figure 4B; Malone et al, 2009). Very similar results were

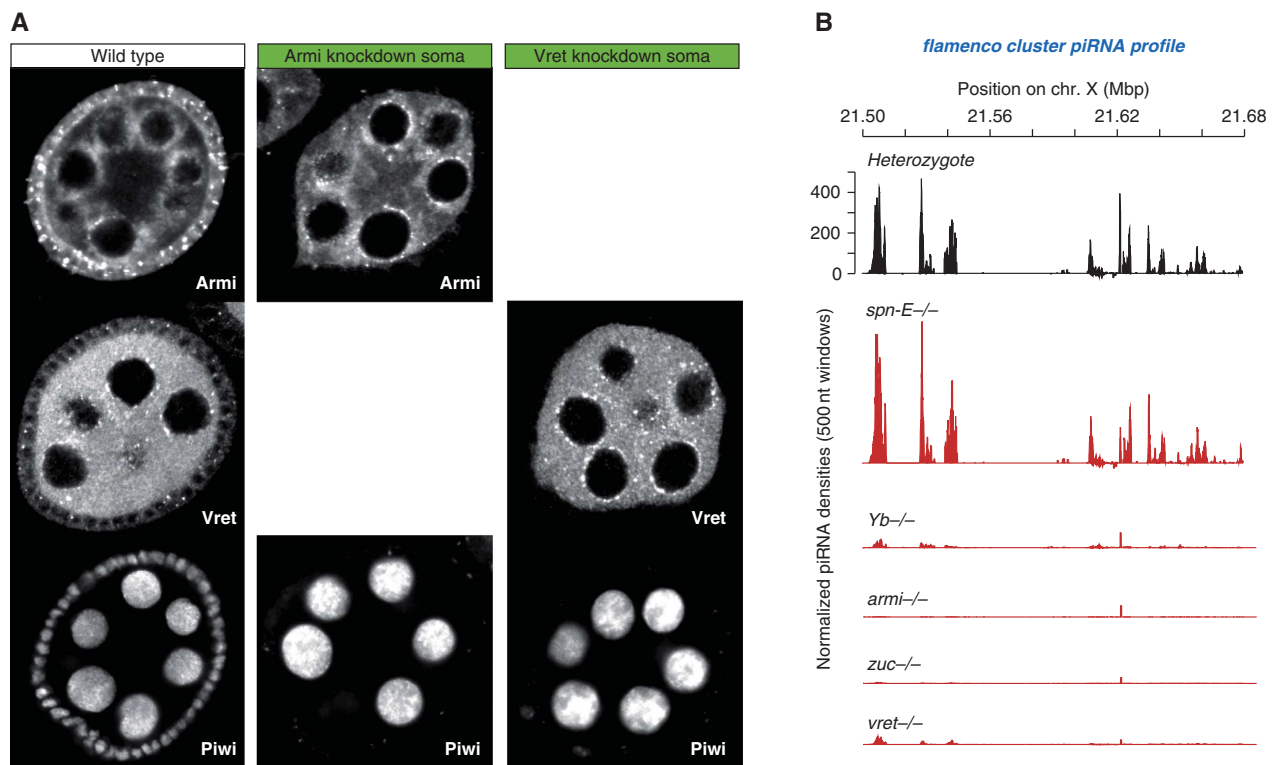


Figure 4 Vreteno is essential for primary piRNA biogenesis in the soma. (A) Immunostaining of Piwi (lower panels) in wild-type egg chambers (left) in comparison to egg chambers expressing hpRNAs against *armitage* (centre) or *vreteno* (right) specifically in somatic cells. Armitage and Vreteno stainings indicate the knockdown efficiency. (B) Normalized piRNA profiles (23–30 nt small RNAs) obtained from control ovaries (*vret* heterozygote; black) in comparison to profiles obtained from indicated mutant ovaries (red) mapping uniquely to the soma-specific piRNA cluster *flamenco*. The y axis for the heterozygote plot is representative for all plots.

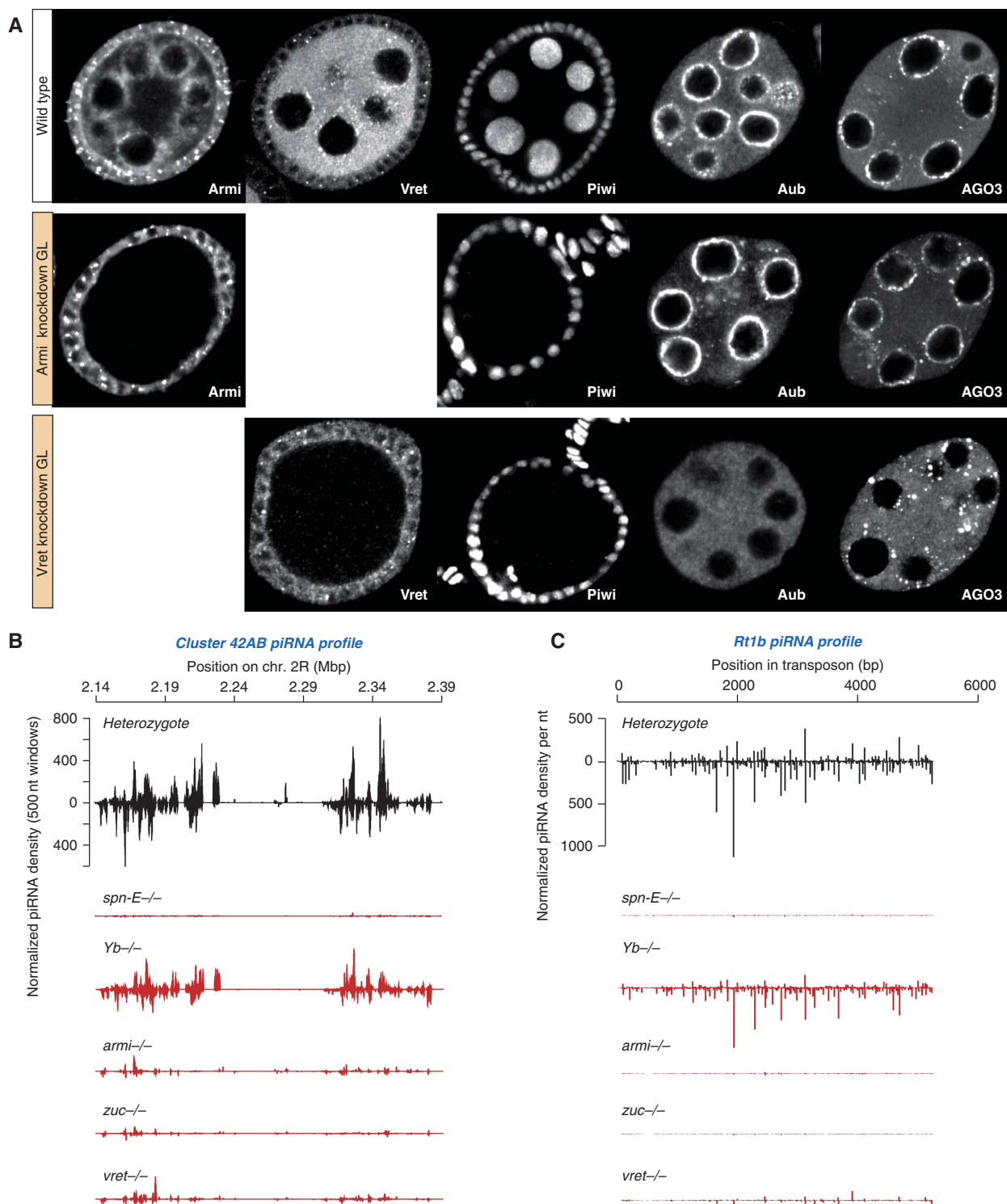


Figure 5 Vreteno is essential for piRNA biogenesis in the germline. **(A)** Immunostaining of Piwi, Aubergine and Ago3 in wild-type egg chambers (top row) in comparison to egg chambers expressing shRNAs against *Armitage* (centre row) or *Vreteno* (lower row) specifically in germline cells. Armitage and Vreteno stainings indicate the knockdown efficiency. **(B)** Normalized piRNA profiles (23–30 nt small RNAs) obtained from control ovaries (*vret* heterozygote; black) in comparison to profiles obtained from indicated mutant ovaries (red) mapping uniquely to the germline-specific piRNA cluster 42AB (sense and antisense piRNAs are indicated with peaks pointing up- and downwards). **(C)** Normalized piRNA profiles obtained from control ovaries (black) in comparison to profiles obtained from indicated mutant ovaries (red) mapping to the germline-dominant *Rt1b* transposon.

obtained when piRNAs mapping to individual soma-specific transposons such as *Tabor* or *ZAM* were analysed (Supplementary Figure S5).

Knockdown of Vreteno in the germline resulted in a near complete loss of Piwi protein, again pheno-copying the Armitage germline knockdown (Figure 5A). In addition,

piRNA populations mapping to the germline-specific piRNA cluster 42AB or to the transposon *Rt1b*, whose piRNA populations depend on primary piRNA biogenesis (Malone et al, 2009), were severely reduced in *vreteno* mutant ovaries as well as in *spindle-E*, *armitage* or *zucchini* mutant ovaries (Figure 5B and C). No significant changes were observed in *Yb* mutant ovaries consistent with this gene acting in somatic follicle cells only. We also observed a severe disruption of the nuage localization of Aubergine and Ago3 in germline cells depleted of *Vreteno* (Figure 5A). This is significant, as loss of the primary biogenesis factor Armitage does not impact the subcellular localization of Aubergine and Ago3 (Figure 5A). Also in *armitage* null mutants, Aubergine and AGO3 localization to nuage was not significantly perturbed (not shown), indicating that the observed differences cannot be attributed to insufficient knockdowns.

To characterize piRNA populations from *vreteno* mutants in more detail, we mapped small RNAs isolated from ovaries mutant for *vreteno*, *armitage*, *zucchini*, *Yb* or *spindle-E* as well as small RNAs isolated from their respective heterozygote controls to annotated transposon families (Jurka et al, 2005; Supplementary Table S2). While the *spindle-E* libraries were taken from the literature (Malone et al, 2009), *armitage*, *zucchini* and *Yb* libraries were prepared for this study (previous libraries were based on unfavourable genetic alleles in the case of *armitage*, or suffered from inaccurate genotyping in the case of *zucchini*; Malone et al, 2009). All heterozygote libraries were scaled to 1 million repeat-derived small RNAs (23–30 nt in length). Each mutant library was then normalized to its respective heterozygote library using non-transposon-derived endo-siRNA populations.

Figure 6A displays the levels of small RNAs that mapped to annotated transposons in sense or antisense orientation as a function of their length. The mutant small RNA profiles could be grouped into three classes: (1) *Yb* is a soma-specific factor essential for primary piRNA biogenesis (Olivieri et al, 2010; Qi et al, 2010; Saito et al, 2010). Accordingly, the majority of ovarian piRNAs were unchanged in this mutant. Only a slight decrease (preferentially in antisense piRNAs) was seen, in agreement with soma piRNAs exhibiting an extreme antisense bias (Malone et al, 2009). (2) *Spindle-E* is required for the germline-specific ping-pong cycle but is dispensable for primary piRNA biogenesis (Malone et al, 2009). Accordingly, *spindle-E* mutant piRNA profiles showed a severe collapse in piRNA populations, with a subpopulation of piRNAs (nearly exclusively antisense) remaining. These showed a pronounced bias towards a 5' terminal Uridine, lacked the characteristic ping-pong signal (see below) and therefore represent primary piRNAs (not shown). (3) The *vreteno*, *armitage* and *zucchini* mutant libraries strongly resembled each other. They exhibited a severe depletion of piRNAs, with more pronounced losses in antisense populations, indicative of these factors being required for primary piRNA biogenesis.

The described classification of the analysed pathway mutants was supported by the changes in transposon-specific piRNA populations (Figure 6B and C). We focused on a set of 37 transposon families that has been well characterized previously (Malone et al, 2009). These transposons can be grouped into three cohorts (Figure 6B) depending on whether piRNAs targeting them are predominantly found in germline cells (germline dominant), in somatic cells (soma dominant)

or in both cell types (intermediate). Transposons shown in Figure 6B were ranked according to their ping-pong signature in the average heterozygote library (blue heatmap). This correlated strongly with the previously reported extent of maternal piRNA inheritance (yellow/red heatmap) as maternal deposition is only possible for germline piRNAs, which typically participate in the ping-pong cycle (Malone et al, 2009).

We mapped piRNAs to the selected transposon families and calculated the log2 fold ratios of the respective heterozygote/mutant pairs. In agreement with the genetic data, piRNA populations from *Yb* and *spindle-E* mutants were essentially anti-correlated (Figure 6B). While *Yb* mutations affected soma-dominant elements, *spindle-E* mutations affected germline-dominant elements. A scatter plot of the log2 fold het/mut ratios underlines this further (Figure 6C; $r = -0.52$; Pearson correlation). In contrast, libraries obtained from *vreteno*, *armitage* or *zucchini* mutants exhibited nearly identical losses of piRNA populations mapping to soma, intermediate or germline-dominant elements (Figure 6B; Pearson correlations: $r(\text{armi/zuc}) = 0.80$; $r(\text{armi/avo}) = 0.76$; $r(\text{zuc/avo}) = 0.87$). The soma-dominant elements were the most consistently affected group, in agreement with all of them depending on primary piRNA biogenesis. Within the germline-dominant group, some transposons showed only very mild losses of piRNAs (e.g. *protoP-A*, *Doc*, *F-element*). This pattern was observed in all three libraries (see e.g. the bottom scatter plot in Figure 6C), further underlining the hypothesis that all three factors participate in a common step of piRNA biogenesis.

We finally analysed the impact of the five mutants on the ping-pong cycle (Figure 6D). While ping-pong collapsed in *spindle-E* mutants (Malone et al, 2009), it was unaffected in *Yb* mutants, consistent with this factor being soma specific. Interestingly, *vreteno*, *armitage* or *zucchini* mutants showed no defects in ping-pong signatures. On the contrary, for many elements, the ping-pong signal was elevated in comparison to the wild-type situation. This trend was particularly evident for the intermediate group of transposons, especially in *armitage* and *zucchini* mutants. Of note, ping-pong was still efficient in *vreteno* mutants despite the fact that the key players Aubergine and Ago3 were delocalized from nuage (Figure 5A). This indicates that nuage localization *per se* is not required for ping-pong.

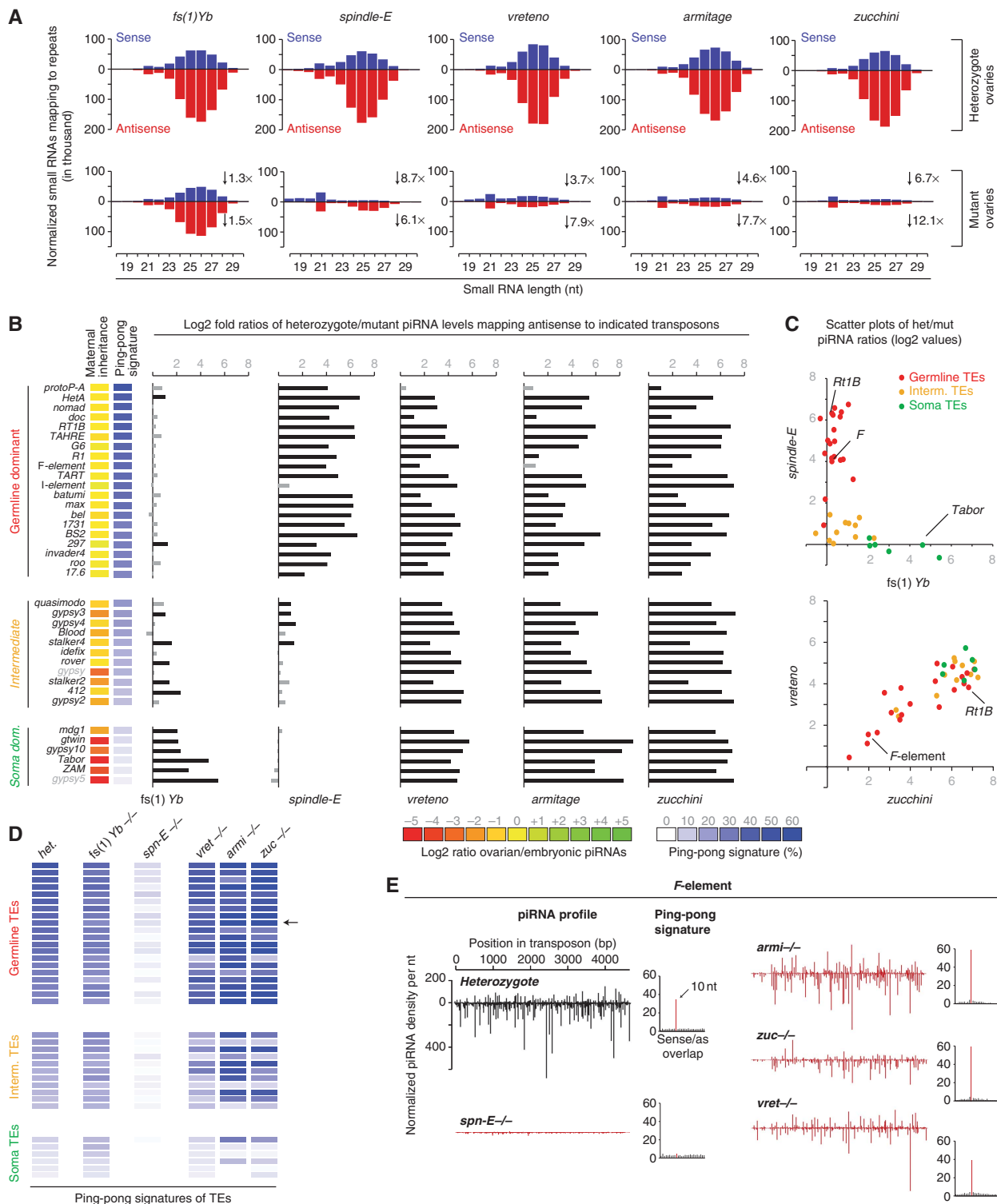
Figure 6E shows sense and antisense piRNA populations mapping to the *F-element*, which is a particularly interesting case. piRNA populations for this element were highly sensitive to perturbations in the ping-pong cycle (*spindle-E* mutant), but not to defects in primary piRNA biogenesis (*vreteno*, *armitage*, *zucchini* mutants). The increased ping-pong signal in primary pathway mutants suggests that in these mutants, the pool of primary piRNAs (lacking ping-pong signatures) is eliminated, thereby leaving only ping-pong pairs. For the *Rt1B* element (also a germline-dominant element), this is different as piRNA populations in this case evidently depend on ping-pong amplification as well as on primary piRNA biogenesis (Figure 5C). Nevertheless, the small pool of remaining *Rt1B* piRNAs in primary pathway mutants still displays strong ping-pong signatures. We speculate that these element-specific differences are related to the maternally transmitted piRNA pool (Brennecke et al, 2008). For unknown reasons, the *F-element* ping-pong might be

maintained by maternally transmitted piRNAs, whereas this is not the case for elements such as *Rt1B*.

In summary, our analysis shows that *Vreteno* is essential for primary piRNA biogenesis and that the ping-pong cycle can operate independently of primary piRNA biogenesis in *Drosophila* as suggested previously (Malone *et al*, 2009).

Vreteno accumulates in Yb bodies and physically interacts with Piwi, Armitage and Yb

In follicle cells and in cultured somatic stem cells (OSCs; Niki *et al*, 2006; Saito *et al*, 2009), Armitage and Yb are enriched in discrete foci termed Yb bodies, which presumably are sites of primary piRNA biogenesis as Piwi transits through these



bodies (Olivieri *et al*, 2010; Qi *et al*, 2010; Saito *et al*, 2010). To better understand at which step Vreteno acts during primary piRNA biogenesis, we performed genetic epistasis experiments in ovarian follicle cells as well as in OSCs. A GFP–Vreteno fusion protein expressed under the endogenous control regions localized to the cytoplasm of somatic and germline cells in the ovary (Figure 7A and B). Intriguingly, Vreteno was enriched in discrete foci in follicle cells and in nuage in germline cells, highly reminiscent of an Armitage–GFP fusion protein. A more detailed analysis of the subcellular localization in follicle cells revealed that the majority of Vreteno and Armitage foci precisely overlapped (Figure 7B). Also in OSCs, Vreteno localized to the cytoplasm and accumulated in distinct foci, which typically were Armitage positive (Supplementary Figure S6). In some cases, however, Vreteno and Armitage foci seemed to directly flank each other and in several cells we observed accumulation of Vreteno in large cytoplasmic domains that showed no Armitage accumulation (Supplementary Figure S6). Typically, these cells lacked discernable Armitage foci, potentially suggesting that Yb bodies undergo remodelling at specific stages.

We next analysed the subcellular localization of Vreteno in cells lacking known primary biogenesis factors using mitotic follicle cell clones in flies or RNAi in OSCs. Vreteno localization to Yb bodies was dependent on Armitage and Yb (Figure 7C). Strikingly, loss of Zucchini, a phospho-lipase or putative nuclease localizing to the outer mitochondrial membrane (Pane *et al*, 2007; Saito *et al*, 2010; Watanabe *et al*, 2011; Huang *et al*, 2011a) resulted in the accumulation of Vreteno in massive Yb bodies, similar to what has been observed previously for Armitage (Figure 7C; Olivieri *et al*, 2010). In contrast, Vreteno localization was unperturbed in cells lacking Piwi (Supplementary Figure S7). Furthermore, Armitage and Yb still localized to Yb bodies in cells lacking Vreteno (Supplementary Figure S7). Comparable results were obtained in OSCs upon knockdown of individual factors (Figure 7D shows Yb and Zucchini knockdowns; Yb and *zuc* mRNA levels were reduced to 7.2% and 13.6%, respectively, resulting in 19-fold (Yb) or eight-fold (*zuc*) elevated *mdg1* transposon levels).

We obtained further support for a role of Vreteno in primary piRNA biogenesis by co-immunoprecipitation (co-IP) experiments from OSC lysate. This indicated that Armitage and Vreteno reside in a common complex (supported by reciprocal co-IP experiments). In addition, both proteins were also found to physically interact with Yb and Piwi (Figure 7E).

Taken together, Vreteno is a novel Yb-body component that interacts with Armitage, Yb and Piwi. Vreteno localization

depends on Armitage and Yb and in *zucchini* mutant cells it accumulates together with Armitage and Yb in large cytoplasmic aggregates.

The Tdrd12 proteins CG11133 and CG31755 interact with Vreteno and are essential primary piRNA pathway factors in the germline

Vreteno is besides Piwi, Armitage and Zucchini the fourth known factor required for primary piRNA biogenesis in the ovarian soma and germline. In contrast, the Tdrd12-like protein Yb is an essential piRNA biogenesis factor but acts only in somatic cells (Olivieri *et al*, 2010; Qi *et al*, 2010; Saito *et al*, 2010). The *Drosophila* genome contains two additional Tdrd12-like proteins (CG11133, CG31755; Figure 8A), both of which are also specifically expressed in gonads (Supplementary Figure S3; FlyAtlas). Since Vreteno physically interacts with Yb, we reasoned that it might also interact with CG11133 and CG31755. We performed IP experiments from ovaries expressing GFP–Vreteno instead of the endogenous protein and from wild-type control ovaries using a monoclonal GFP antibody followed by quantitative mass spectrometry analysis. Strikingly, CG11133 and CG31755 were among the five most enriched proteins in the GFP–Vreteno IP (besides Vreteno, Hsp27 and CG9281; Supplementary Table S3).

To further characterize the fly Tdrd12 family, we analysed expression and localization of GFP-tagged Yb, CG11133 and CG31755 in ovaries (all proteins were expressed under their respective endogenous control regions). While CG31755 was expressed at comparable levels in follicle and germline cells, Yb was expressed specifically in follicle cells and CG11133 predominantly in germline cells (Figure 8B). Remarkably, the subcellular localizations of the three proteins were highly reminiscent of the Armitage or Vreteno localizations. Double labelling experiments confirmed the co-localization of Armitage with Yb and CG31755 in follicle cells (Figure 8B). Also for CG11133, though present only at very low levels, localization to Armitage foci in follicle cells was observed (not shown). In germline cells, CG31755 localized to perinuclear clouds that were also positive for Armitage, while CG11133 was enriched in nuage, where Armitage was also present (Figures 7A and 8B).

Given the follicle cell expression and Yb-body localization of CG31755 and CG11133, we investigated their somatic piRNA pathway involvement in OSCs by knocking them down, individually or in combination with RNAi. Depletion of CG31755, but not CG11133 resulted in a strong increase of *mdg1* transposon levels, similar to what has been observed for *zuc*, *armi* or Yb knockdowns (Figure 8C; Saito *et al*, 2010).

Figure 6 Vreteno, Zucchini and Armitage are essential primary piRNA biogenesis factors but are dispensable for the ping-pong cycle. (A) Length profiles of all repeat-derived (transposon and satellite repeats) small RNAs (18–30 nt) isolated from ovaries of the indicated mutants and their respective heterozygous controls (all heterozygote libraries normalized to 1 million repeat-derived 23–30 nt RNAs). Sense populations are in blue and antisense populations are in red. The fold decrease in the respective populations (23–30 nt only) is indicated. (B) Bar diagram indicating the changes of normalized piRNAs mapping antisense to the indicated transposons (left) in the indicated mutant ovaries compared with the respective heterozygous control ovaries (het/mut ratios are given as log₂ values). Grey bars indicate values below 1 (less than two-fold changes). Identity of the analysed mutants is given at the bottom. Transposons are grouped into germline-dominant (red), intermediate (yellow) and soma-dominant (green) based on Malone *et al* (2009). The heatmaps indicate degree of maternal piRNA inheritance (yellow: strong; red: weak) and ping-pong signature (blue: strong; white: absent) of each individual element. (C) Scatter plots of the log₂ values plotted in (B), where individual transposons are colour coded as in (B). (D) Ping-pong signatures of the individual transposons (classification and order as in (B) in the average heterozygote (het.) and the indicated mutants as a heatmap ranging from strong signals (dark blue) to no signal (white). The F-element is indicated with an arrow. (E) Normalized piRNA profiles (sense and antisense) mapping to the F-element. Compared are populations from heterozygotes (black) to the indicated mutants (red). Ping-pong signatures (basis for the heatmap in (D)) are shown to the right of each plot.

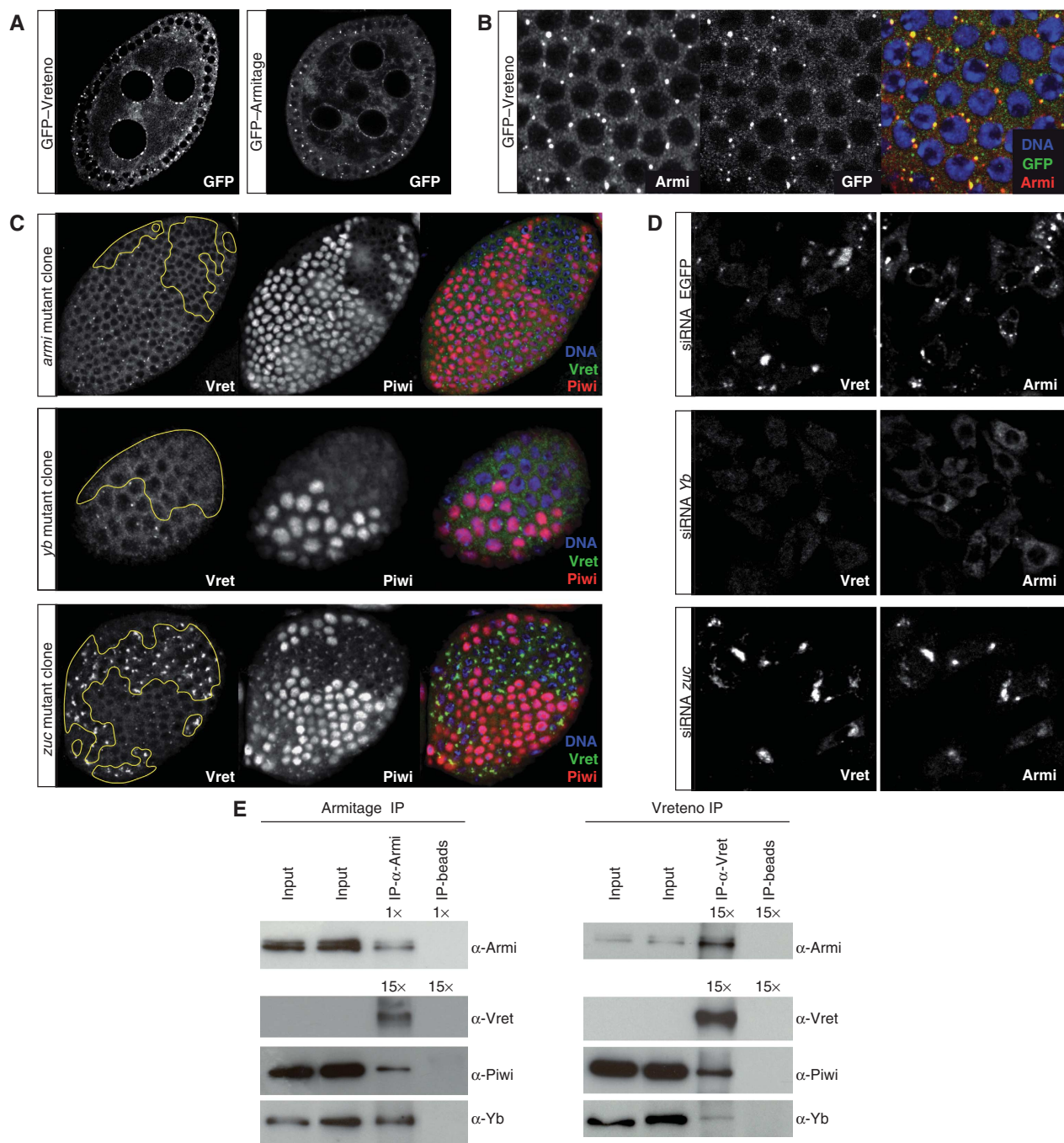


Figure 7 Vreteno is a novel Yb-body component. (A) Subcellular localization of GFP-tagged Vreteno or Armitage (optical section of egg chambers) expressed under the respective endogenous regulatory regions. (B) Confocal section through the follicular epithelium of a GFP-Vreteno (green) expressing egg chamber stained for Armitage (red) and DNA (blue). (Right panel) Merge of all three channels (co-localization of Vreteno and Armitage results in yellow). (C) Immunostaining of Vreteno (green), Piwi (red) and DNA (blue) in egg chambers, where clones of cells mutant for the indicated genes (left) have been induced in the follicular epithelium (clone borders are indicated with a yellow line). (D) Co-immunostaining of Vreteno (left) and Armitage (right) in OSCs transfected with siRNAs against EGFP (top), Yb (middle) or *zucchini* (bottom). (E) Co-IPs of Armitage and Vreteno from OSC lysate. Western blots against Armitage, Vreteno, Piwi and Yb. For the Armitage western blot (Armi-IP), only 1/15th compared with all other blots was loaded in the IP-lanes. Beads lacking antibody (IP-beads) served as control.

Double depletion of CG31755 and CG11133 led to an additional, but small increase in *mdg1* transcript levels, suggesting a minor role for CG11133 in the somatic piRNA pathway.

We performed a similar analysis in germline cells. The initial survey of TUDOR proteins already indicated that CG11133 was critical for transposon silencing in the germline,

while CG31755 was not (Figure 2D and E). We analysed silencing of the *HeT-A* and *blood* elements in ovaries depleted for CG31755 and/or CG11133 via the shRNA system (Figure 8D). While knockdown of CG31755 had no significant impact on *HeT-A/blood* silencing, a nearly 10-fold increase in *blood* levels was measured in CG11133 knockdown ovaries

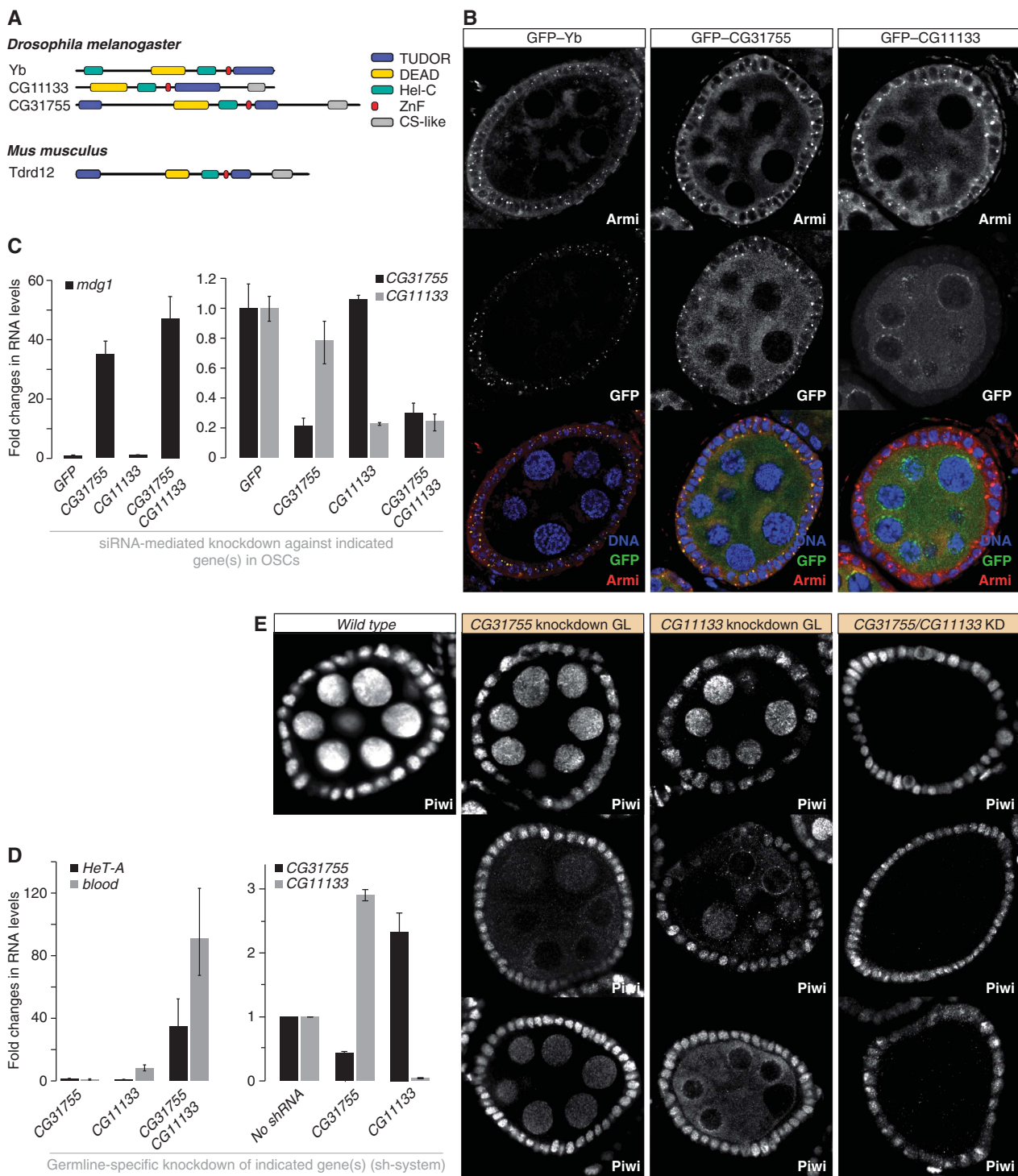


Figure 8 The Tdrd12 orthologues CG11133 and CG31755 are essential primary piRNA pathway factors in the germline. (A) Cartoon overview of the Tdrd12 proteins in fly and mouse (Tudor domains: blue; other domains as indicated). (B) Immunofluorescence staining for Armitage and GFP-tagged Yb, CG31755 or CG11133. (Lower panels) Merge of the three channels (DNA, GFP, Armitage). (C) The left chart shows changes in steady-state levels of the *mdg1* transposon upon knockdown of CG11133, CG31755 or both together in OSCs. Values were normalized to GFP control knockdowns via *rp49*. The right chart indicates the efficiencies of the respective knockdowns ($n = 3$; error bars indicate s.d.). (D) The left chart shows changes in steady-state levels of the *HeT-A* and *blood* transposons upon knockdown of CG11133, CG31755 or both together in germline cells (MTD-GAL4 > shRNA). Values were normalized to 'no-hairpin' control flies via *rp49*. The right chart indicates the efficiencies of the respective knockdowns ($n = 3$; error bars indicate s.d.). (E) Immunofluorescence staining for Piwi in wild-type egg chambers (left) compared with egg chambers where the indicated genes have been knocked down via shRNAs in the germline only. For the CG31755 and CG11133 single knockdowns, most egg chambers displayed a wild-type Piwi pattern (top), while 3–5% displayed the phenotypes of the lower panels. The phenotype for the double knockdown was fully penetrant.

(Figure 8D). Strikingly, double depletion of CG31755 and CG11133 de-silenced *HeT-A* and *blood* to levels comparable to verified piRNA pathway mutants (Figures 2D and 8D). The observed silencing defects correlated with defects in Piwi levels and localization (Figure 8E). In the single knockdowns, most egg chambers displayed wild-type Piwi levels and localization. In about 3–5% of egg chambers, however, we observed reduced Piwi levels or cytoplasmic Piwi localization (representative images in Figure 8E). Strikingly, double depletion of CG11133 and CG31755 resulted in an almost complete loss of germline Piwi (Figure 8E). Based on these observations and the correlation between defects in nuclear Piwi accumulation and defects in primary piRNA biogenesis (Olivieri *et al*, 2010; Saito *et al*, 2010), we conclude that CG11133 and CG31755 function together in primary piRNA biogenesis in germline cells. We therefore named CG11133 ‘Brother of Yb’ and CG31755 ‘Sister of Yb’.

Discussion

The set of TUDOR domain-containing proteins in *Drosophila*

Whereas Tudor, the founding member of TUDOR domain-containing proteins was genetically identified >25 years ago as a grandchild-less gene in *Drosophila* (Boswell and Mahowald, 1985), the link between TUDOR domains and the piRNA pathway has only recently emerged: On the one side, PIWI proteins have been shown to contain sDMA residues, docking sites for TUDOR domains. In selected cases, sDMA-dependent interactions between TUDOR domains and PIWI proteins could indeed be shown (Nishida *et al*, 2009; Vagin *et al*, 2009; Kirino *et al*, 2010; Huang *et al*, 2011b). On the other hand, 10 out of the 17 identified proteins in *Drosophila* that contain an extended TUDOR domain have been directly linked to the piRNA pathway (including the genes identified in this study). We emphasize that our RNAi-based assays could only score those genes, which are individually required for the ovarian piRNA pathway and which perform essential roles in it. Figure 1A suggests that several TUDOR domain-containing genes could act redundantly (in particular *yu* and *papi*, CG9925 and CG9684, *Tejas* and CG8920, CG11133 and CG31755) or specifically in testes (CG15042, CG15930). Double knockdowns of these gene pairs might identify their possible involvements in the pathway.

What could be the role of this surprisingly large set of proteins in the piRNA pathway? The protein domain cartoon of the *Drosophila* TUDOR family (Figure 1A) suggests that numerous effector domains such as helicase domains, RNA-binding domains, zinc-finger domains, etc. are targeted to PIWI proteins via modular sDMA–TUDOR interactions. Several reports indicate that the sDMA–TUDOR affinity is influenced by the peptide sequence harbouring the sDMA residue (Nishida *et al*, 2009; Liu *et al*, 2010a, b; Huang *et al*, 2011b). This might allow for highly specific and ordered interactions between methylated target proteins and subsets of TUDOR proteins. As many TUDOR proteins contain multiple TUDOR domains, a second scenario is that these multi-domain proteins act as scaffolds (Nishida *et al*, 2009; Huang *et al*, 2011b) to bring different effector proteins such as Aubergine and Ago3 into close physical proximity, a likely

prerequisite for the efficient functioning of the ping-pong cycle.

It is, however, critical to mention that many of the identified TUDOR domains in *Drosophila* (similar results emerge in vertebrates) carry mutations in aromatic cage residues, indicating that they lost the ability to interact with sDMA residues. For example, mouse Tdrd12 as well as the three *Drosophila* counterparts Yb, CG11133 and CG31755 lack nearly all aromatic cage residues. Nevertheless, a genetically identified Yb allele that carries a mutation in the ultra-conserved Arginine residue in $\beta 4$ of the extended TUDOR domain is a strong loss of function allele (Szakmary *et al*, 2009). Certain TUDOR domains might thus have evolved to bind alternative residues or post-translational modifications. This might also explain the puzzling observation that loss of the sDMA generating enzyme Csl (PRMT5) results in a surprisingly mild phenotype compared with loss of individual TUDOR domain-containing proteins such as Spindle-E, Krimper or Vreteno (Anne *et al*, 2007).

The role of Vreteno in primary piRNA biogenesis

Four lines of evidence place Vreteno into the process of primary piRNA biogenesis. First, Vreteno is an essential factor for the piRNA pathway in somatic follicle cells, which has been shown to consist only of the primary branch. Second, piRNA profiles from *vreteno* mutant ovaries strongly resemble those from *armitage* and *zucchini* mutant ovaries with primary piRNA populations collapsing, while ping-pong signatures are not affected. Third, Vreteno physically interacts with Armitage, Yb and Piwi, three factors of the primary piRNA pathway. Finally, Vreteno localizes to Yb bodies, the presumed sites of primary piRNA biogenesis in follicle cells.

Our data provide less insight at the mechanistic level. What is the precise role of the TUDOR domains in Vreteno? Do they interact with sDMA residues in PIWI proteins? What are the functions of the MYND domain and the N-terminal RRM domain? Combinations of biochemical and genetic rescue experiments will be crucial to answer these important questions.

An intriguing observation of our studies in germline cells is that Vreteno is required for the nuage localization of Aubergine and Ago3. Nevertheless, secondary piRNA biogenesis via ping-pong was functional. In this, *vreteno* mutants differ significantly from *armitage* mutants. Whether Vreteno merely tethers Aubergine and/or Ago3 to nuage (potentially via sDMA–TUDOR interactions) or whether Vreteno is also important for primary piRNA biogenesis feeding into Aubergine will be an important question for the future. Co-localization experiments in germline cells indicate that Vreteno foci are in very close proximity to Aubergine and especially AGO3 foci (Supplementary Figure S8). sDMA sites in these proteins have been mapped (Nishida *et al*, 2009) and it will be important to determine whether these peptides mediate physical interactions with Vreteno.

The three Tdrd12 orthologues are essential primary piRNA pathway factors in flies

The TUDOR domain-containing protein Yb is an essential factor for primary piRNA biogenesis in somatic follicle cells (Olivieri *et al*, 2010; Qi *et al*, 2010; Saito *et al*, 2010). Genetically, this gene is dispensable for the piRNA pathway in germline cells, despite the fact that other primary piRNA

biogenesis factors such as Armitage or Zucchini are essential factors in soma and germline. As Yb is not expressed in germline cells, it has been suggested that the sequence-related proteins CG11133 and/or CG31755 fulfil Yb's function in the germline (Olivieri *et al*, 2010). All three proteins share a similar domain composition, although CG31755 contains a second TUDOR domain and CG11133 and CG31755 harbour a C-terminal CS-like domain (Figure 8A). The single vertebrate orthologue of these three proteins is the uncharacterized Tdrd12 protein (Tdrd12 likely corresponds to the repro23 locus; repro23 mutant mice are male sterile and resemble piRNA pathway mutants; Asano *et al*, 2009). The molecular role of Tdrd12-family proteins is not known but data from follicle cells suggest that they might be the defining factors of Yb bodies and correspondingly nuage in the germline (Szakmary *et al*, 2009; Olivieri *et al*, 2010; Qi *et al*, 2010; Saito *et al*, 2010).

Our genetic data indicate that Yb and CG31755 are essential piRNA pathway factors in somatic ovarian cells, while CG11133 and CG31755 function in germline cells. Based on the severe impact on Piwi levels and localization in the germline upon double depletion of CG31755 and CG11133, we suggest that these two proteins are indeed the Yb counterparts in germline cells.

Taken together, our study adds functional data for four novel TUDOR proteins in the *Drosophila* piRNA pathway. It becomes increasingly clear that this small RNA silencing pathway is by far more complex than the related miRNA and siRNA pathways. The challenge for the future will be to dissect the precise molecular roles for the multitude of genetically identified factors in this fascinating genome defence pathway.

Materials and methods

Drosophila stocks

tj-GAL4 (DGRC stock 104055); *tj*-Gal4; *gypsy-lacZ* (Olivieri *et al*, 2010); MTD-GAL4 (Ni *et al*, 2011); UAS-Dcr-2; NGT (Bloomington stock 25751); *armi*[Δ 1]/TM3 (Olivieri *et al*, 2010); *fs(1)Yb[72]/FM6* (Swan *et al*, 2001). *zuc*[HM27]/CyO and *Df(2L)PRL/CyO* (Pane *et al*, 2007); *armi*-GFP and GFP-Yb (Olivieri *et al*, 2010); HP36220 (Bloomington stock 22204): P-insertion into the 5'UTR of *vreteno* (CG4771); *vreteno*[Δ 1] is a strong loss of function mutant obtained by mobilization of the HP36220 element via the Δ 2,3 transposase. Molecular analysis of this line revealed an internal deletion of the majority of the P-element without affecting the flanking genomic sequences. We speculate that the enhanced strength of the *vret*[Δ 1] allele compared with the HP36220 allele is due to loss of a cryptic promoter located within the P-element sequence similar to what has been described in Lafave and Sekelsky (2011).

GFP-Vreteno flies carry a genomic rescue construct with an EGFP cassette inserted at the N-terminus of Vreteno (obtained via bacterial recombineering) in the *vreteno*[Δ 1] genetic background.

GFP-CG11133 and GFP-CG31755 flies carry a genomic construct with an EGFP cassette inserted at the respective N-termini (bacterial recombineering).

Lines from the VDRC are listed in Supplementary Table S4.

shRNAi lines were cloned into the Valium-22 vector (Ni *et al*, 2011) modified with a white selection marker and integrated into the *atp2* landing site reported in Markstein *et al* (2008). Hairpin sequences are listed in Supplementary Table S5.

FRT-based mitotic clones were obtained by crossing *armi*[Δ 1], *zuc*[HM27], *Yb*[72], *piwi*[1] or *vret*[Δ 1] alleles with corresponding FRT insertions to respective *hsFlp*¹²²; FRT, *ubi*-GFP flies; clones were induced by heat-shocking freshly enclosed females on 2 consecutive days for 1 h at 37°C; flies were dissected 5 days later.

Antibodies

Antibodies against the N-terminal peptide of Vreteno (MESESSQDD WSAFDP) were raised in rabbits and serum was affinity purified using the same peptide. Other used antibodies were rabbit anti-Armi (peptide sequence in Cook *et al*, 2004), rabbit anti-Yb (peptide sequence in Szakmary *et al*, 2009), mouse anti-Piwi (Saito *et al*, 2006), mouse anti-Armi (Saito *et al*, 2010), rabbit anti-Piwi, anti-Aub and anti-Ago3 (peptide sequences in Brennecke *et al*, 2007).

X-Gal stainings

Ovaries from 5- to 7-day-old flies were dissected into ice-cold PBS (max 30 min), fixed in 0.5% glutaraldehyde/PBS (RT, 15 min) and washed with PBS. The staining reaction was performed with staining solution (10 mM PBS, 1 mM MgCl₂, 150 mM NaCl, 3 mM potassium ferricyanide, 3 mM potassium ferrocyanide, 0.1% Triton, 0.1% X-Gal) at RT overnight.

Immunocytochemistry

IF staining of ovaries and OSCs was as in Olivieri *et al* (2010). All primary antibodies were used at 1:500.

Cell culture

OSCs were cultured as described (Niki *et al*, 2006) and transfected with the Cell Line Nucleofector kit V (Amaxa Biosystems), selecting the programme T-029 (K Saito and M Siomi, personal communication).

Immunoprecipitation

A 15-cm dish of OSCs was harvested, cells were extracted twice with 250 μ l lysis buffer (20 mM Hepes pH 7.5; 150 mM NaCl; 10% glycerol; 1 mM EDTA; 5 mM MgCl₂; 0.5% Triton X-100, complete protease inhibitors (Roche)) and lysates were cleared for 15 min at 16 000 g at 4°C. The supernatant was split and incubated 2 h at 4°C with 20 μ g anti-Vreteno, 20 μ g anti-Armi antibody cross-linked to 100 μ l Invitrogen ProteinG Dynabeads. Empty beads were used in the control reaction. Beads were washed three times for 5 min with wash buffer I (20 mM Hepes pH 7.5; 150 mM NaCl; 10% glycerol; 1 mM EDTA; 5 mM MgCl₂; 0.2% Triton X-100), three times for 5 min with wash buffer II (20 mM Hepes pH 7.5; 150 mM NaCl; 10% glycerol; 1 mM EDTA; 5 mM MgCl₂; 0.04% Triton X-100) and eluted by boiling in 1 \times SDS sample buffer.

Small RNA cloning

Total RNA was isolated from mutant or respective heterozygous ovaries by TRIzol and phenol/chloroform extraction. Small RNA libraries were generated as previously described (Brennecke *et al*, 2007) and sequenced on an Illumina G2 platform.

Transposon QPCR analysis

cDNA was prepared via random priming of 1 μ g total RNA isolated from ovaries of 5- 7-day-old flies. Quantitative PCR was performed using Bio-Rad IQ SYBR Green Super Mix. Each experiment was performed in biological triplicates with technical duplicates. Relative RNA levels were calculated by the 2^{- $\Delta\Delta C_t$} method (Livak and Schmittgen, 2001) normalized to rp49 levels. Fold enrichments were calculated in comparison to respective RNA levels obtained from heterozygote flies or from flies not harbouring a knockdown hairpin. Primers used for QPCR are listed in Supplementary Table S6.

Sterility tests

Ten 3- to 5-day-old female flies were mated with wild-type males overnight. Females were allowed to lay eggs onto apple-juice agar plates for 18 h at 25°C. The number of laid eggs (~200 per experiment) was counted and the number of hatched eggs was counted 48 h later.

TUDOR domain searches and alignment

A subset of the FlyBase FB2010_01 proteome was generated, which included the longest protein for each protein-coding gene. The obtained set was analysed for the presence of TUDOR-clan similarities as described in the following: Each protein sequence was analysed for its presence in a Pfam24 TUDOR-clan domain definition with the aim to identify known and well-classified TUDOR-clan members. Pfamscan, HMMer2 and HMMer3 were used to query the proteome subset for similarities to models of the Pfam24 TUDOR-clan seed and full alignments. Finally, HHpred 1.5

(with PSI-BLAST homologue collection) was used for even more sensitive searches against SMART, Pfam or PDB-based models that can be associated to the Pfam-TUDOR-clan based on domain similarity. For HMMer hits to be considered, an *E*-value of ≤ 0.01 and a hit length of > 10 was required, for HHpred hits the cutoff was set to $E \leq 0.001$ and hit length > 10 .

Bioinformatics analysis

Only sequences matching the *Drosophila* genome (release 5; excluding Uextra) 100% were retained. Libraries from heterozygous flies were scaled to 1 million repeat-derived 23–30 nt small RNAs. Libraries from mutant flies were normalized to their respective heterozygous counterparts using endogenous siRNAs originating from convergent gene pairs and from long hairpin loci as in Malone *et al* (2009). For the piRNA cluster profiles, only piRNAs mapping uniquely to the respective region were considered. For the transposon profiles, all piRNAs mapping with up to three mismatches to the Repbase reference sequence (Jurka *et al*, 2005) were considered. For the calculation of fold changes in piRNA sequences (transposon analysis), only piRNAs mapping antisense to the respective element were considered. This avoids the influence of random degradation products of transposon sense transcripts that are seen for some elements in piRNA pathway mutants. These small RNA populations have a uniform size distribution and lack a clear nucleotide bias as seen for bona fide piRNAs, classifying them as likely degradation products. The ping-pong signatures for individual elements were calculated as described in Malone *et al* (2009).

References

- Anne J, Ollo R, Ephrussi A, Mechler BM (2007) Arginine methyltransferase Capsuleen is essential for methylation of spliceosomal Sm proteins and germ cell formation in *Drosophila*. *Development* **134**: 137–146
- Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, Morris P, Brownstein MJ, Kuramochi-Miyagawa S, Nakano T, Chien M, Russo JJ, Ju J, Sheridan R, Sander C, Zavolan M, Tuschl T (2006) A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442**: 203–207
- Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, Toth KF, Bestor T, Hannon GJ (2008) A piRNA pathway primed by individual transposons is linked to *de novo* DNA methylation in mice. *Mol Cell* **31**: 785–799
- Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ (2007) Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316**: 744–747
- Asano Y, Akiyama K, Tsuji T, Takahashi S, Noguchi J, Kunieda T (2009) Characterization and linkage mapping of an ENU-induced mutant mouse with defective spermatogenesis. *Exp Anim* **58**: 525–532
- Boswell RE, Mahowald AP (1985) Tudor, a gene required for assembly of the germ plasm in *Drosophila melanogaster*. *Cell* **43**: 97–104
- Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ (2007) Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128**: 1089–1103
- Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ (2008) An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* **322**: 1387–1392
- Chen C, Jin J, James DA, Adams-Cioaba MA, Park JG, Guo Y, Tenaglia E, Xu C, Gish G, Min J, Pawson T (2009) Mouse Piwi interactome identifies binding mechanism of Tdrkh Tudor domain to arginine methylated Miwi. *Proc Natl Acad Sci USA* **106**: 20336–20341
- Chen Y, Pane A, Schupbach T (2007) Cutoff and aubergine mutations result in retrotransposon upregulation and checkpoint activation in *Drosophila*. *Curr Biol* **17**: 637–642
- Chintapalli VR, Wang J, Dow JA (2007) Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nat Genet* **39**: 715–720

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (<http://www.embojournal.org>).

Acknowledgements

We thank the members of the Brennecke laboratory for helpful discussions and M Sattler for discussions on Tudor proteins. We are grateful to Sara Farina Lopez for fly injections, to Andreas Sommer for help with deep sequencing and to Mathias Madalinski for peptide synthesis and antibody purification. We thank M Siomi for antibodies and OSCs and the Bloomington stock center and the VDRC for fly lines. We thank the IMBA/IMP Bio-optics Department for excellent support. Work in the Brennecke laboratory is supported by an ERC starting grant from the European Union and a START grant from the Austrian Science Foundation. Small RNA libraries are deposited at GEO (accession no. GSE30955, data sets GSM767594 to GSM767601).

Author contributions: DH, DO and JB conceived the experiments; DH, DO and FSG conducted the experiments; MN, DH and JB performed the TUDOR analysis; KMeixner generated the *vret[Δ1]* allele; KMehtler supervised the mass spectrometry analysis; AS mapped small RNAs to transposons; RS provided the small RNA database; JB performed the computational analysis of small RNA libraries; JB and DH wrote the paper.

Conflict of interest

The authors declare that they have no conflict of interest.

- Cook HA, Koppetsch BS, Wu J, Theurkauf WE (2004) The *Drosophila* SDE3 homolog armitage is required for oskar mRNA silencing and embryonic axis specification. *Cell* **116**: 817–829
- Cote J, Richard S (2005) Tudor domains bind symmetrical dimethylated arginines. *J Biol Chem* **280**: 28476–28483
- Dietzl G, Chen D, Schnorrrer F, Su KC, Barinova Y, Fellner M, Gasser B, Kinsey K, Oppel S, Scheiblaue S, Couto A, Marra V, Keleman K, Dickson BJ (2007) A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature* **448**: 151–156
- Friberg A, Corsini L, Mourao A, Sattler M (2009) Structure and ligand binding of the extended Tudor domain of D. melanogaster Tudor-SN. *J Mol Biol* **387**: 921–934
- Gillespie DE, Berg CA (1995) Homeless is required for RNA localization in *Drosophila* oogenesis and encodes a new member of the DE-H family of RNA-dependent ATPases. *Genes Dev* **9**: 2495–2508
- Girard A, Sachidanandam R, Hannon GJ, Carmell MA (2006) A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**: 199–202
- Grimson A, Srivastava M, Fahey B, Woodcroft BJ, Chiang HR, King N, Degnan BM, Rokhsar DS, Bartel DP (2008) Early origins and evolution of microRNAs and PIWI-interacting RNAs in animals. *Nature* **455**: 1193–1197
- Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi H, Siomi MC (2007) A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* **315**: 1587–1590
- Haase AD, Fenoglio S, Muerdter F, Guzzardo PM, Czech B, Pappin DJ, Chen C, Gordon A, Hannon GJ (2010) Probing the initiation and effector phases of the somatic piRNA pathway in *Drosophila*. *Genes Dev* **24**: 2499–2504
- Haley B, Hendrix D, Trang V, Levine M (2008) A simplified miRNA-based gene silencing method for *Drosophila melanogaster*. *Dev Biol* **321**: 482–490
- Huang H, Gao Q, Peng X, Choi SY, Sarma K, Ren H, Morris AJ, Frohman MA (2011a) piRNA-associated germline nuage formation and spermatogenesis require MitoPLD profusogenic mitochondrial-surface lipid signaling. *Dev Cell* **20**: 376–387
- Huang HY, Houwing S, Kaaij LJ, Meppelink A, Redl S, Gauci S, Vos H, Draper BW, Moens CB, Burgering BM, Ladurner P, Krijgsvelde J, Berezikov E, Ketting RF (2011b) Tdrd1 acts as a molecular scaffold for Piwi proteins and piRNA targets in zebrafish. *EMBO J* **30**: 3298–3308

- Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* **110**: 462–467
- Khurana JS, Theurkauf W (2010) piRNAs, transposon silencing, and *Drosophila* germline development. *J Cell Biol* **191**: 905–913
- Kirino Y, Kim N, de Planell-Saguer M, Khandros E, Chiorean S, Klein PS, Rigoutsos I, Jongens TA, Mourelatos Z (2009) Arginine methylation of Piwi proteins catalysed by dPRMT5 is required for Ago3 and Aub stability. *Nat Cell Biol* **11**: 652–658
- Kirino Y, Vourekas A, Sayed N, de Lima Alves F, Thomson T, Lasko P, Rappsilber J, Jongens TA, Mourelatos Z (2010) Arginine methylation of Aubergine mediates Tudor binding and germ plasm localization. *RNA* **16**: 70–78
- Klattenhoff C, Bratu DP, McGinnis-Schultz N, Koppetsch BS, Cook HA, Theurkauf WE (2007) *Drosophila* rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. *Dev Cell* **12**: 45–55
- Lafave MC, Sekelsky J (2011) Transcription initiation from within P elements generates hypomorphic mutations in *Drosophila* melanogaster. *Genetics* **188**: 749–752
- Lau NC, Robine N, Martin R, Chung WJ, Niki Y, Berezikov E, Lai EC (2009) Abundant primary piRNAs, endo-siRNAs, and microRNAs in a *Drosophila* ovary cell line. *Genome Res* **19**: 1776–1785
- Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, Kingston RE (2006) Characterization of the piRNA complex from rat testes. *Science* **313**: 363–367
- Li C, Vagin VV, Lee S, Xu J, Ma S, Xi H, Seitz H, Horwich MD, Syrzycka M, Honda BM, Kittler EL, Zapp ML, Klattenhoff C, Schulz N, Theurkauf WE, Weng Z, Zamore PD (2009) Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* **137**: 509–521
- Lim AK, Kai T (2007) Unique germ-line organelle, nuage, functions to repress selfish genetic elements in *Drosophila* melanogaster. *Proc Natl Acad Sci USA* **104**: 6714–6719
- Liu H, Wang JY, Huang Y, Li Z, Gong W, Lehmann R, Xu RM (2010a) Structural basis for methylarginine-dependent recognition of Aubergine by Tudor. *Genes Dev* **24**: 1876–1881
- Liu K, Chen C, Guo Y, Lam R, Bian C, Xu C, Zhao DY, Jin J, MacKenzie F, Pawson T, Min J (2010b) Structural basis for recognition of arginine methylated Piwi proteins by the extended Tudor domain. *Proc Natl Acad Sci USA* **107**: 18398–18403
- Liu L, Qi H, Wang J, Lin H (2011) PAPI, a novel TUDOR-domain protein, complexes with AGO3, ME31B and TRAL in the nuage to silence transposition. *Development* **138**: 1863–1873
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2^{−(ΔΔC_T)} method. *Methods* **25**: 402–408
- Malone CD, Brennecke R, Dus M, Stark A, McCombie WR, Sachidanandam R, Hannon GJ (2009) Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell* **137**: 522–535
- Malone CD, Hannon GJ (2009) Small RNAs as guardians of the genome. *Cell* **136**: 656–668
- Markstein M, Pitsouli C, Villalta C, Celniker SE, Perrimon N (2008) Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. *Nat Genet* **40**: 476–483
- Maurer-Stroh S, Dickens NJ, Hughes-Davies L, Kouzarides T, Eisenhaber F, Ponting CP (2003) The Tudor domain 'Royal Family': Tudor, plant Agetet, Chromo, PWWP and MBT domains. *Trends Biochem Sci* **28**: 69–74
- Ni JQ, Zhou R, Czech B, Liu LP, Holderbaum L, Yang-Zhou D, Shim HS, Tao R, Handler D, Karpowicz P, Binari R, Booker M, Brennecke J, Perkins LA, Hannon GJ, Perrimon N (2011) A genome-scale shRNA resource for transgenic RNAi in *Drosophila*. *Nat Methods* **8**: 405–407
- Niki Y, Yamaguchi T, Mahowald AP (2006) Establishment of stable cell lines of *Drosophila* germ-line stem cells. *Proc Natl Acad Sci USA* **103**: 16325–16330
- Nishida KM, Okada TN, Kawamura T, Mituyama T, Kawamura Y, Inagaki S, Huang H, Chen D, Kodama T, Siomi H, Siomi MC (2009) Functional involvement of Tudor and dPRMT5 in the piRNA processing pathway in *Drosophila* germlines. *EMBO J* **28**: 3820–3831
- Olivieri D, Sykora MM, Sachidanandam R, Mechtler K, Brennecke J (2010) An *in vivo* RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in *Drosophila*. *EMBO J* **29**: 3301–3317
- Pane A, Wehr K, Schupbach T (2007) zucchini and squash encode two putative nucleases required for rasiRNA production in the *Drosophila* germline. *Dev Cell* **12**: 851–862
- Patil VS, Kai T (2010) Repression of retroelements in *Drosophila* Germline via piRNA pathway by the Tudor domain protein Tejas. *Curr Biol* **20**: 724–730
- Qi H, Watanabe T, Ku HY, Liu N, Zhong M, Lin H (2010) The Yb body, a major site for Piwi-associated RNA biogenesis and a gateway for Piwi expression and transport to the nucleus in somatic cells. *J Biol Chem* **286**: 3789–3797
- Reuter M, Chuma S, Tanaka T, Franz T, Stark A, Pillai RS (2009) Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile. *Nat Struct Mol Biol* **16**: 639–646
- Robine N, Lau NC, Balla S, Jin Z, Okamura K, Kuramochi-Miyagawa S, Blower MD, Lai EC (2009) A broadly conserved pathway generates 3'UTR-directed primary piRNAs. *Curr Biol* **19**: 2066–2076
- Saito K, Inagaki S, Mituyama T, Kawamura Y, Ono Y, Sakota E, Kotani H, Asai K, Siomi H, Siomi MC (2009) A regulatory circuit for piwi by the large Maf gene traffic jam in *Drosophila*. *Nature* **461**: 1296–1299
- Saito K, Ishizu H, Komai M, Kotani H, Kawamura Y, Nishida KM, Siomi H, Siomi MC (2010) Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in *Drosophila*. *Genes Dev* **24**: 2493–2498
- Saito K, Nishida KM, Mori T, Kawamura Y, Miyoshi K, Nagami T, Siomi H, Siomi MC (2006) Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev* **20**: 2214–2222
- Sarot E, Payen-Groschene G, Bucheton A, Pelisson A (2004) Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the *Drosophila* melanogaster flamenco gene. *Genetics* **166**: 1313–1321
- Selenko P, Sprangers R, Stier G, Buhler D, Fischer U, Sattler M (2001) SMN tudor domain structure and its interaction with the Sm proteins. *Nat Struct Biol* **8**: 27–31
- Senti KA, Brennecke J (2010) The piRNA pathway: a fly's perspective on the guardian of the genome. *Trends Genet* **26**: 499–509
- Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* **33** (Web Server issue): W244–W248
- Sprangers R, Groves MR, Sinning I, Sattler M (2003) High-resolution X-ray and NMR structures of the SMN Tudor domain: conformational variation in the binding site for symmetrically dimethylated arginine residues. *J Mol Biol* **327**: 507–520
- Swan A, Hijal S, Hilfiker A, Suter B (2001) Identification of new X-chromosomal genes required for *Drosophila* oogenesis and novel roles for fs(1)Yb, brainiac and dunce. *Genome Res* **11**: 67–77
- Szakmary A, Reedy M, Qi H, Lin H (2009) The Yb protein defines a novel organelle and regulates male germline stem cell self-renewal in *Drosophila* melanogaster. *J Cell Biol* **185**: 613–627
- Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD (2006) A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**: 320–324
- Vagin VV, Wohlschlegel J, Qu J, Jonsson Z, Huang X, Chuma S, Girard A, Sachidanandam R, Hannon GJ, Aravin AA (2009) Proteomic analysis of murine Piwi proteins reveals a role for arginine methylation in specifying interaction with Tudor family members. *Genes Dev* **23**: 1749–1762
- Watanabe T, Chuma S, Yamamoto Y, Kuramochi-Miyagawa S, Totoki Y, Toyoda A, Hoki Y, Fujiyama A, Shibata T, Sado T, Noce T, Nakano T, Nakatsuji N, Lin H, Sasaki H (2011) MITOPLD is a mitochondrial protein essential for nuage formation and piRNA biogenesis in the mouse germline. *Dev Cell* **20**: 364–375



Tudor.1	FBpp0071508	11	SKVDLXITHVDH-VGPYLKVG	HNVRDAASLISERINNLPTCTFAIEPSWSVERQQALLPGTFCIPKNIINGPAGDVEYRIRVVSADLE	---GQSMRAEIDFVDFGKRTVD-SHDLMPFK
Tudor.2	FBpp0071508	410	LTDDVVISYVEN-GPYLFWHL	---KSSDHLSTMGQIERTKLKAQ	---DGLYRAMVCAVYA
Tudor.3	FBpp0071508	595	NDDAVEIRFIDS-PSNIFYOK	---VANIGFEQLMDMFSTYNANRQVD	---QVIRVHVDFGQONVR-RHLGHIA
Tudor.4	FBpp0071508	1016	SKEAASLWMLS-PPQFYIYP	---KSVSAKYNDRMRREFYRQHQPL	---MRRKRVFCVTGSLTIVT-SEDIWOLE
Tudor.5	FBpp0071508	1214	SSFOALVVTAK-PRVYVQP	---QAIVPSMQTLLDNMYENKAGDSLK	---GQIIRGKFTSIRD
Tudor.6	FBpp0071508	1309	LDLCVVLSHCDN-PAQFYVHP	---IDQLSKNLHLENLOIVSPSLPOLM	---DGNWYRARIISGKDS
Tudor.7	FBpp0071508	1617	SGSCYISHVHG-IDFFIQL	---ERDSKALELLELYLKKDTPLE	---DKCMYRAKIDAE
Tudor.8	FBpp0071508	1795	VTKAIIITHVEN-TSEYIQL	---SEKSDMLDIYCEKLGSKLPKTE	---DELRYRQLKELE
Tudor.9	FBpp0071508	2163	ENSECIIISYGS-PSFYVQM	---KNSADLDIAVTKLSIKELKLI	---DACYRISIKSLVD
Tudor.10	FBpp0071508	2347	ELANCVVVQFG-PHSFYVQM	---ESDVPALQMTDKLLDAEQDLPAFS	---DGYRRAQIRKVL
Tudor.11	FBpp0082239	112	FKCVRTHENDG-PIVHVHA	---NVEAIERTVMQRCHONKMIHNE	---GDKVRGHVLTVCQ
CG9925.1	FBpp0082239	112	FKCVRTHENDG-PIVHVHA	---NVEAIERTVMQRCHONKMIHNE	---GDKVRGHVLTVCQ
CG9925.2	FBpp0082239	271	LNPEVKIVRVFK-ANP	---KHNEPQALLQINVMHVNADLNRIAKPRQFTLP	---KNGYRRAFLDLHINK
CG9925.3	FBpp0082239	461	HDSIYFISHLVS-FREYIST	---PDAKQYAEIKFLRYKCATITKSS	---DVSQGVSLIVVSK
CG9925.4	FBpp0082239	722	KVNLILANADGL-POTGYITAAFY	---KDEKAKEFEKILSTSSQACDHNVPVGY	---VPMVGLCLAIYSE
CG9684.1	FBpp0081342	136	SGAMVITDFTS-SNRCIYRD	---ASESAERAPDEVCKKVTNTATLLPRIK	---NPPNRILCLVHP
CG9684.2	FBpp0081342	477	EGIDLIVVDSTK-KNRGIFGAF	---DSTYASEFSAHSRLSEITDCEPY	---KPVRYEYVLARF
CG4771.1	FBpp0083716	322	SGSKVRITAFEQ-TNVVYVRSR	---DIQIDIAVTVLTVMLGKDKASKLAS	---TPVCGQIVLAKF
CG4771.2	FBpp0083716	524	KINNVVMDNTF-IOCGFIYCTS	---IDLAYEVTKMORDIOEYGEKIAKCATY	---APPINELCIAKY
CG14303.1	FBpp0083074	277	TDILRVRSVQS-PEDFYVQG	---IHAQRLEELDTFAHTLSDSSVPP	---TIIVGQNYIHHKD
CG14303.2	FBpp0083074	487	DVTAVQMLHVED-POEFYVNRH	---DYEKRLMLQFSLQEAADRINISQLQ	---NIFLQGLHLCVLQS
CG14303.3	FBpp0083074	722	KRTVTNIXVRK-POEFYVNRH	---HFQAIINLQSKVQKAAAMYNQMLPRT	---DMQVGMCIYARQVANC
CG14303.4	FBpp0083074	1026	SVFTAIATNVTY-ECCIVYTLA	---SDKPTIEHGNLIVREYKPLMDKOKERSTSY	---TKVQGVAVVYTHM
CG14303.5	FBpp0083074	1404	KEYCTVDNVLS-DTELQIAP	---CLSEFTKHEISLIQETSTLKDAPLEA	---EPKVGDCLVLAISR
Spindle-E.1	FBpp0082637	892	KTISGTCITVFN-PHGFQPOS	---FEECIRNISEIFNAPQOLNRIYVNAS	---AIAGMMVLAQR
Fs_1_Yb.1	FBpp0070462	825	GLIRFLVLACS-PALAVRLSDQFP	---TAIRFLNFMPSDLGERVORHYELEANRHHMP	---NPVPGEMAVVKN
CG11133.1	FBpp0078210	520	GDVKVLQVKYS-PHFVRLLEHLP	---KGTWRMEYSAVEQFRMQLTQIKPRRYM	---PPVAGAICMVHT
CG31755.1	FBpp0079562	22	QESILITHFVN-PHGFVSVR	---CIDVENSAMLRQLEQDLKDCSSERTKO	---VVEDERIVVYQRWS
CG31755.2	FBpp0079562	930	TLIRCKFLKAYD-PAHMAMVPMKY	---KSKDSTIMIDVPYSPSTPLVFRMSGVPKRVHNN	---PKINDVCFVQH
Tejas.1	FBpp0086689	366	QIRIRQVLSLVN-PHNFPMIY	---NDDFKDYEAQFANMOTFYESSKNTYMPLF	---LITTDHLCVVRC
CG8920.1	FBpp0085592	558	HDNMFIISFCDs-KTIVNARM	---IDQIANFEELTKHIGROMESPFRKOVKS	---KPYAQEVILVEM
CG8920.2	FBpp0085592	769	KTPPLVSHIND-DGDLAVLL	---RNDDLKTVERSAQFVADLGEQDRVSYS	---DLLHDRHIFVCDETVDGV
CG8920.3	FBpp0085592	1034	AYFEVRVALSVN-PGHFAVQ	---PYINQLOLTKMLNQHCHCKTAARKQVPS	---QALGAENYAPADS
Krimper.1	FBpp0086333	318	TIVTAVLASVDV-TDNCAYVAK	---WDESSDRICKKQLQRLPLQELDQ	---LPDYGFIFAVLDSI
Krimper.2	FBpp0086333	570	SIVGILITFING-PTEVIGQFLD	---GSPPLVWDKQVPEKNKRTFKS	---KPRLDITVALYS
Papi.1	FBpp0077508	259	KPMEVYVSASVS-PKFKVQLI	---GPQSKLDSMVQETSYSYSSAENRAKHULT	---APVVGQIVAAVFKF
Yu.1	FBpp0070761	417	INNDDVVSASVS-GSHIFIQHP	---HPSHPSLPLLOKQIDYSYSTHEAPLLP	---SLELSACVCIPI
Tudor-SN.1	FBpp0072419	701	NYENVIUTEITE-TLTPFAQSV	---ESGSKLESMLKSHADQSNPIAGSY	---TPKRGDLVAAQFTL
CG15042.1	FBpp0074390	74	IGTKLTGTVLLE-SLPVFTVINGPGS	---KLKCKLMSQVLOLEQ	---LPDYGFIFAFYDKA
CG15042.2	FBpp0074390	266	SIVRVHVTIRVS-HAEFYARFADGPT	---VPTSHKSVMKSGTG	---DFRVMDIVLAPY
CG15930.1	FBpp0291095	285	SIFPIIMSCVFS-PCFEPHIVPPQY	---AKNPVAENTIDLNWYRHTTISYRAELP	---SYFKEGICIAAYS
CG13472.1	FBpp0075441	729	CG13472.1	---AMMQKQGLCMAYMD	---DGYIEAEITGVSE
San.1	FBpp0075153	67	CG13472.1	---PVSKVGVGYARATYV	---DGYIEAEITGVSE
CG17454.1	FBpp0289146	89	CG17454.1	---AKINKKGDCKQAKKE	---DROXYDATIEDISS
CG3251.1	FBpp0077113	300	CG3251.1	---DYNFKVGAQKQVOLDNHR	---RDLISACYIQSIDK
otu.1	FBpp0071181	334	1.....10.....20.....30.....40.....50.....60.....70.....80.....90.....100.....110.....120.....130.....	---DYNFKVGAQKQVOLDNHR	---RDLISACYIQSIDK

C

B

A

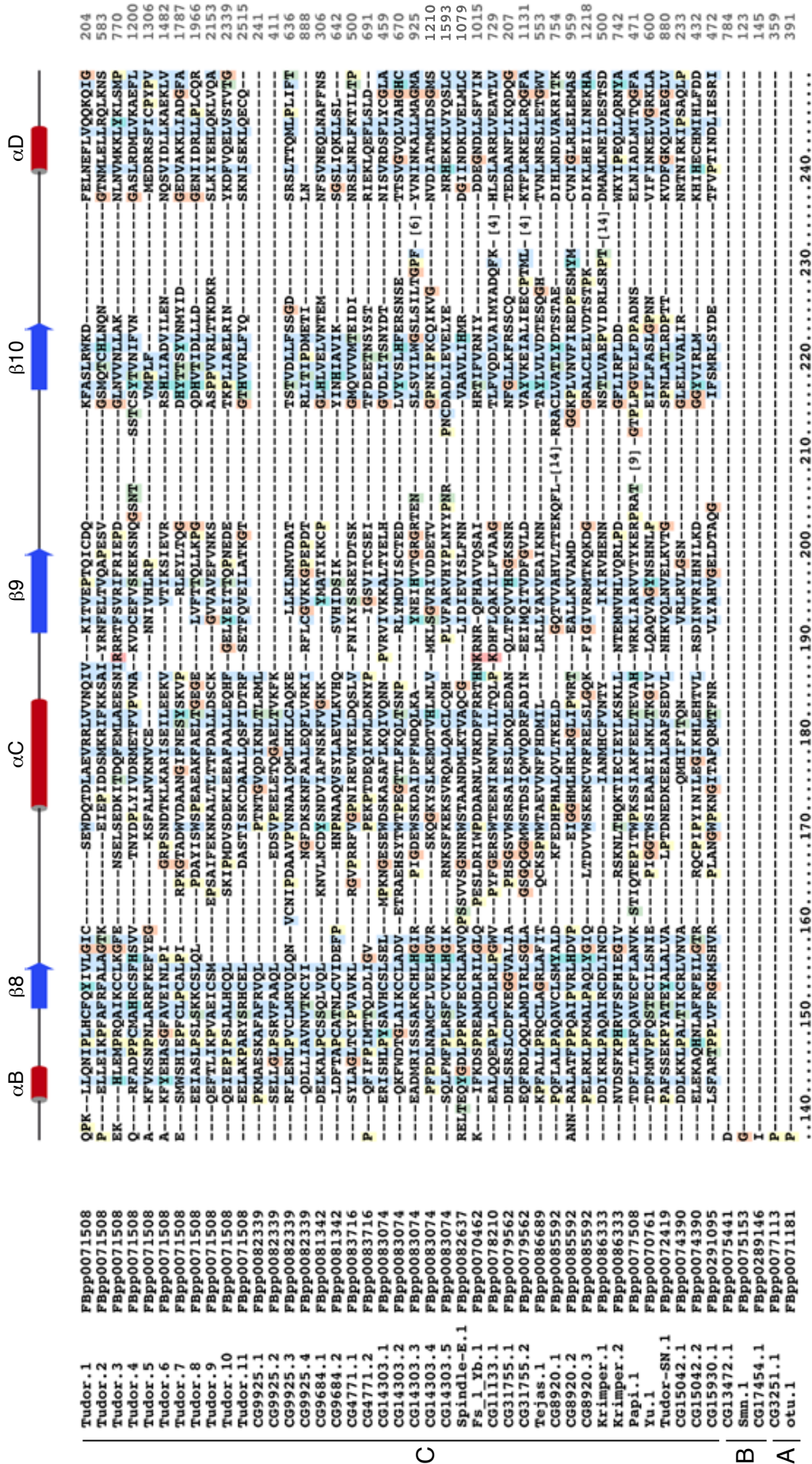


Figure S1. Alignment of all *Drosophila melanogaster* TUDOR and SMN domains including N- and C-terminal extensions

At the top, the secondary structure elements (blue: β -sheets; red: α -helices) as defined by Liu et al. (2011a) are depicted with β -strands in blue and α -helices in red. The alignment covers the entire extended TUDOR domain, including the N- and C-terminal extensions. Start and stop positions of each domain in the respective host protein are indicated at the beginning and at the end of the alignment as well as the corresponding FlyBase Protein ID. Residues were color coded using JalView (color option: ClustalX). To the left, the three groups (A: Otu group; B: Smn group; C: Tudor-extended group) are indicated.

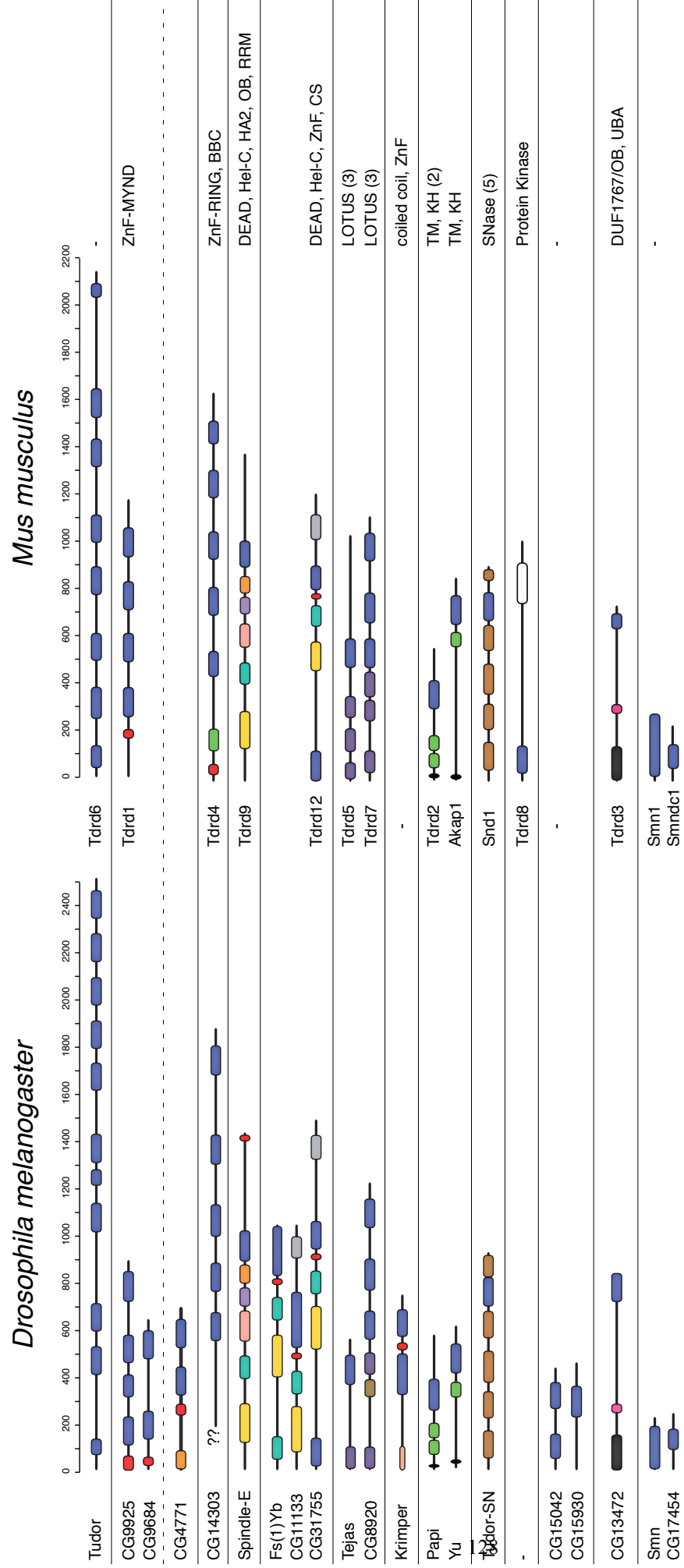


Figure S2. The Fly and Mouse Tudor domain containing proteins

Cartoon depicting all *Drosophila melanogaster* proteins containing Tudor/SMN domains (blue boxes) and their corresponding mouse counterparts. All other significant protein domains identified via HHPRED searches are indicated with colored boxes and their identity (mouse proteins only) is given to the right from N to C-terminus (ZnF: zinc finger; BBC: B-Box C-terminal domain; OB: Oligo-Nucleotide binding domain; RRM: RNA recognition motif; DEAD: DEAD Box RNA Helicase domain; Hel-C: Helicase C-term. domain; HA2: Helicase associated domain; Hydrol.: Hydrolase; CS: HSP20-like domain; TM: transmembrane domain; KH: K homology; SNase: Staphylococcus nuclease; DUF: domain of unknown function; UBA: ubiquitin associated domain). For CG14303, the “??” indicate the non-mapped N-terminus. The scale indicates amino acid positions. Based on the additional contained domains, the proteins were grouped into families, separated by horizontal lines.

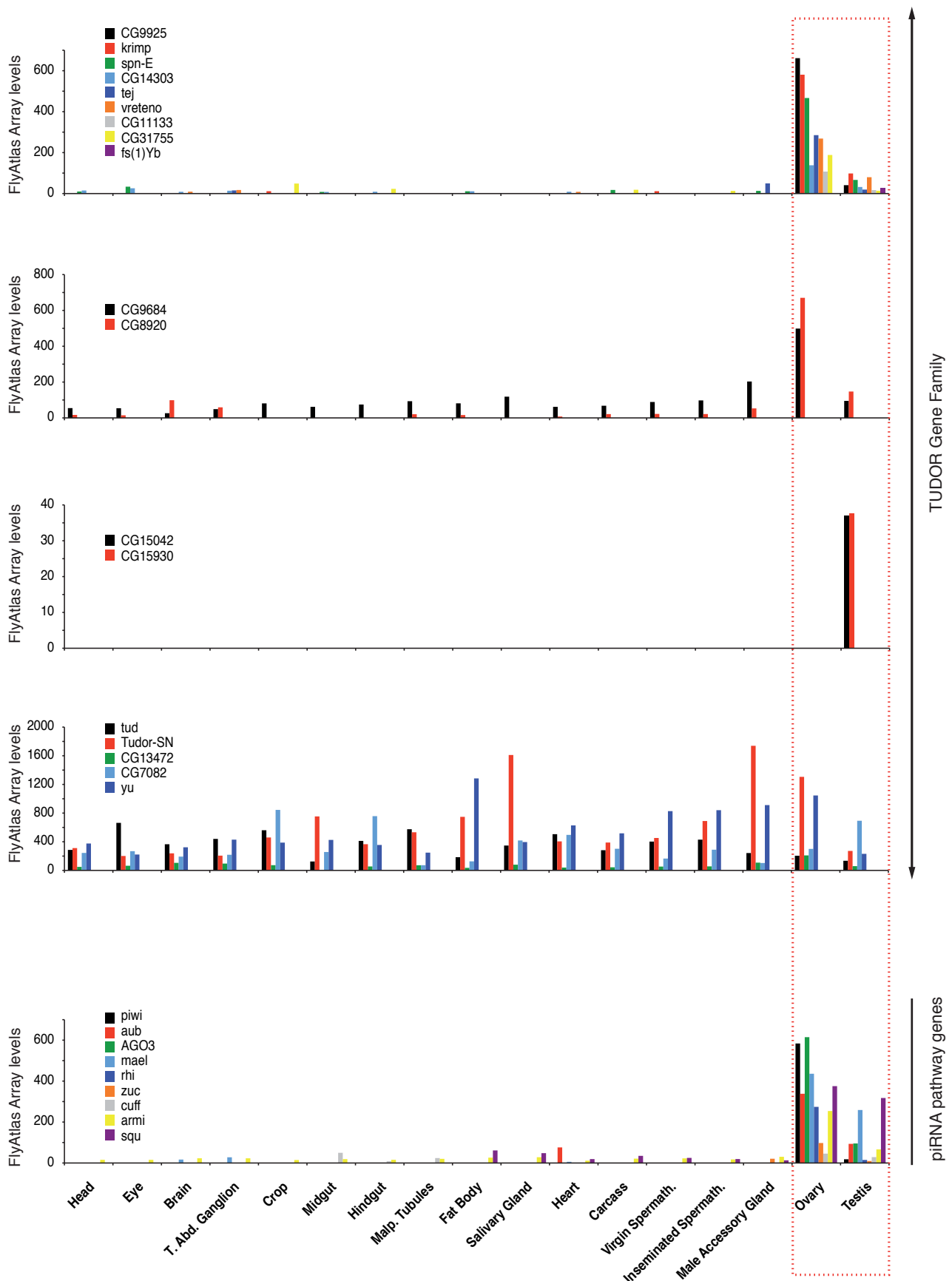


Figure S3. Adult Expression Patterns of TUDOR and piRNA pathway genes in *Drosophila*

Bar Diagrams depicting the detected expression levels of the indicated genes (color coded; legend given for each diagram) in the indicated tissues (bottom). All values are based on data from the adult FlyAtlas (www.flyatlas.org). TUDOR genes were split into gonad specific (first), gonad enriched (second), testes specific (third) and ubiquitous (fourth).

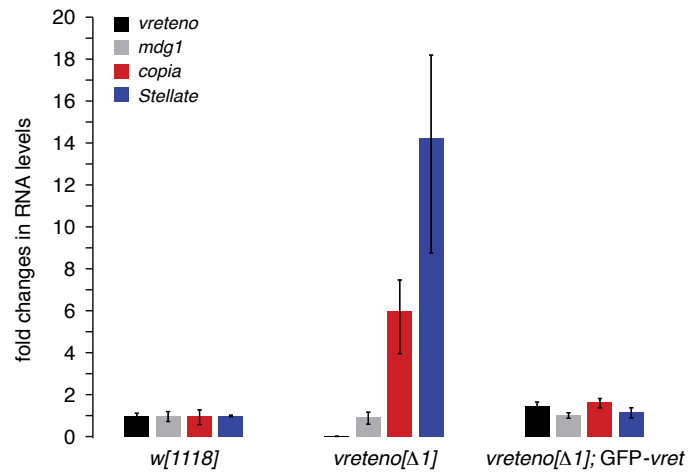
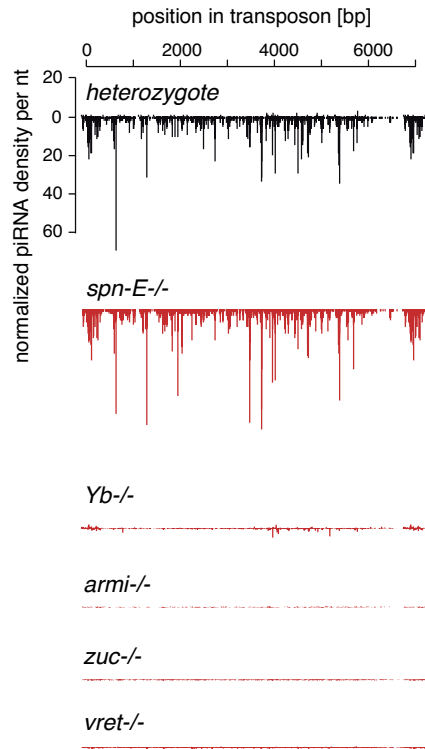


Figure S4. *vreteno* is required for efficient transposon and *Stellate* silencing in testes

Shown are changes in testes steady state RNA levels (normalized to *w[1118]* controls) of *vreteno*, the transposons *mdg1* and *copia* as well as of *Stellate* in *vreteno[Δ1]* flies compared to *vreteno[Δ1]* flies expressing a GFP-*vreteno* rescue construct (n=3; error bars indicate St. dev).

Tabor piRNA profile



ZAM piRNA profile

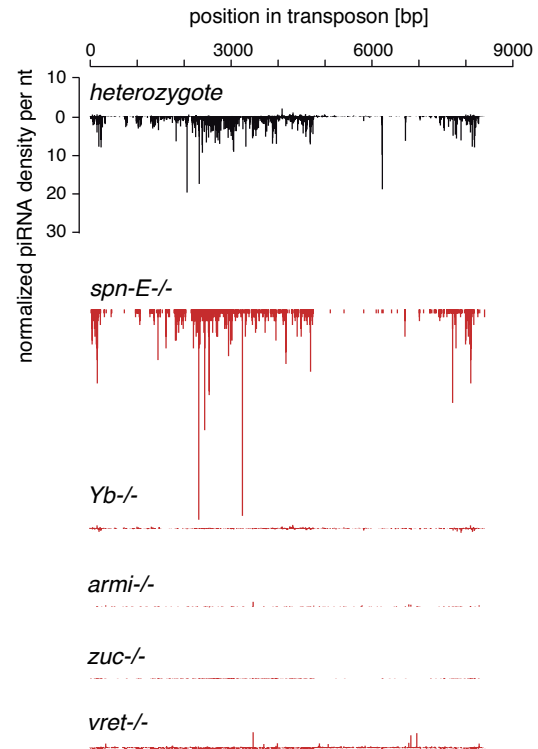


Figure S5. piRNA populations of somatic transposons collapse in *vreteno* mutants

Shown are normalized piRNA profiles obtained from heterozygote ovaries (black) in comparison to profiles obtained from indicated mutant ovaries (red) mapping to the soma dominant transposons *Tabor* and *ZAM* (sense and antisense piRNAs are indicated with peaks pointing up- and downwards).

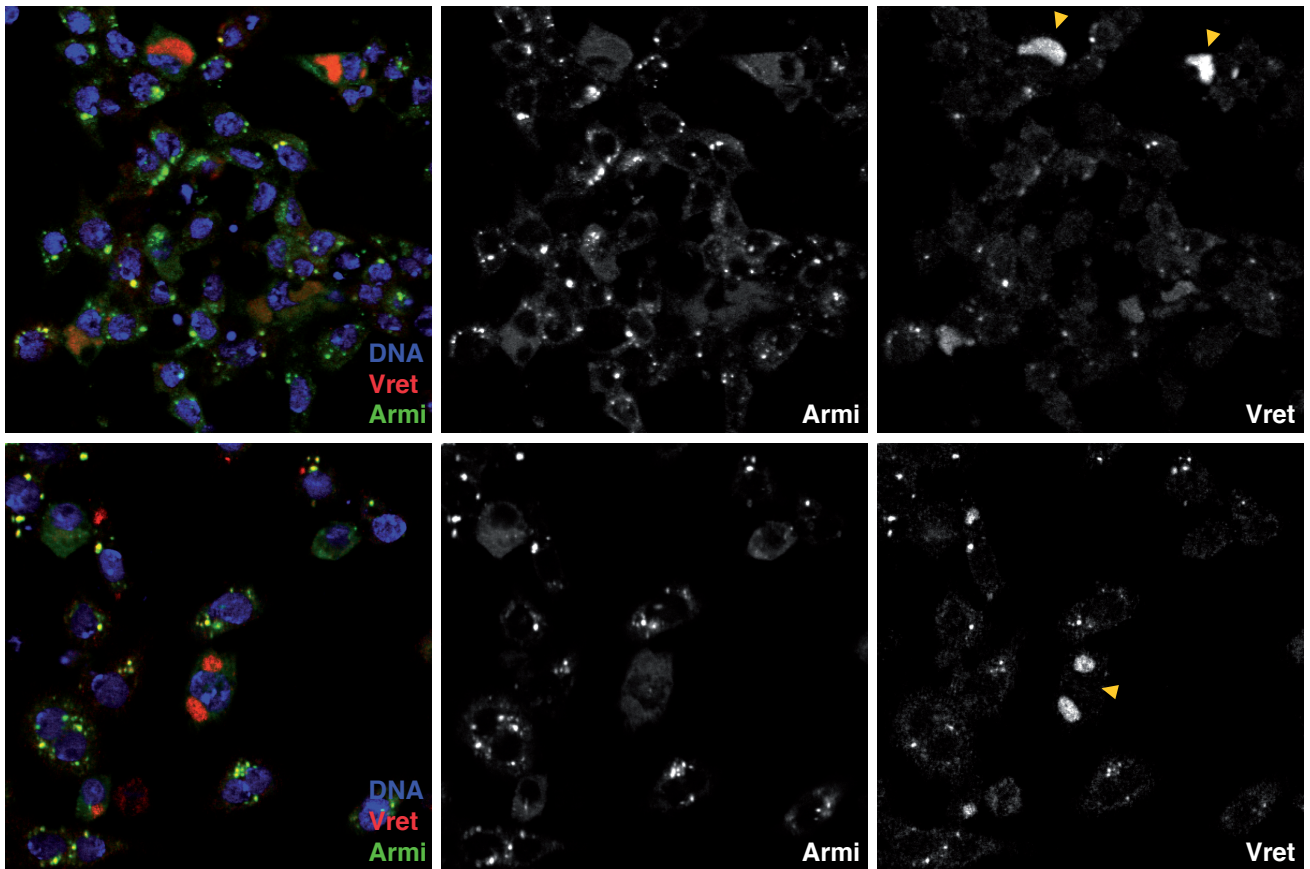


Figure S6. Vreteno localizes to Armi-foci in cultured somatic follicle stem cells

Shown are immunofluorescence images of fixed OSCs stained for Vreteno (red), Armitage (green) and DNA (blue). Two representative images are shown. Yellow arrowheads in the monochromatic Vreteno channel highlight the cells in which Yb-bodies have apparently dispersed, resulting in loss of Armitage foci and large cytoplasmic accumulations of Vreteno.

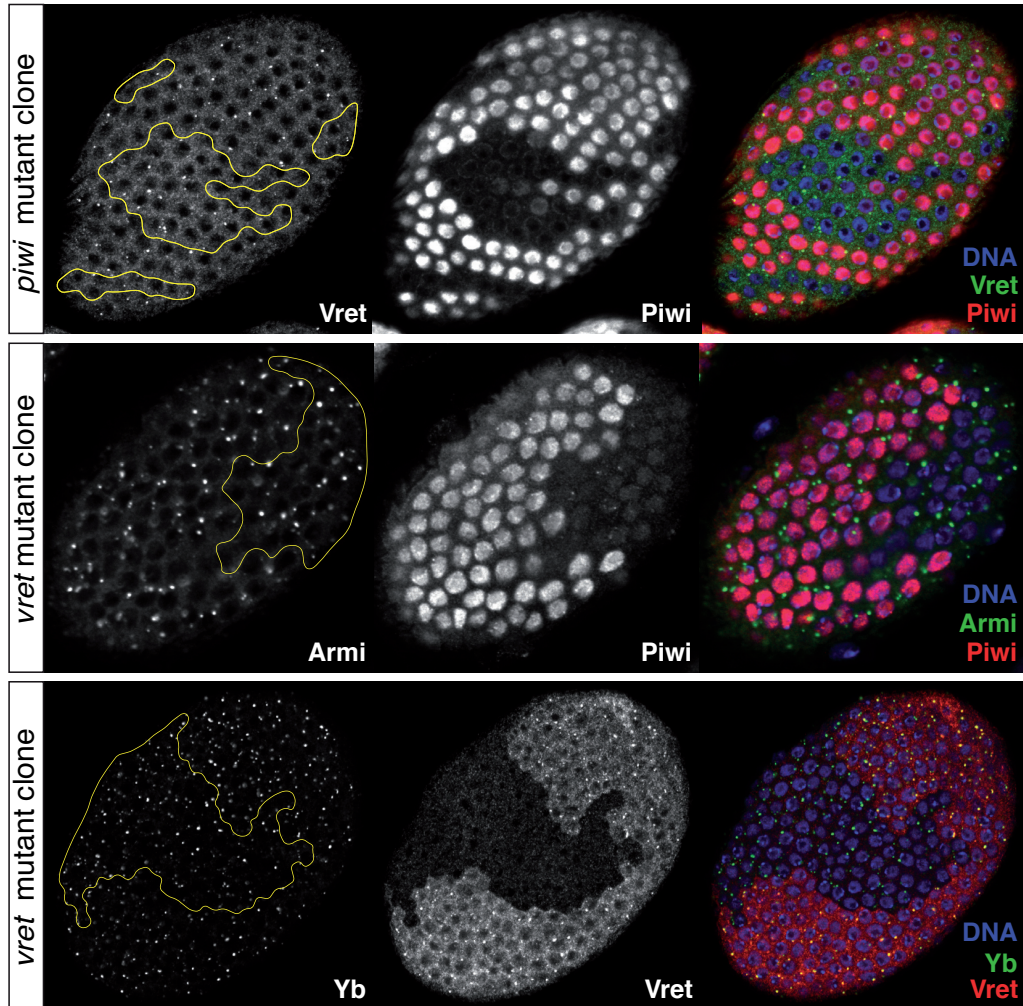


Figure S7. Vreteno localization is independent of Piwi and its levels do not affect Yb-body formation.

Immunostainings of Vret, Armi, Yb, Piwi and DNA (blue) in egg chambers, where clones of cells mutant for the indicated genes (left) have been induced in the follicular epithelium (clone borders indicated by a yellow line). The right panels show the merge of all three channels, the monochrome panels the individual channels as indicated.

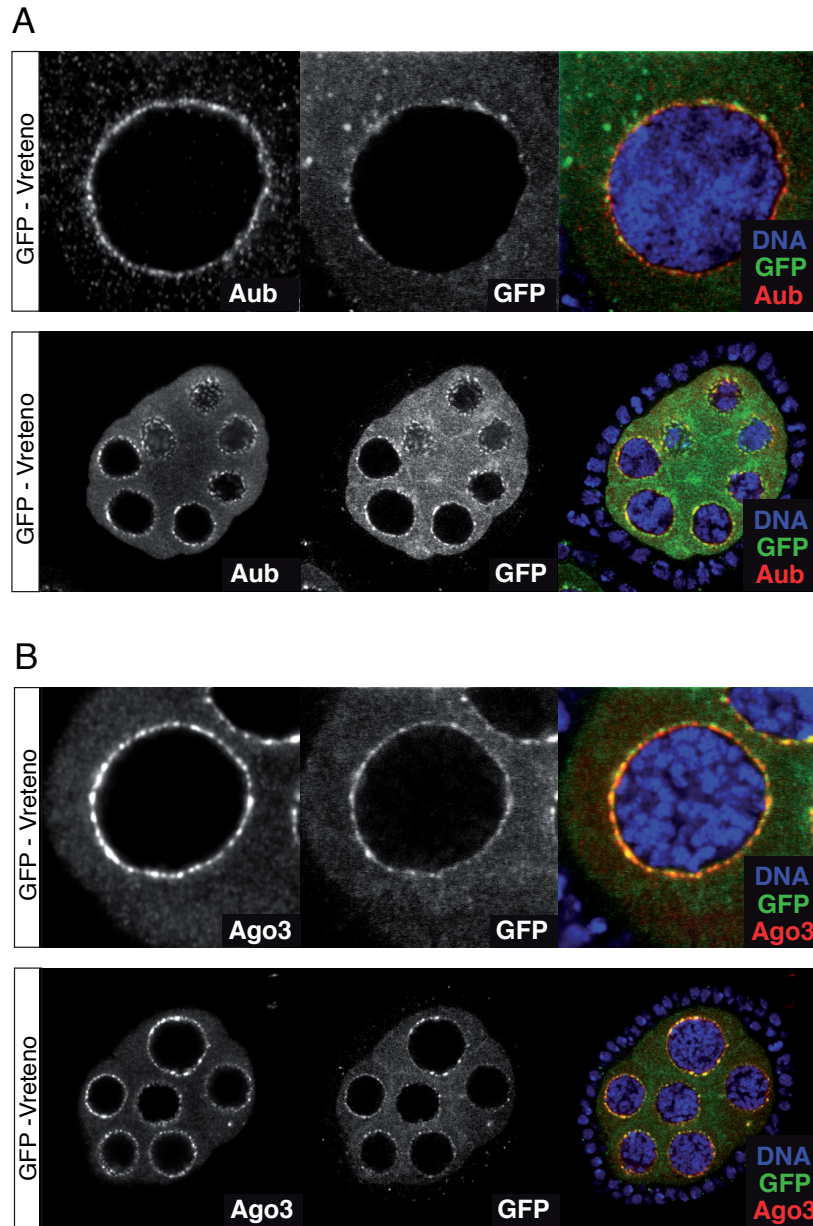


Figure S8. Vreteno localizes to nuage in nurse cells in close proximity to Aubergine and Ago3.

(A) Immunostainings of GFP-Vret (green), Aubergine (red) and DNA (blue) in an entire egg chamber (lower panel) or focused on an individual nurse cell nucleus (upper panels). The right panels show the merge of all three channels where co-localization results in yellow. (B) Immunostainings of GFP-Vret (green), Ago3(red) and DNA (blue) in an entire egg chamber (lower panel) or focused on an individual nurse cell nucleus (upper panels). The right panels show the merge of all three channels where co-localization results in yellow.

Supplementary Table 1

gene	label	name	tudor_clan_domains
FBgn0003044	Pcl	Polycomblake	53-BP1_Tudor
FBgn0250818	CG34360		53-BP1_Tudor
FBgn0038780	CG5060		53-BP1_Tudor
FBgn0051151	wge	winged eye	53-BP1_Tudor
FBgn0023509	mip130	Myb-interacting protein 130	53-BP1_Tudor
FBgn0050390	CG30390		53-BP1_Tudor,DUF1325
FBgn0032169	CG4709		53-BP1_Tudor,SMN
FBgn0028734	Fmr1	Fmr1	Agenet,53-BP1_Tudor
FBgn0034255	CG18186		Chromo
FBgn0030854	CG8289		Chromo
FBgn0044324	Chro	Chromator	Chromo
FBgn0023395	Chd3	Chd3	Chromo
FBgn0013591	Mi-2	Mi-2	Chromo
FBgn0086902	kis	kismet	Chromo
FBgn0250786	Chd1	Chromodomain-helicase-DNA-binding protein 1	Chromo
FBgn0003042	Pc	Polycomb	Chromo
FBgn0003600	Su(var)3-9	Suppressor of variegation 3-9	Chromo
FBgn0028965	A16	A16	Chromo
FBgn0027914	Gen	XPG-like endonuclease	Chromo
FBgn0031613	HP6	Heterochromatin protein 6	Chromo_shadow
FBgn0259922	CG42448		Chromo_shadow
FBgn0030994	CG14193		Chromo_shadow
FBgn0003607	Su(var)205	Suppressor of variegation 205	Chromo_shadow,Chromo
FBgn0030082	HP1b	HP1b	Chromo_shadow,Chromo
FBgn0037675	HP1e	HP1e	Chromo_shadow,Chromo
FBgn0039019	HP1c	HP1c	Chromo_shadow,Chromo
FBgn0004400	rhi	rhino	Chromo_shadow,Chromo
FBgn0032475	Sfmbt	Scm-related gene containing four mbt domains	MBT
FBgn0003334	Scm	Sex comb on midleg	MBT
FBgn0038016	MBD-R2	MBD-R2	MBT,53-BP1_Tudor
FBgn0086908	egg	eggless	MBT,53-BP1_Tudor
FBgn0002441	l(3)mbt	lethal (3) malignant brain tumor	MBT,Tudor-knot
FBgn0039743	CG7946		PWWP
FBgn0033155	CG1845		PWWP
FBgn0043456	CG4747		PWWP
FBgn0031023	CG14200		PWWP
FBgn0038190	CG9926		PWWP
FBgn0039559	Mes-4	Mes-4	PWWP
FBgn0016754	sba	six-banded	PWWP
FBgn0033752	CG8569		PWWP
FBgn0039863	CG1815		PWWP
FBgn0036882	CG9279		PWWP
FBgn0020503	CLIP-190	Cytoplasmic linker protein 190	PWWP
FBgn0085451	CG34422		PWWP,53-BP1_Tudor,Tudor-knot,Chromo
FBgn0011785	BRWD3	BRWD3	SMN
FBgn0036467	CG12310		SMN
FBgn0051268	CG31268		SMN
FBgn0259241	CG42339		SMN
FBgn0038191	CG9925		TUDOR
FBgn0037583	CG9684		TUDOR
FBgn0039018	CG4771		TUDOR
FBgn0033921	CG8589		TUDOR
FBgn0031401	CG7082		TUDOR
FBgn0029764	yu	yu	TUDOR
FBgn0035121	Tudor-SN	Tudor-SN	TUDOR
FBgn0003483	spn-E	spindle E	TUDOR
FBgn0029754	CG15930		TUDOR
FBgn0034098	krimp	krimper	TUDOR
FBgn0030937	CG15042		TUDOR
FBgn0037205	CG11133		TUDOR
FBgn0051755	CG31755		TUDOR
FBgn0000928	fs(1)Yb	female sterile (1) Yb	TUDOR
FBgn0003891	tud	tudor	TUDOR
FBgn0038633	CG14303		TUDOR
FBgn0027529	CG8920		TUDOR

FBgn0003023	otu	ovarian tumor	TUDOR
FBgn0031622	CG3251		TUDOR
FBgn0036450	CG13472		TUDOR,SMN
FBgn0039977	CG17454		TUDOR,SMN
FBgn0036641	Smn	survival motor neuron	TUDOR,SMN
FBgn0034975	enok	enoki mushroom	Tudor-knot
FBgn0026080	Tip60	Tip60	Tudor-knot
FBgn0014340	mof	males absent on the first	Tudor-knot
FBgn0027378	MRG15	MRG15	Tudor-knot
FBgn0002775	msl-3	male-specific lethal 3	Tudor-knot
FBgn0039585	CG1894		Tudor-knot

Supplementary Table 2A

this table lists the raw counts of 23-30mer small RNAs that map with up to 3 mismatches to the indicated transposons (from Repbase) in antisense orientation.

#TP	ZucMut	ZucHet	ArmiMut	ArmiHet	Fs1YbMut	Fs1YbHet	VretMut	VretHet
GYPSY6A_LTR	97	2521	42	1855	947	2143	802	4476
DM412B_LTR	1378	32722	1527	48441	6182	19121	2035	58834
STALKER3_LTR	34	1880	12	1727	518	1024	139	4206
GYPSY12A_LTR	17	438	6	449	346	515	27	510
DMTOM1_LTR	3	56	1	63	31	41	14	77
BATUMI	43581	108401	21970	108450	105682	161040	59856	128518
DMSAT6	1	12	3	11	13	11	1	10
ACCORD	799	28448	840	40484	27183	34841	3159	62863
TLD3	19	230	14	122	74	95	34	209
GTWIN	1587	75721	136	92222	4748	19474	4657	168692
PEN2	0	0	0	0	1	0	0	0
INVADER3	3574	35224	8770	23828	43761	51587	1303	13088
HETRP	4144	29908	3852	16439	15304	73977	4033	11782
TART	2566	116067	2608	91599	112408	142211	8669	95882
LTRMDG3	445	5033	711	2440	3436	7542	333	2154
PROTOP_A	23501	22911	10835	22749	11919	20057	30847	28848
QUASIMODO2-I	571	44515	164	11247	9821	9220	1013	14357
GYPSY11	13	176	2	136	77	154	29	264
MARINER2	2	49	0	22	31	32	10	33
DOC2	2010	28340	1331	16483	22688	18080	2431	23040
COPIA2LTR	335	609	142	711	558	1219	1954	1814
S	2943	9717	1068	5354	3091	4870	848	8944
STALKER4	60984	309121	34776	363296	143910	417014	87495	322716
PEN1	0	0	0	0	1	0	0	0
DIVER2	319	8386	198	7468	6320	5635	1207	12167
SAR	204	96	73	225	79	316	187	141
TC1-2	118	4768	42	4418	3036	3642	151	2597
ACCORD2	487	29104	192	20781	18556	23209	1153	33485
FROGGER	5	204	2	116	106	137	18	198
INVADER6	2659	65074	884	45323	24189	49730	2238	52300
MDG3	10989	41569	8276	19762	41292	60944	1387	20294
DNAREP1	55	3756	19	2361	1413	2200	169	2211
ROO	77745	412727	43983	405566	312266	480554	96228	317238
Transib5	97	2795	43	2269	3349	3780	380	3942
DMRPR	0	0	0	0	0	0	0	0
INVADER4	576	9983	1112	9543	10593	9650	778	9301
DMRT1C	22	1524	7	709	777	913	49	1888
MDG1	7031	155867	4158	162877	29219	118023	18868	286447
Transib1	46	3039	17	1579	783	1547	67	2085
HELENA_RT	26	740	9	1052	672	954	178	1140
RSP	2874	695	76	96	265	274	552	650
LOOPER1	69	4492	23	2781	1472	2574	189	2943
S2	7	92	2	86	64	94	19	161
DIVER	541	34067	734	23497	17485	11185	1039	24270
LINEJ1	1413	61586	606	37437	36915	42340	2625	47357
ALA	0	0	0	0	0	0	0	0
FW3	26	389	22	375	341	390	81	586
DOC3	4336	40993	2663	30692	35865	36331	5447	34662
DM412	8321	270760	3328	337464	39605	187076	18113	469794
INVADER1	4015	21972	4586	36423	31300	50284	7957	22514
INVADER5	4	48	2	33	63	52	5	40
TLD1	0	11	0	27	19	19	0	25
G7	3	51	1	43	43	38	4	33
PEN4	0	0	0	0	0	0	0	0
Transib4	5	65	2	40	25	43	7	78
GYPSY2	478	30560	189	20703	18139	24554	1639	37980
JOCKEY2	592	2568	821	3243	1251	2585	556	4014
DM1731	1342	24641	4106	31713	30464	37698	1123	24255
XDMR	27	218	6	143	51	86	46	75
ZAM	453	10528	124	9261	747	5798	821	16933
TC1	8	388	3	393	162	251	17	447

M4DM	1007	1344	557	2107	1899	2887	1223	2729
QUASIMODO2-LTR	189	3476	139	2191	2000	2519	228	2899
DM297	38812	213721	7138	288019	162021	380655	34943	333093
BURDOCK	728	46273	296	26774	26904	30477	2974	40870
BS	1345	53920	359	41004	34089	50471	2041	44540
COPIA	9499	15762	20532	44219	28185	60125	7054	36067
PROTOP_B	15078	16023	6992	14876	8962	13893	22175	18416
TAHRE	1879	61358	746	36881	30077	47691	6976	62452
DOC6	1670	31487	1016	22146	8515	14824	1250	19501
G5	109	5041	74	5214	10160	7148	286	7928
FW	180343	332458	127402	308490	204087	235846	149840	302042
DMRT1A	3465	165599	1379	131193	135617	157803	13404	143919
TLD2	11	704	1	470	134	326	39	737
DOC4	16	732	2	377	451	448	81	967
DOC5	86	4029	39	3881	3513	3159	272	5446
ROVER	1143	38223	1270	57502	11307	28729	4184	97115
DMCR1A	33215	74290	20850	78918	59635	80445	41028	83065
TABOR	3126	142029	915	66046	5252	125609	7686	94074
MICROPIA	6724	24440	6376	29067	40087	63692	3268	22360
ROOA	1305	75828	493	53944	38370	65337	2568	46756
DMHMR2	623	12227	152	5104	4653	6486	1860	8670
R1-2	7	215	4	155	187	160	16	179
GYPSY8	221	19327	49	5606	6444	6342	282	10472
ARS406	1785	670	1139	441	343	1736	893	755
IVK	1101	38364	761	24938	27568	30862	1761	25968
MAX	72102	286753	16063	212047	195127	233856	49357	206279
BS2	2389	101346	807	82456	79428	100581	6285	87017
PROTOP	43793	61404	20580	52411	44750	82868	39255	58297
BLOOD	2604	111437	3298	97950	241547	162272	5883	119228
FB4	546	490	169	420	383	1891	1463	1724
G4	181	9798	81	8670	5024	7596	621	8628
I	635	40335	596	25739	23111	20342	1933	34365
BARI	25306	18046	8467	16634	16929	10421	7619	26591
DMHMR1	14	222	10	193	242	160	145	567
COPIA2I	6254	15085	2494	11491	12077	10348	5660	12619
FUSHI	0	0	0	0	0	0	0	0
ALUII	3	4	3	5	4	5	5	6
BAGGINS1	5790	55230	5015	33337	36033	38331	8808	39345
QUASIMODO	1654	29995	5531	55306	52596	95512	6745	49666
G5A	21	925	5	676	704	772	51	1084
POGON1	75	309	50	373	199	312	65	460
Transib-N1	0	2	0	5	3	4	0	3
G2	2930	42007	3063	50130	31785	33079	5972	50000
DMRP1	0	0	0	1	0	0	0	2
IDEFIX	645	22790	1098	20493	23656	28790	2355	29259
R2	350	29321	129	9711	9577	8746	628	15087
TRANSPAC	281	22259	109	10662	7192	8320	882	12035
BLASTOPIA	3307	46589	5737	48428	74833	104142	1454	29453
DM176	22632	71658	18910	95940	66147	61209	13913	113089
GYPSY3	323	23268	153	13268	11213	22802	1372	18623
TRANSIB2	3545	18219	1314	12845	16653	20695	1178	13107
NTS	6222	22425	2599	3855	2763	5786	3266	7179
GYPSY9	3	308	3	245	216	212	16	234
GYPSY12	1742	41350	2161	39436	27541	52156	4632	49803
DOC	219113	390869	75007	190911	208756	260108	129523	193104
G6	1144	35443	714	20462	74708	85817	1174	22809
GYPSY7	58	3849	18	1806	1615	1858	98	1611
TART_B1	1196	70136	2064	45495	184495	231569	5617	84953
GYPSY	826	46772	368	22393	18658	19784	2323	34844
NOMAD	41353	309011	2538	89759	155151	156313	16040	89768
R1	51022	286739	40397	116622	166842	180826	48810	189292
BS4	0	19	1	9	17	21	0	10
HETA	4778	94328	1307	70882	26633	53380	15765	79338
G	975	64037	2488	45424	47731	39336	12943	79748
FW2	14	448	12	385	485	517	70	467
POGO	7125	13818	3630	16938	8288	12200	1367	12117
STALKER2	14139	66003	7194	75972	35137	88825	20186	92249

BEL	983	49215	3024	35552	53716	42046	2220	34789
PLACW	1	4	0	5	5	7	3	12
GYPSY6	1290	40096	543	26140	20379	20652	3901	41936
HOBO	121	1748	77	2245	4393	5535	366	2070
INVADER2	8534	42497	11419	31943	22462	21460	2741	34058
CIRCE	2844	24163	4740	36527	29734	63183	11952	38875
DMRT1B	5150	271702	2493	188024	168162	193999	20495	199542
GYPSY5	974	63388	97	37495	1316	55096	3512	62182
G3	4	457	4	298	348	450	50	421
TRANSIB3	3	160	1	242	60	92	12	440
GYPSY10	190	11350	25	8966	1676	8190	592	14464
DMLTR5	192	1693	202	3700	3978	8193	129	2039
FTZ	0	0	0	0	0	0	0	0
HMSBEAGLE_I	6455	36692	14892	57164	96395	125998	2642	33230
GYPSY4	2501	59095	2335	57480	33732	36478	4628	68495
HELITRON1	0	4	0	5	7	12	1	10
BS3	51	3221	32	1992	3190	3304	99	2846
TIRANT	8	1125	1	106	65	79	441	133

Supplementary Table 2B

all heterozygote libraries were scaled to 1 million repeat derived 23-30mer small RNAs
each mutant library was normalized too its respective net library via non-TIE derived endogenous siRNAs (see Malone et al. 2009)
the spn-E data was taken from Malone et al. 2009
elements colored in red are germline dominant, in yellow are intermediate and in green are soma dominant (according to Malone et al. 2009)

#TP	ZucMut	Zuchet	zuc_log 2 ratio	ArmiMut	ArmiHet	armi_log 2 ratio	FstYbMut	FstYbHet	yb_log2_ratio	VretMut	VretHet	vret-log2_ratio	spnE-log2_ratio
PROTOP_A	680	1413	1.054244095	927	1593	0.781545759	715	1232	0.78488102	1254	1712	0.449316679	4.10
HETA	138	5816	5.394133133	112	4962	5.472530049	1598	3279	1.037122746	641	4707	2.877262425	6.75
NOMAD	1197	19052	3.992520108	217	6284	4.855734413	9309	9603	0.044803158	652	5326	3.030502676	5.01
DOC	6342	24099	1.925935899	6414	13365	1.059244869	12525	15979	0.351331495	5263	11457	1.122145307	4.23
DMRT1B	149	16752	6.812234755	213	13163	5.948331915	10090	11918	0.240235869	833	11839	3.82932411	6.33
TAHRE	54	3783	6.120140299	64	2582	5.339000117	1805	2930	0.699094746	283	3705	3.708251394	6.38
GG_DM	33	2185	6.044267578	61	1433	4.552321063	4482	5272	0.234039114	48	1353	4.826074449	4.16
R1_DM	1477	17679	3.581472597	3454	8165	1.24096169	10010	11108	0.150158005	1983	11231	2.501340348	4.86
FW_DM	5220	20498	1.97335448	10894	21597	0.987277582	12245	14488	0.242699231	6089	17920	1.557301936	4.00
TART	74	7156	6.590218677	223	6413	4.845757857	6744	8736	0.373326733	352	5689	4.013298079	4.97
L_DM	18	2487	7.080057666	51	1802	5.143941647	1387	1250	-0.150079769	79	2039	4.698001982	0.93
BATUMI	1261	6684	2.405532672	1879	7592	2.014865062	6341	9893	0.641727789	2432	7625	1.648378224	6.16
MAX	2087	17680	3.082623209	1374	14845	3.434012588	11707	14366	0.295245411	2006	12238	2.609245909	6.24
BEL	28	3034	6.73668868	259	2489	3.266842461	3223	2583	-0.319344933	90	2064	4.515975193	6.07
DM1731	39	1519	5.289530042	351	2220	2.660710612	1828	2316	0.341420989	46	1439	4.978827912	5.53
B52	69	6249	6.497664297	69	5773	6.386353777	4766	6179	0.374676607	255	5163	4.337288195	6.58
DM297	1123	13177	3.552079802	610	20164	5.045942059	9721	23384	1.266341598	1420	19762	3.798825221	3.14
INVADER4	17	616	5.206258522	95	668	2.812727868	636	593	-0.100471993	32	552	4.125519341	4.39
ROO	2250	25447	3.499291833	3761	28393	2.916360579	18736	29521	0.655961448	3910	18821	2.267012509	4.07
DM176	655	4418	2.753689509	1617	6717	2.05442486	3969	3760	-0.077893542	565	6709	3.568928732	2.18
QUASIMODO	48	1849	5.271616704	473	3872	3.03326554	3156	5867	0.894767305	274	2947	3.446343777	1.05
GYPYS3	9	1435	7.261595047	13	929	6.149717238	673	1401	1.05802644	56	1105	4.308708676	1.02
GYPYS4	72	3644	5.65338492	200	4024	4.333007391	2024	2241	0.146947159	188	4064	4.433513474	1.43
BLOOD	75	6871	6.510280878	282	6857	4.603823972	14493	9969	-0.539851464	239	7074	4.887002803	0.54
STALKER4	1765	19059	3.432594793	2974	25434	3.096423359	8634	25618	1.56896736	3555	19146	2.428968192	1.29
IDEFIX	19	1405	6.233883766	94	1435	3.933623022	1419	1769	0.317400027	96	1736	4.181057063	0.04
ROVER_DM	33	2357	6.154469519	109	4026	5.212155318	1119	1765	1.379329962	170	5762	5.082714954	0.36
GYPYS	24	2884	6.914285341	31	1568	5.638640027	1119	1215	0.118578183	94	2067	4.452824757	0.36
STALKER2	409	4069	3.313777325	615	5319	3.112043521	2108	5457	1.372013151	820	5473	2.738153779	0.92
DM412	241	16694	6.115039705	285	23625	6.375376147	2376	11492	2.273908391	736	27872	5.242906433	0.31
GYPYS2	14	1884	7.089415927	16	1449	6.486751578	1088	1508	0.470901532	67	2253	5.080327711	0.58
MDG1	204	9610	5.561367622	356	11403	5.003191109	1753	7250	2.048127713	767	16995	4.470230382	0.31
GTWIN	46	4669	6.667245243	12	6456	9.116802276	285	1196	2.070195793	189	10008	5.72482243	-0.06
GYPYS10	5	700	6.991474881	2	628	8.197834378	101	503	2.322879689	24	858	5.156701162	-0.01
TABOR	90	8757	6.596649758	78	4624	5.884997419	315	7716	4.613967498	312	5581	4.159466732	-0.03
ZAM	13	649	5.629502346	11	648	5.934197742	45	356	2.990413572	33	1005	4.912287129	-0.33
GYPYS5	28	3908	7.115069979	8	2625	8.305939545	79	3385	5.421754574	143	3689	4.692107835	-0.64

Supplementary Table 3

Identified Proteins that were more than 10fold enriched (quantitative mass-spec analysis) in Vreteno-GFP immuno-precipitates compared to a control sample (wildtype ovaries). Ribosomal proteins were excluded.

protein ID	gene name	peptide coverage [%]	# unique peptides	# of peptides	iTRAQ enrichment
FBpp0076182	Hsp27	26.3	2	5	16x
FBpp0083716	CG4771	54.9	30	37	13x
FBpp0073872	CG9281	5.4	3	3	12x
FBpp0079562	CG31755	33.0	35	42	10x
FBpp0078210	CG11133	35.0	26	34	10x

Supplementary Table 4

VDRC lines used in this study

Tudor-SN	19011GD
yu	48006GD
CG13472	32193GD
tej	24181GD
CG8920	28998GD
CG4771	34897GD
CG9925	29329GD
CG9684	24090GD
CG7082	2553GD
spn-E	21374GD
CG11133	18149GD
tud	24031GD
CG31755	100190KK
CG14303	17474GD
fs(1)Yb	25437GD
armi	16205GD
zuc	48764GD

Supplementary Table 5

Primers for shRNA construct cloning

CG9925_top	ctagcagtAAAGTGTATGTTCAAACATTAtagttatattcaagcataTAATGTTTGAACATACACTTTGcg
CG9925_bottom	aattcgcAAAGTGTATGTTCAAACATTAtatgcttgaatataactaTAATGTTTGAACATACACTTTactg
CG31755_top	ctagcagtAAGGTTCAAAGTATCAGCGAAtagttatattcaagcataATTCGCTGATACTTTGAACCTTgcg
CG31755_bottom	aattcgcAAGGTTCAAAGTATCAGCGAAatgcttgaatataactaATTCGCTGATACTTTGAACCTTactg
CG9684_top	ctagcagtAAGGATATCAATGATGAGTTAtagttatattcaagcataTAACTCATCATTGATATCCTTgcg
CG9684_bottom	aattcgcAAGGATATCAATGATGAGTTAtatgcttgaatataactaTAACTCATCATTGATATCCTTactg
yu_top	ctagcagtCAGCAAGTCGATGAACATCAAtagttatattcaagcataTTGATGTTCATCGACTTGCTGcg
yu_bottom	aattcgcCAGCAAGTCGATGAACATCAAatgcttgaatataactaTTGATGTTCATCGACTTGCTGactg
CG13472_top	ctagcagtCTGCAGCGAAATCGACAAATAtagttatattcaagcataTATTTGTCGATTTGCTGCAGcg
CG13472_bottom	aattcgcCTGCAGCGAAATCGACAAATAtatgcttgaatataactaTATTTGTCGATTTGCTGCAGactg
Tudor-SN_top	ctagcagtTAGAAGAAGTGCCTAAAGAAAtagttatattcaagcataTTTCTTTAGGCACTTCTTCTAgcg
Tudor-SN_bottom	aattcgcTAGAAGAAGTGCCTAAAGAAAtatgcttgaatataactaTTTCTTTAGGCACTTCTTCTAactg
CG8920_top	ctagcagtCAGCATGTATGCGTTGGATAAtagttatattcaagcataTTATCCAA GCATACATGCTGcg
CG8920_bottom	aattcgcCAGCATGTATGCGTTGGATAAtatgcttgaatataactaTTATCCAACGCATACATGCTGactg
CG7082_top	ctagcagtCACCCACAACAAGTTAATCAAtagttatattcaagcataTTGATTAACTTGTGTGGGTGcg
CG7082_bottom	aattcgcCACCCACAACAAGTTAATCAAatgcttgaatataactaTTGATTAACTTGTGTGGGTGactg
tud_top	ctagcagtCAGATTGACTACTAAAGATAAtagttatattcaagcataTTATCTTTAGTAGTCAATCTGcg
tud_bottom	aattcgcCAGATTGACTACTAAAGATAAtatgcttgaatataactaTTATCTTTAGTAGTCAATCTGactg
CG4771_top	ctagcagtCAGCTGGAAGACTGTATGAAtagttatattcaagcataTTCATACAGTCTTTCCAGCTGcg
CG4771_bottom	aattcgcCAGCTGGAAGACTGTATGAAatgcttgaatataactaTTCATACAGTCTTTCCAGCTGactg
CG11133_bottom	aattcgcCAGCTGGAAGATGAAAGTAAAtatgcttgaatataactaTTTACTTTTATCTTCCAGCTGactg
CG11133_top	ctagcagtCAGCTGGAAGATGAAAGTAAAtagttatattcaagcataTTTACTTTTATCTTCCAGCTGcg
CG8589_top	ctagcagtCTCCAAGTCATTGAAAGTAAtagttatattcaagcataTTAACTTTCAATGACTTGGAGcg

CG8589_bottom	aattcgcCTCCAAGTCATTGAAAGTTAAatgcttgaatataactaTTAACTTTC AATGACTTGGAGactg
CG14303_top	ctagcagtCCGGAGGATTTCTATGTTCAAatagttatattcaagcataTTGAACAT AGAAATCCTCCGGgcg
CG14303_bottom	aattcgcCCGGAGGATTTCTATGTTCAAatgcttgaatataactaTTGAACATA GAAATCCTCCGGactg
spn-E_top	ctagcagtCTCGAAGAAGCTATTATTATAtagttatattcaagcataTATAATAA TAGCTTCTTCGAGgcg
spn-E_bottom	aattcgcCTCGAAGAAGCTATTATTATAtatgcttgaatataactaTATAATAAT AGCTTCTTCGAGactg
krimp_top	ctagcagtCAGATTGGGAGACTACGAATAtagttatattcaagcataTATTCGTA GTCTCCCAATCTGgcg
krimp_bottom	aattcgcCAGATTGGGAGACTACGAATAtatgcttgaatataactaTATTCGTAG TCTCCCAATCTGactg
fs(1)yb_top	ctagcagtCAGCTGCGATAAGATCTTCAAatagttatattcaagcataTTGAAGAT CTTATCGCAGCTGgcg
fs(1)yb_bottom	aattcgcCAGCTGCGATAAGATCTTCAAatgcttgaatataactaTTGAAGATC TTATCGCAGCTGactg

Supplementary Table 6

QPCR primers used in this study:

rp49_for	CCGCTTCAAGGGACAGTATCTG
rp49_rev	ATCTCGCCGCAGTAAACGC
HeT-A_for	CGCGCGGAACCCATCTTCAGA
HeT-A_rev	CGCCGCAGTCGTTTGGTGAGT
blood_for	AACAATAGAAAGAAGCCACCGAAC
blood_rev	AGTCATGGACTATTGAGGGTGTTG
ZAM_for	ACTTGACCTGGATACACTCACAAC
ZAM_rev	GAGTATTACGGCGACTAGGGATAC
HP1c_for	GTGCGAAGAGATCCAGAAGC
HP1c_rev	AGTCGAACTCGTGCGAGAAC
CG6985_for	ACCGCATTTGGAAATTAGCC
CG6985_rev	ATTCCGATTGGGTGAACTCC
vret_for	TGGCCAACAATGAACCTCTT
vret_rev	GACTTCCACTGAGCCAATGC
CG31755_for	TTCAAACAAACACTTGGCTTCC
CG31755_rev	AAAAACCCATCCGAAAGAAGTG
CG11133_for	CAGATTTTCCCAACTGTATGAGTGT
CG11133_rev	AATGCAGTCTTCTCCTGAGTATGG

9.2 Abbreviations

Acn	Acinus
AGO1	Argonaute-1
AGO2	Argonaute-2
AGO3	Argonaute-3
Aly	always early
Armi	Armitage
Atms	Antimeros
Atu	Another Transcription Unit
Aub	Aubergine
Bm	Bombyx Mori
Btz	Barentz
Cdc73	Cell Division Cycle 73
cDNA	complementary DNA
Ctr9	Clb Trhee (CLN3) Requiring
DNA	Deoxyribonucleic Acid
EJC	Exon Junction Complex
flam	flamenco
Fs(1)Yb	Female sterile (1) Yb
GO	Gene Ontology
GPAT	glycerol-3-phosphate O-acyltransferase
H2Av	Histone H2A variant
H3K9me	Histone 3 Lysine 9 methylation
HE	Homing Endonuclease
hSki8	human Super Killer
Hsp90	Heat shock protein 90
Hyx	Hyrax
KEGG	Kyoto Encyclopedia of Genes and Genomes
lacZ	β -Galactosidase
Leo1	Left Open Reading Frame 1
LINE	Long Interspersed Elements
LTR	Long Terminal Repeat
Mago	Mago nashi
Mino	Minotaur
miRNA	Micro RNA
mRNA	messenger RNA
Nup	Nucleoporin
Nxf2	Nuclear RNA export factor 2
Nxf3	Nuclear RNA export factor 3
Nxt1	NTF2-related export protein 1
ORF	Open Reading Frame
OSC	Ovarian Somatic Cells
OSS	Ovarian Somatic Sheet
PA	Phosphatidic Acid
Paf1	RNA polymerase II associated factor 1

PAPI	Partner of PIWIs
piRNA	PIWI-interacting RNA
Piwi	P-element induced wimpy testis
PIWI	PIWI-clade Argonaute proteins
PSD	Post segregation distorters
PTGS	post transcriptional gene silencing
RDC	Rhino Deadlock Cutoff
RISC	RNA Induced Silencing Complex
RNA	Ribonucleic Acid
RNA Pol II	RNA Polymerase II
RNAi	RNA-interference
RnpS1	RNA-binding protein S1
RPKM	reads per 1 KB per 1 million sequenced reads
RT-qPCR	Real Time Quantitative PCR
Rtf1	Restores TBP function
SAM	Sterile Alpha Motive
Sbr	Small bristles
shRNA	short hairpin RNA
SINE	Sort Interspersed Elements
siRNA	Small interfering RNA
TDRKH	tudor and KH domain containing
TE	Transposable Element (Transposon)
TFIIS	RNA polymerase II elongation factor
TGS	transcriptional gene silencing
Tsu	Tsunagi
UAP56	Helicase at 25E
UTR	untranslated region
VDRC	Vienna Drosophila RNAi Center
Vret	Vreteno
Wde	Windei
Zuc	Zucchini

9.3 Curriculum Vitae

Personal Information

Name	Dominik Handler
Address	Dr. Bohr Gasse 3 1030 Vienna
Email	dominik.handler@imba.oeaw.ac.at
Nationality	Austrian
Date of Birth	09. March 1986

Education

Dates	10.2010 – 12.2014 (expected)
Institute	University of Vienna – Vienna Biocenter PhD-Program
Dates	09.2006 – 10.2010
Institute	University of Applied Science, FH Campus Wien, Vienna, Austria
Title awarded	Diplomingenieur (FH) für technisch-wissenschaftliche Berufe (equivalent to Master of Science in Natural Sciences – MSc)
Dates	09.2000 – 06.2005
Institute	Secondary College for Food Technology, Hollabrunn, Austria
Title awarded	

Research Experience

Dates	10.2010 – 12.2014
Supervisor	Dr. Julius Brennecke
Position	PHD student
Institute	IMBA, Institute of Molecular Biotechnology of the Austrian Academy of Sciences Vienna, Austria
Dates	09.2009 – 10.2010
Supervisor	Dr. Julius Brennecke
Position	Diploma student
Institute	IMBA, Institute of Molecular Biotechnology of the Austrian Academy of Sciences Vienna, Austria
Dates	09.2008 – 09.2009
Supervisor	Dr. Leonie Ringrose
Position	Technician + 2 Month internship
Institute	IMBA, Institute of Molecular Biotechnology of the Austrian Academy of Sciences Vienna, Austria

Publications

- 2014** **The exon junction complex is required for definition and excision of neighboring introns in *Drosophila***
Rippe Hayashi*, Dominik Handler*, David Ish-Horowicz, and Julius Brennecke
* Authors contributed equally to this work
Genes Dev. 2014 Aug 15;28(16):1772-85.
- The Rhino-Deadlock-Cutoff Complex Licenses Noncanonical Transcription of Dual-Strand piRNA Clusters in *Drosophila*.**
Fabio Mohn, Grzegorz Sienski, Dominik Handler, and Julius Brennecke
Cell. 2014 Jun 5;157(6):1364-79
- 2013** **The Genetic Makeup of the *Drosophila* piRNA Pathway**
Dominik Handler, Katharina Meixner, Manfred Pizka, Kathrin Lauss, Christopher Schmied, Franz Sebastian Gruber, and Julius Brennecke
Mol Cell. 2013 Jun 6;50(5):762-77.
- 2011** **A systematic analysis of *Drosophila* TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors**
Dominik Handler, Daniel Olivieri, Maria Novatchkova, Franz Sebastian Gruber, Katharina Meixner, Karl Mechtler, Alexander Stark, Ravi Sachidanandam and Julius Brennecke
EMBO J. 2011 Aug 23;30(19):3977-93.
- A genome-scale shRNA resource for transgenic RNAi in *Drosophila***
Jian-Quan Ni, Rui Zhou, Benjamin Czech, Lu-Ping Liu, Laura Holderbaum, Donghui Yang-Zhou, Hye-Seok Shim, Rong Tao, Dominik Handler, Phillip Karpowicz, Richard Binari, Matthew Booker, Julius Brennecke, Elizabeth A Perkins, Gregory J Hannon & Norbert Perrimon
Nat Methods. 2011 May;8(5):405-7.

Meetings

- 2014** **Cell-Symposia – Regulatory RNAs**
Berkeley, USA
Poster presentation
- Keystone-Symposia – RNA Silencing**
Seattle, USA
Poster presentation
- 2012** **EMBO/EMBL Symposium – The Complex Life of mRNA**
Heidelberg, Germany
Poster presentation

- | | |
|-------------|--|
| 2011 | Cell-Symposia – Regulatory RNAs
Chicago, USA
Poster presentation |
| 2010 | EMBO/EMBL Symposium – The Non-Coding Genome
Heidelberg, Germany
Poster presentation |

Language skills

German	Basic
English	Advanced

10 Acknowledgements

First of all I want to thank Julius for his constant support and extraordinary mentoring. He was the perfect supervisor for me. I enjoyed the endless discussions and the appreciation he had for my way of doing research.

I also want to thank the whole Brennecke Lab. You were the bunch of people that made this place so enjoyable. Particularly I want to thank Fabio and Rippei for the very interesting and fruitful collaborations we had.

I want to thank my family. It would not have been possible for me to get here without their great support. I want to thank all of my friends for distracting me from research sometimes.

Especially I want to thank Sandra for the patience she had with my work and me and for being the best girlfriend I can imagine.