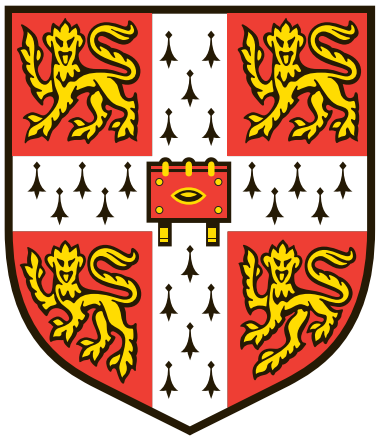
Mechanisms of piRNA biogenesis and co-transcriptional silencing of transposable elements in *Drosophila*



Marzia Munafò

Supervisors:
Prof. Gregory J. Hannon
Dr. Benjamin Czech

Cancer Research UK – Cambridge Institute
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Darwin College

February 2020

Ai miei genitori.

"Ergo vivida vis animi pervicit, et extra
processit longe flammantia moenia mundi"

Lucrezio, "De rerum natura"; Lib.I v.v. 72-73

PREFACE

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text and authors contributions at the beginning of each Chapter.

This thesis is the result of work carried out at the Cancer Research UK – Cambridge Institute between September 2016 and February 2020. It is not substantially the same as any that I have submitted or is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted or is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

This thesis does not exceed the prescribed word limit of 60,000 words for the Clinical Medicine and Veterinary Medicine Degree Committee, excluding tables, appendices and bibliography.

Marzia Munafò

February 2020

ABSTRACT

A large fraction of eukaryotic genomes consists of mobile, repetitive elements called transposons. Since their uncontrolled mobilisation is a potentially harmful event, several molecular mechanisms have evolved to counteract transposon activation and thus safeguard genome integrity. Among these is the piRNA pathway, a gonad-specific system based on small non-coding RNAs that recognise active transposons and instruct their silencing. Ultimately, piRNAs trigger epigenetic silencing of transposon loci. The work in this thesis investigates the molecular mechanisms by which piRNAs are produced from the correct substrates and how piRNAs can induce silencing of target loci in *Drosophila melanogaster*.

First, I investigated how piRNA precursors are selected for processing and how they are transported to mitochondria, where piRNA production occurs. I explored the role of an uncharacterised *Drosophila* gene previously implicated in germline transposon control: CG10880/Daedalus (Daed). I found that Daed is an essential component of the mitochondrial piRNA biogenesis machinery and that it recruits the RNA helicase Armitage (Armi) to mitochondria. If Armi fails to be recruited, piRNA biogenesis cannot occur since Armi's role is that of delivering piRNA precursors to the mitochondrial processing machinery. Secondly, I investigated how the major piRNA precursor in somatic cells, namely *flamenco* (*flam*) transcript, is exported and specified for downstream piRNA production. I uncovered that *flam* export is closely linked to the assembly of peri-nuclear condensates of the helicase fs(1)Yb (Yb). Furthermore, some subunits of the Nuclear Pore Complex (NPC) are also required for the production *flam*-derived piRNAs, thus suggesting the evolution of a specialised machinery that couples nuclear export and processing of this transcript. Finally, I set out to understand how piRNAs trigger silencing of active transposons. I found that Panoramix (Panx), the central effector of piRNA-guided epigenetic silencing, assembles into a complex with two other proteins: Nxf2 and Nxt1. We characterised the dependencies within the complex and found that all three components are essential to initiate silencing. Intriguingly, Nxf2 and Nxt1 belong to the family of nuclear export factors, thus suggesting that the piRNA pathway has co-opted proteins involved in RNA export and repurposed them for transposon control.

Overall, this work provides new insights on the molecular mechanisms of piRNA-guided transposon silencing in *Drosophila* and shows evidence that transposon control pathways can exploit cellular factors for novel functions.

Contents

ACKNOWLEDGEMENTS

Those who know me well, know that I have always struggled with words. Nonetheless, after this long journey I want to find the right ones to thank those that I believe have contributed to make me the person I am today.

First of all, I owe a great thank you to my supervisors.
Thanks to Greg, for giving me the opportunity to be a PhD student in his lab in the first place. But, most importantly, thank you for being an inspiring mentor, for challenging me to see things from a different perspective and for the most intellectually stimulating scientific conversations. Thanks to Ben, for being always there and for the enormous support throughout my entire PhD. Thanks for caring about me and about my project and being always up for discussing science and beyond. I have grown so much as a scientist and it would have not been possible without your guidance. I will always be grateful for this.

Thanks to my labmates, as I've been lucky enough to work with incredibly talented and fun people who have filled these years with wonderful memories. Thanks to the piRNA Spice Girls, Emma and Evelyn, because I would have literally lost my mind without you. Thanks for all the time spent "chinwagging" in the fly room, sharing epic fails (and occasionally even good results), wearing ridiculous group costumes at lab parties and being grannies together most of the time. Thanks to Cristina, Martin, Sophia, Tatjana and everyone else in the Hannon lab for all the fun, the summer parties, the lab lunches, “spoons” and every great moment spent together. You have been the best crew to share this adventure with.

Thanks to my housemates, Miriam and Mara, for becoming my family here. Thanks for caring, for listening, for being there during the highest and the lowest moments and for sharing life, laughs, music and wine. You truly deserve one of the biggest acknowledgements of this thesis, as without you I would have never been able to call Cambridge "home".

Thanks, in random order, to many people that have accompanied me throughout these years. Thanks to the DarMates, for being the best bunch of friends I could hope for, and to DarBar and BarComm, because let's be honest nobody makes it to the finish line without a few too many pints. Thanks to Richard, for giving me the very first chance to prove myself in a lab, and to Mehdi, for all the lessons and lab tricks I have learnt with you. Thanks to all the students and Research Assistants that I have had the opportunity to work with and from whom I have learnt

a great deal. To Alessandro, for some of the most hilarious moments in the lab and for accurately establishing that the correct concentration of Piwi antibody is 1:2,236; to Fede and Vera, for being an amazing team and for the unforgettable pipetting music recommendations; to Serena and Tori, for their hard work and patience dealing with me while I was writing this thesis. Thanks to Boehringer Ingelheim Fonds for funding my PhD and giving me the opportunity to meet wonderful people and talk science even while climbing up a mountain. Thanks to the CRUK-CI people and Core Facilities for all their help and support.

Thanks to my best friends, Giulia, Giusi and Vale, because it takes a lot of care and commitment to be so close when you live in different countries. I am incredibly lucky to have you by my side, even when it means spending half of my life sending and listening to endless voice messages.

Thanks to my parents, for supporting me in every single thing I ever did. Thanks for giving me the freedom to follow my dreams without feeling any pressure or expectation. For being patient and understanding, even when my phone calls accounted to less than once a week because "I was busy". To my mom, who won her fear of flying just to come and see me. Thanks for giving me this insatiable thirst for knowledge. And to my dad, who taught me to be patient and firm. This thesis is entirely dedicated to you.

Finally, I owe an infinite thank you to the person that has been holding my hand all this time. There are so many things I am grateful for and that I cannot express. Above all, thank you for not being afraid to let me go.
Ora che siamo liberi così
di sceglierci ogni volta invece che
lasciare troppe cose già decise
scegliere per noi

Grazie.

PUBLICATIONS

Czech, B., **Munafò, M.**, Ciabrelli, F., Eastwood, E.L., Fabry, M.H., Kneuss, E., and Hannon, G.J. (2018). piRNA-Guided Genome Defense: From Biogenesis to Silencing. Annu Rev Genet *52*, 131-157.

Munafò, M., Manelli, V., Falconio, F.A., Sawle, A., Kneuss, E., Eastwood, E.L., Seah, J.W.E., Czech, B., and Hannon, G.J. (2019). Daedalus and Gasz recruit Armitage to mitochondria, bringing piRNA precursors to the biogenesis machinery. Genes Dev *33*, 844-856.

Fabry, M.H.*, Ciabrelli, F. *, **Munafò, M.** *, Eastwood, E.L., Kneuss, E., Falciatori, I., Falconio, F.A., Hannon, G.J., and Czech, B. (2019). piRNA-guided co-transcriptional silencing coopts nuclear export factors. Elife *8*.
***equal contribution**

Kneuss, E., **Munafò, M.**, Eastwood, E.L., Deumer, U.S., Preall, J.B., Hannon, G.J., and Czech, B. (2019). Specialization of the *Drosophila* nuclear export family protein Nxf3 for piRNA precursor export. Genes Dev *33*, 1208-1220.

Munafò M., MacMillan S., Passera A., Lawless T., Czech B., Hannon G. J.. Requirement of Nuclear Pore Complex subunits for transposon silencing in *Drosophila* (manuscript in preparation)

LIST OF ABBREVIATIONS

5’-P	5-prime monophosphate
ANK	Ankyrin repeats
bp	base pair
CC	Coiled-Coil
CDS	Coding Sequence
CLIP-seq	Cross-Linking Immuno-Precipitation and sequencing
CoIP	Co-immunoprecipitation
FBS	Fetal Bovine Serum
FL	full-length
gDNA	genomic DNA
GLKD	Germline-knockdown
HCR	Hybridization Chain Reaction
hr(s)	hour(s)
IP-MS	Immunoprecipitation and Mass Spectrometry
kb	kilobase
min(s)	minute(s)
miRNA	microRNA
mRNA	messenger RNA
NLS	Nuclear Localisation Signal
NPC	Nuclear Pore Complex
nt	nucleotide
Nup	Nucleoporin
ORF	Open Reading Frame
OSCs	Ovarian Somatic Cells
PCR	Polymerase Chain Reaction
piRNA	Piwi-interacting RNA
PL-MS	Proximity Labelling and Mass Spectrometry
POI	Protein of interest
pol II	RNA polymerase II
PTGS	Post-transcriptional gene silencing
RBP	RNA-binding protein
RdRP	RNA-dependent RNA Polymerase
RIP-seq	RNA Immuno-Precipitation and sequencing

RISC	RNA-induced silencing complex
RNAi	RNA interference
RPM	Reads Per Million
rRNA	ribosomal RNA
RT	Reverse Transcription
S2	Schneider 2 cells
SAM	Sterile Alpha Motif
sDMA	symmetric di-methyl Arginine
sec	seconds
shRNA	short hairpin RNA
siRNA	short-interfering RNAs
TE	Transposable Element
TEV	Tobacco Etch Virus
TGS	co-transcriptional gene silencing
TMM	Trans-membrane domain
tRNA	transfer RNA
TSS	Transcriptional Start Site

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

I.1 Transposable elements: genomic threats and drivers for evolution

Transposable Elements (TEs) are mobile repetitive genomic elements that widely populate eukaryotic genomes, constituting up to 40% of the mouse genetic material and 30% of that of the fruit fly, *Drosophila melanogaster* (Adams et al., 2000; Mouse Genome Sequencing et al., 2002). These genomic "parasites" feature the ability to mobilise and colonize novel locations of the genome they reside in. TEs are often called "selfish" genomic elements (Doolittle, 1980; Orgel, 1980) because they aim to spread within the host, both horizontally (from one organism to another) (Gilbert and Feschotte, 2018) and vertically (across generations) (Blumenstiel, 2019). Vertical propagation of TEs necessitates their mobilisation in germ cells, whose genome will give rise to an entire new organism, and indeed transposition is frequent during germline development in animals. However, uncontrolled TE mobilisation poses a severe threat to genomic integrity, since it could disrupt essential genes and eventually impair the fitness of the host (Payer and Burns, 2019). This is especially critical in the germline, where any *de novo* mutation would be inherited by the offspring. Hence, several defence mechanisms have emerged to tame transposon activity. Such mechanisms must be easily adaptable to novel transposon challenges and must co-evolve with the "parasites" they fight, as in a genomic arms race. This can be formalised by the Red Queen model (Van Valen, 1973), which postulates that in any host-pathogen interaction, one must evolve as fast as possible to escape the other and survive. In the context of transposon-host interactions, this implies that the host will be under selective pressure to evolve mechanisms that dampen TE activity. On the other hand, if a host defence mechanism completely suppresses a particular transposon, this will force the TE to evolve new ways to evade it (Cosby et al., 2019). In fact, many transposon-response mechanisms show signs of rapid diversification and evolution, implying that they are engaged in an active conflict against TEs (Daugherty and Malik, 2012; Jacobs et al., 2014; Vermaak et al., 2005). Genes involved in transposon control often manifest signs of positive selection, which consists of an insurgence of many more nonsynonymous mutations compared to synonymous over time, thus leading to a faster adaptation.

Despite their seemingly threatening nature, active TEs have persisted within eukaryotic genomes during evolution. This represents an intriguing paradox, suggesting that TEs somehow provide benefits to the host and therefore have been maintained and allowed activation at specific times and places. Indeed, transposon sequences are a unique source of genetic variation that has often been co-opted by the host into endogenous regulatory networks (Cosby et al., 2019; Fedoroff, 2012). An example of this is the maintenance of telomeres in *Drosophilids*. Unlike most eukaryotes, *Drosophila* species lack a telomerase, therefore an alternative mechanism must be in place to ensure continuous replication of chromosomal ends. Notably, telomeric repeats in these species consist of arrays of active retroelements (*HeT-A, TAHRE* and *TART* in *D. melanogaster*) which continuously retro-transpose head to tail to prevent the shortening of chromosome ends (Pardue and DeBaryshe, 2008). In this case, controlled TE mobilisation provides an unequivocal advantage to the host genome and can be seen as a representative example of host-TE mutualism. Another example can be found in mammalian pre-implantation embryogenesis, where a burst in the expression of various TE families has been documented (Rodriguez-Terrones and Torres-Padilla, 2018). This TE upregulation is concomitant with the activation of zygotic transcription and with the epigenetic reprogramming of the embryo, which involves global erasure of repressive marks. Emerging evidence suggests that, in this time window, TE expression is not a mere by-product of epigenetic reprogramming but instead actively contributes to shape chromatin architecture and totipotency of the embryo (J. Yuyang Lu, 2019; Kruse, 2019; Zhang et al., 2019).

TE expression is overall tightly regulated throughout the life of animals and their overactivation must be prevented. The arms race of TEs against their host genomes is often at an equilibrium, whereby TEs inhabit the host without compromising its survival, as this would impair their own propagation, and the host exploits controlled TE expression for cellular functions (Cosby et al., 2019). Understanding transposon defence mechanisms is therefore primarily important because of their role in safeguarding genomic integrity. Furthermore, it advances our understanding of how genomes evolve in response to external cues.

I.2 RNA-interference

I.2.1 The discovery of RNAi

The biological phenomenon known as RNA interference (RNAi) is a molecular response to double-stranded RNA that modulates the expression of genes or mediates the defence against endogenous and exogenous parasitic nucleic acids. RNAi was first discovered in the nematode *Caenorhabditis elegans* by Fire and Mello, who observed that injection of dsRNA in the worm elicited potent and specific silencing of the corresponding sequences (Fire et al., 1998). RNAi soon emerged as a remarkably conserved mechanism, and not a peculiarity of nematodes, and other gene silencing events previously observed in different organisms were recognised as manifestations of the same phenomenon (Baulcombe, 1996; Ecker and Davis, 1986; Napoli et al., 1990; Romano and Macino, 1992). In fact, small RNAs can trigger gene silencing also in mammals without causing any unwanted immune response (Caplen et al., 2001; Elbashir et al., 2001a). The discovery of RNAi has had enormous consequences on our understanding of how gene expression is regulated as well as on how cellular machineries can be harnessed as research tools. Over the past two decades, RNAi has unlocked a vast array of possibilities to modulate gene expression, thus revolutionising approaches to study gene function and paving the way for novel therapeutic applications (Hannon and Rossi, 2004).

I.2.2 The Argonaute protein family

RNAi is widely spread across the plant and animal kingdoms and requires two components: a small RNA, which provides the sequence specificity, and an effector protein, which executes the silencing. Small RNAs differ in length, protein partners and biogenesis modes and are classified accordingly into small-interfering RNAs (siRNAs), microRNAs (miRNAs) and Piwi-interacting RNAs (piRNAs). While siRNAs and miRNAs are ubiquitous, piRNAs are predominantly expressed in animal gonads. The effectors of RNAi belong to the family of Argonaute proteins, which are highly conserved and can be found in almost all eukaryotes (Meister, 2013), with the notable exception of the budding yeast *Saccaromyces cerevisiae* (Drinnenberg et al., 2009). Argonaute proteins serve as binding modules for the small RNAs, which in turn guide them towards their targets. Argonaute proteins act on substrate RNAs either directly via their endonuclease activity (hereafter referred to as “slicing”) or indirectly by recruiting other effector proteins (Meister, 2013). Structural studies have revealed specific domains of the Argonaute family proteins that are necessary to bind small RNAs and to perform

slicing (Jinek and Doudna, 2009). The PAZ and the MID domains accommodate the 3' and 5' end of the small RNA, respectively, leaving the central nucleotides free to base-pair with the target (Lingel et al., 2003; Ma et al., 2004; Ma et al., 2005; Parker et al., 2005; Song et al., 2003; Yan et al., 2003). An un-structured amino-terminus participates to the loading of the small RNA while the PIWI domain contains a tetrad of amino acids (DEDX, where X can be D or H) that confers the slicing activity to some members of the family (Liu et al., 2004; Meister et al., 2004; Nakanishi et al., 2012). Several Argonaute proteins have lost the catalytic activity or do not require it for gene silencing (e.g. *D. melanogaster* Piwi, which will be discussed later) (Darricarrere et al., 2013), whilst for others it is the main mode of action (e.g. mammalian Ago2) (Liu et al., 2004). The Argonaute family comprises two main clades: AGO proteins, which resemble *Arabidopsis thaliana* AGO1, and PIWI proteins, which are related to *D. melanogaster* Piwi (P-element-induced wimpy testes) (hereafter PIWI will indicate the protein clade whereas Piwi will refer to the *Drosophila* protein). While AGO proteins predominantly bind to siRNAs and miRNAs and are ubiquitously expressed, PIWI proteins appear to be mostly restricted to the gonads and only load piRNAs. Furthermore, *C. elegans* possesses an amazingly expanded array of 27 Argonaute proteins, most of which belong to a third clade known as WAGO and differ significantly from both AGO and PIWI proteins (Yigit et al., 2006; Youngman and Claycomb, 2014).

I.2.3 RNAi in *Drosophila melanogaster*

The genome of *D. melanogaster* encodes five members of the Argonaute family: two AGO proteins (Argonaute-1 [Ago1] and Argonaute-2 [Ago2]) and three PIWI proteins (Piwi, Aubergine [Aub] and Argonaute-3 [Ago3]), each bound to a specific subset of small RNAs (Brennecke et al., 2007; Czech et al., 2008; Forstemann et al., 2007; Gunawardane et al., 2007; Okamura et al., 2008; Tomari et al., 2007). As characteristic of the respective clade, *Drosophila* Ago1 and Ago2 are ubiquitously expressed throughout the fly lifespan, whereas the three PIWIs are restricted to the reproductive tissues.

Ago1 predominantly binds to ~22- to 23-nt long miRNAs originated from long RNA polymerase II (pol II) transcripts (pri-miRNAs), which are temporally and spatially regulated similarly to protein-coding genes (Kim, 2005). Pri-miRNAs fold into characteristic hairpin structures recognised by the nuclear miRNA processing machinery, namely the Microprocessor complex, which consists of the RNase III endonuclease Drosha and its cofactor DGCR8, also known as Pasha (Denli et al., 2004; Gregory et al., 2004). The Microprocessor cleaves the pri-miRNA to release an approximately 70-nt long hairpin (pre-miRNA, **Figure 1.1A top**) which

is exported to the cytoplasm by Exportin-5 and RanGTP (Lund et al., 2004). A second conserved RNase III-type nuclease, Dicer (Dicer-1 in *Drosophila*), together with its co-factor Loquacious (Loqs), performs the cytoplasmic maturation of pre-miRNAs by cleaving off the hairpin loop and releasing a ~22- to 23-nt RNA duplex (Bernstein et al., 2001; Ketting et al., 2001; Lee et al., 2004; Macrae et al., 2006)(**Figure 1.1A, middle**). Only one strand of this duplex (the "miR" or "guide" strand) is loaded into an Argonaute protein, a decision guided by well-defined thermodynamic rules (Czech et al., 2009; Khvorova et al., 2003; Okamura et al., 2009). The other strand, called "miR*" or "passenger", is usually degraded (**Figure 1.1A bottom**). Ultimately, the sequence of the mature miRNA that is released from each pri-miRNA is determined by the distinctive cleavage patterns of Drosha and Dicer-1 (Park et al., 2011). Mature miRNAs loaded into Ago1 constitute an active RNA-induced silencing complex (RISC, generally referred to an Argonaute protein in complex with a small RNA). miRNA-RISCs typically recognise their targets via imperfect base-pairing which leads to slicer-independent mRNA destabilisation (Bagga et al., 2005; Eichhorn et al., 2014; Filipowicz et al., 2008; Liu et al., 2005; Rehwinkel et al., 2005). The base-pairing between the miRNA and target transcripts engages only the nucleotides between position 2 and 7 of the miRNA, known as the "seed" sequence (Lewis et al., 2003; Parker et al., 2005). miRNAs processed from different precursors but sharing the same "seed" are grouped into families. Some miRNA families appear to be extremely conserved across evolution, such as the let-7 family of key developmental regulators (Pasquinelli et al., 2000; Roush and Slack, 2008), further supporting a fundamental regulatory role for these small molecules. Altogether, miRNAs act as critical regulators of cellular functions by fine-tuning the expression of protein-coding genes at the post-transcriptional level. *Drosophila* Ago2 mainly associates with 21-nt siRNAs, which are processed from long dsRNAs arising either from convergent transcriptional units or from genomic loci that contain long inverted repeats (**Figure 1.1B top**). siRNA precursors can originate from endogenous loci transcribed into long stem-loops (endo-siRNAs) or from ectopically introduced dsRNAs of various lengths (exo-siRNAs). siRNAs are generated via consecutive cleavages of dsRNA precursors by Dicer-2 together with its co-factor Loqs-PD, an isoform different from the one required in miRNA biogenesis (Czech et al., 2008; Ghildiyal et al., 2008; Kawamura et al., 2008; Lee et al., 2004; Liu et al., 2003; Okamura et al., 2008; Zhang et al., 2004; Zhou et al., 2009) (**Figure 1.1B middle**).

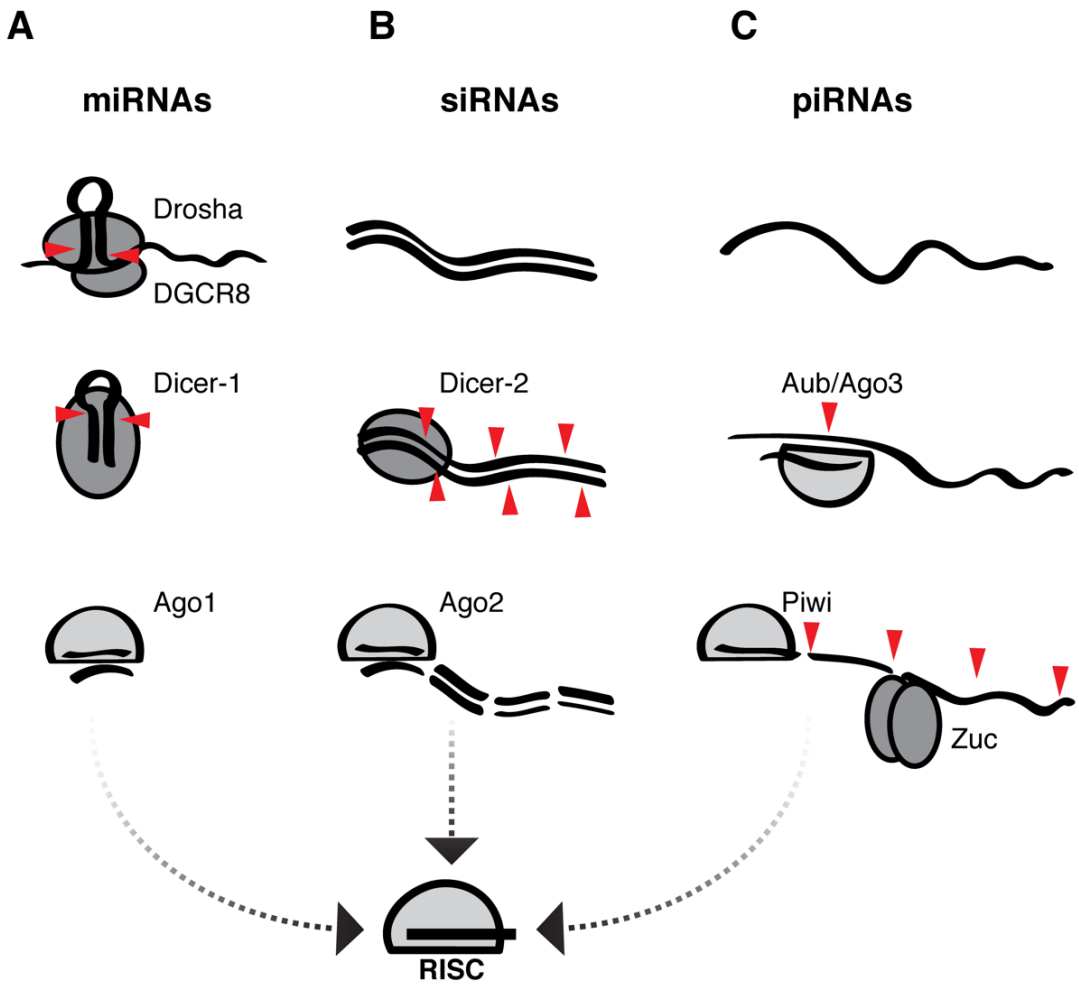


Figure 1.1. Simplified model of small RNA biogenesis pathways in *Drosophila*
A) Pri-miRNA transcripts are processed into pre-miRNAs by the Microprocessor complex (Drosha and DGCR8). Pre-miRNA hairpins are cleaved by Dicer-1 and loaded by Ago1. **B)** Long dsRNAs are cleaved by Dicer-2 into strings of siRNA duplexes, loaded by Ago2. **C)** Long ssRNA precursors are cleaved by Aub or Ago3 during the ping-pong cycle and then consumed by Zuc, resulting in Piwi-loaded piRNAs. All small RNA biogenesis pathways produce a functional RNA-induced Silencing Complex (RISC, at the bottom).

Similarly to miRNAs, siRNA duplexes are asymmetrically loaded into a silencing-competent RISC in a process that requires Dicer-2 and the dsRNA-binding protein R2D2 (Liu et al., 2003) (**Figure 1.1B bottom**). However, unlike miRNAs, siRNAs usually bind their target with a (near) perfect match and do not possess any seed sequence preferences (Elbashir et al., 2001b; Hutvagner and Zamore, 2002). In fact, siRNAs are produced as pools by subsequent Dicer cleavages and are not identifiable by a unique sequence nor by families. The perfect match between siRNAs and complementary transcripts releases the slicing activity of Ago2, hence siRNA silencing happens via substrate cleavage and subsequent degradation (Hutvagner and Zamore, 2002; Matranga et al., 2005). Endo-siRNAs in *Drosophila* mostly target genes and transposable elements whereas exo-siRNAs provide an anti-viral defence by targeting exogenous viral RNAs (Czech et al., 2008; Ghildiyal et al., 2008; Kawamura et al., 2008; Okamura et al., 2008; Wang et al., 2006).

Contrary to microRNAs and siRNAs, the third class of fly small RNAs, piRNAs, follows a different and unique biogenesis route, which will be extensively discussed in **Chapter I.3** (**Figure 1.1C**). Among the main differences with the other small RNA pathways, piRNA production does not involve any dsRNA intermediate nor structured transcript. Secondly, Argonaute PIWI proteins are directly involved in a self-amplification loop that can trigger and sustain piRNA production. Thirdly, piRNA biogenesis only initiates after precursor RNAs leave the nucleus and requires a wide range of factors, some of which share a conserved mitochondrial localisation.

The extensive characterisation of the molecular mechanisms underlying small RNA production and silencing activity has been fundamental to enable their exploitation as research tools. Notably, artificial expression of both miRNAs and siRNAs is now routinely used to down-regulate the expression of genes and infer their function. The introduction of RNAi as a research tool has paved the way to genome-wide loss-of-function genetic screens (Boutros and Ahringer, 2008), which have uncovered key players of the most diverse cellular processes and aided drug discovery. The RNAi toolkit available to *Drosophila* geneticists is quite remarkable and includes several strategies to achieve down-regulation of any gene of interest in the desired tissue. A first generation of transgenic fly lines has been built by the Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA) and by the *Drosophila* Transgenic RNAi Project (TRiP) at Harvard Medical School and is based on long dsRNAs (Dietzl et al., 2007; Ni et al., 2009; Ni et al., 2008; Perkins et al., 2015). These collections contain genome-wide libraries of extended inverted repeats that are transcribed into ~300 to 500 nt long double-

stranded RNAs recognised by the endogenous siRNA processing machinery. Such strategy achieved successful knockdowns in somatic tissues, but proved less effective in the germline (Ni et al., 2011). Subsequent strategies involved a short hairpin (shRNA) structure mimicking a miRNA precursor which is efficiently processed by the Microprocessor complex and gives rise to potent sequence-specific silencers (Ni et al., 2011). A different vector design combined with the over-expression of Dicer-2 have enabled knockdowns also in the fly germline (Wang and Elgin, 2011). The advancements in target prediction algorithms allow efficient design of shRNAs against any desired gene, thus constituting an extremely powerful resource to study gene function.

I.3 Piwi-interacting RNAs: guardians of the genome

piRNAs are the third class of animal small RNAs and are generally defined by their length (23-30nt) and association with PIWI clade Argonaute proteins. In flies, piRNAs constitute the core of a germline-specific molecular system that counteracts TE mobilisation: the piRNA pathway. TEs mobilise frequently in the germline, where they have lower chances of immediately affecting the host fitness while increasing their probability of propagating vertically across generations. However, from the host perspective, TE propagation in the germline is most catastrophic, since genomic information must be faithfully transmitted to offspring. Hence, the piRNA pathway is in place to safeguard the genomic integrity of germ cells. piRNAs can be found across the animal kingdom, from Cnidarians to humans (Aravin et al., 2006; Brennecke et al., 2007; Gainetdinov et al., 2018; Houwing et al., 2007) and, in most organisms in which they have been studied, failure in the piRNA pathway causes sterility. Nonetheless, piRNAs are also found in the somatic tissues of many animals. An example of this is the arthropod phylum, whose phylogenetic analysis shows that the last common ancestor likely had piRNAs in both somatic and gonadal tissues (Lewis et al., 2018). Most of the studied arthropods still possess both but a few have lost somatic piRNAs, as is the case in *D. melanogaster*. The nematode *C. elegans* does not have a direct correspondent of the fly piRNA pathway, but does recapitulate some of its features in various small RNA pathways. Nematode piRNAs are therefore mainly defined by their size profile and association with PIWI-clade Argonautes (Almeida et al., 2018). PRG-1 is a worm PIWI protein that binds to 21U RNAs and has been historically referred to as the piRNA pathway in *C.elegans.* PRG-1 is expressed in germ cells and mutants for this gene show progressive fertility defects across generations (Wang and Reinke, 2008). Mammalian piRNAs are mostly restricted to male gonads and mutants in murine PIWI homologs fail to complete spermatogenesis. Despite some components of the piRNA

pathway being detected in the brain of flies and mice (Ross et al., 2014), there is yet no clear evidence that a fully functional pathway exists in those tissues. Although the specificity to the germline cannot be invoked as a universal feature of the piRNA pathway, there are many examples beyond the fruit fly where piRNA-mediated genome defence is inextricably linked to the ability to generate offspring. Throughout this thesis, I will primarily discuss the function of piRNAs in relation to transposon control in the germline of *D. melanogaster*.

The piRNA pathway is an example of the arms race between transposons and their host genome. The system has evolved to be flexible and easily adaptable to new transposon challenges, and in fact signs of positive selection can be found in many of the involved genes. For instance, several *Drosophila* genes involved in piRNA precursor transcription appear to be evolving fast, leading to functional incompatibilities between proteins of two species as closely related as *D. melanogaster* and *D. simulans* (Parhad et al., 2017). Furthermore, inter-species hybrids between *D. simulans* and *D. melanogaster* phenocopy mutations in the piRNA pathway (Kelleher et al., 2012). Although no evidence of transposon mechanisms that evade silencing has been found yet, it is hypothesized that the quick evolutionary pace of fly piRNA pathway genes depends on the selective pressure posed by transposon mobilisation. Notably, the piRNA pathway is often described as a "molecular immune system" because of two of its features: firstly, it can detect and neutralise *non-self,* unwanted nucleic acids; secondly, it comprises an innate and an adaptive branch, like mammalian systemic immunity. The two branches cooperate to dampen the activity of transposons that are already present within the genome ("innate immunity") but also of new invading elements, which are rapidly recognised and integrated into the silencing process ("adaptive immunity"), though the latter is much less understood.

The piRNA pathway has been extensively characterised in *D. melanogaster*, where it predominantly acts in the ovary. PIWI proteins are also expressed in testes, but their role in this tissue has not been deeply investigated (Nagao et al., 2010). A female fruit fly has two ovaries, each comprising approximately 18 ovarioles. The ovariole represents the functional unit of the ovary and consists of a series of interconnected egg chambers, each stemming from a few stem cells and ultimately giving rise to a mature egg (Bastock and St Johnston, 2008) (**Figure 1.2**). At the anterior tip of the ovariole lies the germarium (**Figure 1.2, left**), which contains a pool of actively dividing germline stem cells (GSCs) that bud off into individual egg chambers. Egg chambers move towards the posterior pole as they undergo maturation, with the mature egg at the posterior tip of the ovariole (**Figure 1.2, right**). At the inside of each egg chamber is a syncytium of germline cells, one of which will become the egg (also known as oocyte) while

the others (called nurse cells) will remain connected through cytoplasmic bridges and provide nutrients, proteins and mRNAs, many of which are deposited in the developing oocyte. Each egg chamber is enclosed by a layer of epithelial "follicle cells", of somatic origin, which also support oocyte development and form the egg shell and the extra-embryonic dorsal appendages. One key process occurring during oogenesis is the establishment of the Anterior-Posterior axis of the oocyte, which in turn determines the body axes of the embryo. This axis is mainly established via asymmetric deposition in the oocyte of key mRNAs required for embryonic patterning. Overall, an intricate signalling network between soma and germline and between different populations of somatic cells sustains ovarian development and oocyte pattern formation (Bastock and St Johnston, 2008; Roth and Lynch, 2009). The piRNA pathway operates in both compartments of the ovary, although some features are specific to either nurse or follicle cells (Li et al., 2009a; Malone et al., 2009).

The three PIWI-clade Argonaute proteins expressed in the *Drosophila* ovary (Piwi, Aub and Ago3) differ in their subcellular localisation, tissue specificity and small RNA partners (Brennecke et al., 2007). Piwi is predominantly localised to the nuclei and is expressed in both germline and somatic cells (Cox et al., 2000). Piwi is also detectable in the cytosol of the developing oocyte with a slight enrichment at the posterior pole, where the germline of the future embryo will form (Brennecke et al., 2007; Brennecke et al., 2008). Piwi binds to piRNAs that are on average 25 nt long, have a pronounced bias for a uridine at their 5' end (namely, "1U bias") and are mostly in antisense orientation to transposon sequences. On the other hand, Aub and Ago3 are present exclusively in germ cells, where they localise to peri-nuclear, ring-like structures known as "nuage" (Harris and Macdonald, 2001; Lim and Kai, 2007). Aub, but not Ago3, also accumulates at the posterior pole of the egg (Brennecke et al., 2007; Brennecke et al., 2008). Both of them bind to piRNAs that are slightly shorter than Piwi's (on average 24 nt) but only Aub piRNAs show the same preference for a uridine at position 1. This sequence bias is absent from the piRNA population bound to Ago3, which instead is enriched for adenine at position 10. Finally, Aub and Ago3 preferentially incorporate piRNAs that match the antisense and sense sequence of transposable elements, respectively.

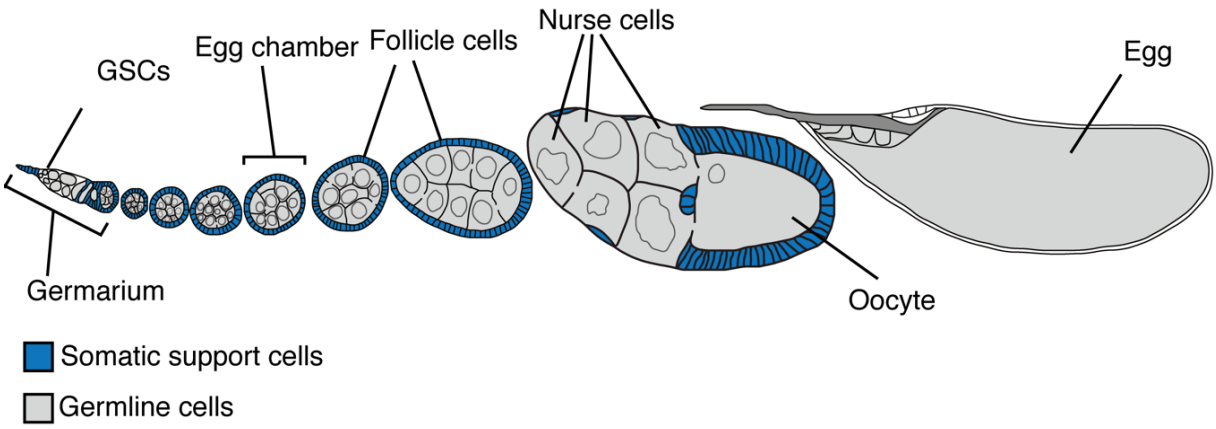


Figure 1.2. The fly ovary
Cartoon of a *Drosophila* ovariole with somatic cells in blue and germline cells in grey. Modified from (Handler et al., 2013)

I.3.1 piRNA biogenesis

The biogenesis of piRNAs is fundamentally different from that of other classes of small RNAs. The main features of piRNA biogenesis stem directly from the necessity of regulating transposons, which are a diverse pool and rapidly evolving. For instance, the piRNA pathway targets a wide range of transposon families with variable sequence content, therefore it cannot rely on discrete small RNA guides of a fixed sequence. This is substantially different from miRNAs, whose activity is directed towards well-defined targets and for which sequence changes would profoundly alter target specificity. Hence, unlike miRNAs, piRNAs cannot be collapsed to well-defined sequences nor be classified as families sharing the same seed. Furthermore, whenever a novel transposon invades the germline, the piRNA pathway must rapidly integrate it into the system and produce piRNAs against it. As a consequence, piRNA precursors (namely piRNA cluster transcripts) must have the capability to accommodate new transposon sequences while still retaining the ability to be recognised and processed. However, the molecular details of this process remain obscure.

piRNAs are typically produced as "pools" against the whole sequence of a TE transcript. This fuels PIWI proteins with a large reservoir of transposon-targeting small RNA guides that mount a robust silencing response (Han et al., 2015; Homolka et al., 2015; Mohn et al., 2015; Senti et al., 2015). Furthermore, PIWI proteins themselves sustain a feed forward amplification loop that produces more piRNAs against active elements (Brennecke et al., 2007; Gunawardane et al., 2007). Although different species adopt slightly dissimilar amplification methods, sequence diversification and signal amplification constitute a unifying feature of most known piRNA pathways. Fly piRNA biogenesis can be divided into two interconnected branches, each relying on a separate subset of proteins. While initially considered independent and dubbed "primary" and "secondary" biogenesis, recent work has shown that these two routes are indeed tightly interconnected (Han et al., 2015; Homolka et al., 2015; Mohn et al., 2015; Senti et al., 2015). Notably, piRNA biogenesis is orchestrated in a way that, if one branch fails, the other will try to compensate (Hayashi et al., 2016), further underlining the importance of effective piRNA production in the germline.

A first difference with other small RNA production pathways lies in the nature of piRNA precursor transcripts, which arise from discrete genomic loci known as "piRNA clusters". The products of piRNA clusters can range from a few to hundreds of kilobases (kb), although the precise size has not been determined yet, do not form characteristic secondary structures (e.g.

the hairpins of pri-miRNAs or the long dsRNA of siRNA precursors; compare **Figure 1.1 A-C, top**) and are processed as single-stranded RNAs (ssRNAs) (**Figure 1.1C**). The recognition of a piRNA precursor does not appear to depend on genetically-encoded features that are universally shared among all piRNA precursors, such as secondary structures or sequence motifs. Secondly, processing of ssRNA precursors starts after they have left the nucleus and mostly occurs on the surface of mitochondria, a very conserved feature in all animals in which piRNA biogenesis has been characterised thus far. Thirdly, piRNA biogenesis requires a large number of factors (Olivieri et al., 2012), is independent from Dicer proteins (Vagin et al., 2006) and, most importantly, involves Argonaute proteins themselves (Brennecke et al., 2007; Gunawardane et al., 2007). Therefore, PIWI-clade Argonautes are not only the effectors of RNAi but can also participate directly to the production of the small RNAs they associate with.

It has recently become apparent that piRNA production in *Drosophilids* is deeply connected with heterochromatin biology. Eukaryotic DNA is wrapped around histone proteins to form a higher-order structure called chromatin, which regulates the accessibility of the genetic information. Chromatin is classified into two different states, based on its cytological staining properties using basophilic aniline dyes and compaction levels: euchromatin, which appears loosely packed and weakly stained, and heterochromatin, which instead appears more compacted and strongly stained (Heitz, 1928; Kouzarides, 2007; Trojer and Reinberg, 2007). Euchromatin is typically associated with gene-rich, transcriptionally active genomic regions, preferentially located at the interior of the nucleus. Hallmarks of euchromatin are high levels of RNA pol II, Histone 3 Lysine 4 tri-methylation (H3K4me3) and histone acetylation (Kouzarides, 2007). Contrastingly, heterochromatin predominantly encompasses areas that contain fewer genes but are rich in repetitive sequences and is often found at the nuclear periphery. Heterochromatin is generally hypo-acetylated and can be further classified into constitutive heterochromatin, marked by H3K9me2/me3 and mostly covering centromeres, telomeres as well as other highly repetitive regions, and facultative heterochromatin, marked by H3K27me3 and often associated with developmentally regulated loci (Trojer and Reinberg, 2007; Wang et al., 2016). Although historically considered as a transcriptionally inert domain, heterochromatin has emerged to be dynamically regulated and active (Grewal and Jia, 2007; Wang et al., 2016), although often it does not follow the same rules of euchromatic regions. As will be discussed in the following paragraphs, flies have evolved a mechanism involving several factors not found outside *Drosophilids* that allows specific transcription from within heterochromatin and export of resulting RNA precursors.

I.3.1.1 Uni-strand piRNA clusters

piRNA clusters are discrete genomic loci that encode for the vast majority of piRNA precursor molecules. These loci are predominantly composed of inactive transposon copies and fragments rearranged in various orientations. Whenever a transposon sequence is inserted into a piRNA cluster, it can act *in trans* to silence other copies throughout the genome (Muerdter et al., 2012). Hence, these loci encode a genomic long-term memory of past transposon challenges and are fundamental components of an "innate", genetically-encoded immunity against TEs (Brennecke et al., 2007). piRNA clusters in *Drosophila* can be divided into two groups primarily based on their mode of transcription: uni-strand clusters and dual-strand clusters. Uni-strand clusters give rise to piRNAs from only one genomic strand and seem to be the default type of clusters in most species (Goriaux et al., 2014; Li et al., 2013). Only two uni-strand clusters exist in *D. melanogaster*: cluster *20A* and *flamenco* (*flam*), the latter being expressed exclusively in somatic follicle cells. Unlike dual-strand clusters, fly uni-strand loci resemble canonical RNA pol II transcriptional units in which they possess promoter regions marked by H3K4me3 and their transcripts undergo canonical mRNA processing, including capping, splicing and poly-adenylation (Goriaux et al., 2014). This allows their export via the canonical Nxf1(sbr)/Nxt1 pathway, which only acts on correctly processed transcripts (Dennis et al., 2016; Dennis et al., 2013; Goriaux et al., 2014; Kohler and Hurt, 2007; Zanni et al., 2013).

Mice and humans possess similar uni-strand clusters bearing the features of RNA pol II-transcribed genes that are predominantly expressed during the pachytene stage of male meiosis. Although some of their promoters can initiate divergent transcription, these clusters are still considered as "uni-strand" because the resulting transcripts do not overlap. Intriguingly, all these meiotic piRNA clusters are controlled by one conserved transcription factor, namely A-MYB (Li et al., 2013; Ozata et al., 2020). Since A-MYB promotes its own expression as well as that of several proteins participating to the piRNA pathway and that of piRNA clusters, its activation generates a feed-forward loop that sustains piRNA production during spermatogenesis. No such "master regulator" has been identified in flies thus far. Nematode clusters differ from flies and mammals, though they can still be classified as "uni-strand", strictly speaking. In *C. elegans*, 21U RNAs, which bind to the PIWI-clade Argonaute PRG-1, originate from very short transcription units clustered on two regions of chromosome IV. These clusters are driven by a conserved sequence motif (the "Ruby motif"), bound by Forkhead transcription factors, and are transcribed by RNA pol II with the aid of some specialised factors (Weick and Miska, 2014; Almeida et al., 2019). Notably, these piRNA precursor molecules are

only ~26nt long, have a conventional 5' cap but do not possess a 3' poly-A tail. Processing of their 5' end, including removal of the cap, enables loading into PRG-1, which is followed by resection of their 3' end up to their mature size of 21 nt by the exonuclease PARN-1.

Flam is the main source of piRNAs in fly somatic cells and was genetically characterized in the 1990s (before piRNAs were discovered), as the master regulator of *gypsy* retroelements (**Figure 1.3A**) (Desset et al., 2008; Mevel-Ninio et al., 2007; Pelisson et al., 1994; Prud'homme et al., 1995; Sarot et al., 2004). *Gypsy* is one of the few retroviruses active in flies (Kim et al., 1994) and, in permissive strains carrying deficiencies in *flam*, it can assemble into viral-like particles that are capable of infecting the oocyte, leading to catastrophic effects on germline development (Lecher et al., 1997; Pelisson et al., 1994). In fact, flies carrying mutations that disrupt *flam* only develop rudimentary ovaries (Mevel-Ninio et al., 2007). *Flam* is located on the peri-centromeric heterochromatin of chromosome X and carries several remarkable features. Firstly, the locus is entirely covered by the repressive mark H3K9me3, deposited by the methyl-transferase dSETDB1/Eggless (Egg) (Rangan et al., 2011). Lack of H3K9me3 deposition abolishes piRNA production from *flam*, implying a role for this mark in cluster definition (Rangan et al., 2011). Secondly, *flam* appears to produce an unusually long transcript of at least 180 kb, which is well above the average transcript length of fly genes (3,058 bp) (Adams et al., 2000) and that of some of the longest processed transcripts in flies (e.g. ~58 kb of *sls* mRNA). Due to the poor mappability of this highly repetitive locus, we cannot presently rule out that *flam* transcription encompasses the entire ~650 kb of transposon-rich heterochromatin downstream of the DIP1 gene. Whether a locus-specific mechanism sustains RNA pol II elongation to produce such a long RNA has not been investigated yet. Thirdly, like all piRNA clusters *flam* is entirely composed of transposon fragments, mostly *gypsy* family elements and other somatic transposons (that include *gypsy* itself, *ZAM*, *idefix*, *mdg1* and *Stalker*). However, nearly all transposon insertions within the *flam* locus are located on the minus strand, meaning that all piRNAs produced from this cluster will be antisense to active TEs (Brennecke et al., 2007; Zanni et al., 2013). This peculiar feature raises the question of how such sequence arrangement has evolved and whether the locus acts somehow as a strand-specific "trap" for active elements. The *flam* locus bears a canonical RNA pol II promoter whose expression was reported to be activated by the transcription factor Cubitus interruptus (ci) (Goriaux et al., 2014). The resulting transcript is capped, (alternatively) spliced and poly-adenylated. The splicing events occurring close to the 5' end recruit the Nxf1/Nxt1 heterodimer, which mediates export of *flam* transcripts

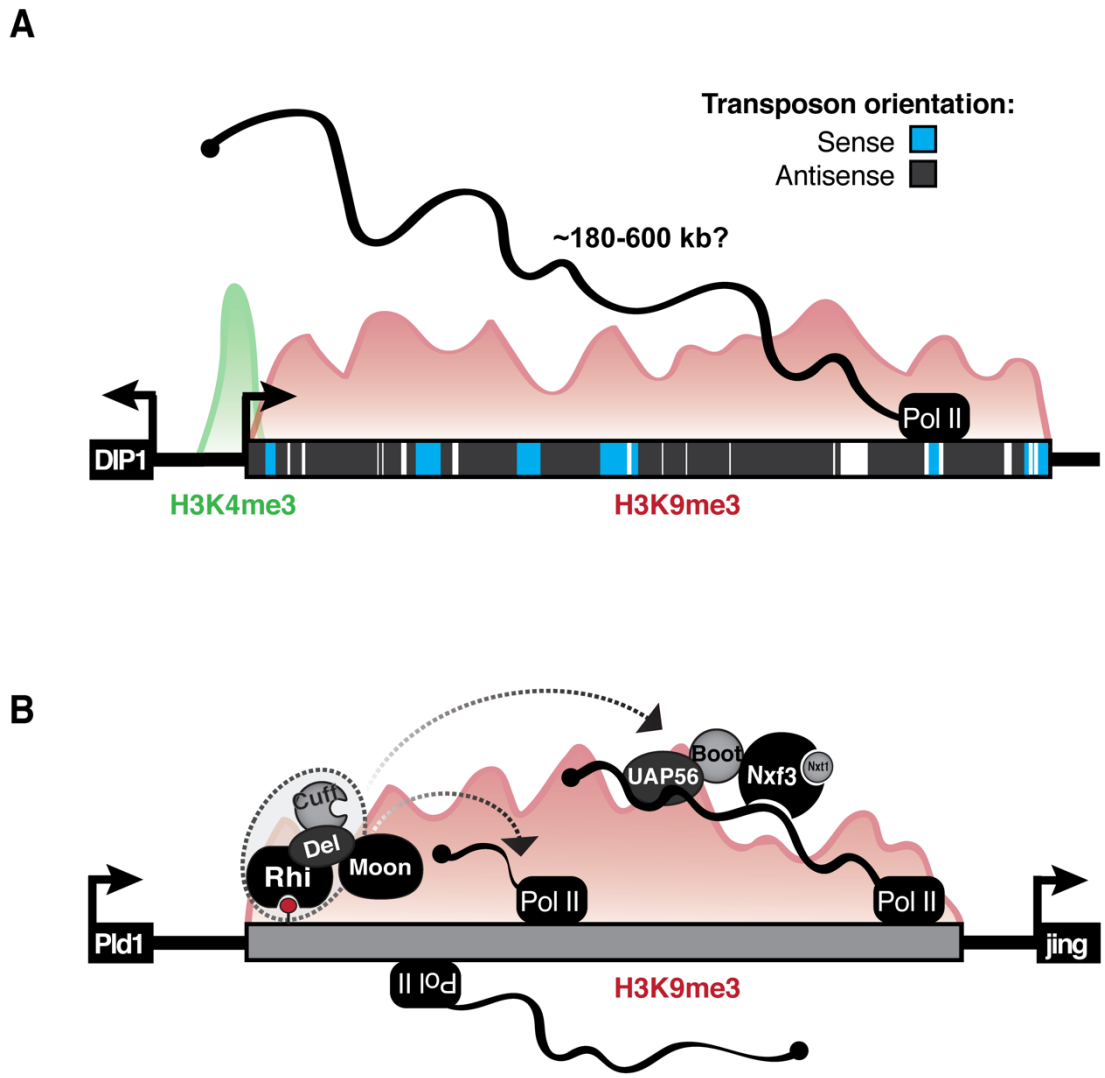


Figure 1.3. piRNA clusters are major piRNA source loci

A) *Flam*, the major uni-strand piRNA cluster in Drosophila, encompasses up to 600 kb of peri-centromeric heterochromatin and is transcribed by RNA pol II. It features a canonical promoter marked by H3K4me3, whilst the body of the cluster is decorated by H3K9me3. The vast majority of transposon fragments within the cluster are inserted in an antisense orientation. **B)** Dual-strand piRNA clusters (exemplified by *42AB*) are decorated by H3K9me3 and are expressed via a specialised machinery. The RDC (Rhino-Deadlock-Cutoff) complex licenses the clusters and recruits Moon, which in turn triggers non-canonical transcription via pol II. Export of these transcripts relies on UAP56, Boot, Nxf3 and Nxt1.

out of the nucleus, via the same route followed by mRNAs (Dennis et al., 2016). Exported *flam* transcripts accumulate in peri-nuclear foci (namely, *flam* bodies or Dot COM), which are adjacent to and partially overlapping with Yb bodies (Dennis et al., 2013; Murota et al., 2014). The transcript seems to localise asymmetrically within these foci, with the 5' end anchored to the nuclear lamina and the remainder of the RNA protruding towards the Yb bodies (Sokolova et al., 2019). Recent work has established a link between the Yb bodies and *flam* export (Dennis et al., 2019; Hirakata et al., 2019; Sokolova et al., 2019), although a detailed molecular understanding of this specificity remains unknown.

I.3.1.2 Dual-strand piRNA clusters

Dual-strand piRNA clusters are the major source of piRNAs in *Drosophila* germ cells and are defined by the production of piRNAs from both genomic strands (**Figure 1.3B**). Dual-strand transcription of small RNA precursors is not unique to *Drosophilids* and is found for instance in yeast and ciliates, where it produces dsRNA substrates for Dicer cleavage (Holoch and Moazed, 2015; Noto and Mochizuki, 2017). In other cases, such as plants, nematodes and yeast, a double-stranded intermediate is produced by RNA-dependent RNA polymerases (RdRPs) (Holoch and Moazed, 2015; Almeida et al. 2019) acting on ssRNA targets. In contrast with other small RNA pathways, flies rely neither on RdRPs to produce dsRNA precursors nor on Dicer proteins to process them. Nonetheless, dual-strand transcription has a prominent role in licensing piRNA fly precursors and this occurs via a dedicated molecular machinery. Unlike uni-strand clusters, small RNAs originating from dual-strand precursors can base-pair with the original transcript. This property provides the basis for a positive feedback loop that specifies the cluster as such and at the same time enables production of additional small RNAs (Holoch and Moazed, 2015; Ozata et al., 2019).

Drosophila dual-strand clusters are embedded in heterochromatic regions of the genome and are decorated by H3K9me3 (Mohn et al., 2014; Rangan et al., 2011), an environment which is typically associated with repeat-rich, highly compacted and often transcriptionally inert areas (Wang et al., 2016). Furthermore, dual-strand clusters lack canonical RNA pol II promoters marked by H3K4me3. Nonetheless, they are actively transcribed by RNA pol II, thanks to an extremely specialised machinery that licenses their transcription and export outside of the nucleus. It is interesting to note that key genes involved in dual-strand cluster biology are unique to *Drosophilids* and have undergone strong positive selection, likely due to the selective pressure of TE mobilisation (Parhad et al., 2017; Vermaak et al., 2005). A similar intriguing

connection between heterochromatin biology and transcription of small RNA precursors can also be found in plants and fission yeast. In *S. pombe,* heterochromatic centromeric repeats are transcribed into dual-strand precursors that can be processed by Dicer. This transcription does not seem to rely on a specialised machinery but rather exploits global chromatin de-condensation during G1/S phase of the cell cycle (Chen et al., 2008). On the other hand, plant small RNA precursors are transcribed from heterochromatic loci by RNA polymerase IV (pol IV), which is recruited by the H3K9me reader SHH1 (Holoch and Moazed, 2015). Small RNAs originating from these loci help maintaining the heterochromatic marks, which in turn recruit more SHH1 and pol IV. A similar positive feedback loop ensures continuous dual-strand cluster licensing and processing in *Drosophila.*

The specification of dual-strand piRNA cluster loci and transcript licensing centre on HP1D/Rhino (hereafter referred to as Rhi), a germline-specific homolog of the ubiquitous Heterochromatin Protein 1A (HP1A) (Klattenhoff et al., 2009; Vermaak et al., 2005; Vermaak and Malik, 2009; Zhang et al., 2014b). Like all members of the HP1 family, Rhi has an amino-terminal chromodomain, responsible for binding to H3K9me3 (Bannister et al., 2001; Lachner et al., 2001), a central hinge region, and a carboxy-terminal chromoshadow domain, that typically mediates homo- and hetero-dimerization (Vermaak and Malik, 2009). Rhi has some unique features compared to the other HP1 paralogs in *Drosophila* (HP1A, B, C and E). It is exclusively expressed in the germline and localises to discrete nuclear foci within heterochromatin (Klattenhoff et al., 2009; Mohn et al., 2014; Vermaak et al., 2005; Zhang et al., 2014b). The amino acid sequence of Rhi shows strong signs of positive selection and thus appears to be evolving at an unusually fast pace. This is especially true in its hinge region, which cannot be uniquely aligned between different *Drosophila* species groups (Vermaak et al., 2005). Because of these peculiar characteristics, it had been suggested that Rhi is involved in the genetic conflict against transposable elements (Vermaak et al., 2005). In fact, Rhi is a fundamental player of piRNA-mediated genome defence, since its binding to H3K9me3 is necessary and sufficient to identify a certain locus as a dual-strand piRNA cluster (Klattenhoff et al., 2009; Mohn et al., 2014; Vermaak et al., 2005; Zhang et al., 2014b). Despite heterochromatin representing up to 20% of the fly genome (Hoskins et al., 2002), Rhi is only present at very discrete genomic loci (making up ~1-2% of the genome) (Mohn et al., 2014), but what drives this specificity remains a mystery.

Rhi is the chromatin-bound core component of a multi-protein complex (namely RDC complex, **Figure 1.3B**), which also includes Deadlock (Del) and Cutoff (Cuff) (Chen et al., 2016; Le

Thomas et al., 2014; Mohn et al., 2014; Pane et al., 2011; Zhang et al., 2014b). Within RDC, Del appears to function as an adaptor protein recruiting other effectors to Rhi foci. Notably, *D. simulans* Del has been shown to be functionally incompatible with *D. melanogaster* Rhi, and vice-versa, providing another example of the rapid evolutionary turnover of these genes (Daugherty and Malik, 2012; Parhad et al., 2017; Yu et al., 2018). A critical effector brought to dual strand clusters via RDC is Moonshiner (Moon), an ovary-specific paralog of the canonical RNA pol II initiation factor TFIIA-L (Andersen et al., 2017). Moon interacts with Del and in turn recruits the TATA-box binding protein-related factor 2 (TRF-2) to dual-strand cluster loci, thus allowing RNA pol II transcription initiation within those inaccessible heterochromatic loci that lack canonical promoters. Transcription initiation driven by Moon can occur from any pyrimidine/purine (YR) dinucleotide within the piRNA cluster and is not biased to either genomic strand, thus explaining the bi-directionality of these clusters (Andersen et al., 2017). The peculiarities of dual-strand piRNA clusters do not end at transcription initiation, as they give rise to 'non-canonical' transcripts that are capped but neither undergo splicing nor polyadenylation. Cap, splicing and poly-A tails are three hallmarks of correctly processed RNA pol II transcripts that are export-competent, hence their lack typically leads to RNA clearance via nuclear surveillance systems (Tutucci and Stutz, 2011). However, dual-strand clusters evade nuclear decay mechanisms via yet another specialised protein complex recruited by Del: Nxf3, Nxt1 and Bootlegger (Boot) (ElMaghraby et al., 2019; Kneuss et al., 2019) (**Figure 1.3B**). Nxf3 is a homolog of the general mRNA export factor Nxf1 and has been co-opted to escort non-canonical piRNA precursor transcripts out of the nucleus, escaping the canonical RNA quality-control pathways. Nxt1 binds to the NTF2 fold of NXF family proteins, such as Nxf1 and Nxf3, and, at least in the case of Nxf1, mediates their interaction with the nuclear pore complex (NPC) (Braun et al., 2001; Fribourg et al., 2001). Boot lacks readily identifiable domains and interacts with Del, thus bridging the Nxf3-driven export complex to Rhi-bound loci (ElMaghraby et al., 2019; Kneuss et al., 2019). Export of dual-strand piRNA clusters also relies on other ubiquitous factors of the mRNA export machinery, such as the DEAD-box helicase UAP56 and the THO complex (Zhang et al., 2012; Zhang et al., 2018a). Interestingly, all these proteins display the same punctate localisation to Rhi foci within nurse cell nuclei. The third component of RDC, Cuff, is a homolog of the yeast de-capping factor Rai1 (Pane et al., 2011) and is also recruited via Del. Cuff is thought to perform various functions, such as cap binding to shield nascent transcripts from the degradation machinery, suppressing the recruitment of the cleavage and poly-adenylation complex as well as that of the TREX complex (Chen et al., 2016; Mohn et al., 2014; Pane et al., 2011; Zhang et al., 2012; Zhang et al., 2014b). It is noteworthy that the stunning specialisation of the dual-strand cluster

transcription machinery has often been achieved via repurposing paralogs of genes involved in other cellular pathways. Strikingly, Rhi, Cuff, Moon and Nxf3 are all related to proteins involved in the canonical mRNA transcription and export processes but have specialised to allow licensing of heterochromatic sites otherwise inaccessible.

Mutants in RDC components are unable to produce piRNAs from dual-strand clusters. Nonetheless, in these mutant backgrounds another class of ectopic piRNAs coming from mRNAs becomes apparent (Mohn et al., 2015). Studies in Ovarian Somatic Cells (OSCs), a cell line derived from the somatic cells of the ovary that possesses a functional piRNA pathway (Niki et al., 2006; Saito et al., 2009), have also identified this additional class of piRNA precursors. In fact, a fraction of somatic piRNAs derives from the 3'UTR of protein-coding genes, among which the most abundant is *traffic-jam* (*tj*) (Robine et al., 2009; Saito et al., 2009). Interestingly, piRNA production from protein-coding genes does not seem to correlate with their expression levels, implying that this is not a spurious processing of abundant transcripts but instead is somehow regulated. A sequence in the 3'UTR of *tj* has been implicated in triggering piRNA production (Ishizu et al., 2015), however, no systematic studies have been performed to generalise this finding to other transcripts. Furthermore, these so-called "genic piRNAs" are not complementary to transposable elements, therefore leaving open the question of their functional role. Production of piRNAs from the 3'UTR of coding transcript has been reported in other animals, such as mouse, but their function remains equally mysterious (Robine et al., 2009).

I.3.1.3 piRNA precursor specification

Recognising *self*- from *non-self* nucleic acids and selecting the correct substrates for piRNA production is a task of greatest importance. Studies so far have failed to identify one secondary structure or sequence motif that is shared among all piRNA precursors and could explain their recognition via a unified mechanism. One can reason that a conserved, genetically-encoded feature within piRNA clusters can easily be disrupted by novel transposon insertions and is therefore unlikely to be an evolutionarily successful strategy. Even though theoretically any cellular transcript can be processed into piRNAs, promiscuous processing is relatively rare, which implies that a selection step operates somewhere within the piRNA system. Trigger sequences able to drive piRNA production from reporters have been identified for *flam* and *tj* (Homolka et al., 2015; Ishizu et al., 2015; Pandey et al., 2017). However, these signals do not

seem to be present in other piRNA source loci such as *20A* or dual-strand clusters, suggesting that they might be locus-specific features.

The cell's ability to distinguish piRNA precursors from other cellular transcripts can be partly explained by their preferential localisation to specialised peri-nuclear, electron-dense structures: the nuage (in germ cells) and the Yb bodies (in somatic cells) (**Figure 1.4**). The nuage is a ring-like structure surrounding nurse cell nuclei that is marked by the presence of the PIWI proteins Aub and Ago3 and the germline-specific DEAD-box helicase Vasa (Hay et al., 1990; Lasko and Ashburner, 1990; Liang et al., 1994; Lim and Kai, 2007; Malone et al., 2009). Interestingly, Vasa accumulates in foci on the cytoplasmic side of the nuclear envelope that directly face nuclear foci of UAP56, Nxf3 and Boot (where cluster transcripts are exported) and overlaps with the cytosolic foci of Nxf3 and Boot (ElMaghraby et al., 2019; Kneuss et al., 2019; Zhang et al., 2012). It is noteworthy that mutants for dual-strand cluster factors disrupt Aub and Ago3 localisation to nuage (Czech et al., 2013; Klattenhoff et al., 2009; Pane et al., 2011; Zhang et al., 2012), implying a direct link between cluster transcription and export and nuage assembly. The somatic Yb bodies are instead named after their main component, the DEAD box helicase Yb, and are adjacent to the *flam* bodies (Ishizu et al., 2015; King et al., 2001; Murota et al., 2014; Qi et al., 2011; Saito et al., 2010; Szakmary et al., 2009). Similarly to nuage, perturbations of *flam* transcription and export impair the formation of Yb bodies (see **Chapter IV**) (Dennis et al., 2016). The importance of correct subcellular localisation of piRNA precursors is confirmed by the fact that tethering of artificial reporter transcripts to nuage/Yb bodies components is sufficient to induce processing into *bona-fide* piRNAs (Pandey et al., 2017; Rogers et al., 2017). Phase-separated granules of RNA and protein are also found in many small RNA pathways and likely improve the efficiency of the associated molecular processes. Examples of this are nuage in mice and zebrafish germ cells and P granules in worms (Ozata et al. 2019).

The molecular mechanism that links precursor localisation to piRNA production has been mostly elucidated in nuage (**Figure 1.4, left**). Here, Aub and Ago3 in complex with a piRNA are thought to scan all transcripts leaving the nucleus, cleaving any with complementary sequence (Han et al., 2015; Mohn et al., 2015; Senti et al., 2015; Wang et al., 2015). In this way, when a transposable element is actively transcribed and exported through the nuclear envelope, it can be readily identified by a complementary piRNA. Similarly, cluster transcripts can be immediately recognised via sequence complementarity. This base-pairing triggers the slicing activity of target-engaged PIWI proteins (Aub and Ago3) that cleave the transcript and

release 3' products carrying a monophosphate at their 5' end (5'-P). Generation of a 5'-P intermediate seems to be the first step of nearly all piRNA production routes, since only 5'-P RNAs can be bound by PIWI proteins. The so-called "ping-pong" looping centred on PIWI proteins operates this initial cleavage in mice and flies, and likely in most animals showing a ping-pong signature, which will be discussed later. An exception to this is *C.elegans*, where the mechanism removing the 5' cap from 21-U precursor RNAs to produce 5'-P intermediates is not yet fully clarified (Ozata et al., 2019).

Aub- and Ago3-mediated target slicing has two consequences: first, transposon mRNAs are cleaved and therefore neutralised immediately after leaving the nucleus (Brennecke et al., 2007; Gunawardane et al., 2007). Second, the 5'-P RNA fragment is consumed in the production of more piRNAs via the ping-pong cycle and the mitochondrial biogenesis machinery, both discussed later (Brennecke et al., 2007; Gunawardane et al., 2007; Han et al., 2015; Homolka et al., 2015; Mohn et al., 2015). This specification mechanism has the substantial advantage that anything partially complementary to an Aub-/Ago3-bound piRNA can be recognised and cleaved. Since transposon sequences are largely repetitive and fast evolving, this confers the ability to target even a new invading element that has not yet integrated into a piRNA cluster but carries sufficient sequence complementarity to a pre-existing one. The active element itself can become a substrate for further piRNA production, thus reinforcing the silencing response. This can be defined as an "adaptive" branch of piRNA-mediated immunity, because it can identify unknown *non-self* sequences and mount a silencing response against them. Precursor specification in somatic cells is still poorly understood (**Figure 1.4, right**). Although recent work (including that described in **Chapter IV**) has assigned a pivotal role to Yb, the molecular mechanisms are still largely unclear (Hirakata et al., 2019; Ishizu et al., 2015; Ishizu et al., 2019; Pandey et al., 2017). Since somatic cells lack Aub and Ago3 (Saito et al., 2009), and Piwi's catalytic activity is dispensable for transposon silencing in these cells (Darricarrere et al., 2013; Saito et al., 2010; Sienski et al., 2012), PIWI-mediated precursor slicing cannot explain substrate selection in this compartment of the ovary. It is likely that 5'-P RNA intermediates are equally present in Yb bodies but are generated via a slicing-independent mechanism, for example by a yet to be identified endonuclease. Since *flam* is the predominant source of somatic piRNAs, one can hypothesize that a distinct specification mechanism has evolved to enable processing of this unique gene.

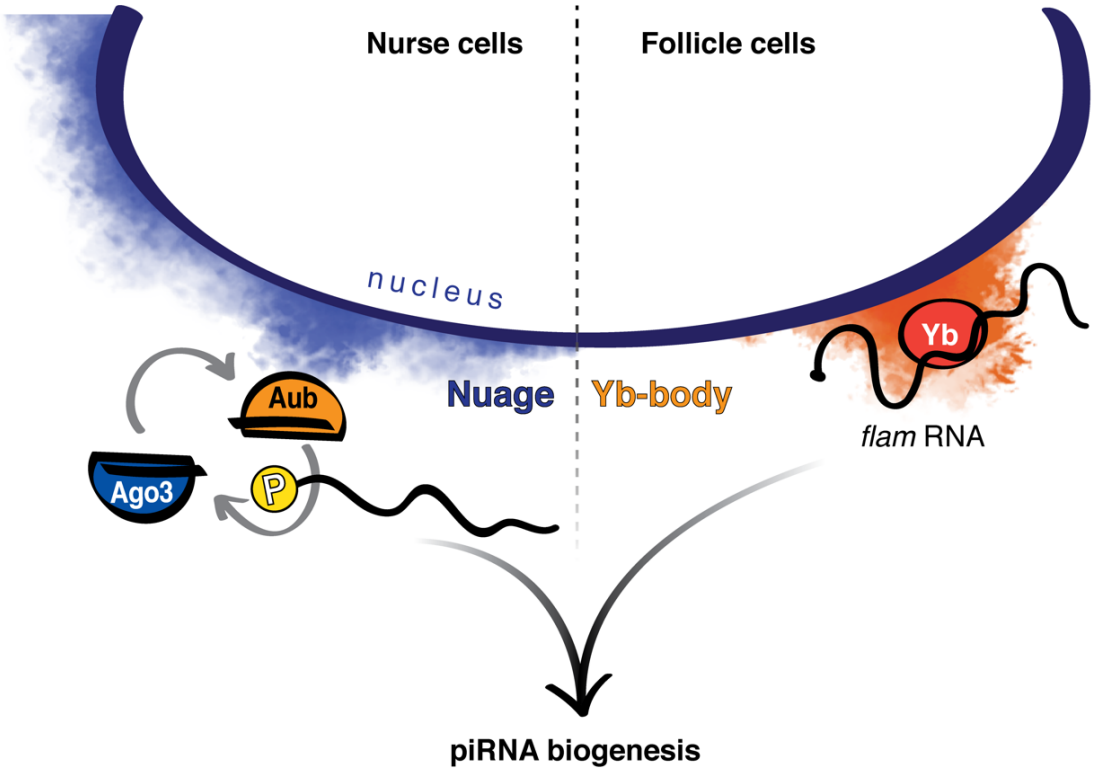


Figure 1.4. Precursor specification in nuage/Yb bodies
Long, single-stranded piRNA precursors are specified for piRNA biogenesis in peri-nuclear subcellular compartments: nuage (in nurse cells) or Yb bodies (in somatic follicle cells). While the Aub/Ago3 ping-pong cycle generates 5'-P precursors in nuage, the molecular mechanisms of precursor specification in Yb bodies are still unknown.

I.3.1.4 Mitochondrial processing of pre-piRNAs

Once recognised in nuage/Yb bodies, 5'-P piRNA precursors (pre-piRNAs) undergo processing into ~25-nt long, silencing-competent piRNAs. Processing of long ssRNA precursors into mature piRNAs occurs on the outer surface of mitochondria. This distinctive subcellular localisation of the piRNA biogenesis machinery is remarkably conserved across animals, ranging from the fruit fly, to mosquitoes, silkworm and even mouse (Ipsaro et al., 2012; Nishida et al., 2018; Nishimasu et al., 2012; Shiromoto et al., 2013; Watanabe et al., 2011; Zhang et al., 2016). In *Drosophila,* several proteins implicated in piRNA biogenesis are anchored to the outer membrane of mitochondria and some of them share similarities with mitochondrial lipid-biosynthesis factors, further underscoring the link between these organelles and transposon defence (Czech et al., 2013; Handler et al., 2013; Munafò et al., 2019; Vagin et al., 2013). However, it remains unclear why mitochondria have been preferentially chosen for piRNA biogenesis and whether they are functionally involved in this process or simply act as physical platforms. Mitochondria are physically separated from nuage, where piRNA precursor specification occurs (Ge et al., 2019), thus implying that precursor transcripts must be actively transported from one site to the other.

Genetic and structural studies have placed the mitochondria-anchored endonuclease Zucchini (Zuc) at the heart of the piRNA biogenesis machinery (Haase et al., 2010; Ipsaro et al., 2012; Nishimasu et al., 2012; Pane et al., 2007). The crystal structure of mouse and fly Zuc shows remarkable similarities to the Phospholipase D (PLD) family of phosphodiesterases, some of which can act as nucleases (Selvy et al., 2011). *In vitro* recombinant Zuc can indeed cleave single-stranded nucleic acids, either RNA or DNA, but does not hydrolyse double-stranded nucleic acids or lipids (Ipsaro et al., 2012; Nishimasu et al., 2012). PLD family nucleases feature two HKD (Histidine-Lysine-Aspartic Acid) motifs that are brought together to assemble the catalytic centre. Zuc has one HKD motif and forms a stable dimer *in vivo*, thus the catalytic core is formed inter-molecularly. Mutations in the Histidine (H) of Zuc's HKD domain abolish its catalytic activity and cause a collapse in piRNA production (Haase et al., 2010; Ipsaro et al., 2012; Nishimasu et al., 2012; Pane et al., 2007). Altogether, these data point towards a two-step reaction mechanism for dimeric Zuc, resembling that of PLD-family nucleases (Gottlin et al., 1998; Stuckey and Dixon, 1999). According to the proposed mechanism, the RNA substrate forms a short-lived covalent bond with the Histidine of one Zuc monomer, which is subsequently hydrolysed by the Histidine of the other monomer. *In vivo*, Zuc shows a remarkable specificity and processivity, cleaving at regular intervals of ~25 nt and generating

strings of adjacent piRNAs (known as "trail piRNAs", **Figure 1.5A-B**) from each substrate molecule (Han et al., 2015; Mohn et al., 2015). Trail piRNAs are arranged in a tail-to-head fashion, meaning that the 3' end of one piRNA and the 5' end of the following one are generated via the same cleavage event, a signature known as "phasing" (see **Figure 1.5B**). Since this signature has been identified in a broad range of animals, it is likely that coupling of PIWI-mediated slicing to phased piRNA production is evolutionarily conserved (Gainetdinov et al., 2018). Furthermore, the vast majority of Zuc-generated piRNAs start with a uridine, a property referred to as "1U bias". Processing of reporter constructs shows that the position of uridines within the transcript is predictive of where Zuc cleavages will occur (**Figure 1.5B**) (Mohn et al., 2015). In contrast with what is observed *in vivo*, cleavage assays with recombinant Zuc fail to recapitulate this specificity and instead only produce a ladder of products of variable size, lacking any 5' nucleotide bias (Ipsaro et al., 2012; Nishimasu et al., 2012). A recent study in silkworm reports that Zuc-mediated production of piRNAs can be recapitulated *in vitro* when using purified mitochondria extracts, rather than recombinant protein, and is enhanced upon over-expression of Armi (Izumi et al., 2020). This strongly suggests the existence of one or several "co-factor(s)" that confer the distinctive cleavage periodicity and 1U bias *in vivo*. Various recent studies, including **Chapter III** of this thesis, have investigated whether PIWI proteins or other factors play a role in guiding Zuc cleavages (Gainetdinov et al., 2018; Ge et al., 2019; Ishizu et al., 2019; Munafò et al., 2019; Rogers et al., 2017; Stein et al., 2019; Yamashiro et al., 2019). The first nucleotide of a small RNA resides in a conserved binding pocket of the MID domain of PIWI proteins, known as the Specificity Loop ("SL"). Piwi carrying various mutations in the SL displays similar loading of 1U piRNAs (Stein et al., 2019), implying that the nucleotide bias is determined upstream of Piwi binding. Another recent study has shown that binding of PIWI proteins to the 5'end of a pre-piRNA positions Zuc to cleave at the first accessible uridine (Gainetdinov et al., 2018). This implies that the footprint of a PIWI protein is predictive of the size of the corresponding piRNAs, and indeed that is the case in mice and flies. Likewise in *C.elegans* generation of mature 21-U RNAs involves trimming of their 3' ends based on the footprint of the PIWI homolog PRG-1 (Tang et al., 2016). Furthermore, the work presented in this thesis shows that the RNA helicase Armitage (Armi) localises to mitochondria in close proximity to dimeric Zuc, arguing for Armi's key contribution to Zuc cleavage pattern, perhaps via its ATP-dependent helicase activity (Ge et al., 2019; Ishizu et al., 2019; Izumi et al., 2020; Munafò et al., 2019; Yamashiro et al., 2019). These data provide an explanation for the regular spacing of Zuc cleavages, whereas the origin of the 1U bias remains unclear.

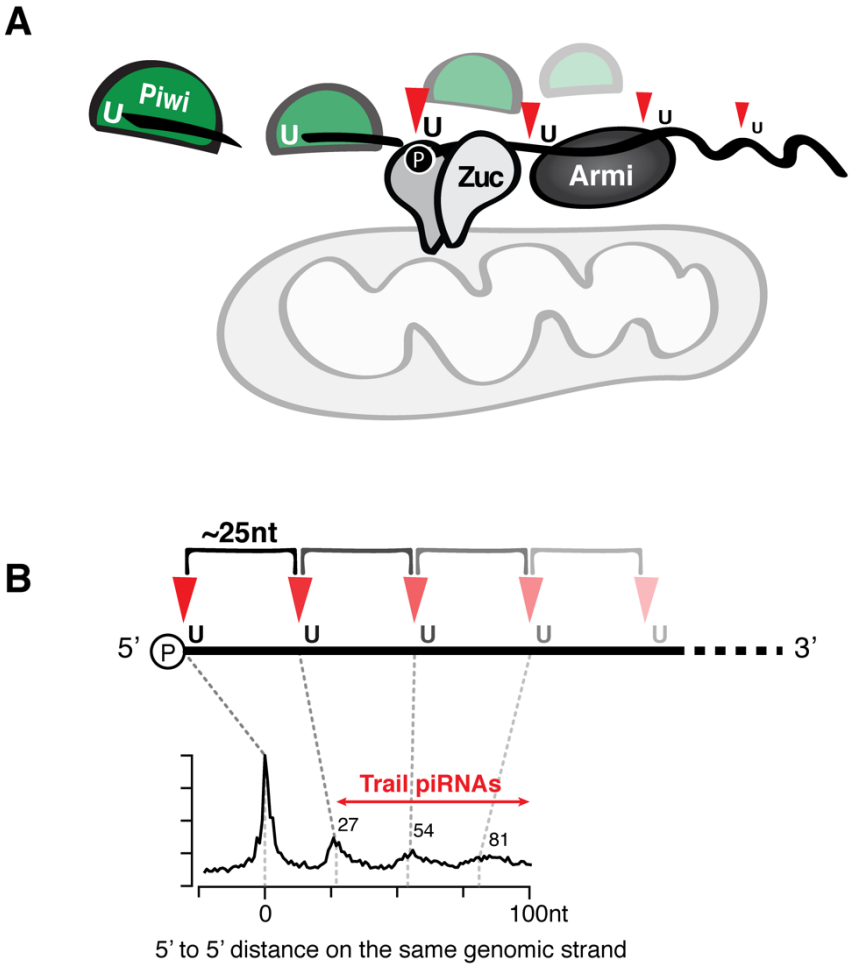


Figure 1.5. Zuc-mediated phased piRNA biogenesis

A) piRNA precursors are bound by Armi and translocated to mitochondria, where the transmembrane nuclease Zuc processes them into mature piRNAs. Piwi binds to the 5'-P of the precursors and Zuc cleaves at the first accessible uridine, thus generating a characteristic footprint of ~25nt and 1U bias. **B)** From a single precursor, Zuc releases strings of consecutive tail-to-head piRNAs that are ~25nt long and mostly begin with a uridine (trail piRNAs). This characteristic cleavage pattern is referred to as "phasing" of piRNA biogenesis. Modified from (Mohn et al., 2015).

Genetic screens performed in *Drosophila* have provided a general framework for piRNA biogenesis and have identified most of the involved players (Czech et al., 2013; Handler et al., 2013; Handler et al., 2011; Muerdter et al., 2013; Olivieri et al., 2010). Nonetheless, the mechanistic understanding of the molecular events leading to piRNA production is not complete yet. Several piRNA biogenesis factors share Zuc's localisation to the outer membrane of mitochondria. Among these, Gasz, a germ cell protein with ankyrin repeats, a sterile-alpha motif and a leucine zipper, and its partner Daedalus (Daed), also carrying a sterile alpha motif and a coiled-coil, are implicated in the recruitment of Armi to mitochondria (Handler et al., 2013; Munafò et al., 2019; Yamashiro et al., 2019) (see **Chapter III**). Intriguingly, loss of either Gasz or Daed affects mitochondrial morphology, thus strengthening the link between the piRNA biogenesis machinery and mitochondria. Minotaur is a glycerol-3-phosphate acyltransferase whose catalytic activity is dispensable for piRNA biogenesis, yet absence of the protein itself abolishes piRNA production (Vagin et al., 2013). Partner of PIWIs (Papi) and TDRKH are instead involved in 3' trimming of piRNAs in mouse and silkworm, yet their activity seems neglectable in flies (Hayashi et al., 2016; Honda et al., 2013; Liu et al., 2011; Nishida et al., 2018; Zhang et al., 2018b).

In addition to the mitochondrial proteins mentioned above, piRNA production requires a plethora of cytosolic factors. Among these is Armi, an RNA helicase of the Upf1 family, which binds to piRNA precursors in nuage/Yb bodies and delivers them to mitochondria (Ge et al., 2019; Ishizu et al., 2019; Munafò et al., 2019; Vourekas et al., 2015). Armi typically shuttles between the two compartments and is anchored on the mitochondrial surface via Gasz and Daed (Munafò et al., 2019). Armi's function requires ATP binding and hydrolysis, as mutants defective for either activity lose their sub-cellular localisation and are unable to sustain piRNA production (Ge et al., 2019; Ishizu et al., 2019; Pandey et al., 2017). Most importantly, ATPase mutants show promiscuous RNA binding and inability to recognise the correct substrates. Nonetheless, Armi binding to a transcript is necessary and sufficient to feed an RNA to the mitochondrial piRNA biogenesis machinery, as shown *in vivo* and via reporter assays (Ishizu et al., 2019; Pandey et al., 2017; Rogers et al., 2017). Overall, Armi is essential both in somatic and in germ cells and Zuc-mediated piRNA production collapses in its absence. This, together with its mitochondrial localisation in close proximity to the biogenesis machinery, places Armi at a critical juncture of piRNA production, possibly contributing to substrate selection and to the processivity of Zuc cleavages.

In contrast to Armi, Yb, another RNA helicase of the DEAD box family carrying a Tudor domain, is specifically required in the soma. It has been initially identified as an essential gene

for germline and somatic stem cell division (King et al., 2001; Szakmary et al., 2009) and later on linked to transposon regulation (Olivieri et al., 2010; Qi et al., 2011; Saito et al., 2010). Yb is required for somatic piRNA production and functions upstream of all known mitochondrial and cytosolic biogenesis factors (Hirakata et al., 2019). Upon loss of Yb, Zuc-/Armi-driven piRNA production still occurs to some extent, but also involves promiscuously selected substrates, since Armi binding specificity is lost (**Chapter III** and **IV**) (Ishizu et al., 2019). This suggests that Yb selects the correct substrates for downstream processing, whereas Armi's role is that of participating in the actual production step in close association to Zuc. Yb typically forms discrete peri-nuclear foci, namely Yb bodies, which have originally been identified as electron-dense spheres of ~1.5 μm in diameter, often proximal to mitochondria and to RNA-rich foci of comparable size (Szakmary et al., 2009). More recently, Yb bodies have been show to form dynamic, phase-separated condensates thanks to protein-protein and protein-RNA interactions (Hirakata et al., 2019). The adjacent RNA-rich foci have instead been shown to correspond to *flam* RNAs, tightly bound by Yb, and therefore termed "*flam* bodies" (Ishizu et al., 2015; Murota et al., 2014). Various lines of evidence point towards an intimate relationship between the Yb bodies and the export of *flam* precursors, which happens via the canonical Nxf1/Nxt1 route (Dennis et al., 2016; Dennis et al., 2019; Dennis et al., 2013; Hirakata et al., 2019; Sokolova et al., 2019). *Flam* processing specifically depends on the helicase domain (HelC) of Yb, whereas all somatic precursors, including *flam*, require the RNA-binding domain and the Tudor domain, with the latter involved in binding to Armi (Hirakata et al., 2019).

Two proteins sharing similar domain structures to Yb, including the Tudor domain, have been implicated in somatic and germline piRNA production, namely Sister of Yb and Brother of Yb (SoYb, BoYb) (Handler et al., 2011). The Tudor domain, which facilitates protein-protein interactions often via the post-translational modification symmetric di-methyl Arginine (sDMA), is frequently found in piRNA biogenesis factors. Some fly PIWI-family proteins possess sDMA residues (Kirino et al., 2009), therefore Tudor proteins are thought to act as a "scaffold" that coordinates molecular interactions within the piRNA pathway (Vagin et al., 2009a; Vagin et al., 2009b). In addition to the previously cited Tudor domain-containing proteins (Papi, Yb, SoYb and BoYb), there are also the germline-specific Spindle-E, Krimper, Tapas and Tejas (Kennerdell et al., 2002; Lim and Kai, 2007; Patil et al., 2014; Patil and Kai, 2010; Sato et al., 2015; Webster et al., 2015), which are involved in the ping-pong cycle. Vreteno, also carrying a Tudor domain, is instead active in both compartments of the ovary (Handler et al., 2011; Zamparini et al., 2011).

The vast majority of Zuc-generated piRNAs are loaded into the effector protein Piwi (Han et al., 2015; Mohn et al., 2015; Senti et al., 2015; Wang et al., 2015). Piwi binding to piRNAs requires several factors, such as the co-chaperone Shutdown (Shu) (Olivieri et al., 2012; Preall et al., 2012), and is necessary to stabilise the protein itself. If mitochondrial piRNA production is impaired, unloaded Piwi is unstable and targeted for degradation. Indeed, a hallmark shared among all piRNA biogenesis mutants is a striking loss of nuclear Piwi (Malone et al., 2009; Olivieri et al., 2012). Once a piRNA is associated with Piwi, the protein is thought to undergo a conformational change that exposes a Nuclear Localisation Signal (NLS) and allows translocation into the nucleus via Importin-α (Yashiro et al., 2018). Although no crystal structure of *Drosophila* Piwi is available yet, this is the most accredited model to explain Piwi shuttling between the two subcellular compartments upon piRNA loading. Once in the nucleus, Piwi silences transposons by triggering the deposition of repressive histone marks, as will be discussed in the next paragraph. Because Piwi-RISC recognises nascent TE transcripts via base-pairing, Zuc-generated piRNAs ultimately determine the potency of the transcriptional silencing response. The generation of antisense trail piRNAs in a 5' to 3' directed fashion results in a pool of small RNAs that span a large portion of the locus and are progressively closer to the Transcriptional Start Site (TSS). Hence, the sole mechanism of phased biogenesis enforces Piwi-guided transposon silencing via diversifying the pool of transposon-targeting sequences (Senti et al., 2015; Wang et al., 2015).

I.3.2 piRNA-guided transposon silencing

piRNA-guided silencing targets transposable elements at the transcriptional and post-transcriptional levels via two different mechanisms. Post-transcriptional gene silencing (or PTGS) is a cytosolic event whereby Aub and Ago3 cleave active transposon mRNAs and thus prevent their translation into proteins that could drive TE mobilisation. Because Aub and Ago3 are only expressed in the germline, PTGS is restricted to nurse cells (Li et al., 2009a; Malone et al., 2009). As explained earlier, the cleavage products of PTGS directly stream into Zuc-mediated biogenesis of piRNAs bound to Piwi, which in turn is the key effector of nuclear co-transcriptional silencing (TGS). TGS is in place in both cell types of the ovary and initiates with Piwi-RISC recognising a nascent transposon RNA and inducing a cascade of events that culminates in the epigenetic silencing of the locus. Since the production of the germline Piwi-bound piRNA pool directly depends on Aub and Ago3, PTGS and TGS robustly wire transposon sequences into a self-amplified silencing loop.

I.3.2.1 The ping-pong cycle

Upon sequencing of *D. melanogaster* small RNAs bound to each PIWI protein, it was immediately apparent that each bound a distinct subset of piRNAs. Aub- and Piwi-loaded piRNAs are predominantly antisense to transposon sequences and exhibit a preference for uridine at their 5'end. Ago3 instead binds to sense piRNAs and favours an adenine at position 10 ("10A bias"). Furthermore, Aub and Ago3 piRNAs typically originate from opposite genomic strands and overlap by exactly 10-nt, a feature known as the "ping-pong signature" (Brennecke et al., 2007; Gunawardane et al., 2007). These characteristics are explained by the ping-pong amplification loop, a PTGS mechanism in which a transcript is first cleaved via piRNA-guided slicing and then used for piRNA production (**Figure 1.6**). This self-amplification loop is a well-conserved hallmark of the piRNA pathway, as a ping-pong signature can be readily identified across the animal kingdom (Aravin et al., 2007; Brennecke et al., 2007; Gainetdinov et al., 2018; Gunawardane et al., 2007; Houwing et al., 2007). An exception is *C.elegans*, where PRG-1-mediated target repression does not involve slicing but instead induces production of more small RNAs via RdRPs. Upon base-pairing with the target, PRG-1 stimulates the recruitment of the RdRPs RRF-1 and EGO-1 which, together with the Mutator complex, synthesize "secondary" small RNAs complementary to the target (Almeida et al., 2019). These RNAs begin with a guanosine and are typically 22 nt-long, hence they are called 22G RNAs. Once loaded by cytosolic (such as WAGO-1) or nuclear Argonaute proteins (such as HRDE-1 and NRDE-1), the 22-G RNAs elicit PTGS or TGS, respectively. In some instances, sustained production of 22-G RNAs becomes independent of the initial 21-U RNA trigger and is able to stably maintain silencing across generations (Almeida et al., 2019). While mechanistically different, the process of 22-G RNA synthesis will eventually lead to a signal amplification similar to that observed in the ping-pong cycle (Ozata et al., 2019). Especially in flies, where no RdRP-mediated amplification exists, the ping-pong cycle is key to enhance the silencing response against actively invading TEs.

Ping-pong takes place in nuage and starts with Aub bound to a maternally-deposited or cluster-derived antisense piRNA. This directs Aub to cleave a complementary sense transposon transcript between nucleotide 10 and 11 of the piRNA guide, which in turn defines the 5'end of a new, sense piRNA. The Aub-generated 3' cleavage product is readily loaded by Ago3 and its 3' end processed into a mature piRNA predominantly via the exonuclease Nibbler (Nbr) (Hayashi et al., 2016). This new Ago3-piRNA complex will then recognise a complementary cluster transcript exiting from the nucleus and cleave it, again generating the 5' end of an Aub piRNA, while the remaining transcript is funnelled into the Zuc-mediated biogenesis route

(**Figure 1.6**). Loss of either Aub or Ago3 individually can be partially compensated by a homotypic ping-pong looping centred on the remaining PIWI protein while still feeding some substrates to Zuc (Senti et al., 2015). Simultaneous absence of Aub and Ago3 instead causes a complete collapse of transposon-targeting piRNA production in the germline and promiscuous 5'P RNAs become substrate for mitochondrial processing (Senti et al., 2015; Wang et al., 2015). Conversely, impaired mitochondrial processing typically results in the appearance of novel ping-pong pairs that sustain a certain degree of piRNA production (Hayashi et al., 2016; Mohn et al., 2015). Altogether, these two biogenesis pathways are finely orchestrated to provide the highest robustness to transposon silencing, especially for elements that have colonised the *Drosophila* genome in ancient times (Senti et al., 2015). Notably, the evolutionarily age of transposon insertions correlates with a higher dependency on either TGS or PTGS. The sequences of old transposon insertions are robustly integrated into the piRNA-driven silencing machinery. As a consequence, old elements are predominantly targeted by Zuc-generated piRNAs and are the most robustly silenced by TGS and PTGS. Conversely, young elements rely more on heterotypic ping-pong, possibly because they are not yet present within piRNA clusters (Senti et al., 2015). Ping-pong looping is maintained by several nuage factors, such as Qin, that prevent homotypic ping-pong and smoothly funnel each intermediate into the right acceptor protein (Wang et al., 2015; Zhang et al., 2014a; Zhang et al., 2011).

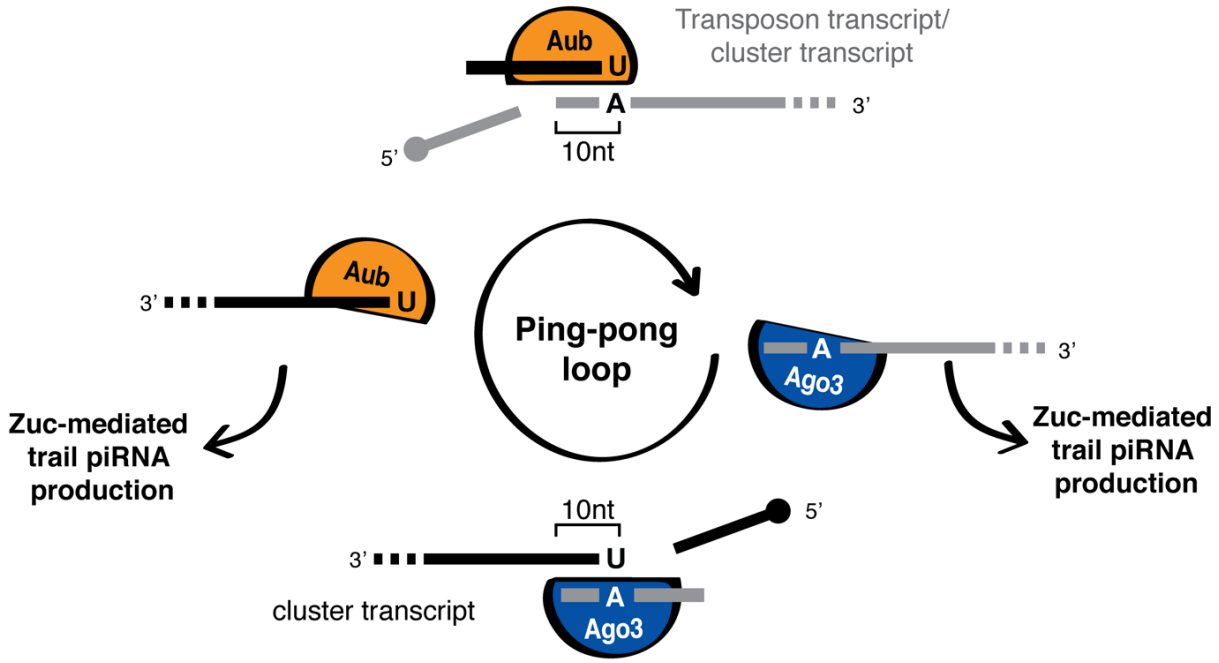


Figure 1.6. The ping pong cycle

Aub bound to an antisense piRNA recognizes and cleaves a transposon mRNA or a cluster transcript. The resulting 3′ cleavage product is converted into a new sense piRNA that associates with Ago3 and is trimmed by Nbr, whilst the rest is consumed by Zuc-mediated phased piRNA production. Ago3 associated with a sense piRNA can in turn recognize and cleave cluster transcripts. The product of this slicing event reinitiates the cycle, becoming an Aub-bound piRNA, whilst the remaining 3′ slicing product is consumed by Zuc-mediated phased piRNA production. Specialized loading complexes ensure that slicing products are loaded into the correct PIWI protein throughout the cycle.

I.3.2.2 Co-transcriptional silencing

The ultimate step of piRNA-guided transposon control occurs in the nucleus, where active transposon insertions become epigenetically silenced. Chromatin silencing by nuclear PIWI proteins depends on RNA, as opposed to DNA, aiming to specifically target actively transcribed TEs (Klenov et al., 2014; Le Thomas et al., 2013; Rozhkov et al., 2013; Sienski et al., 2012). Transcription-dependent heterochromatin formation is strikingly conserved throughout the eukaryotic domain of life. From fission yeast to mammals, many different organisms exploit small RNAs to direct heterochromatin formation towards discrete loci. Although the molecular mechanisms may differ, the core concepts of RNAi-mediated epigenetic silencing are conserved. TGS typically entails a small RNA guide, either a piRNA or an siRNA, bound to a nuclear Argonaute protein. Base-pairing between the small RNA and the target is thought to trigger recruitment of epigenetic modifiers to the locus, primarily histone methyl-transferases which add the repressive mark tri-methylation of Lysine 9 of Histone 3 (H3K9me3).

Several pioneering studies on TGS have been conducted in fission yeast. In *Schizosaccharomyces pombe,* the Argonaute protein Ago1 bound to an siRNA assembles the RNA-induced Transcriptional Silencing complex (RITS) which recognises and silences transcripts arising from centromeric repeats (Martienssen and Moazed, 2015; Duempelmann et al., 2020). RITS engages with the transcribing RNAs and recruits a protein complex containing Clr4, a histone methyl-transferase. Clr4 in turn deposits H3K9me3 which, in a positive feed-forward loop, promotes recruitment of more RITS, spreading of H3K9me3 and heterochromatin formation. This silencing response is further reinforced via the generation of secondary siRNAs by an RdRP and the Dicer homolog Dcr1. Studies in fission yeast have demonstrated that RITS targeting occurs via siRNA-mediated base-pairing with RNA, thus contributing to the paradigm that nuclear Argonautes recognise nascent transcripts via their associated small RNAs (Martienssen and Moazed, 2015). In a similar fashion, some of *C.elegans* nuclear WAGOs, such as HRDE-1, bind to small RNAs and trigger transcriptional silencing of complementary transcripts. WAGO-associated small RNAs, the so-called 22-G RNAs, are produced in response to 21-U RNAs and are continuously amplified by RdRPs, thus reinforcing the silencing response (Almeida et al., 2019; Duempelmann et al., 2020). Epigenetic silencing in *C.elegans* can be maintained after the initial trigger (the 21-U RNAs) ceases to be expressed thanks to the continuous production of 22-G RNAs, a process known as 'RNA-induced epigenetic silencing' (RNAe; Duempelmann et al., 2020). Initiation and maintenance of silencing in nematodes thus rely on separate Argonaute proteins, whereas how these two

processes are orchestrated in other model organisms is still largely unclear. TGS in nematodes also entails H3K9 methyltransferases, which play a conserved essential role in initiating the silencing cascade (Duempelmann et al., 2020). Mouse TGS instead results in DNA methylation of target loci, a prominent form of epigenetic silencing in mammals, along with other repressive histone modifications that include H3K9me3 (Aravin et al., 2008; Molaro et al., 2014). In mouse, MIWI2 is the nuclear PIWI protein that triggers *de-novo* DNA methylation of retrotransposons during male germ cell development. The factor(s) that connect murine PIWI proteins to the epigenetic machinery is not yet known. Small RNA-guided heterochromatin formation also extends beyond the animal kingdom and has been extensively characterised in *Arabidopsis thaliana*, where nuclear Argonaute proteins bound to siRNAs trigger *de-novo* DNA methylation to silence TEs. Following DNA methylation, the histone methyl-transferases SUVH4/5/6 deposit H3K9me2 and enable a feedback loop that maintains a repressive epigenetic state (Duempelmann et al., 2020). Lastly, ciliates also employ small RNA-directed heterochromatinisation to mark areas of the genome that undergo rearrangements during sexual reproduction (Noto and Mochizuki, 2017). This mechanism has been investigated in *Tetrahymena termophila*, which has a somatic macronucleus (MAC) and a germline micronucleus (MIC). Following sexual reproduction, small RNAs produced from the MIC direct heterochromatin formation on complementary sequences within the MAC which, as a result, are excised and eliminated. In line with the other mechanisms discussed above, *Tetrahymena* also utilises H3K9me3 and a feedback loop producing more "secondary" small RNAs, thus providing highest robustness to the system (Noto and Mochizuki, 2017). Similar mechanisms exist in other ciliates, such as *Paramecium tetraurelia*, where small RNAs produced from the germline nucleus mark sequences for elimination, or *Oxytricha trifallax*, where instead small RNAs originated from the somatic nucleus mark genomic regions that must be retained (Lepère et al., 2008; Fang et al., 2012). Overall, small RNA-guided heterochromatin formation employs highly conserved proteins, such as the Argonaute family and H3K9 methyl-transferases, and conceptually similar mechanisms, involving RNA recognition, priming of the silenced state and maintenance across cell division and/or generations, often via an amplification loop.

In line with what reported in other organisms, piRNA-dependent TGS in *Drosophila* occurs via the deposition of histone modifications typically associated with heterochromatin and a transcriptionally inert state, namely H3K9me3, and the concomitant removal of H3K4me2, mostly found at active loci (Le Thomas et al., 2013; Rozhkov et al., 2013; Sienski et al., 2012; Wang and Elgin, 2011). Heterochromatin establishment in other organisms, such as yeast, also

involves di-methylation of H3K9 (H3K9me2) (Jih et al., 2017) but, since the difference and interplay between di- and tri-methylation has not been investigated in the context of the fly piRNA pathway, we will hereafter refer to H3K9me3 only. piRNA-guided heterochromatin assembly is a step-wise process (Wang et al., 2016) similar to that described earlier for *S. pombe* (**Figure 1.7**). First, an active transposon insertion is recognised via its nascent RNA (**Figure 1.7A**) and chromatin modifiers are recruited to the locus to initiate the deposition of repressive marks and concomitant removal of active ones, though the precise hierarchy is yet to be established (**Figure 1.7B**). Heterochromatin then typically spreads from nucleation centres into neighbouring regions, until the transposon insertion is entirely covered (**Figure 1.7B**). The newly-deposited histone marks ultimately serve as binding platforms for heterochromatin-specific proteins that will consolidate the silenced state and compact the chromatin (**Figure 1.7C**). At the heart of the *Drosophila* TGS machinery lies Piwi bound to a Zuc-generated piRNA (Sienski et al., 2012). In contrast with slicing-mediated PTGS, the catalytic activity of Piwi is dispensable for TGS, since mutations of key residues in the catalytic tetrad do not impact silencing nor cause fertility defects (Saito et al., 2010; Sienski et al., 2012). The current model of TGS postulates that Piwi scans all nascent transcripts and, upon recognition of transposon RNA, primes the corresponding locus for silencing. Although several crucial players cooperating with Piwi have been recently unmasked, the precise sequence of events leading from target engagement to the establishment of a repressive chromatin state is not entirely understood.

The initial event that triggers piRNA-guided TGS is Piwi-RISC binding to a transcript arising from an active transposable element (Klenov et al., 2014; Le Thomas et al., 2013; Rozhkov et al., 2013; Sienski et al., 2012). Re-activated transposon loci are typically transcribed by RNA pol II and carry active histone marks, such as H3K4me3 near the TSS and H3K4me2 at more distal regions (**Figure 1.7A**). Piwi-RISC target engagement is thought to cause a second conformational change of Piwi that is permissive for recruitment of downstream silencing factors (Yu et al., 2015). In fact, artificial tethering of Piwi to a reporter transcript does not induce TGS, thus supporting the hypothesis that only a trimeric Piwi-piRNA-target interaction can instruct repression of a locus (Sienski et al., 2015; Yu et al., 2015) (**Figure 1.7A**). Indirect evidence suggests that this is a transient event, since Piwi can hardly be isolated in complex with any silencing factor (Batki et al., 2019; Fabry et al., 2019). Because Piwi-piRNA complexes are very abundant in the nucleus, the event that initiates TGS has to be tightly regulated to avoid promiscuous silencing.

Several proteins cooperate with Piwi to convey the silencing response, namely Asterix (Arx), Panoramix (Panx) and Maelstrom (Mael). Among these, Panx stands out for being necessary and sufficient for TGS (Sienski et al., 2015; Yu et al., 2015). Its artificial recruitment to either RNA or DNA elicits potent epigenetic silencing, a property which is in striking contrast with Piwi and had initially been ascribed to Panx alone (Sienski et al., 2015; Yu et al., 2015). Recent work has uncovered two more proteins that show similar silencing capability and, together with Panx, form the PICTS (Panx-induced co-transcriptional silencing) complex: Nxf2 and Nxt1 (Batki et al., 2019; Fabry et al., 2019; Murano et al., 2019; Zhao et al., 2019) (see **Chapter V**). Absence of any PICTS component causes dramatic transposon re-activation due to the loss of repressive histone marks, despite Piwi still localising to the nucleus (Batki et al., 2019; Fabry et al., 2019; Murano et al., 2019; Zhao et al., 2019). Hence, PICTS acts downstream of Piwi at a critical node of TGS, bridging Piwi-piRNA complexes to the machinery responsible for epigenetic modification of target loci. Notably, the silencing activity of PICTS can be narrowed down to the amino-terminus of Panx, but how this links to the silencing machinery is still unknown. In addition to PICTS, Arx, a small zinc-finger protein with a putative RNA-binding domain, also participates in TGS (Donertas et al., 2013; Muerdter et al., 2013; Ohtani et al., 2013). Arx localises to the nucleus in a Piwi-dependent manner and possibly stabilises its target-engaged state (Donertas et al., 2013; Ohtani et al., 2013). Finally, Mael is a high-mobility group protein that shuttles between the nucleus and nuage thanks to Embargoed (also known as CRM-1; Chromosomal region maintenance 1), which mediates its nuclear export, and nuage components, such as Vasa and Aub, that recruit it to the cytoplasmic side of the nuclear envelope (Findley et al., 2003). Mael has a domain that adopts an RNase H-like fold but lacks the conserved catalytic residues. Nonetheless, this nuclease activity is dispensable for transposon silencing hence Mael likely acts via binding to structured RNAs (Genzor and Bortvin, 2015; Matsumoto et al., 2015). Mael has been implicated as a key effector of nuclear TGS acting downstream of Piwi and H3K9me3 deposition (Sienski et al., 2012). Its loss only modestly affects H3K9me3 levels at transposon loci, but seems to cause its spreading into downstream regions (Sienski et al., 2012). Recent work also suggests that Mael plays a role in the transcription of dual-strand clusters via suppressing canonical RNA pol II transcription of transposon insertions (Chang et al., 2019).

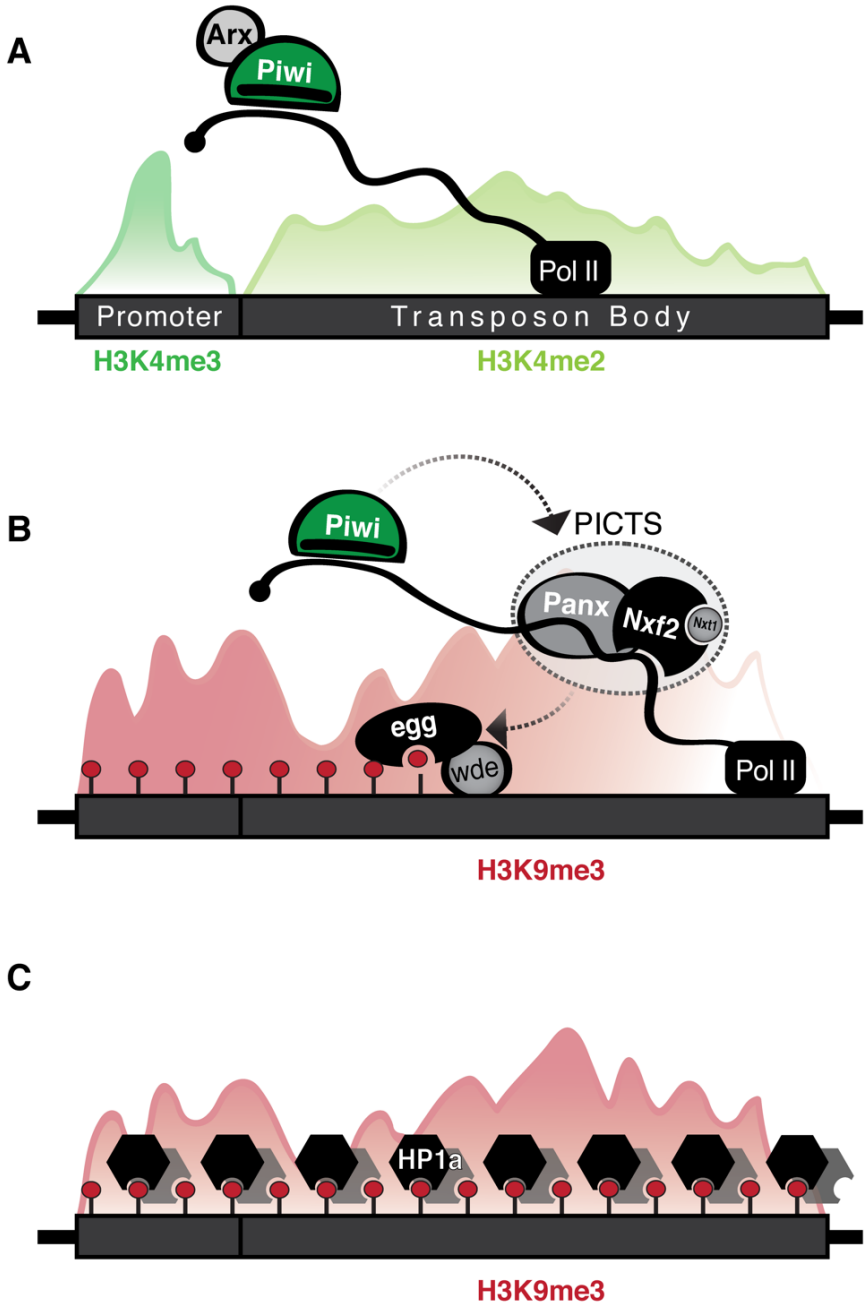


Figure 1.7. piRNA-guided Transcriptional Gene Silencing

A) Piwi-piRNA complexes engage with nascent transposon transcripts arising from active loci, marked by H3K4me3 on the promoter and H3K4me2 along the gene body. **B)** Once engaged with the target, Piwi recruits the PICTS complex (Panx, Nxf2, Nxt1), which in turn instructs silencing of the locus via recruitment of the histone methyl-transferase Egg and its co-factor Wde. The transposon locus becomes progressively covered by H3K9me3. **C)** Upon establishment of silencing, HP1a covers the entire locus and maintains its repressed status.

Once an active locus has been primed for repression, it loses H3K4me3/2 and becomes covered by H3K9me3 (Le Thomas et al., 2013; Rozhkov et al., 2013; Sienski et al., 2012) (**Figure 1.7B**). The histone methyl-transferase Egg is the most downstream factor of the co-transcriptional silencing cascade and, together with its co-factor Windei (Wde), deposits H3K9me3 on Piwi-targeted loci (Huang et al., 2013; Rangan et al., 2011; Sienski et al., 2015; Yu et al., 2015). Genetic studies have implicated Egg in piRNA-guided heterochromatin formation, however, what factor(s) directly link Piwi and PICTS to Egg is unknown. Egg is predominantly nuclear and becomes monoubiquitinated on conserved residues, a modification that is in turn important for its interaction with Wde and its transposon silencing capability (Osumi et al., 2019). Wde is instead required for Egg nuclear retention and recruitment onto chromatin (Osumi et al., 2019). In addition to transposon silencing, Egg participates in both uni- and dual-strand piRNA cluster establishment, as they are embedded into heterochromatic regions and mutations in Egg abolish piRNA production (Rangan et al., 2011). In contrast with that, studies in OSCs report that piRNA production from *flam* is not affected by knockdown of *egg* (Sienski et al., 2015), leaving its role on uni-strand cluster establishment open for future investigation. Recent reports also found an involvement of the SUMO E3 ligase Su(var)2-10 in the TGS process. Su(var)2-10, in addition to being required for global heterochromatin formation, directly interacts with components of the piRNA-guided TGS machinery and induces SUMO-dependent recruitment of Egg and Wde (Ninova et al., 2019a; Ninova et al., 2019b). TGS also entails Lsd1/Su(var)3-3 and its partner CoRest, responsible for removing H3K4me2/3 marks (Yu et al., 2015), and ultimately HP1a. HP1a acts downstream of Piwi and is likely required to compact H3K9me3-marked transposon loci into heterochromatic domains (Sienski et al., 2015; Yu et al., 2015) (**Figure 7C**). Heterochromatin formation in *Drosophila* occurs via aggregation of liquid-like, phase separated compartments centered on HP1a (Larson et al., 2017; Strom et al., 2017). The absence of either Egg, Wde, Lsd1 or HP1a severely compromises the ability of PICTS to induce heterochromatin formation (Sienski et al., 2015; Yu et al., 2015), further highlighting the connection between the piRNA pathway and the general silencing machinery. However, how transposon silencing is maintained when PICTS ceases to be expressed remains to be investigated. Some organisms, such as yeast and nematodes, possess positive feed-forward loops involving RdRPs to maintain a repressed chromatin state after the initial trigger is terminated (Duempelmann et al., 2020). Flies, which do not possess RdRPs, amplify the silencing response against TEs via the ping-pong cycle and Zuc-mediated phased piRNA biogenesis, ultimately fueling Piwi with a wide range of piRNA sequences. If this is implicated in silencing maintenance, similarly to RNAe in *C.elegans*, is yet to be uncovered.

CHAPTER II

MATERIALS AND METHODS

2.1. Fly stocks and handling

All flies were kept at 25°C on standard cornmeal or propionic (malt extract, molasses, cornmeal, yeast, agar, soya powder, water, Propionic acid, Nipagin) food. Control w^{1118} flies were a gift from the University of Cambridge Department of Genetics Fly Facility. For GLKD we used a stock containing a UAS::Dcr2 transgene and a nos::GAL4 driver (Czech et al., 2013) and shRNA lines from the Bloomington Drosophila Stock Center and Vienna Drosophila Resource Center.

2.1.1. CRISPR/Cas9 mutant generation

Frameshift mutant alleles of *daed* were generated by injecting pCFD4 (Addgene plasmid # 49411; (Port et al., 2014)) containing two gRNAs against *CG10880* into embryos expressing vas-Cas9 (Bloomington stock 51323). The $gasz^{KO}$ allele was generated by injecting a plasmid containing two gRNAs against *gasz* and a donor construct with ~1 kb homology arms flanking a 3xP3-RFP cassette into vas-Cas9 flies. *nxf2* mutant alleles were generated by injecting pCFD4 containing two gRNAs against *nxf2* into embryos expressing vas-Cas9. Microinjection and fly stock generation was carried out by the University of Cambridge Department of Genetics Fly Facility. Mutant flies were identified via genotyping PCR and confirmed by Sanger sequencing (carried out by GATC/Eurofins).

2.1.2. Transgenic and knock-in fly lines generation

The shRNA against *daed* was cloned into pVALIUM20 (Ni et al., 2011), GFP-Daed and Zuc-GFP were cloned into an in-house generated transgenesis vector for phiC31-mediated integration and expressed under the *D. melanogaster* Ubiquitin promoter (pUBI). All plasmids were integrated into the attP40 landing site on chromosome 2 (Stock 13-20). To generate GFP-Nxf2 fusion knock-in flies, the GFP insert flanked by homology arms of approximately 1 kb were cloned into pUC19 by Gibson assembly and co-injected with pCFD3 (Addgene plasmid # 49410; (Port et al., 2014)) containing

a single guide RNA into embryos expressing vas-Cas9. Microinjection and fly stock generation was carried out by the University of Cambridge Department of Genetics Fly Facility. Transgenic and knock-in flies were identified via genotyping PCR and confirmed by Sanger sequencing (carried out by GATC/Eurofins).

2.1.3. Fly genotyping

Genomic DNA isolation for genotyping PCR was carried out as follows. Individual flies were placed in a 96-well plate and lysed in 100 µl of Squishing buffer (10 mM Tris HCl pH8.2, 1 mM EDTA, 25 mM NaCl and 200 µg/ml of Proteinase K) using a Qiagen TissueLyzer for 10 min at 28 Hz. The plate was incubated for 2 hrs at 50°C followed by 10 min at 95°C.

Genotyping PCR was performed on 1 µl of gDNA template mixed with 2 µl KOD buffer, 2 µl dNTPs, 1.2 µl $MgSO_4$, 0.6 µl DMSO, 0.6 µl of each forward and reverse primer, 0.2 µl of KOD polymerase and 11.8 µl of nuclease-free water. Obtained bands were analysed on a 1.5% agarose gel or using a QIAxcel instrument (Qiagen) and confirmed by Sanger sequencing (carried out by GATC/Eurofins).

2.1.4. Egg laying

Fertility was scored by crossing ten freshly hatched mutant females to five w^{1118} males and counting the number of eggs laid in 12 hr periods as well as pupae that developed following 7 days.

2.2. Cell Culture

2.2.1. S2 cell culture

Gibco® Drosophila Schneider 2 (S2) cells were purchased from Thermo Fisher Scientific (catalog number R69007) and were grown at 26°C in Schneider's Drosophila Media (Gibco) supplemented with 10% heat-inactivated FBS. Cell identity was characterized by Thermo Fisher Scientific through isozyme and karyotype analysis (see product description). S2 cells tested negative for mycoplasma contamination in-house.

2.2.2. S2 cell transfection

S2 cells were transfected using Effectene (Qiagen), according to the manufacturer's instructions. For one well of a 6-well plate, $2x10^6$ cells were plated ~30 min prior to transfection. 2 µg of plasmid were mixed with 182 µl of EC Buffer and 16 µl of

Enhancer, vortexed and incubated for 1 min at room temperature. 20 µl of Effectene (1:10 ratio between plasmid and Effectene) were added to the samples, vortexed and incubated for 15 min at room temperature. The reaction mixture was added dropwise to 2 ml of cell suspension and cells were harvested for analysis 48 hrs later.

2.2.3. OSC culture

Drosophila Ovarian Somatic Cells (OSCs) were a gift from Dr Mikiko Siomi and were cultured at 26°C in Shields and Sang M3 Insect Medium (Sigma Aldrich) supplemented with 0.6 mg/ml Glutathione, 10% FBS, 10 mU/ml insulin and 10% fly extract (purchased from DGRC) as described (Niki et al., 2006; Saito, 2014; Saito et al., 2009). Cell identity was authenticated by whole genome DNA sequencing in-house. OSCs tested negative for mycoplasma contamination in-house.

2.2.3.1. OSC transfections

OSCs were transfected using Xfect transfection reagent (Takara Bio 631317). For one well of a 6-well plate, $5x10^5$ cells were plated one day in advance. On the day of transfection, 5 µg of plasmid were diluted with 100 µl of Xfect buffer, vortexed and mixed with 100 µl of Xfect buffer supplemented with 1.5 µl of Xfect polymer (0.3 µg of Xfect polymer for each µg of plasmid). The transfection mix was vortexed and incubated for 10 min at room temperature. Meanwhile, the cells were washed once with M3 media and left in 1 ml of M3 media. The transfection mixture was then added to the cells drop-wise and incubated for 2 hrs at 26°C. Subsequently the M3 media was aspirated and replaced with 2 ml of complete media. Cells were harvested and analysed 48 hrs after transfection.

2.2.3.2. OSC nucleofections

OSCs nucleofection was carried out using the Cell Line Nucleofector™ Kit V (Lonza VVCA-1003), according to a published protocol (Saito, 2014). $10x10^6$ cells were pelleted and resuspended in 100 µl of Nucleofection solution containing the desired plasmid or siRNA. The resuspended mixture was transferred to cuvettes and subjected to program T-029 in a Nucleofector II device. Cells were then resuspended in 10 ml of OSC media and grown at 26°C. In cases where a second pulse of nucleofection was required, after 48 hrs $10x10^6$ cells were pelleted and treated exactly as described above.

2.2.3.3. OSC knockdowns

siRNAs against the desired genes were designed using DSIR (http://biodev.cea.fr/DSIR/DSIR.html) with standard parameters (siRNA 21 nt; score threshold 90) and off-targeting was analysed with BLAT. siRNAs were purchased from IDT as ssRNAs (100 nm scale, standard desalting) and sense and anti-sense siRNAs were annealed prior to nucleofection. The annealing was performed by mixing 25 µl of each sense and antisense RNA oligonucleotides (resuspended at 400 µM in RNase-free water) with 50 µl of 2x annealing buffer (1x annealing buffer is 100mM HEPES pH7.5 and 30 mM Potassium Acetate). The siRNAs were boiled for 5 min at 75°C then slowly cooled down to 25°C (-0.1°C/second). Annealed siRNAs were stored at -80°C.

For knockdowns, $10x10^6$ OSCs were nucleofected with 2 µl of annealed siRNAs as described in 2.2.2.4. Following 48 hrs, $10x10^6$ OSCs were nucleofected again with 2 µl of the same siRNA. Cells were harvested and analysed after an additional 48 hrs. For knockdown in combination with an expression construct, 2 µl of siRNA were mixed with 5 µg of plasmid.

2.2.3.4. OSC stable line generation

One well of a 6-well plate was transfected as described in 2.2.2.1 with 2.5 µg of the desired plasmid and 2.5 µg of a helper plasmid carrying a Puromycin resistance gene (pMT-Puro). Cells that have stably integrated the plasmid have been selected by adding Puromycin to the media (1:2000 dilution from a 10mg/ml stock). Stable expression of the construct was confirmed via western blot.

2.3. RNA isolation and RT-qPCR

OSC or ovary samples were lysed in 1 ml Trizol and RNA was extracted according to manufacturer's instructions. 200 µl of Chloroform were added and, following mixing, the samples were spun for 15 min at 12,000g at 4°C. The aqueous phase was collected and precipitated for 15 min at room temperature with 500 µl of Isopropanol. Precipitated RNA was pelleted via centrifugation for 20 min at top speed at 4°C and the pellet washed once with 80% Ethanol. After Ethanol removal, the pellet was air dried for 3 min and resuspended in RNase-free water and quantified on a Nanodrop.

To remove genomic DNA contamination prior to reverse transcription, 1 µg of total RNA was incubated with 1 µl of DNaseI (Thermo Fisher Scientific) in 1x DNase buffer for 20 min at 37°C. The reaction was stopped by adding 1 µl of EDTA. The samples were used

directly for reverse transcription with the Superscript III First Strand Synthesis Kit (Thermo Fisher Scientific). The DNase-treated RNA was incubated with 1 µl of dNTP Mix (10 mM each) and 1 µl of oligo$(dT)_{20}$ (50 µM) at 65°C for 5 min, then cooled on ice for 2 min. The samples were then incubated with 4 µl of 5X First-Strand Buffer, 1 µl of 0.1 M DTT and 1 µl of RNAse OUT for 5 min at 25°C, 1 hr at 50°C and 15 min at 75°C. The obtained cDNA was resuspended in 100 µl final volume with RNase-free water and used directly for quantitative PCR (qPCR).

The qPCR reaction was carried out in a 384-well plate and, for each well, the reaction was set up as follows: 2 µl of cDNA, 0.2 µl of pre-mixed oligo pair (10 µM), 5 µl of Fast SYBR™ Green Master Mix (Applied Biosystems) and 2.8 µl of water. qPCR experiments were performed with a QuantStudio Real-Time PCR Light Cycler (Thermo Fisher Scientific). Transposon levels were quantified using the ΔΔCT method (Livak and Schmittgen, 2001), normalised to *rp49* and fold changes were calculated relative to the indicated controls.

2.4. Immunofluorescence

2.4.1. Immunofluorescence in cells

OSCs were plated one day in advance on Fibronectin-coated coverslips, whereas S2 cells were plated 1 hr in advance on Concanavalin-A-coated coverslips. Cells were fixed for 15 min in 4% PFA, permeabilized for 10 min in PBS, 0.2% Triton and blocked for 30 min in PBS, 0.1% Tween-20 (PBST) and 1% BSA. Primary antibodies were diluted 1:500 in PBST and 0.1% BSA and incubated overnight at 4°C. After 3x5 min washes in PBST, secondary antibodies were incubated for 1 hr at room temperature. After 3x5 min washes in PBST, DAPI (1:10,000 dilution in PBST) was incubated for 10 min at room temperature and washed twice in PBST. Coverslips were mounted with ProLong Diamond Antifade Mountant (Thermo Fisher Scientific #P36961) and imaged on a Leica SP8 confocal microscope (100x Oil objective).

2.4.2. Immunofluorescence in cells for STED

For STED, the same protocol was used with the following modifications: cells were plated on Fibronectin-coated 1.5H coverslips, blocking was for 1.5 hrs in PBS, 0.1% Tween-20 (PBST) and 1% BSA. Primary and secondary antibodies were diluted 1:150 in PBST and 1% BSA. Coverslips were mounted using ProLong Glass Antifade Mountant (Thermo Fisher Scientific # P36982) and imaged on a Leica SP8 confocal

microscope (100x Oil objective). The images were deconvolved using Huygens Professional.

2.4.3. Immunofluroescence in fly ovaries

Fly ovaries were dissected in ice-cold PBS, fixed for 15 min in 4% PFA at room temperature and permeabilized with 3x10min washes in PBS with 0.3% Triton (PBS-Tr). Samples were blocked in PBS-Tr with 1% BSA for 2 hrs at room temperature and incubated overnight at 4°C with primary antibodies in PBS-Tr and 1% BSA. After 3x10 min washes at room temperature in PBS-Tr, secondary antibodies were incubated overnight at 4°C in PBS-Tr and 1% BSA. After 4x10min washes in PBS-Tr at room temperature (DAPI was added during the third wash; 1:1 000) and 2x5 min washes in PBS, samples were mounted with ProLong Diamond Antifade Mountant (Thermo Fisher Scientific #P36961) and imaged on a Leica SP8 confocal microscope. Images were deconvolved using Huygens Professional.

2.5. RNA FISH

2.5.1. HCR RNA FISH Probe design

Probes and hairpins for HCR were designed according to the description in (Ang and Yung, 2016; Choi et al., 2014). Probes targeting *flamenco* (7 probes) or *ZsGreen* reporter mRNA (10 probes) were designed as 52-nt long antisense oligos with a GC content between 40 and 70%. Off-targeting effects were assessed using BLAST. The HCR initiator sequence and a poly-A linker (GCCCTTACTCCCAATTCCaaaaa) were added to the 5' end of each probe sequence. HCR probes were purchased from IDT as 100 nmoles oligonucleotides with HPLC purification. The hairpins are described in Ang and Yung, 2016 and have been purchased from IDT conjugated with AlexaFluor-647.

2.5.2. HCR RNA FISH in OSCs

RNA FISH was performed with Hybridization Chain Reaction (HCR), similar as reported in Ang and Yung, 2016 and Choi et al., 2014. OSCs were seeded on Fibronectin-coated coverslips, fixed for 15 min in 4% PFA, washed 2x5 min with PBS and permeabilized for at least 24 hrs in 70% Ethanol at -20°C. Ethanol was removed and slides were washed 2x5 min in 2x Saline-Sodium Citrate buffer (SSC). Priming for hybridization was done by incubating for 10 min in 15% formamide in 2x SSC. HCR probes were diluted to 1 nM each in hybridization buffer (15% formamide, 10% dextran

sulfate in 2x SSC) and incubated overnight at 37°C in a humidified chamber. Excess probes were removed by rinsing twice in 2x SSC and washing once in 30% formamide for 10 min at 37°C. HCR hairpins conjugated to AlexaFluor-647 (Hall et al.) were heat-denatured and diluted to 120 nM in 5x SSC and 0.1% Tween-20 (SSCT). HCR amplification was carried out for 2 hrs at room temperature in the dark and washed 3 x 10 min with 5x SSCT. Nuclei were stained with DAPI (1:10,000 in SSCT) for 10 min, followed by 3x10 min washes in 5x SSC. Slides were mounted with ProLong Diamond Antifade Mountant (Thermo Fisher Scientific) and imaged on a Leica SP8 confocal microscope (100x Oil objective).

2.5.3. Immunofluorescence and HCR RNA FISH in OSCs

For co-staining of RNA and protein, the immunofluorescence was carried out first as described in 2.4.1. After secondary antibody incubation, cells were washed 3x4 min in PBST and fixed in 4% PFA for 10 min. Following 2x5 min washes in PBS, cells were permeabilised for at least 24 hrs in 70% Ethanol at -20°C. HCR-RNA FISH was then carried out as described in 2.5.2.

2.6. DNA FISH

2.6.1. DNA FISH Probe design

DNA FISH probes were designed against a 10 kb region spanning the *DIP1* and *flamenco* genomic loci (chrX:21,624,796-21,634,619) using Oligominer (Beliveau et al., 2018). Probes with the desired characteristics were first identified using blockparse (settings used -l 31 -L 35 -g 35 -G 80 -S 10) and then aligned to the genome using bowtie2 to keep only those mapping to a unique location. Probes were further screened for the presence of highly abundant k-mers using jellyfish. The final set of 74 probes (half against the sense strand and half against the antisense) were completed with the addition of the SABER primer sequences at their 3'ends (tttCAACTTAAC). Each final probe is thus structured as follows: 5'-[probe sequence] - TTT - [9-mer primer sequence]-3'. All oligo probes were ordered from IDT (25 nmole DNA Plate Oligo; Standard Desalting; 200 µM resuspended in IDTE Buffer pH 7.5 (10 mM Tris-HCl, 0.1 mM EDTA)) as well as the PER hairpin, the imager strand conjugated to ATTO-647 fluorescent dye and the 'clean G' hairpin (as described in Kishi et al., 2019).

PER amplification was carried out prior to each DNA FISH experiment as described in (Kishi et al., 2019). A reaction mix containing 10 µl of 10x PBS, 10 µl of $MgSO_4$ (100

mM; NEB), 5 µl of dNTP mix (A,C,T only at 6 µM each; NEB), 10 µl of clean.G haipin (1 µM), 0.5 µl of Bst LF Polymerase (NEB), 10 µl of PER hairpin (5 µM) and 44.5 of water was incubated for 15 min at 37°C. 10 µl of probe oligo pool (10 µM) were added to the mix and the reaction was incubated for 3 hrs at 37°C, followed by 20 min at 80°C to inactivate the polymerase. The extended probes were used directly for DNA FISH.

2.6.2. SABER DNA FISH in OSCs

DNA FISH was carried out as described in (Kishi et al., 2019). OSCs were plated on Fibronectin-coated slides and fixed for 10 min in 4% PFA, permeabilized for 10 min in PBS, 0.5% TritonX-100 and washed twice in PBS with 0.1% Tween-20 (PBST). If necessary, DNase treatment was carried out at this stage by incubation with 4 µl of Turbo DNase in 100 µl of 1x Turbo DNase buffer for 30 min at 37°C. Cells were incubated 5 min in 0.1 N HCl, washed twice in PBST and incubated in 2x SSCT (2x SSC with 0.1% Tween-20) with 50% formamide for 2 hrs at 60°C. Cells were hybridised in 80 µl of ISH solution consisting of 2× SSCT, 50% formamide, 10% dextran sulfate, 400 ng/µl RNase A and each PER extension at a final concentration of ~67 nM (1:15 dilution from 1 µM PER). After denaturation for 3 min at 80°C, cells were incubated overnight at 44°C in a humidified incubator. Hybridised samples were washed 4x5 min in prewarmed 2x SSCT at 60°C and then twice at room temperature. 80 µl of fluorescent hybridisation solution consisting of 1× PBS and 1 µM fluorescent imager strands were added to the samples and incubated for 1 hr at 37°C. Cells were washed 3x5 min in prewarmed PBS at 37°C, stained for 10 min at room temperature with DAPI (1:1,000 dilution in PBS) and mounted using ProLong Diamond Antifade Mountant (Thermo Fisher Scientific #P36961). Samples were imaged on a Leica SP8 confocal microscope (100x Oil objective).

2.6.3. Immunofluorescence and SABER DNA FISH in OSCs

For co-staining of DNA and protein, the DNA FISH was carried out first as described in 2.6.2. After fluorescent imager strands hybridisation, cells were washed 3x 5 min in PBS and then transferred to the blocking solution (PBST+1%BSA). Primary antibodies were diluted 1:200 in PBST and 0.1% BSA and incubated 2 hrs at room temperature. After 3x5 min washes in PBST, secondary antibodies were incubated for 1 hr at room temperature. After 3x5 min washes in PBST, DAPI was incubated for 10 min at room temperature (1:1,000 dilution in PBS) and washed twice in PBST. Coverslips were

mounted with ProLong Diamond Antifade Mountant (Thermo Fisher Scientific #P36961) and imaged on a Leica SP8 confocal microscope (100x Oil objective).

2.7. Image analysis on Fiji

Acquired images were analysed on Fiji using custom scripts. A representative script is provided as **Appendix VIII.1.** Briefly, each channel was converted to a monochromatic inverted grey image, the same dynamic range was applied for each channel across different samples and then converted to RGB. For the composite image, the same dynamic range was applied as in the monochromatic image and the channels merged and converted to RGB. The nuclear outline was drawn based on the DAPI signal and merged with the other channels using Photoshop.

2.8. RNA-seq

2.8.1. RNA isolation for RNA seq

RNA samples used for library preparation were extracted using the RNeasy Mini Kit (Qiagen 74106). Samples from cells or ovaries were lysed in 1 ml Trizol, then 200 µl of Chloroform were added, followed by 15 min centrifugation at 12,000g at 4°C. The upper aqueous phase was mixed with 1 volume of 70% Ethanol and directly transferred to the RNeasy column, followed by 15 sec spin at 8,000g at room temperature. The resin was washed once with 350 µl of RW1 buffer. DNase digestion was performed on the column by adding 10 µl of DNase I mixed with 70 µl of RDD1 buffer and incubating for 15 min at room temperature. 350 µl of RW1 buffer were added to the resin, incubated 5 min, followed by 15 sec spin at 8,000g. The resin was washed twice by adding 500 µl of RPE buffer followed by 15 sec spin at 8,000g. To elute RNA, 30 µl of nuclease-free water were added to the column, incubated for 2 min at room temperature and recovered by centrifugation for 1 min at top speed. RNA concentration was measured on a NanoDrop Spectrophotometer.

2.8.2. Library preparation

Ribosomal RNAs were depleted using RiboPOOLs against *Drosophila melanogaster* rRNAs (siTOOLs Biotech). riboPOOLs were first hybridised to 1 µg of RNA by adding 1 µl of resuspended riboPOOLS (100 µM), 5 µl of Hybridisation buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 2 M NaCl) and 1 µl of RNAse Inhibitor Plus (Promega),

incubating for 10 min at 68°C and cooling down slowly to 37°C. Meanwhile, 80 µl of MyOne Streptavidin C1 beads (Thermo Fisher 65001) for each sample were washed twice in 100 µl of Beads Resuspension Buffer (0.1 M NaOH, 0.05 M NaCl) and twice in 100 µl of Beads Wash Buffer (0.1 M NaCl). The beads were resuspended in 160 µl of Depletion buffer (5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl) and divided into two 80 µl aliquots for two consecutive rounds of rRNA depletion. Hybridised riboPOOLs were incubated with 80 µl of washed beads, mixed well and incubated for 15 min at 37°C, followed by 5 min at 50°C. The supernatant was added to the second tube containing 80 µl of washed beads and incubated for 15 min at 37°C, followed by 5 min at 50°C. The rRNA-depleted RNA samples were then transferred to a fresh tube and purified using Agencourt RNAClean XP beads (Beckman Coulter A63987). 2.2x Agencourt RNAClean XP beads were added to the sample, vortexed and incubated on ice for 15 min. the supernatant was discarded and the beads washed twice in 300 µl of freshly prepared 80% Ethanol. The beads were air-dried for 3 min and the RNA resuspended in 8 µl of nuclease-free water. 5 µl of cleaned-up RNA were used directly for library preparation.

RNAseq libraries were prepared using the NEBNext Ultra Directional Library Prep Kit for Illumina (NEB #E7760), according to the manufacturer's instructions for ribosome-depleted RNA. 5 µl of rRNA-depleted RNA (see above) were fragmented by mixing with 1 µl of Random Primers and 4 µl of NEBNext First Strand Synthesis Reaction Buffer and incubated 15 min at 94°C. First Strand cDNA synthesis was performed by adding 0.5 µl of Murine RNase Inhibitor, 5 µl of Actinomycin D (0.1 µg/µl), 1 µl of ProtoScript II Reverse Transcriptase and 3.5 µl of nuclease-free water followed by incubation for 10 min at 25°C, 15 min at 42°C and 15 min at 70°C. 48 µl of nuclease-free water, 8 µl of Second Strand Synthesis Reaction Buffer and 4 µl of Second Strand Synthesis Enzyme mix were added to the sample and second strand cDNA synthesis was performed for 1 hr at 16°C. cDNA was purified using 1.8X Agencourt AMPure XP beads (Beckman). 1.8X beads were mixed with 20 µl of cDNA, vortexed and incubated for 5 min at room temperature. Samples were placed on a magnetic rack, the supernatant discarded and the beads washed twice in freshly-prepared 80% Ethanol. Beads were air-dried for 5 min at room temperature and the cDNA eluted in 60 µl of 0.1X TE Buffer. 55.5 µl of eluted cDNA were subjected to End-Prep reaction by adding 6.5 µl of NEBNext End Repair Reaction Buffer and 3 µl of NEBNext End Prep Enzyme mix and incubating for 30 min at 20°C and 30 min at 65°C. Adaptor ligation was performed by

adding to 65 µl of End-Prep reaction mix 15 µl of Blunt T/A Ligase Master Mix, 1 µl of NEBNext Adaptor (1.5 µM) and 2.5 µl of nuclease-free water and incubating for 15 min at 20°C. The ligation reaction was purified using 1.0X Agencourt AMPure XP Beads (Beckman) as described above and resuspended in 52 µl of 0.1X TE. 50 µl of the supernatant were subjected to a second round of clean-up using 1.0X Agencourt AMPure XP Beads (Beckman) and resuspended in 19 µl of 0.1X TE. 17 µl of eluted library were used for PCR enrichment by adding 3 µl of NEBNext USER enzyme, 25 µl of NEBNext Q5 Hot Start HiFi PCR Master Mix, 2.5 µl of Universal forward i5 Primer and 2.5 µl of Indexed Reverse i7 primer. Library enrichment PCR was carried out as follows:

15 min at 37°C

30 sec at 98°C

13 cycles:

10 sec at 98°C

75 sec at 65°C

5 min at 65°C

The amplified library was purified using 45 µl of Agencourt AMPure XP Beads, eluted in 23 µl of 0.1X TE and 20 µl were transferred to a new tube. A 2x dilution was analysed on a Tapestation instrument (Agilent). DNA libraries were quantified with the KAPA Library Quantification Kit for Illumina (Kapa Biosystems) and deep-sequenced with Illumina HiSeq 4000 (Illumina).

2.8.3. Data analysis

Raw fastq files generated by Illumina sequencing were analysed by a pipeline developed in-house. In short, for CLIP-seq the first 5 bases of each 50 bp read were removed using fastx trimmer (http://hannonlab.cshl.edu/fastx_toolkit/). After removal of reads mapping to Drosophila rRNA using STAR, high-quality reads were aligned to the Drosophila melanogaster genome release 6 (dm6; downloaded from Flybase) (Hoskins et al., 2015) using STAR (Dobin et al., 2013). For transposon-wide analysis, genome multi-mapping reads were randomly assigned to one location using option '--outFilterMultimapNmax 1000 --outMultimapperOrder Random' and non-mapping reads were removed. Alignment files were then converted back to fastq format with samtools (Li et al., 2009) and re-aligned to the transposon consensus sequences allowing multi-mappers that were assigned to a random position. Generated bam alignment files were indexed using samtools index. For genome-wide analyses, multi-mapping reads

were removed to ensure unique locations of reads. Normalization was achieved by calculating rpm (reads per million) using the deepTools2 bamCoverage function with 10 bp bin sizes (Ramirez et al., 2016). Reads mapping to genes were counted with htseq (Anders et al., 2015) and transposon derived reads were calculated using a custom script (**Appendix VIII.2**). Differential expression analysis was performed using custom built R scripts (**Appendix VIII.3-4**). Coordinates used for the piRNA clusters are chrX:21631891-22282863 (*flam*) and chrX:21520428-21556793 (*20A*).

2.9. small RNA-seq

2.9.1. Library preparation

15 µg of RNA from OSCs or ovaries were used for small RNA library preparation as described in (McGinn and Czech, 2014). RNA samples were mixed with radioactively labelled 19- and 28-nt RNA oligos and run on a 15% denaturing polyacrylamide gel for size selection. The gel fragment between the 19- and 28-nt markers was excised and the RNA eluted by soaking overnight in 0.4 M NaCl. RNA samples were cleaned up from residual gel fragments using a Micropore 0.22 µm filter column (Millipore) and precipitated for 2 hrs at -20°C with 1.2 ml Ethanol and 1 µl Glycoblue (Ambion). RNA was pelleted by centrifugation for 30 min at top speed and resuspended in 10 µl of nuclease-free water. 3' adaptor ligation was carried out by adding 4 µl of PEG, 2 µl of 10x ATP-free T4 RNA ligase buffer, 1 µl of RNase Inhibitor, 1 µl of 100 µM 3' adaptor (oLig3), and 2 µl of T4 RNA ligase 2 truncated K227Q and incubating 6 hrs at 16°C. Ligated RNAs were separated on a 10% denaturing polyacrylamide gel. The gel fragment between the 57- and 66-nt size markers (ligated to the 3' adaptor) was excised and the RNA eluted by soaking overnight in 0.4 M NaCl. RNA samples were cleaned up from residual gel fragments using a Micropore 0.22 µm filter column and precipitated for 2 hrs at -20°C with 1.2 ml EtOH and 1 µl Glycoblue. RNA was pelleted by centrifugation for 30 min at top speed and resuspended in 12 µl of nuclease-free water. 5' adaptor ligation was carried out by adding 2 µl of DMSO, 2 µl of 10x T4 RNA ligase buffer, 1 µl of RNase inhibitor, 1 µl of 100 µM 5' adaptor (oLig5), and 2 µl of T4 RNA ligase and incubating for 2.5 hrs at 37°C. Ligated RNAs were separated on a 8% denaturing polyacrylamide gel. The gel fragment between the 90- and 99-nt size markers (ligated to the 5' adaptor) was excised and the RNA eluted by soaking overnight in 0.4 M NaCl. RNA samples were cleaned up from residual gel fragments using a Micropore 0.22 µm filter column and precipitated for 2 hrs at -20°C with 1.2 ml

EtOH and 1 µl Glycoblue. RNA was pelleted by centrifugation for 30 min at top speed and resuspended in 11 µl of nuclease-free water. Reverse transcription was carried out using Superscript III First Strand Synthesis Kit (Thermo Fisher Scientific). Ligated RNAs were mixed with 1 µl of 10mM dNTPmix and 1 µl of 100 µM oRT primer and incubated for 3 min at 65°C, then cooled on ice. Reverse Transcription was performed by adding 4 µl of 5x First-Strand Buffer, 1 µl of 0.1 M DTT, 1 µl of RNase inhibitor, and 1 µl of SuperScript III reverse transcriptase and incubating for 60 min at 50°C, then at 70°C for 15 min. 2.5 µl of library cDNA were amplified via PCR by adding 5 µl of 10X PCR buffer, 5 µl of 2 mM dNTP mix, 3 µl of $MgSO_4$ (25 mM), 0.5 µl of KOD Polymerase, 34.25 µl of nuclease-free water, 0.5 µl of 100 µM oPCR5 and 1 µl of 100 µM indexed oPCR3_x (containing different indices for each libraries). The PCR reaction was run as follows:

95°C for 2 min followed by 5 cycles of:

95°C for 15 sec

54°C for 30 sec

72°C for 15 sec

Run an additional 17 cycles of:

95°C for 15 sec

60°C for 30 sec

72°C for 15 sec

followed by: 72°C for 2 min

4°C forever

The PCR reaction was purified using the Wizard® SV Gel and PCR Clean-Up System (Promega A9285) according to the manufacturer's instructions and resuspended in 27 µl of nuclease-free water. Non-ligated oligos were removed by PmeI digestion, via adding 3 µl of Cutsmart Buffer (NEB) and 2 µl of PmeI enzyme and incubating for 2 hrs at 37°C. The samples were run on a 2% agarose gel and the band corresponding to ~145-154bp was excised and purified using the Wizard® SV Gel and PCR Clean-Up System (Promega A9285), then eluted in 30 µl of nuclease-free water. The libraries were quantified using a NanoDrop Spectrophotometer, quantified with KAPA Library Quantification Kit for Illumina (Kapa Biosystems) and deep-sequenced with Illumina HiSeq 4000 (Illumina).

2.9.2. Data analysis

For small RNA-seq, adapters were clipped from raw fastq files with fastx_clipper (adapter sequence AGATCGGAAGAGCACACGTCTGAACTCCAGTCA) keeping only reads with at least 23 bp length. The first and last 4 bases were trimmed using seqtk (https://github.com/lh3/seqtk). After removal of cloning markers and reads mapping to rRNAs and tRNAs using STAR, high-quality reads were aligned to the *Drosophila melanogaster* genome release 6 (dm6; downloaded from Flybase) using STAR (Dobin et al., 2013). For transposon-wide analysis, genome multi-mapping reads were randomly assigned to one location using option '--outFilterMultimapNmax 1000 --outMultimapperOrder Random' and non-mapping reads were removed. Alignment files were then converted back to fastq format with samtools (Li et al., 2009) and re-aligned to the transposon consensus sequences allowing multi-mappers that were assigned to a random position. Generated bam alignment files were indexed using samtools index. For genome-wide analyses, multi-mapping reads were removed to ensure unique locations of reads. Small RNA-seq reads were normalised to miRNA reads in the control library (set to rpm). Only high-quality small RNA reads with a length between 23 and 29 bp were used for further analysis of small RNA profiles. piRNA distribution was calculated and plotted in R. The ping-pong signature was calculated using piPipes (Han et al., 2015).

2.10. CLIP-seq

2.10.1. HALO-CLIP

$1x10^7$ OSCs were nucleofected first with 2 µl of siRNA only and, 48 hrs later, with 2 µl of siRNA and 5 µg of the desired plasmid. 96 hrs later, cells were crosslinked on ice with 150 mJ/cm^2 at 254 nm. Cell pellets were lysed in 300 µl of Lysis Buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% Triton® X-100, 0.1% deoxycholate, Protease Inhibitor [1:50 Promega] and RNasin Plus [1:500, Promega]), for 30 min at 4°C. For CLIP of nuclear proteins, DNase digestion was performed by adding 2 ul (4U) of Turbo DNase to the cell lysate and immediately placing the samples at 37°C for 3 min, shaking at 1,100 rpm. Samples were transferred to and kept on ice for >3 min, then cleared by centrifugation at top speed for 20 min at 4°C. Protein concentration was measured with a Direct Detect instrument and equal amounts of cell lysates were diluted up to 1 ml with 100 mM Tris-HCl pH 7.5, 150 mM NaCl and incubated with 200 µl of Magne-HaloTag® (Promega G7282) beads overnight at 4°C. Beads were washed 2x in Wash Buffer A (100 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05% IGEPAL® CA-630), 3x in

Wash Buffer B (PBS, 500 mM NaCl, 0.1 % Triton X-100, RNasin Plus 1:2,000), 3x in PBS, 0.1% Triton X-100, and rinsed in Wash Buffer A. For release of the bait protein from the tag, beads were resuspended in 100 µl of 1X ProTEV Buffer, 1 mM DTT and RNasin Plus (1:50) and 25 units of ProTEV Plus Protease (Promega V6101) and incubated 2 hrs at 30°C, shaking at 1,300 rpm. The supernatant containing the eluted protein and the crosslinked RNA was transferred to a fresh tube and 15 µl Proteinase K in 300 µl PK/SDS buffer (100 mM Tris, pH 7.5; 50 mM NaCl; 1 mM EDTA; 0.2% SDS) were added to the eluate and incubated 1 hr at 50°C. 400 µl of Phenol:Chloroform:Isoamyl alcohol were added to the tube, vortexed thoroughly and spun for 10 min at 12,000g at room temperature. The upper aqueous phase was precipitated overnight at -20°C with 40 µl of 3M sodium acetate pH 5.5, 1 µl of Glycoblue and 1 ml of Ethanol. RNA was pelleted by 30 min centrifugation at 4°C, washed once with 80% Ethanol and resuspended in 8 µl of nuclease-free water. The recovered RNA was used directly for library preparation.

2.10.2. Library preparation

Library preparation for deep sequencing was carried out with the SMARTer® Stranded RNAseq kit (Takara Bio 634839), according to the manufacturer's instructions. Prior to the first-strand cDNA synthesis, 8 µl of CLIP RNA were fragmented by mixing with 1 µl of SMART stranded N6 primer and 4 µl of 5x First Strand Buffer and incubating for 4 min at 94°C, then cooled on ice for 2 min. The first strand reaction master mix (containing 0.5 µl DTT, 0.5 µl of RNase Inhibitor, 2 µl of dNTP mix, 2 µl of SMARTer Stranded oligo and 2 µl of SMARTScribe Reverse Transcriptase) was immediately added to the fragmented RNA and the reaction was incubated for 90 min at 42°C followed by 10 min at 70°C. The cDNA was purified AMPure XP beads as follows. 20 µl of beads were mixed with 20 µl of cDNA, vortexed and incubated for 8 min at room temperature. Samples were placed on a magnetic rack, the supernatant discarded and the beads washed twice in freshly-prepared 70% Ethanol. Beads were air-dried for 5 min at room temperature and the cDNA eluted in 20 µl of Nuclease-free water. Library PCR amplification reaction was set up by adding to the purified cDNA 25 µl of 2X SeqAmp PCR Buffer, 2 µl of pre-mixed Forward and Reverse SMARTer RNA Dual Indexes (Clontech 634451), 1 µl of SeqAmp DNA Polymerase and 2 µl of Nuclease-free water. The PCR reaction was carried out as follows:

1 min at 94°C

18 cycles:

15 sec at 98°C

15 sec at 55°C

30 sec at 68°C

The amplified library was purified with 50 µl of AMPure XP beads as described above. The purified library was eluted in 20 µl of Stranded Elution Buffer and a 1:4 dilution was analysed on a Bioanalyser or Tapestation instrument (Agilent). DNA libraries were quantified with KAPA Library Quantification Kit for Illumina (Kapa Biosystems) and deep-sequenced with Illumina HiSeq 4000 (Illumina).

2.10.3. Data analysis

Raw fastq files generated by Illumina sequencing were analysed by a pipeline developed in-house. In short, for CLIP-seq the first 5 bases of each 50 bp read were removed using fastx trimmer (http://hannonlab.cshl.edu/fastx_toolkit/). After removal of reads mapping to *Drosophila* rRNA using STAR, high-quality reads were aligned to the *Drosophila melanogaster* genome release 6 (dm6; downloaded from Flybase) using STAR (Dobin et al., 2013). For transposon-wide analysis, genome multi-mapping reads were randomly assigned to one location using option '--outFilterMultimapNmax 1000 --outMultimapperOrder Random' and non-mapping reads were removed. Alignment files were then converted back to fastq format with samtools (Li et al., 2009) and re-aligned to the transposon consensus sequences allowing multi-mappers that were assigned to a random position. Generated bam alignment files were indexed using samtools index. For genome-wide analyses, multi-mapping reads were removed to ensure unique locations of reads. Normalization was achieved by calculating rpm (reads per million) using the deepTools2 bamCoverage function with 10 bp bin sizes (Ramirez et al., 2016). Reads mapping to genes were counted with htseq (Anders et al., 2015) and transposon derived reads were calculated using a custom script (**Appendix VIII.2**). Differential expression analysis was performed using custom built R scripts (**Appendix VIII.3-4**). Coordinates used for the piRNA clusters are chrX:21631891-22282863 (*flam*) and chrX:21520428-21556793 (*20A*).

2.11. RIP-seq

2.11.1. GFP RIP from fly ovaries

Ovaries from ~100 GFP-Panx or GFP-Nxf2 flies (3-5 days old) were dissected in ice-cold PBS and fixed with 0.1% PFA for 20 min, followed by quenching with equal

volumes of 125 mM Glycine. Fixed ovaries were lysed in 200 µl of RIPA Buffer (supplemented with complete protease inhibitors (Roche) and RNasin Plus 40 U/ml) and homogenized using a motorized pestle. Lysates were incubated 3 min at 37 °C with 4 µl of Turbo DNase, incubated 20 min at 4°C on a tube rotator and sonicated with a Bioruptor® Pico (Diagenode; 3 cycles of 30 sec on/30 sec off). Lysates were pre-cleared using 40 µl of Pierce Protein A/G beads for 1 hr at 4°C and GFP-tagged proteins were immunoprecipitated by incubation with 50 µl of GFP-Trap magnetic agarose beads (Chromotek) overnight at 4°C. An aliquot of pre-cleared input lysate was saved for RNA isolation and library preparation. Following 3 washes in 150 mM KCl, 25 mM Tris (pH 7.5), 5 mM EDTA, 0.5% NP40, 0.5 mM DTT (supplemented with protease inhibitors and RNasin Plus 1:1,000), IP and input samples were reverse crosslinked in 1x Reverse Crosslinking buffer (PBS, 2% Nlauroyl sarcosine, 10 mM EDTA, 5 mM DTT) and Proteinase K. RNA isolation was performed using Trizol and quantified using NanoDrop.

2.11.2. Library preparation

100 ng of input or IP RNA were used for library preparation using the SMARTer® stranded RNA-seq Kit (Takara Bio 634839), as described in **2.10.2**. DNA libraries were quantified with KAPA Library Quantification Kit for Illumina (Kapa Biosystems) and deep-sequenced with Illumina HiSeq 4000 (Illumina).

2.11.3. Data analysis

Analysis was carried out as described for CLIP-seq in **2.10.3**.

2.12. Proximity labelling and Mass Spectrometry (PL-MS)

2.12.1. BASU PL and pulldown

$4x10^6$ OSCs were transfected with 20 µg of plasmid expressing an HA-BASU fusion or HA-ZsGreen. After 48 hrs, the media was supplemented with 200 µM Biotin for 1 hr. Cell pellets were lysed in 1.8 ml Lysis buffer (50 mM Tris, pH 7.4, 500 mM NaCl, 0.4% SDS, 1 mM DTT, 2% Triton-100 with protease inhibitors) and sonicated using a Bioruptor Pico (Diagenode, 3x cycles 30 sec on / 30 sec off). Sonicated lysates were diluted 2x in 50 mM Tris, pH 7.4 and cleared for 10 min at 16,500g. Following pre-clearing of the lysate with 100 µl of Protein A/G Dynabeads (Thermo Fischer Scientific 10015D), biotinylated proteins were isolated by incubation with 200 µl of Dynabeads

(MyOne Streptavidin C1; Life Technologies) overnight at 4°C. The beads were washed 2x in 2% SDS, 2x in Wash Buffer 1 (0.1% deoxycholate, 1% Triton X-100, 500 mM NaCl, 1 mM EDTA, and 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid, pH 7.5), 2x with Wash Buffer 2 (250 mM LiCl, 0.5% NP-40, 0.5% deoxycholate, 1 mM EDTA, and 10 mM Tris, pH 8), and 2x with 50 mM Tris. Beads were rinsed twice with 100 mM Ammonium Bicarbonate and submitted for Mass Spectrometry by the CRUK CI proteomics core facility. HA-BASU-Daed, Yb, Nup54 and Nup58 pulldowns were subjected to TMT-labelling followed by quantitative Mass Spectrometry on a nano-ESI Fusion Lumos mass spectrometer (Thermo Fisher Scientific). BASU-Gasz, Armi-BASU, Zuc-BASU and Zuc-SplitBioID pulldowns were analysed on a Q-Exactive HF mass spectrometer (Thermo Fisher Scientific). On bead Trypsin digestion and TMT chemical isobaric labelling were performed as described (Papachristou et al., 2018).

2.12.2. Split-BioID PL and pulldown

$4x10^6$ OSCs were transfected with 10 µg of each plasmid expressing Zuc-CBirA*-6xHis and Zuc-NBirA*-HA or 20 µg of HA-ZsGreen. After 36 hrs, the growth media was supplemented overnight (~18 hrs) with 50 mM Biotin. Harvesting and pulldown of biotinylated proteins were performed as stated in **2.12.2**.

2.12.3. Data analysis

Spectral .raw files from PL-MS of BASU-Daed, Yb, Nup54 and Nup58 were processed with the SequestHT search engine on Thermo ScientificTM Proteome Discoverer™ 2.1. Data was searched against a custom FlyBase database ("dmel-all-translation-r6.24") at 1% spectrum level FDR criteria using Percolator (University of Washington). MS1 mass tolerance was constrained to 20 ppm and the fragment ion mass tolerance was set to 0.5 Da. TMT tags on lysine residues and peptide N termini (+229.163 Da) were set as static modifications. Oxidation of methionine residues (+15.995 Da), deamidation (+0.984) of asparagine and glutamine residues, and biotinylation of lysines and protein N-terminus (+226.078) were included as dynamic modifications. For TMT-based reporter ion quantitation, we extracted the signal-to-noise ratio for each TMT channel. Parsimony principle was applied for protein grouping and the level of confidence for peptide identifications was estimated using the Percolator node with decoy database search. Strict FDR was set at q-value < 0.01. Downstream data analysis was performed on R using the qPLEXanalyzer package (https://doi.org/10.5281/zenodo.1237825) as

described (Papachristou et al., 2018). Only proteins with more than one unique peptide were plotted.

Spectral .raw files from PL-MS of BASU-Gasz, Armi-BASU, Zuc-BASU, Zuc Split-BioID were processed with the SequestHT search engine on Thermo ScientificTM Proteome Discoverer™ 2.2. The node for SequestHT included the same parameters as above (except Fragment Mass Tolerance set to 0.02 Da) and modifications. The Precursor Ion Quantifier node (Minora Feature Detector) included a Minimum Trace Length of 5, Max. ΔRT of Isotope Pattern 0.2 minutes. For calculation of Precursor ion intensities, Feature mapper was set True for RT alignment (mass tolerance of 10 ppm). Precursor abundance was quantified based on intensity and the level of confidence for peptide identifications was estimated using the Percolator node with a Strict FDR at q-value < 0.01. Analysis of label-free quantification protein intensity data was carried out in R (v 3.5.1) using the qPLEXanalyzer package (v 1.0.3) (Papachristou et al., 2018). Peptides for which all control (ZsGreen) samples lacked measurements or for which more than 1 target protein sample lacked measurements were discarded. Remaining missing values were then imputed using the nearest neighbour averaging (knn) imputation method provided in the R package MSnbase (v 2.8.3) (Gatto and Lilley, 2012). Differential analysis was carried out by linear modelling using limma based methods provided by the qPLEXanalyzer package. Multiple testing correction of p-values was applied using the Benjamini & Yekutieli method to control FDR (Benjamini et al., 2001). Only proteins with more than one unique peptide were plotted.

2.13. Immunoprecipitation and Mass Spectrometry (IP-MS)

2.13.1. GFP IP-MS from fly ovaries

Ovaries from ~170 GFP-Panx, GFP-Nxf2 and control flies (3-5 days old) were dissected in ice-cold PBS and lysed in 300 µl of CoIP Lysis Buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM MgCl2, 10% glycerol, 1 mM DTT, 0.1 mM PMSF, 0.2% NP-40 supplemented with complete protease inhibitors [Roche]) and homogenized using a motorized pestle. Lysates were cleared for 5 min at 16,000g and the residual pellet re-extracted with the same procedure. GFP-tagged proteins were immunoprecipitated by incubation with 30 µl of GFP-Trap magnetic beads (Chromotek) for 3 hrs at 4°C on a tube rotator. The beads were washed 6x with Lysis Buffer and 2x with 100 mM Ammonium Bicarbonate, before TMT-labelling followed by quantitative Mass

Spectrometry. TMT chemical isobaric labelling was performed as described (Papachristou et al., 2018).

2.13.2. FLAG IP-MS from OSCs

$10x10^6$ OSCs were nucleofected with 2 µl of an siRNA targeting Yb 3'UTR and, two days later, again with 2 µl of the same siRNA and 5 µg of plasmid expressing 3xFLAG-tagged constructs, as described in 2.2.2.3. Cell pellets were lysed in 250 µl of CoIP Lysis Buffer (Choi et al.) with cOmplete™ Mini EDTA-free Protease Inhibitor Cocktail (Roche; 11836170001) and rotated at 4°C for 30 min. Cell lysates were cleared with centrifugation at top speed for 15 min at 4°C and protein concentration measured with a Direct Detect Spectrometer (Merck Millipore; DDHW00010-WW). 400 µg of protein for each sample were diluted to 1 ml with CoIP Lysis Buffer and incubated with 50 µl of anti-FLAG M2 Magnetic Beads (Sigma M8823) overnight at 4°C. The beads were washed 3x15 min in TBS with protease inhibitors and a 10% aliquot was saved for western blot analysis. The remaining beads were rinsed 2x in PBS and 2x in 100 mM Ammonium Bicarbonate, prior to submission for TMT labelling and quantitative Mass Spectrometry. TMT chemical isobaric labelling was performed as described (Papachristou et al., 2018).

2.13.3. Data analysis

Spectral .raw files were processed with the SequestHT search engine on Thermo ScientificTM Proteome Discoverer™ 2.1. Data was searched against a custom FlyBase database ("dmel-all-translation-r6.24") at a 1% spectrum level FDR criteria using Percolator (University of Washington). MS1 mass tolerance was constrained to 20 ppm and the fragment ion mass tolerance was set to 0.5 Da. TMT tags on lysine residues and peptide N termini (+229.163 Da) were set as static modifications. Oxidation of methionine residues (+15.995 Da), deamidation (+0.984) of asparagine and glutamine residues, and biotinylation of lysines and protein N-terminus (+226.078) were included as dynamic modifications. For TMT-based reporter ion quantitation, we extracted the signal-to-noise ratio for each TMT channel. Parsimony principle was applied for protein grouping and the level of confidence for peptide identifications was estimated using the Percolator node with decoy database search. Strict FDR was set at q-value < 0.01. Downstream data analysis was performed on R using the qPLEXanalyzer package (https://doi.org/10.5281/zenodo.1237825) as described (Papachristou et al., 2018). Only proteins with more than one unique peptide were plotted.

2.14. Co-IP from cells

S2 cells or OSCs were transfected with 3xFLAG- and HA-tagged constructs. After 48 hrs, cells were lysed in 250 µl of CoIP Lysis Buffer (Choi et al.) with cOmplete™ Mini EDTA-free Protease Inhibitor Cocktail (Roche; 11836170001). For cross-linking experiments, cell pellets were incubated with disuccinimidyl sulfoxide at 1 mM final concentration (diluted in PBS) for 10 min at room temperature and 20 min at 4°C followed by lysis in 50 mM Tris, pH 7.4, 500 mM NaCl, 0.4% SDS, 1 mM dithiothreitol, 2% Triton-100 with protease inhibitors and sonication using a Bioruptor Pico (Diagenode; 3x cycles 30 sec on / 30 sec off). 200 µg of proteins for each sample were diluted to 1 ml with CoIP Lysis Buffer and incubated with 30 µl of anti-FLAG M2 Magnetic Beads (Sigma M8823) for 2 hrs at 4°C. The beads were washed 3x15 min in TBS with protease inhibitors, then resuspended in 2x NuPAGE LDS Sample Buffer (Thermo Fisher Scientific) and boiled for 3 min at 90°C to elute immunoprecipitated proteins. 2% of input and unbound fractions and 25% of immunoprecipitate were used for Western Blot analysis as described in 2.15.

2.15. Western Blot

Ovaries or cell pellets were lysed in Pierce RIPA Buffer (Thermo Fisher Scientific, 89900) supplemented with cOmplete™ Mini EDTA-free Protease Inhibitor Cocktail (Roche; 11836170001). Protein concentration was measured using a Direct Detect Spectrometer (Merck Millipore; DDHW00010-WW). 20 µg of protein were mixed with NuPAGE™ LDS Sample Buffer (4X; Thermo Fisher Scientific NP0007) and NuPAGE™ Sample Reducing Agent (10X; Thermo Fisher Scientific NP0004) and boiled for 10 min at 70°C. proteins were separated on NuPAGE™ 4-12% Bis-Tris Protein Gels (Thermo Fisher Scientific) and transferred to a nitrocellulose membrane using a dry transfer system (iBlot™ 2 Dry Blotting System; Invitrogen IB21001) with program P0 (20 V for 1 minute, 23 V for 4 minutes, 25 V for 2 minutes). The membrane was blocked for 1hr at room temperature in 1x Odyssey Blocking Buffer TBS (LiCor) and incubated overnight at 4°C with primary antibodies diluted in 1x Odyssey Blocking Buffer TBS + 0.1% Tween-20 (TBST). Following 3x5 min washes in TBST, the membranes were incubated for 45 min at room temperature with the infra-red dye-conjugated secondary antibodies (LiCor; 1: 10,000 dilution in 1x Odyssey Blocking Buffer TBST) or Streptavidin (LiCor; 1:4,000 dilution in 1x Odyssey Blocking Buffer TBST). Following 3x5 min washes at room temperature with TBS, images were acquired on an Odyssey CLx scanner (LiCor) Images were processed using Image Studio.

2.16. RNA tethering

For RNA tethering, OSCs with a stable integration of the sensor plasmid (pBC392 and pBC468) were generated as described in 2.2.2.4. $4x10^6$ cells were nucleofected with 5 µg of plasmid expressing λN-tagged constructs, as described in 2.2.2.2. After 48 hrs, $4x10^6$ cells were nucleofected again with 5 µg of the same plasmid and allowed to grow for an additional 48 hrs before the relative expression of the sensor was analyzed. Western Blot and qPCR analysis were carried out as described in 2.15 and 2.3, respectively.

CHAPTER III

Daedalus and Gasz recruit Armitage to mitochondria, bringing piRNA precursors to the biogenesis machinery

Author contributions

This chapter has been published in a similar form as: **Munafò M**., Manelli V., Falconio F.A., Sawle A., Kneuss E., Eastwood E.L., Seah J.W.E., Czech B., and Hannon G.J. "Daedalus and Gasz recruit Armitage to mitochondria, bringing piRNA precursors to the biogenesis machinery." Genes Dev. 2019;33(13-14):844–856. doi:10.1101/gad.325662.119

I performed all experiments with help from V.M., F.A.F. and B.C.. I analysed all sequencing data. A.S. analysed the proteomics data, E.L.E. generated *daed* mutant flies and shRNA lines, E.K. contributed to the initial characterisation of their phenotype and J.W.E.S. generated *gasz* mutant flies. I, B.C. and G.J.H. designed the experiments, analysed and interpreted the data. I wrote the first draft of the manuscript with inputs from B.C. and G.J.H.. All authors proofread the final version of the manuscript.

BACKGROUND

piRNAs originate from long, single-stranded precursors transcribed from piRNA clusters or transposon loci (Brennecke et al., 2007). These precursors are exported to specialised peri-nuclear structures, namely nuage in germ cells and Yb bodies in somatic cells, where they become licensed for downstream processing. Their specification involves the formation of a 5'-P end, via the ping-pong cycle in nuage or a currently unknown mechanism in Yb bodies (Gainetdinov et al., 2018; Han et al., 2015; Mohn et al., 2015; Senti et al., 2015; Wang et al., 2015). 5'-P precursors are processed on the surface of mitochondria, where the conserved endonuclease Zuc cleaves them to release phased ~25 nt long piRNAs (Han et al., 2015; Homolka et al., 2015; Ipsaro et al., 2012; Mohn et al., 2015; Nishimasu et al., 2012). The mitochondrial localisation of the piRNA biogenesis machinery is a remarkably conserved feature, found in many animals in which this process has been characterised so far (Ipsaro et al., 2012; Nishida et al., 2018; Nishimasu et al., 2012; Watanabe et al., 2011). Several other mitochondrial proteins are required for successful piRNA production but, in many cases, lack an ascribed function (Czech et al., 2018). Finally, this process also requires cytosolic factors, among which is the Upf1-like RNA helicase Armi, previously identified as necessary for piRNA production in somatic and germ cells of the fly ovary (Malone et al., 2009; Olivieri et al., 2010; Pandey et al., 2017; Rogers et al., 2017; Saito et al., 2010).

Our current model of piRNA production identifies two subcellular compartments as being critical for piRNA production: the nuage/Yb bodies, where piRNA precursors are exported and specified for processing, and mitochondria, where piRNAs are produced. Nonetheless, these two compartments are spatially separated and it is thus unclear how piRNA precursors are specifically recognized and directionally trafficked towards the mitochondrial surface.

RESULTS

III.1 CG10880/Daedalus is a mitochondrially localised protein required for piRNA biogenesis

Several genetic screens have provided a general framework of how the production of piRNAs is orchestrated (Czech et al., 2013; Handler et al., 2013; Muerdter et al., 2013). Yet, a detailed understanding of the role of each of the identified factors is still lacking. Among screen hits involved in germline transposon control in *Drosophila* was the uncharacterised gene *CG10880*

(Czech et al., 2013). The *CG10880* locus is on the left arm of chromosome 2 and the corresponding transcript shows the highest expression levels in the adult fly ovary, similar to other piRNA pathway factors (e.g. *armi* and *zuc* in **Figure 3.1A**) (Chintapalli et al., 2007). Depletion of *CG10880* in the fly germline (hereafter referred to as GLKD) using shRNAs leads to transposon de-repression at levels comparable to that caused by loss of *zuc* or *gasz* (**Figure 3.1B**) (Czech et al., 2013). Furthermore, Piwi is equally de-localised from germline nuclei upon *zuc, gasz* and *CG10880* knockdown (**Figure 3.1C**). This is a hallmark of impaired piRNA biogenesis, since Piwi is unstable without associated piRNAs (Olivieri et al., 2012). Hence, I set out to investigate the role of the uncharacterised gene *CG10880*.

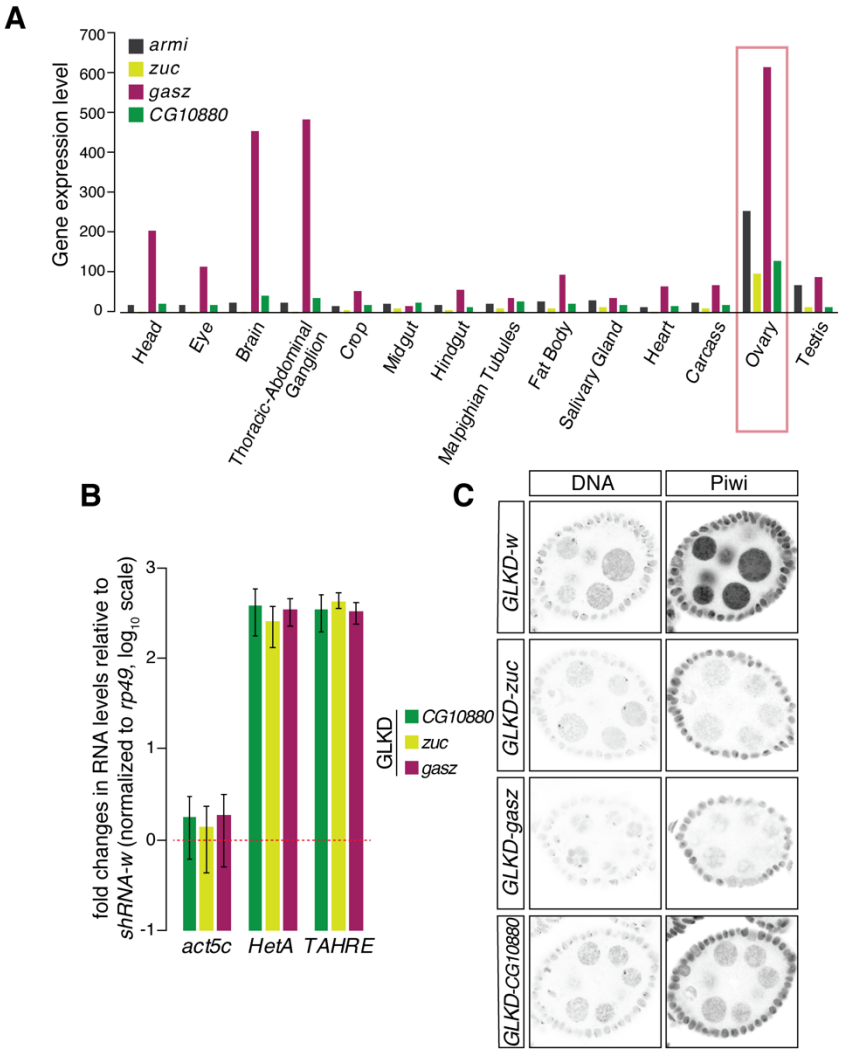


Figure 3.1 CG10880 is implicated in germline transposon control. A) Expression levels of *armi*, *zuc*, *gasz* and *CG10880* in various tissues of the adult fly. **B)** Fold changes in the steady-state RNA levels of transposons from Germline knockdown (GLKD) ovaries. Values are relative to *white* (*w*) knockdown and normalised to *rp49*. Error bars indicate standard deviation (n=2). **C)** Confocal images of Piwi in ovaries upon indicated GLKD.

CG10880 encodes a 272 amino acid polypeptide which harbours a Sterile Alpha Motif (SAM, possibly involved in protein-protein or protein-RNA interaction), a Coiled Coil (CC, often involved in protein oligomerization and linked to various cellular functions) and a predicted transmembrane domain at its carboxy-terminus (TMM) (cartoon in **Figure 3.2A**). Interestingly, the CG10880 domain structure resembles that of the established piRNA biogenesis factor Gasz (Czech et al., 2013; Handler et al., 2013), which carries two series of ankyrin repeats (ANK, mostly involved in protein-protein interactions) in addition to a SAM and a TMM, the latter mediating its mitochondria localisation (cartoon in **Figure 3.2B**). To determine the subcellular localisation of CG10880 in the fly ovary, I generated flies expressing a GFP-CG10880 fusion under the control of the constitutive *Drosophila* Ubiquitin promoter. The GFP tag was inserted at the amino-terminus to preserve the putative TMM domain. As shown in **Figure 3.2C**, GFP-CG10880 co-localises with the mitochondrial marker Atp5a, and is adjacent to but separate from nuage, marked by Ago3. This pattern is virtually indistinguishable from that of Zuc-GFP (also ubiquitously expressed) or GFP-Gasz (driven by its endogenous regulatory elements) (Handler et al., 2013).

Triggered by these results, we hypothesized that CG10880 is implicated in mitochondrial piRNA biogenesis and decided to generate null mutants using CRISPR/Cas9. We used a pair of gRNAs targeting the 5' proximal region of the CG10880 Open Reading Frame (ORF) to generate indels and obtained two mutant alleles (cartoon in **Figure 3.3A**). One allele harbours a deletion disrupting the ORF of CG10880, thus leading to an aberrant out-of-frame polypeptide ($CG10880^{oof1}$), whilst the other carries a deletion causing a premature stop codon that only produces a 4 amino acid peptide ($CG10880^{\Delta 2*}$) (**Figure 3.3A**). Unless otherwise specified, all the following experiments have been carried out by crossing the two alleles and analysing the trans-heterozygous offspring ($CG10880^{oof1/\Delta 2*}$, hereafter referred to as "homozygous mutants"). These are compared to their heterozygous sibling, which are a mixed population of the two alleles. The phenotype of *CG10880* null mutants was also compared to that of a *gasz* mutant line previously generated in the lab and depicted in **Figure 3.3B**. The Gasz ORF is replaced through knock-in of an RFP cassette under the control of the 3xP3-Hsp70 promoter, hereafter referred to as $gasz^{KO}$.

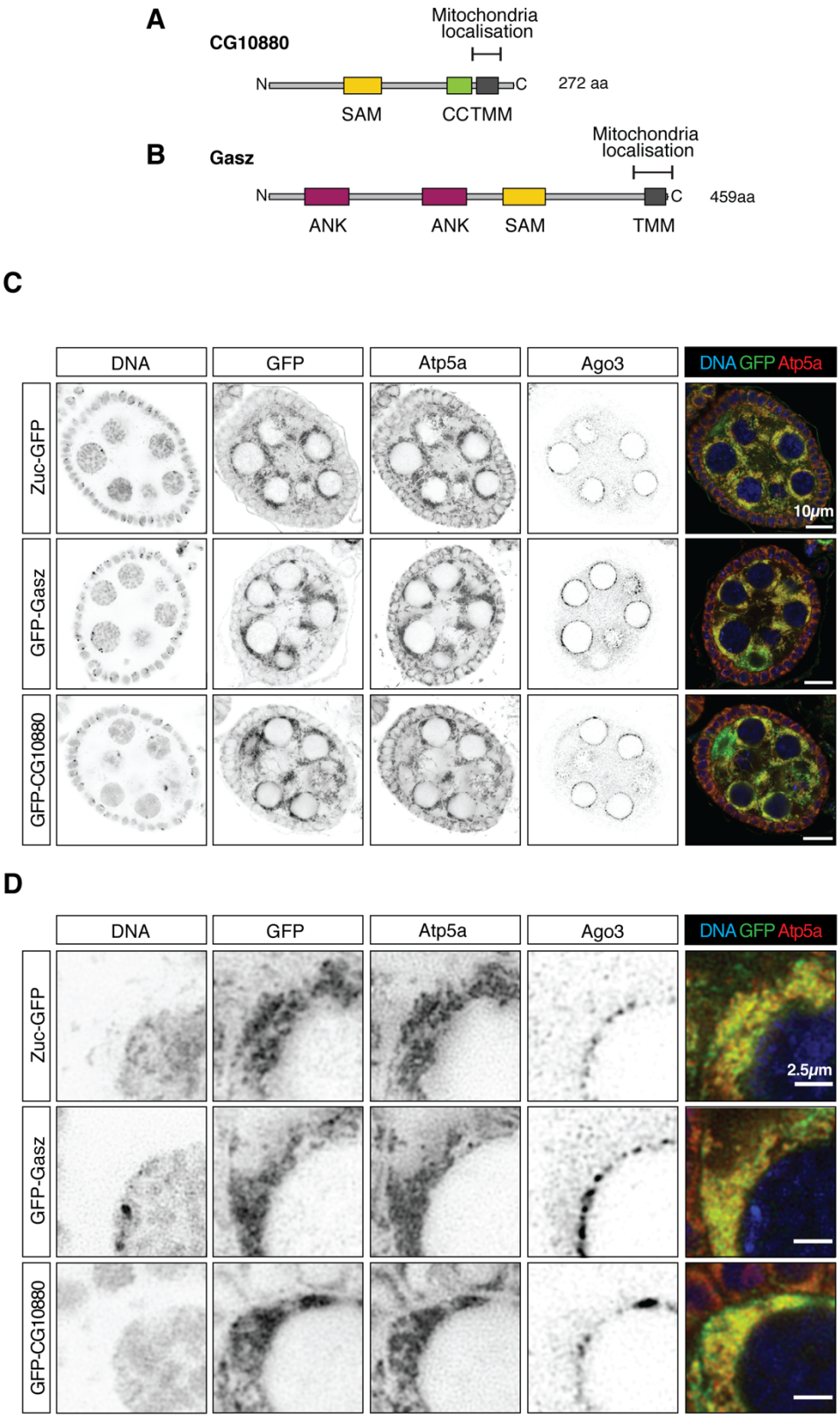


Figure 3.2 **CG10880 is a mitochondrial protein. A-B)** Cartoon representing CG10880 and Gasz domain structure. ANK = Ankyrin repeats, SAM = Sterile Alpha Motif, CC = Coiled Coil, TMM = transmembrane domain. **C)** Confocal images of Zuc-GFP, GFP-Gasz or GFP-CG10880 with Atp5a and Ago3 in ovaries. Scale bar, 10µm.

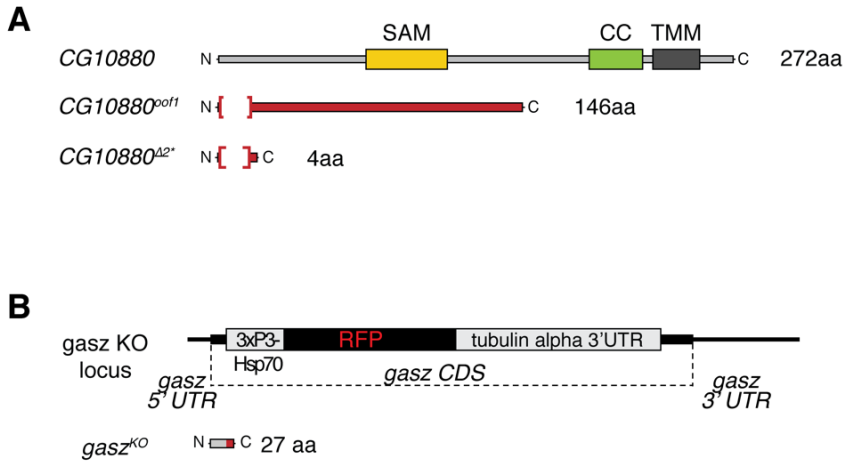


Figure 3.3 Mutant alleles of *CG10880* and *gasz*. A) Schematic representation of the CG10880 domain structure and of the two null alleles (*CG10880*^{oof1} and *CG10880*$^{\Delta 2*}$). Highlighted in red is the portion of coding sequence that is out of frame in the mutants. **B)** Schematic representation of the *gasz*^{KO} allele.

I first asked whether mutations in *CG10880* affect female fertility. To do so, I crossed control flies (*w*^{1118}), heterozygous or homozygous mutant females with control males and counted the number of eggs laid and how many of those developed into adult flies. As expected from a mutant defective in the piRNA pathway, *CG10880*$^{\Delta 2*/\Delta 2*}$ homozygous mutant females laid fewer eggs than controls and none of them hatched. Likewise, *gasz*^{KO} eggs also failed to hatch (**Figure 3.4A**). I then isolated RNA from heterozygous and homozygous mutant ovaries and carried out RT-qPCR analysis to probe for transposon levels. Transposon repression was indeed impaired in *gasz* and *CG10880* homozygous mutants, with both soma- and germline-specific elements being significantly upregulated compared to controls (**Figure 3.4B**). Piwi nuclear localisation was dramatically lost from both somatic and germ cells of the homozygous mutants (**Figure 3.4C**, extended panel in **Figure 3.5A**), arguing that CG10880, like Gasz, is required for piRNA biogenesis in both compartments of the fly ovary. Interestingly, I observed that mutations in *CG10880* also impacted the localisation of the RNA helicase Armi (**Figure 3.4C**, extended panel in **Figure 3.5A**), which is a known piRNA biogenesis factor typically localised to nuage and mitochondria. *CG10880* homozygous mutants showed that Armi was aberrantly dispersed in the cytosol (**Figure 3.4C**, extended panel in **Figure 3.5A**). Interestingly, the same was true upon *gasz* loss, both in our mutants and in a previously reported germline knockdown (Handler et al., 2013) (**Figure 3.4C**, extended panel in **Figure 3.5A**), raising the possibility that CG10880 and Gasz are performing similar functions. To further characterise the *CG10880*

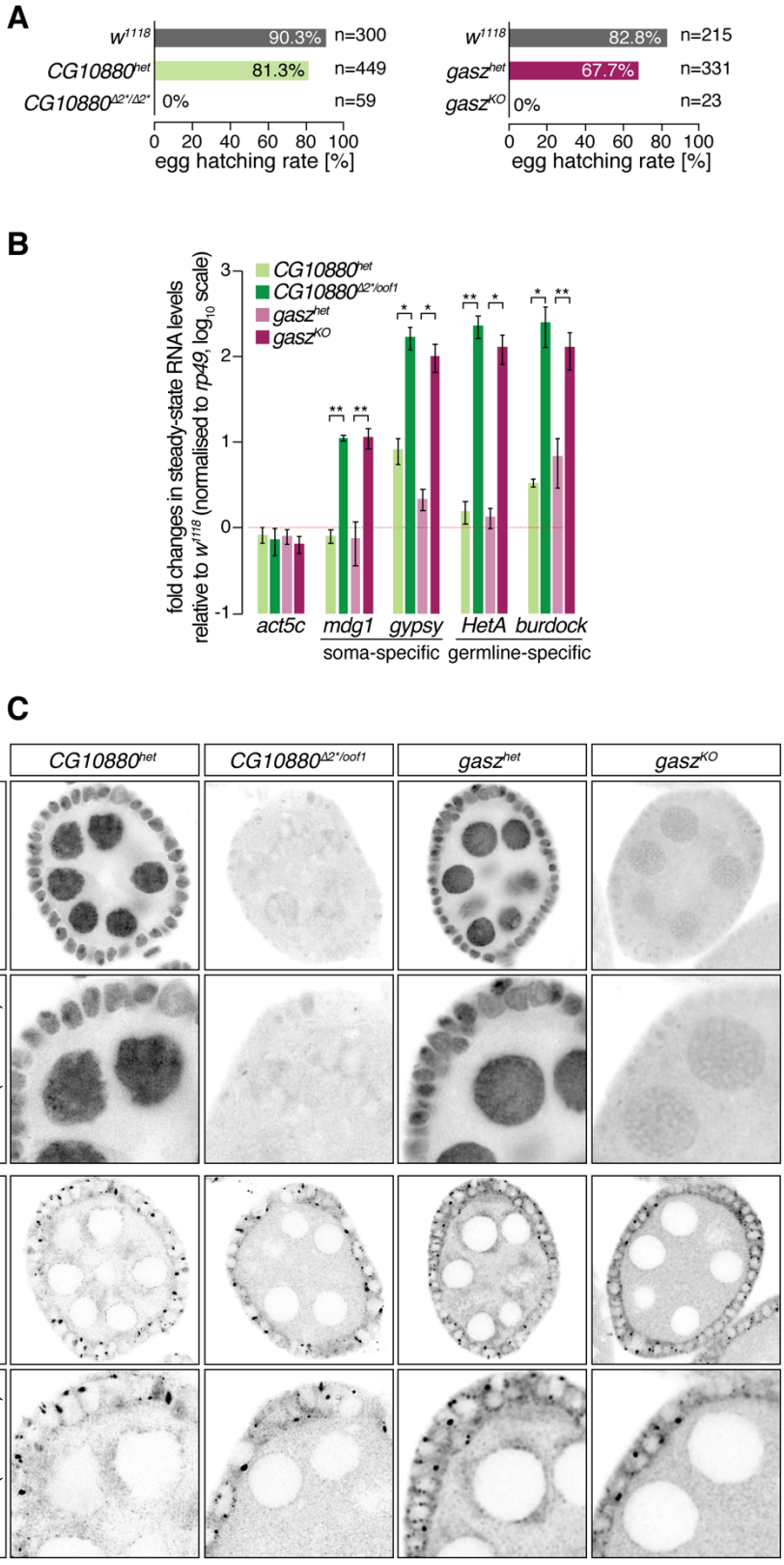


Figure 3.4 CG10880/Daedalus is a *bona-fide* piRNA biogenesis factor acting in soma and germline. A) Bar graphs showing the egg hatching rates of female flies of the indicated genotypes. **B)** Fold changes in steady-state RNA levels of the indicated soma- and germline-specific transposons from ovaries. Values are relative to w^{1118} flies and normalized to *rp49*. * = P value < 0.05; ** = P < 0.001 (unpaired t-test). Error bars indicate standard deviation (n=4). **C)** Confocal images of Piwi and Armi in ovaries of the indicated genotypes (see also **Figure 3.5A-C**)

mutants, I also probed for the localisation of the nuage proteins Aub and Vasa. Neither was affected upon loss of *gasz* or *CG10880* (**Figure 3.5B**), consistent with a function at the level of Zuc-mediated piRNA production and downstream of the ping-pong cycle (Olivieri et al., 2012). Finally, *CG10880* mutants had a highly altered mitochondrial morphology, again resembling what was observed in the absence of *gasz* (Handler et al., 2013) (**Figure 3.5C**). Taken together, these results suggest that CG10880 is a *bona-fide* piRNA biogenesis factor involved in Zuc-mediated processing of phased piRNAs on mitochondria and, since it is required for the correct assembly of the mitochondrial "labyrinth" in germ cells, I named it Daedalus (Daed).

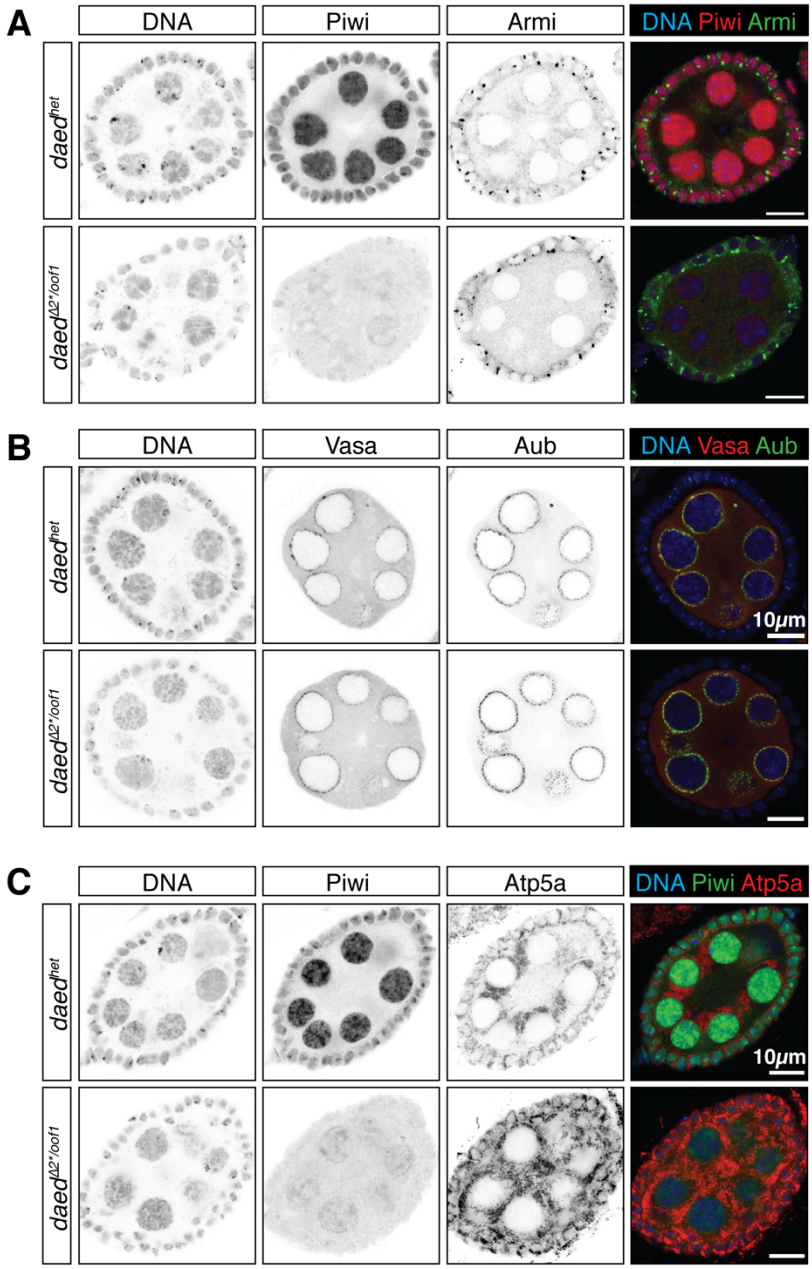


Figure 3.5 **Characterisation of *daed* and *gasz* mutants. A-C)** Confocal images of Piwi, Armi, Vasa, Aub and Atp5a in ovaries. Scale bars, 10µm.[continued to the next page]

Given the resemblance between *daed* and *gasz* mutant phenotypes and the relatively similar domain structure, we also asked whether overexpression of one can rescue loss of the other. Toward this goal, we expressed the GFP-Gasz transgene in a $daed^{\Delta 2*}$ homozygous mutant background. As shown in **Figure 3.6**, over-expression of GFP-Gasz did not alleviate either Piwi nuclear loss nor the distorted mitochondrial morphology caused by *daed* knockout. Therefore, we concluded that both Gasz and Daed are individually required for piRNA production and mitochondria assembly.

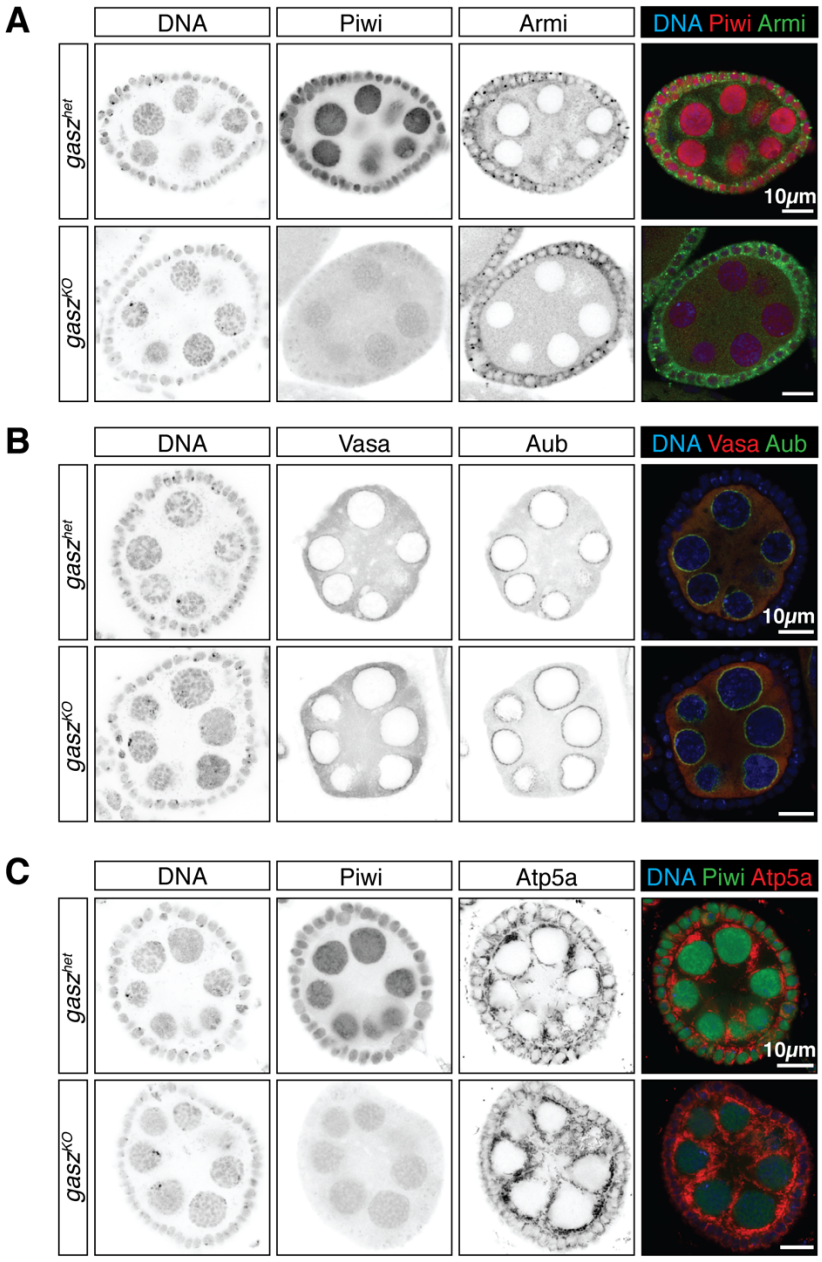


Figure 3.5 **Characterisation of *daed* and *gasz* mutants.**

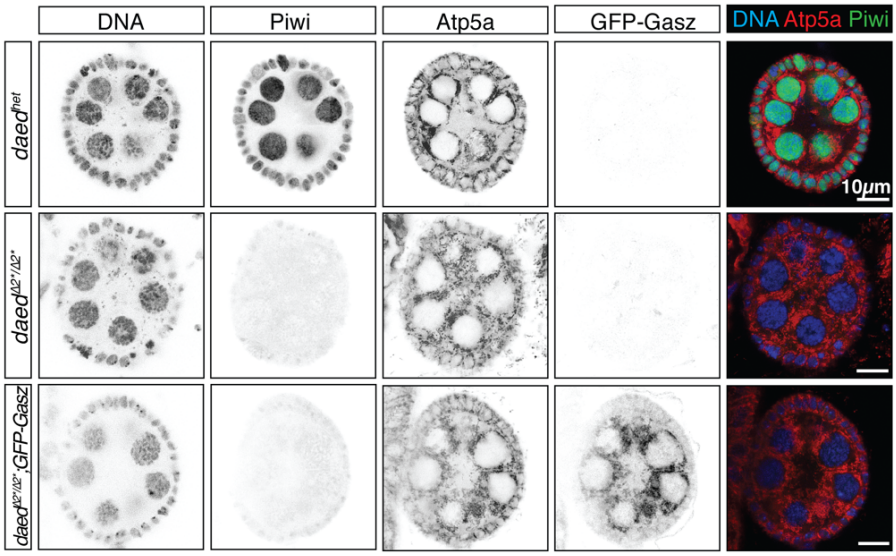


Figure 3.6 Gasz overexpression does not rescue *daed* mutants. Confocal images of Piwi, Atp5a and GFP in ovaries of indicated genotypes. Scale bars, 10µm.

piRNA biogenesis involves the ping-pong amplification loop in nuage and the downstream production of phased piRNAs on mitochondria, catalysed by Zuc (Han et al., 2015; Mohn et al., 2015). Aiming to demonstrate which step of piRNA biogenesis is affected by *daed* loss, I cloned and sequenced small RNAs from ovaries of *daed* and *gasz* mutant flies. The global levels of transposon-mapping piRNA populations was dramatically reduced in *daed* and *gasz* homozygous mutants compared to controls (**Figure 3.7A** and **3.8A**), whilst the 21-nt long siRNAs appeared unaltered, in line with them being processed via an orthogonal pathway (**Figure 1.1**). Closer inspection of piRNA source loci showed that piRNA production was compromised from all types of precursors, namely dual-strand clusters (e.g. *42AB* in **Figure 3.7B** and **3.8B**), uni-strand clusters (e.g. *flam* in **Figure 3.7C** and **3.8C**) and coding genes (e.g. *traffic-jam* [*tj*] in **Figure 3.7D** and **3.8D**). On the other hand, the characteristic 10 nucleotide overlap between piRNAs mapping to opposite genomic strands, known as the ping-pong signature (Brennecke et al., 2007), was unaffected, despite globally diminished levels of piRNA ping-pong pairs (**Figure 3.7E** and **3.8E**). This, together with the unperturbed nuage localisation of Aub and Vasa, demonstrates that Daed is acting downstream of the ping-pong cycle during the production of phased piRNAs on mitochondria.

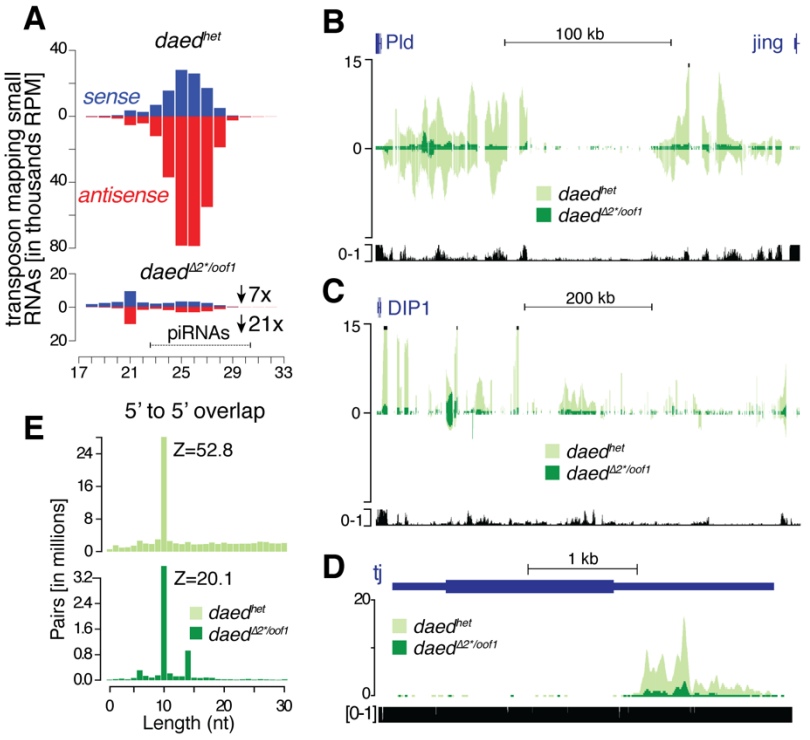


Figure 3.7 ***Daed*** **mutants show a collapse in Zuc-mediated piRNA production. A)** Size distribution of transposon-mapping small RNAs from ovaries. Sense reads are shown in blue, antisense in red. **B-D)** Coverage plots of small RNA reads uniquely mapped to the dual-strand cluster *42AB* **(B)**, the uni-strand cluster *flamenco* (*flam*) **(C)** and the coding gene *traffic-jam* (*tj*) **(D)**. Shown are normalised reads per million (RPM). The mappability for an average 25 bp read length is shown at the bottom. **E)** Ping-pong analysis of transposon-mapping small RNAs from ovaries.

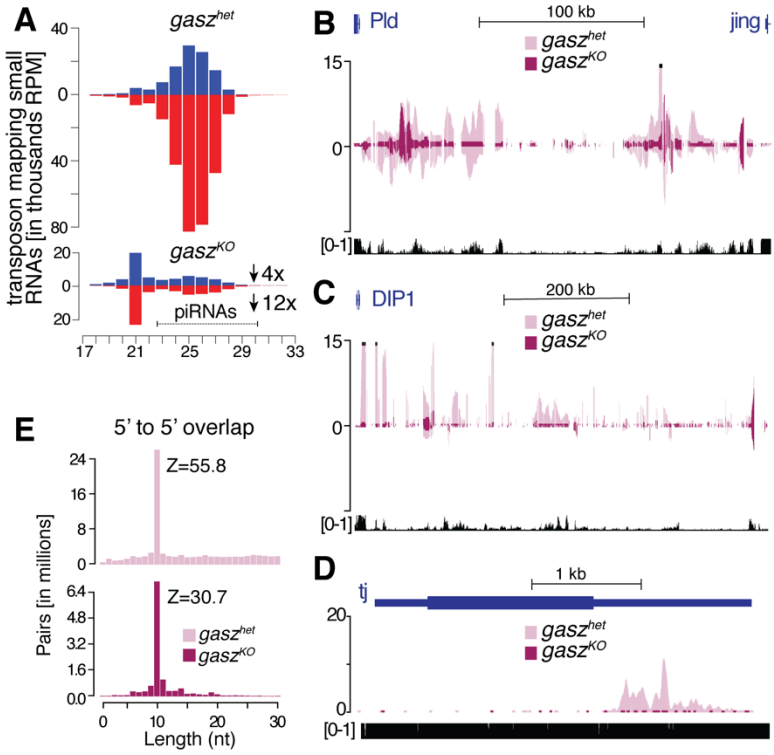


Figure 3.8 ***Gasz* mutants show a collapse in Zuc-mediated piRNA production. A)** Size distribution of transposon-mapping small RNAs from ovaries. Sense reads are shown in blue, antisense in red. **B-D)** Coverage plots of small RNA reads uniquely mapped to the dual-strand cluster *42AB* **(B)**, the uni-strand cluster *flam* **(C)** and the coding gene *tj*. **(D)**. Shown are normalised RPM. The mappability for an average 25 bp read length is shown at the bottom. **E)** Ping-pong analysis of transposon-mapping small RNAs from ovaries.

III.2 Establishment of proximity biotinylation assays

To understand the role of Daed in the production of piRNAs, I decided to exploit a cell culture system derived from the somatic compartment of the fly ovary, namely Ovarian Somatic Cells (OSCs). OSCs harbour a functional somatic piRNA pathway and represent an excellent system to investigate the molecular mechanisms of piRNA biogenesis and co-transcriptional silencing (Niki et al., 2006; Saito et al., 2009). 3xFLAG-Daed expressed in OSCs co-localises with the mitochondrial marker Atp5a, whereas deletion of the TMM domain causes its re-distribution throughout the cell (**Figure 3.9A**). Similarly, 3xFLAG-Daed co-localises with HA-tagged Zuc and Gasz (**Figure 3.9B**). I thus set out to investigate which cellular proteins interact with Daed in OSCs. Daed is a transmembrane protein inserted in the outer envelope of mitochondria, hence identification of its protein partners via canonical IP-MS (Immunoprecipitation and Mass Spectrometry) poses the challenge of successfully extracting Daed from the membrane while

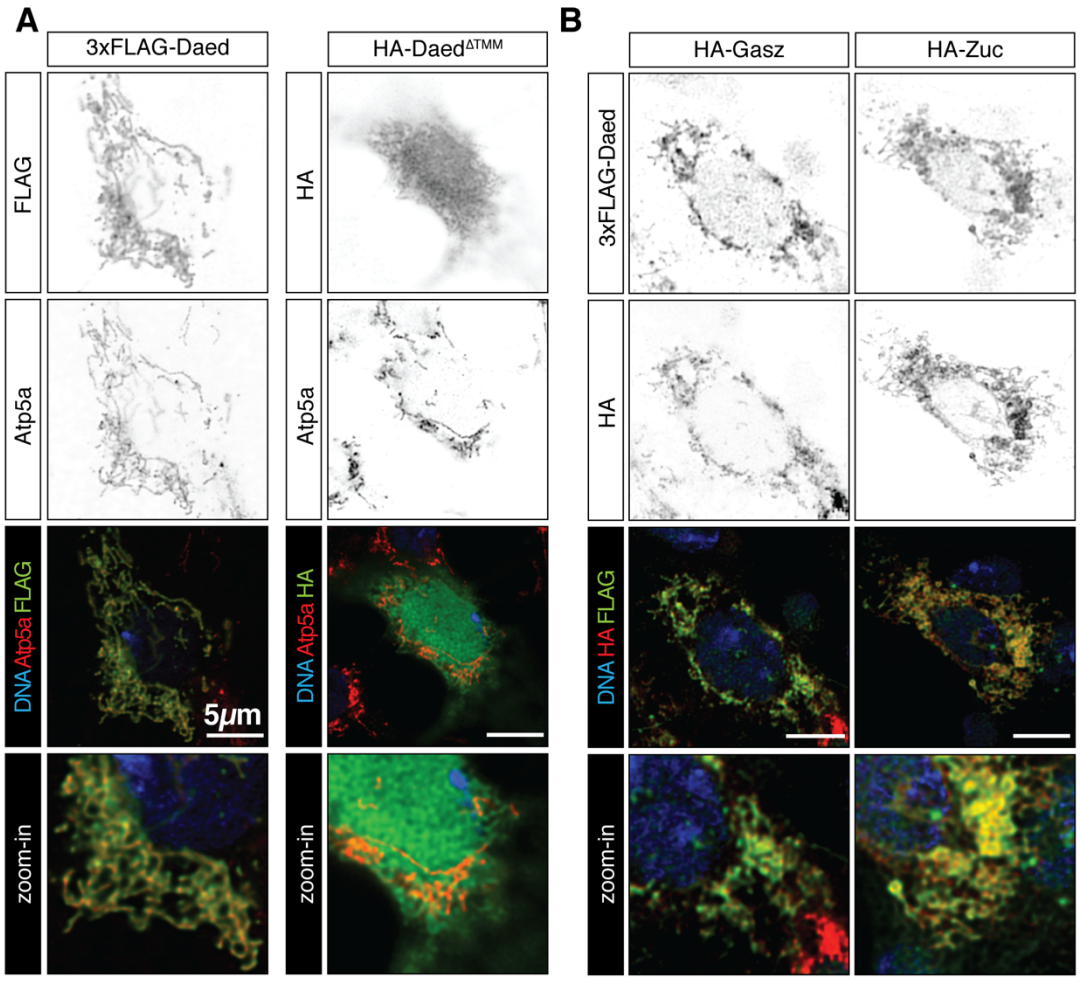


Figure 3.9 Daed localises to mitochondria via its TMM A-B) Confocal images of fusion constructs and the mitochondrial marker Atp5a in OSCs. Scale bar, 5µm.

retaining its endogenous interactors. This is experimentally challenging, as the high detergent conditions necessary to solubilise mitochondrial membranes can disrupt native interactions. Since several piRNA biogenesis factors are transmembrane proteins, we decided to establish an alternative approach that could be broadly applied to study the interaction within the biogenesis machinery.

Proximity biotinylation has been successfully used to study protein-protein interactions in different cellular contexts and has been applied to transmembrane proteins and components of the Nuclear Pore Complex (NPC) (Kim et al., 2014; Kim et al., 2016; Roux et al., 2012). In this assay, originally named BioID, a promiscuous prokaryotic biotin ligase, namely BirA* (or BASU in **Figure 3.10A**) generates activated biotin-AMP intermediates which covalently attach to the free amine groups of accessible lysines in neighbouring proteins (Roux et al., 2012). Fusion of BirA* to a protein of interest (the bait in **Figure 3.10A**) leads to the preferential biotinylation of proteins in close proximity to the bait, including transient interactors. Such proteins can be isolated via pull-down with streptavidin beads, taking advantage of the high affinity between biotin and streptavidin to remove all contaminants (**Figure 3.10A**). The interaction between biotin and streptavidin resists stringent and denaturing lysis conditions and therefore enables solubilisation of membrane proteins without losing native interactors marked with biotin. Isolated proteins are subjected to trypsin digestion and are identified by Mass Spectrometry. Because biotin is a relatively rare modification only present on a few carboxylases, this approach enables selective enrichment and identification of the bait's vicinal proteins. In addition, it also allows the capture of weak or transient interactions that would escape detection with standard IP-MS techniques. Hereafter, I will refer to this method as Proximity Labelling Mass Spectrometry (PL-MS).

Since the first development of BioID, several improved biotin ligases have been described, such as BioID2, TurboID or BASU (Branon et al., 2018; Kim et al., 2016; Ramanathan et al., 2018). However, most of them have been optimised in mammalian cells, which grow at 37°C, as opposed to fly cells which are kept at 26°C. Therefore, I first tested different biotin ligases (BirA*, BioID2 and BASU) for their activity in *Drosophila* cells and found BASU (the biotin ligase from *Bacillus subtilis*) (Ramanathan et al., 2018) to yield robust results at 26°C. I also tried to grow fly cells in the absence of biotin and instead supplementing it during a brief time window to selectively capture a snapshot of interactions. However, neither S2 cells nor OSCs seemed to tolerate biotin-depleted media conditions, therefore all the following BASU PL-MS

experiments have been performed in standard media supplemented with 200µM biotin for 1 hour prior to harvesting.

A representative PL-MS experiment is shown in **Figure 3.10B-D**, using Daed as bait protein. OSCs were transiently transfected with the desired construct and those expressing HA-BASU-Daed showed a specific pattern of biotinylated proteins (detected with streptavidin, green in **Figure 3.10B**) when compared to the control expressing ZsGreen. Biotinylated proteins include the BASU-Daed fusion (yellow in **Figure 3.10B**) as well as endogenously-biotinylated

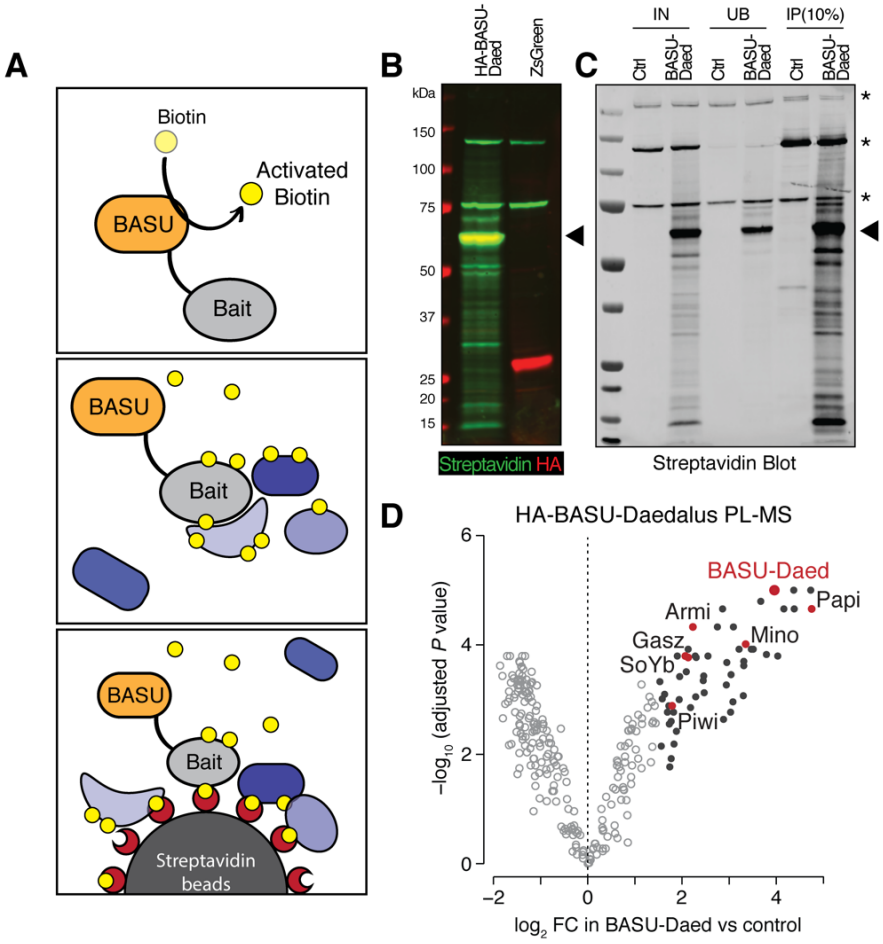


Figure 3.10 Establishment of PL-MS and identification of Daed interactors. A) Cartoon showing the principle of proximity biotinylation experiments. **B)** Western blot showing biotinylated proteins (in green, detected with streptavidin) upon expression of HA-BASU-Daed compared to the HA-ZsGreen control. The arrowhead indicates the size of HA-BASU-Daed fusion. Note that HA-BASU-Daed fusion biotinylates itself (green and red signal overlap). **C)** Western blot showing efficient isolation of biotinylated proteins. Asterisks indicate endogenously biotinylated proteins, the arrowhead indicates the size of HA-BASU-Daed fusion. IN=input; UB=unbound; IP=immunoprecipitate. **D)** Volcano plot showing enrichment and corresponding significance of biotinylated proteins identified via quantitative Mass Spectrometry from OSCs expressing BASU-Daed (n=3) versus control (n=2). Black dots indicate proteins showing a $\log_2$FC > 1.5 and adjusted P value < 0.05 in BASU-Daed. Highlighted in red are piRNA pathway factors. See also supplementary **Table 1.**

carboxylases that are also present in the control sample (asterisks in **Figure 3.10C**). Biotinylated proteins were efficiently isolated via streptavidin pulldown, as shown in **Figure 3.10C**, and subjected to MS (a detailed protocol is provided in **Materials and Methods II.12**). Quantitative MS analysis showed robust enrichment of BASU-Daed itself, consistent with the self-biotinylation of the bait protein, along with that of other proteins known to participate to the piRNA pathway (**Figure 3.10D**). Strikingly, Daed PL-MS enriched for both mitochondrial (Papi, Mino and Gasz) and cytosolic (SoYb, Piwi and Armi) piRNA biogenesis factors. Similar results were also obtained with BASU-Gasz PL-MS, with Armi being one of the most enriched proteins (**Figure 3.11**). These data combined with the delocalisation of Armi in both *gasz* and *daed* mutants led us to hypothesize the existence of a Daed-Gasz transmembrane complex anchoring Armi to the surface of mitochondria. Hence, I proceeded to further dissect the interaction(s) between Daed, Gasz and Armi.

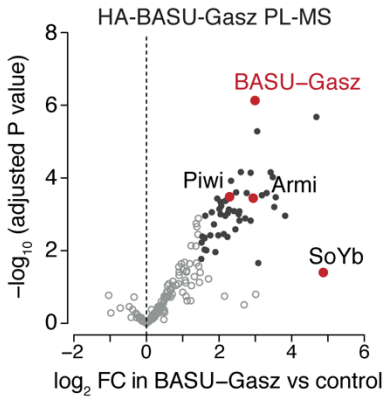


Figure 3.11 Identification of Gasz interactors via PL-MS. Volcano plot showing enrichment and corresponding significance of biotinylated proteins identified via PL-MS from OSCs expressing BASU-Gasz versus control (n=3). Black dots indicate proteins showing a $\log_2$FC > 1.5 and adjusted P value < 0.05 in BASU-Gasz. Highlighted in red are piRNA pathway factors. See also supplementary **Table 2.**

III.3 Investigation of the molecular function of Daed and Gasz domains

Whilst allowing the detection of both stable and transient interactors, proximity labelling data does not necessarily imply a direct association between proteins. Aiming to address whether Daed, Gasz and Armi interact directly or are simply in close spatial proximity, we generated HA- and 3xFLAG-tagged constructs, either full-length or carrying domain deletions, and expressed them in various combination in S2 cells. Since S2 cells lack a functional piRNA pathway, they only inform on direct protein-protein interactions or those mediated by an

ubiquitously expressed adaptor. Armi-HA co-immunoprecipitated with both 3xFLAG-tagged Daed and Gasz, but not the negative control ZsGreen (**Figure 3.12A**). Furthermore, 3xFLAG-tagged Daed co-immunoprecipitates with Daed itself as well as with Gasz, and vice-versa did 3xFLAG-tagged Gasz, suggesting both homo- and heterodimer formation on the mitochondrial surface (**Figure 3.12B**), though we cannot rule out the possibility that other proteins act as bridges in a larger complex. The mitochondrial marker Atp5a was not enriched in the IPs, indicating that this result was not due to isolation of intact mitochondria. To further investigate the existence of homodimers of Daed as well as Gasz, I expressed each factor in S2 cells and performed chemical crosslinking of the lysates to stabilise protein complexes prior to immunoprecipitation. As a positive control, I used a Zuc-3xFLAG construct. Zuc is known to exist as a homodimer *in vivo*, therefore its IP from crosslinked lysates should recover an additional band at approximately double the size of the protein itself. Indeed, **Figure 3.12C** shows the presence of a band corresponding to the Zuc:Zuc homodimer in the Zuc-3xFLAG immunoprecipitate (light green arrowheads in **Figure 3.12C**) but not the ZsGreen control. Similarly, Daed and Gasz IPs also showed a second band corresponding to the size of a homodimer (dark green and purple arrowheads in **Figure 3.12C**). This interaction appears to require their mitochondria localisation, as Gasz and Daed lacking the TMM domain fail to form homodimers while retaining the capability to interact as heterodimers (**Figure 3.12D**).

Next, I aimed to understand which domains of Daed mediate the interaction with Gasz and are required for transposon control. To this purpose, I generated expression constructs of HA-tagged Daed lacking each domain individually (schematic in **Figure 3.13A**) and tested their ability to copurify with 3xFLAG-Gasz in S2 cells. Notably, only Daed lacking the CC domain failed to interact with Gasz (**Figure 3.13B**). Furthermore, I tested whether these same mutants can rescue the transposon up-regulation caused by knockdown of *daed* in OSCs (**Figure 3.13C-D**). In line with the loss of interaction with Gasz, Daed$^{\Delta CC}$ failed to rescue transposon levels. Interestingly, deletion of the SAM domain had the same effect, possibly implicating this domain in the interaction with Armi or with RNA. Daed lacking the TMM domain was instead able to restore transposon expression at baseline levels, thus indicating that this mutant can still perform its function to some extent, probably thanks to the interaction with Gasz (**Figure 3.12D** and **3.13B**). Finally, I also asked which domains of Gasz mediate the interaction with Daed and expressed various HA-tagged Gasz mutants (schematic in **Figure 3.13E**) in combination with 3xFLAG-Daed in S2 cells. However, all domain mutants equally co-purified with Daed (**Figure 3.13F**), thus indicating that this interaction either involves more than one individual domain of Gasz (e.g. both ANK repeats) or one of the other regions.

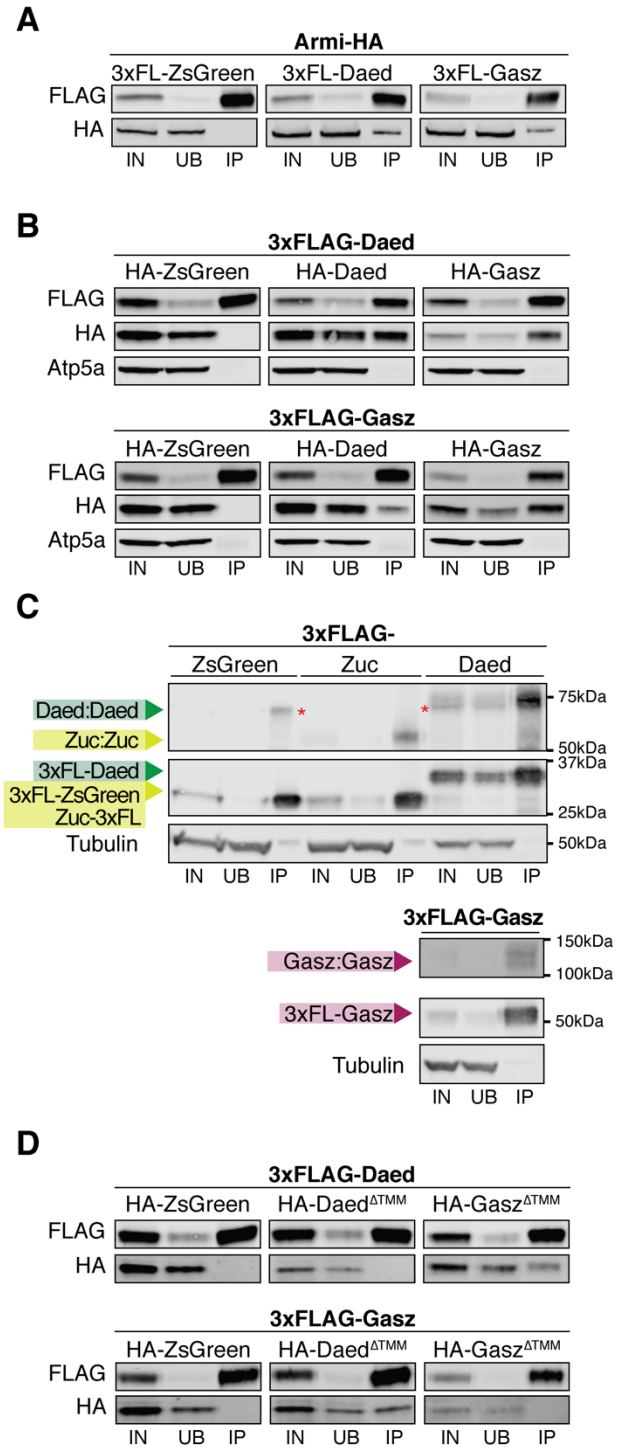


Figure 3.12 Daed and Gasz interact as homo- and heterodimers. A-B, D) Western blots of FLAG-tag co-immunoprecipitation from lysates of S2 cells transfected with the indicated constructs. **C)** Western blot of FLAG-tag co-immunoprecipitates from lysates of chemically crosslinked S2 cells. Red asterisk indicates an unspecific band from anti-FLAG antibody. Coloured arrowheads indicate the size of monomeric and dimeric proteins.

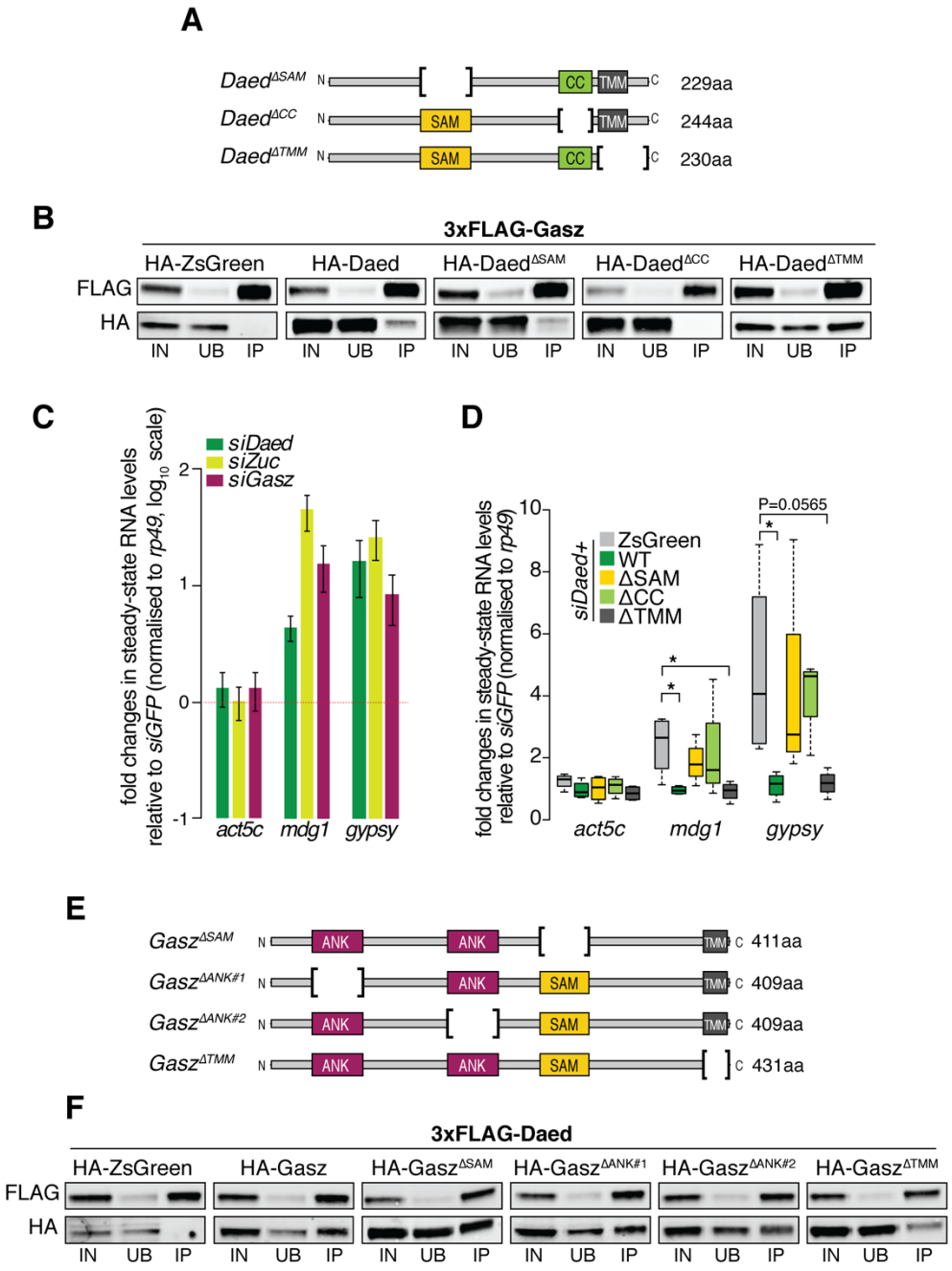


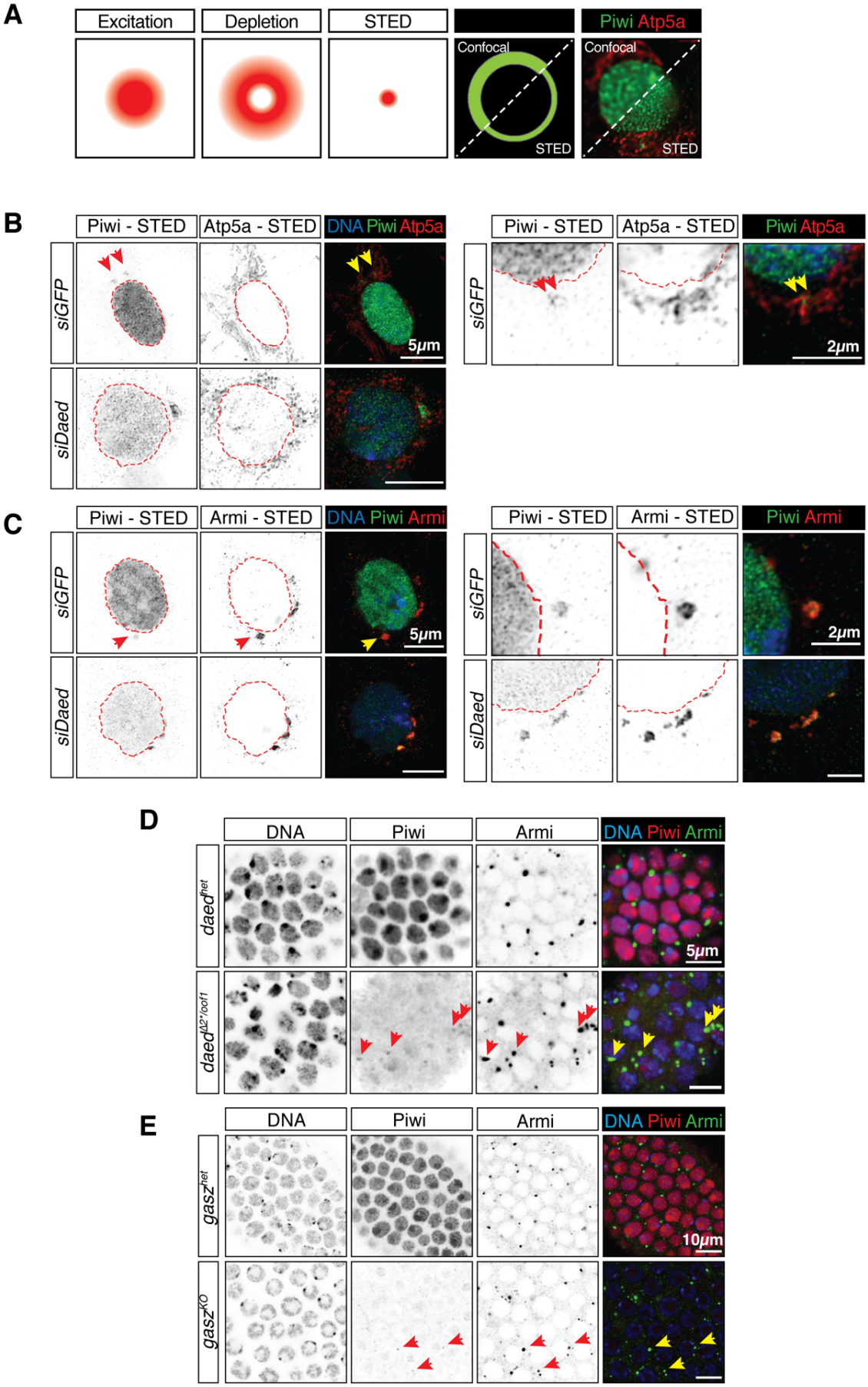
Figure 3.13 Daed CC and SAM domain are required for transposon control. A) Schematic representation of Daed mutant constructs. **B)** Western blots of FLAG-tag co-immunoprecipitations from lysates of S2 cells transfected with the indicated constructs. **C)** Fold changes in steady-state RNA levels of somatic transposons in OSC knockdown cells. Values are relative to *GFP* knockdown and normalised to *rp49*. Error bars indicate standard deviation (n=3). **D)** Fold changes in the steady-state RNA levels of somatic transposons in OSCs nucleofected with siRNAs and various rescue constructs. Values are relative to *GFP* control knockdown and normalised to *rp49*. * = P value < 0.05; ** = P < 0.001 (unpaired t-test) (n=4). **E)** Schematic representation of Gasz mutant constructs. **F)** Western blots of FLAG-tag co-immunoprecipitations from lysates of S2 cells transfected with the indicated constructs.

III.4 Daedalus is essential for recruitment of Armi to mitochondria

Having established that knockdown of *daed* in OSCs leads to transposon up-regulation at levels comparable to that of *zuc* and *gasz* (**Figure 3.13B**), I decided to use this system to investigate its consequences on the subcellular localisation of other piRNA pathway proteins. Light microscopy is typically hampered by the phenomenon of diffraction, which makes it impossible to distinguish objects closer than ~200nm (Huang et al., 2010). In a cellular context, this is approximately the size of a mitochondrion, whereas a single protein is only a few nanometres. This means that anything smaller than 200nm cannot be adequately resolved by conventional light microscopy techniques. Over the last years, several "super-resolution" imaging methods have been devised to overcome the diffraction limit and achieve a better resolution (Huang et al., 2010; Sydor et al., 2015). Among these is Stimulated Emission Depletion (STED) microscopy (Hell and Wichmann, 1994; Vicidomini et al., 2018) (schematic in **Figure 3.14A**). This system employs two laser beams to illuminate the specimen: the excitation laser and a doughnut-shaped depletion laser. The intense depletion laser reverts the excitation of the fluorophores to ground state, leaving only that in the centre of the doughnut. As a result, when the two lasers are combined the emission is depleted everywhere except from the centre of the excitation focus, thus producing sharp, highly-resolved images (**Figure 3.14A**; the last panel compares the same image taken with a standard confocal and with STED). The maximum lateral resolution achievable with STED is ~20nm, thus significantly outperforming standard confocal imaging (Huang et al., 2010; Vicidomini et al., 2018).

Aiming to get a better understanding of the consequences of *daed* depletion in OSCs, I exploited STED microscopy to investigate the localisation of Piwi and Armi with respect to mitochondria (marked with Atp5a). In control cells, treated with *siGFP*, Piwi was often detected in close association with mitochondria (arrows in **Figure 3.14B**, top panel), whereas upon *daed* knockdown most of the remaining Piwi became confined to discrete Yb-bodies, surrounded by morphologically altered mitochondria (**Figure 3.14B**, bottom panel).

Figure 3.14 In the absence of *daed*, Armi is retained in Yb bodies. A) Cartoon showing the principle of Stimulated Emission Depletion (STED) microscopy. The last panel shows the same picture taken with confocal and STED. **B-C)** STED microscopy of Piwi and Atp5a **(B)** or Piwi and Armi **(C)** in OSCs from indicated knockdowns. Scale bar, 5µm or 2µm in the zoom-ins (right panels). **D-E)** Confocal images of Piwi and Armi in follicle cells of indicated genotypes. Arrows point at sites of Piwi and Armi co-localization. Scale bar, 10µm.


A
Excitation
Depletion
STED
Piwi Atp5a
Confocal
STED
Confocal
STED
B
Piwi - STED
Atp5a - STED
DNA Piwi Atp5a
siGFP
siDaed
5µm
Piwi - STED
Atp5a - STED
Piwi Atp5a
siGFP
2µm
C
Piwi - STED
Armi - STED
DNA Piwi Armi
siGFP
siDaed
5µm
Piwi - STED
Armi - STED
Piwi Armi
siGFP
siDaed
2µm
D
DNA
Piwi
Armi
DNA Piwi Armi
daedhet
daedΔ2*1oof1
5µm
E
DNA
Piwi
Armi
DNA Piwi Armi
gaszhet
gaszKO
10µm

Co-staining of Piwi and Armi showed that, when outside of the nucleus, Piwi was generally observed in close association with Armi, and this colocalization was enhanced upon loss of *daed*, when most of the remaining Piwi was trapped into Yb bodies (**Figure 3.14C**). The same was observed in the follicle cells of *daed* and *gasz* homozygous mutant flies, where residual Piwi signal overlapped with Armi in the Yb bodies (arrows in **Figure 3.14D-E**). Taken together, these data suggest a model in which Piwi moves together with Armi onto the mitochondrial surface, where the latter is held by Daed and Gasz. In the absence of either *daed* or *gasz*, Piwi is instead unable to reach the processing centres and remains in Yb bodies.

III.5 Armi shuttles from Yb-bodies to mitochondria where it associates with dimeric Zuc

Aiming to gain a better understanding of the protein-protein interactions occurring among piRNA biogenesis factors, I also carried out PL-MS experiments for Armi-BASU (**Figure 3.15A-B**) and Zuc-BASU (**Figure 3.15C-D**). Armi PL-MS led to enrichment of a large number of proteins, in line with it being cytosolic and coming close to more proteins than a mitochondrial bait. Among the piRNA pathway factors significantly enriched were Piwi and Yb, which have been previously described to interact with Armi (Olivieri et al., 2010; Saito et al., 2009), thus validating the specificity of my PL-MS approach. Armi PL-MS also identified some mitochondrial factors, in particular Daed and Gasz, thus confirming previous data, but also Mino and Papi. Conversely, Zuc PL-MS enriched for all known mitochondrial piRNA pathway factors (Daed, Gasz, Mino and Papi) and only two cytosolic ones (Armi and SoYb), among which Armi showed the strongest enrichment. Notably, Yb was not significantly enriched, indicating that it does not leave Yb bodies to relocate on mitochondria, in contrast with Armi (and potentially SoYb).

Zuc is a PLD-family ribonuclease that can cleave its substrates with a remarkable processivity and specificity (Ipsaro et al., 2012; Nishimasu et al., 2012). Strikingly, *in vitro* assays with recombinant Zuc fail to recapitulate the cleavage pattern observed *in vivo*, thus hinting to the presence of co-factor(s) that assist Zuc during piRNA precursor processing. I therefore aimed to pin-point whether any of the proteins found in Zuc PL-MS was more closely associated to Zuc than others and decided to exploit a variant of proximity biotinylation, namely Split BioID (Schopp et al., 2017) (cartoon in **Figure 3.16A**). In this assay, the bacterial biotin ligase BirA*

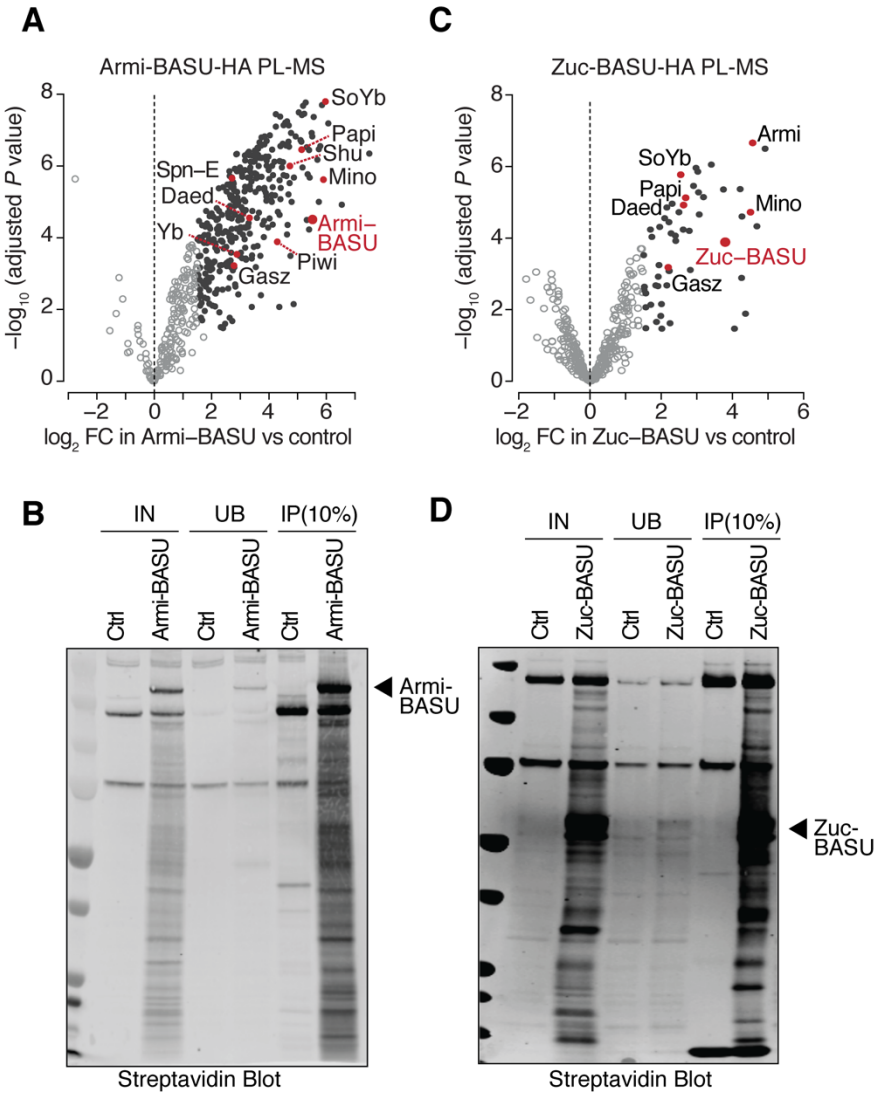


Figure 3.15 Identification of Armi and Zuc interactors via PL-MS. A, C) Volcano plots showing enrichment and corresponding significance of biotinylated proteins identified via PL-MS from OSCs expressing the indicated constructs against control (n=3). Black dots indicate proteins showing a $\log_2$FC > 1.5 and adjusted P value < 0.05. Highlighted in red are piRNA pathway factors. See also supplementary **Tables 3-4**. **B, D)** Western blots showing pulldown of biotinylated proteins from OSCs transfected with the indicated constructs. Arrowheads indicate the size of the fusion protein.

is split into two halves, N-BirA* and C-BirA*, each inactive on their own. N-BirA* and C-BirA* reconstitute the active enzyme only if fused to two proteins that interact with each other *in vivo*. The reconstituted BirA* can then biotinylate vicinal proteins upon biotin supplementation, albeit to a lower extent than the full-length enzyme. Structural studies indicate that Zuc cleaves RNA as a dimer (Ipsaro et al., 2012; Nishimasu et al., 2012),

thus we decided to exploit this feature to refine further our proteomics analysis and attempt to pinpoint which piRNA pathway factor is more closely associated with the "cleavage-competent" Zuc dimer. Expression of Zuc fused to N- and C-BirA* in OSCs led to protein biotinylation only when the two constructs were present together and biotin was supplemented in the media (**Figure 3.16B**). This strategy generated a much smaller number of biotinylated proteins compared to BASU (compare **3.15D** and **3.16B**), in line with a lower activity of the split enzyme. Strikingly, Zuc Split-BioID PL-MS revealed enrichment of a limited set of proteins, yet still readily identified Armi (**Figure 3.16C**). This could indicate that Armi has a closer association with Zuc than, for instance, SoYb, which was only identified by BASU PL-MS. In this scenario Armi, once translocated to mitochondria, could be proximal to the site where piRNA precursors are being processed by Zuc and could possibly contribute to its distinctive cleavage pattern. Interestingly, *zuc* and *armi* knockdowns showed a higher correlation in the levels of genome-mapped small RNAs (r^2=0.782; **Figure 3.16D**) than *zuc* and *yb* knockdowns (r^2=0.406; **Figure 3.16E**), further supporting a role for Armi as a putative Zuc co-factor.

Based upon this, we envisioned a model in which a Piwi-Armi-pre-piRNA complex is released from the sites of precursor specification (nuage in germ cells and Yb-body in follicle cells/OSCs) to translocate on mitochondria. There, Armi is held in place by Gasz/Daed and engages in piRNA production in close association with Zuc. Consistent with this model, in both fly germline and OSCs, loss of *zuc* caused a dramatic accumulation of Piwi and Armi on mitochondria, whereas loss of *gasz* and *daed* led to their dispersal in the cytosol or concentration in Yb-bodies (**Figure 3.17A-B**). Based on my proximity labelling results, this behaviour appears to be a unique feature of Armi and not, for instance, Yb, which also plays a critical role in somatic piRNA production but remains confined to Yb bodies. We therefore suggest that Armi shuttles from nuage/Yb-bodies to mitochondria and hypothesize that it is involved in the presentation of piRNA precursors to Zuc for downstream processing into phased piRNAs.

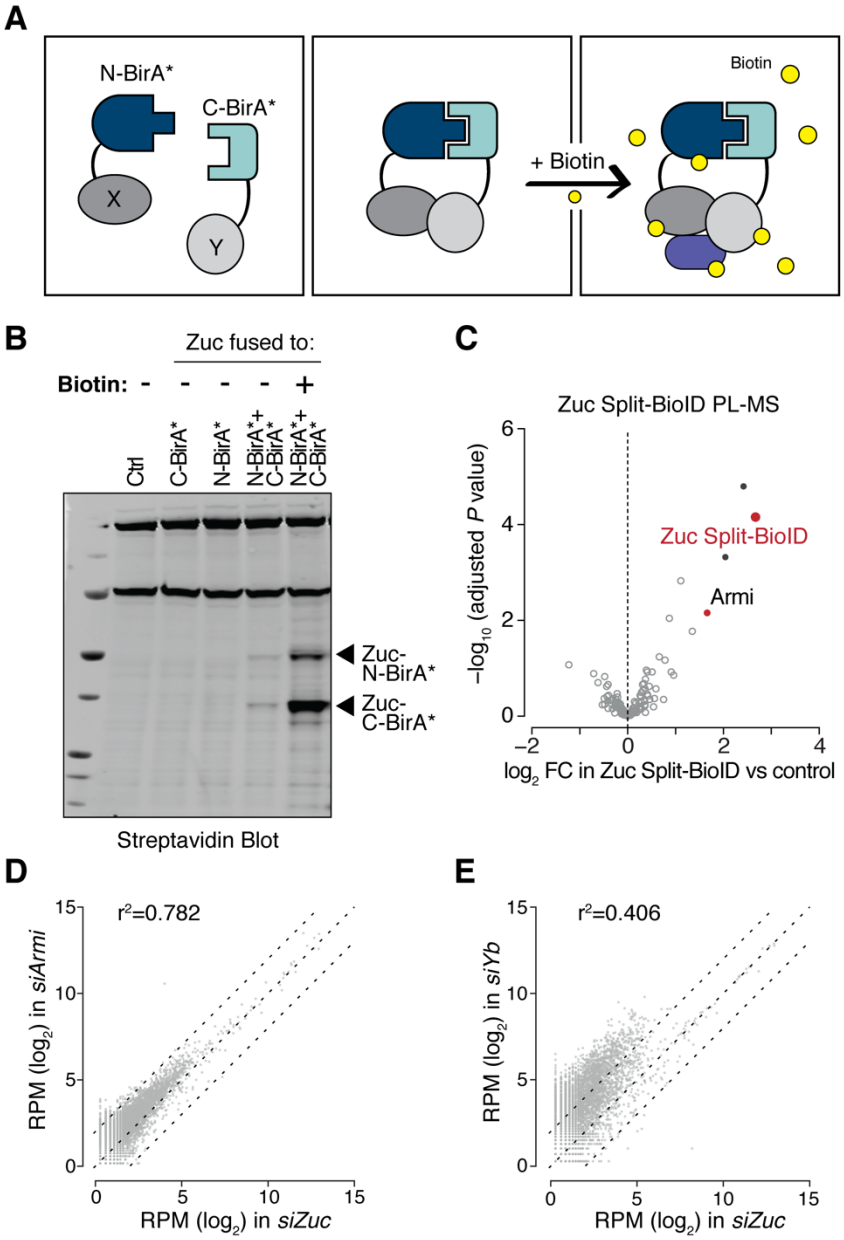


Figure 3.16 **Armi closely associates with Zuc. A)** Cartoon showing the principle of the Split BioID proximity biotinylation assay. **B)** Western blot showing pulldown of biotinylated proteins from OSCs transfected with the Zuc Split-BioID constructs. Arrowheads indicate the size of the fusion protein. **C)** Volcano plots showing enrichment and corresponding significance of biotinylated proteins identified via PL-MS from OSCs expressing the indicated constructs against control (n=3). Black dots indicate proteins showing a $\log_2$FC > 1.5 and adjusted P value < 0.05. Highlighted in red are piRNA pathway factors. See also supplementary **Table 5. D, E)** Scatter plots showing expression levels (normalised RPM) of genome-mapped small RNAs in OSCs upon indicated knockdown.

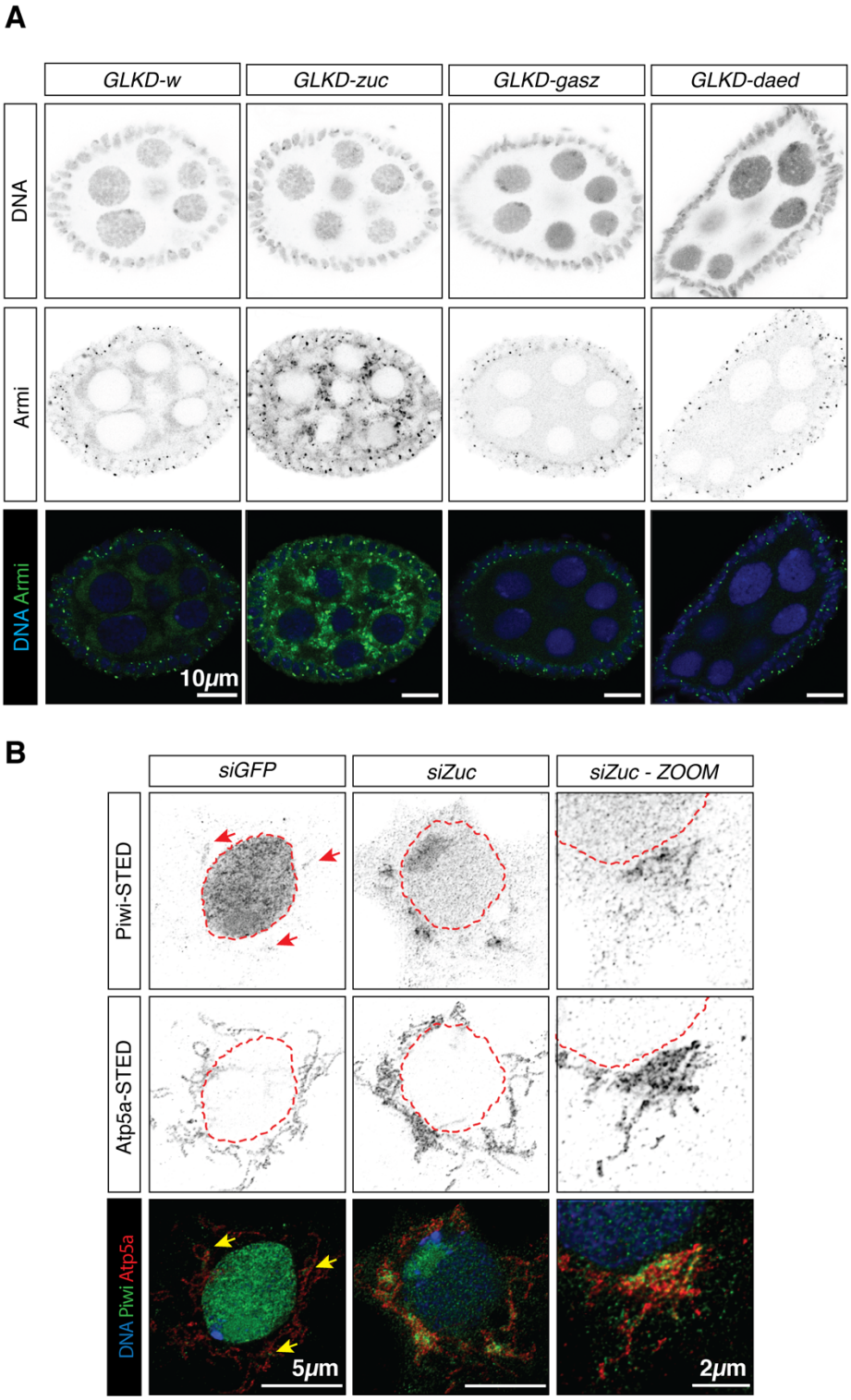


Figure 3.17 Armi localisation to mitochondria close to Zuc depends on Gasz and Daed. A) Confocal images of Armi in ovaries from indicated GLKD. Scale bar, 10µm. **B)** STED microscopy of Piwi and Atp5a in OSCs. Scale bar, 5µm.

III.6 Establishment of CLIP-seq

A broadly used method to study the interactions between protein and RNA is Cross-Linking Immuno-Precipitation, followed by sequencing (CLIP-seq) (Lee and Ule, 2018). Although several variations of CLIP exist, each with their own specific features, the core concept is shared amongst all. The method is based on treating cells or tissue with UV light which induces the formation of covalent bonds between protein and RNA. Unlike chemical crosslinking, UV-C light only crosslinks protein and RNAs that are in direct contact but does not crosslink proteins to each other. The RNA-binding protein (RBP) of interest is then immunoprecipitated in stringent conditions, to avoid co-purification of other RBPs, and the crosslinked RNAs are isolated and identified via high-throughput sequencing. One of the first methods to combine CLIP with sequencing, namely iCLIP, developed a specific library preparation procedure to identify crosslink sites at single nucleotide resolution, therefore unveiling sequence motifs preferentially bound by a certain protein (Huppertz et al., 2014). Other methods omit some steps or provide variations that increase the overall efficiency and/or the convenience of the protocol. Regardless of the protocol of choice, there are two key requirements for a successful CLIP-seq experiment: a clean immunoprecipitation of the RBP of interest and a library preparation procedure that either overcomes or takes advantage of the crosslink sites to achieve single-nucleotide resolution.

Armi belongs to the family of Upf1-like RNA helicases and its mouse homolog (MOV10L1) has been shown to bind to piRNA precursors (Vourekas et al., 2015). I thus sought to determine whether *Drosophila* Armi performs a similar function and, to do so, I set out to establish a CLIP-seq protocol in the lab. Due to the requirement of large quantities of antibody against endogenous Armi, which is not commercially available, I decided to carry out the immunoprecipitation against a tag and tested various options to find the most suitable one. I first tried 3xFLAG, a synthetic peptide of 22 amino acids which is recognised by the corresponding antibody pre-immobilised on magnetic agarose beads. 3xFLAG-tagged Armi could be robustly immunoprecipitated, however, I also detected a large amount of non-specific RNA bound to the beads, which would negatively affect the signal-to-noise ratio of a CLIP-seq experiment. Aiming to reduce the background, I tested the StrepTagII system which has been developed based on the biotin-streptavidin interaction. It employs an 8 amino acid peptide (the StrepTagII) which binds to StrepTactin beads with a nearly covalent affinity. Despite showing a much lower background RNA binding, immunoprecipitation with this tag gave scarcely reproducible results. This was likely due to the high levels of free biotin present in fly cells,

which can saturate the beads and interfere with the capture of the StrepTagII fusion protein. Finally, I tested the HALO tag, which has been derived from a bacterial haloalkane dehalogenase and can covalently bind to synthetic ligands (Encell et al., 2012). The HALO ligands can be conjugated to magnetic beads or functionalised with fluorophores or other chemical groups, thus providing a versatile tool for many applications (England et al., 2015). This system provides two substantial advantages. First, the irreversible attachment between the HALO tag and the ligand resists any type of denaturing condition, thus avoiding co-purification of undesired contaminants (Gu et al., 2018). Secondly, the purified bait protein can be specifically released in solution by cleaving off the HALO tag, which instead remains covalently bound to the beads. This has the advantage of leaving behind any RNA unspecifically purified because it crosslinked to the tag rather than the POI. To do this, a TEV (Tobacco Etch Virus) protease cleavage site can be placed between the tag and the bait protein. An overview of the resulting CLIP-seq workflow is shown in **Figure 3.18A-D** and a detailed protocol is provided in **Materials and Methods II.10**. The expression constructs used for CLIP-seq carry a fusion between a 3xFLAG-tagged bait (e.g. Armi) and the HALO tag, linked by three copies of the TEV protease cleavage site (cartoon in **Figure 3.18E**). In this manner, cleavage by TEV protease releases the 3xFLAG-tagged bait which can be detected on a western blot using anti-FLAG antibody (**Figure 3.18F**). OSCs nucleofected with an Armi-3xFLAG-3xTEV-HALO fusion protein are first crosslinked under UV-C light (**Figure 3.18A**) and the RBP complexes purified with a HALO resin. Following TEV protease-mediated release of the bait protein (**Figure 3.18B**) and proteinase K digestion to only leave the RNA with a small peptide at the crosslinking sites (**Figure 3.18C**), the samples are subjected to library preparation for high-throughput sequencing. piRNA precursors typically lack readily identifiable sequence motifs, therefore Armi is not expected to have a preferential binding pattern to certain sequences. This means that Armi CLIP-seq does not necessarily require identification of the crosslinking sites at single-nucleotide resolution. Hence, I decided to use the SMARTer (Switching Mechanism At 5' end of PCR Template) technology to prepare libraries from my crosslinked and released RNA samples (**Figure 3.18D**). This is based on a template-switching mechanism during the reverse transcription step, which enables simultaneous addition of the 3' and 5' sequencing adaptors. When the Reverse Transcriptase (RT) reaches the 5' end of the template molecule, the enzyme adds a few additional nucleotides to the 3' end of the cDNA. The SMARTer Stranded oligo can then base-pair with the non-templated nucleotide stretch, creating an extended template that enables the RT to continue

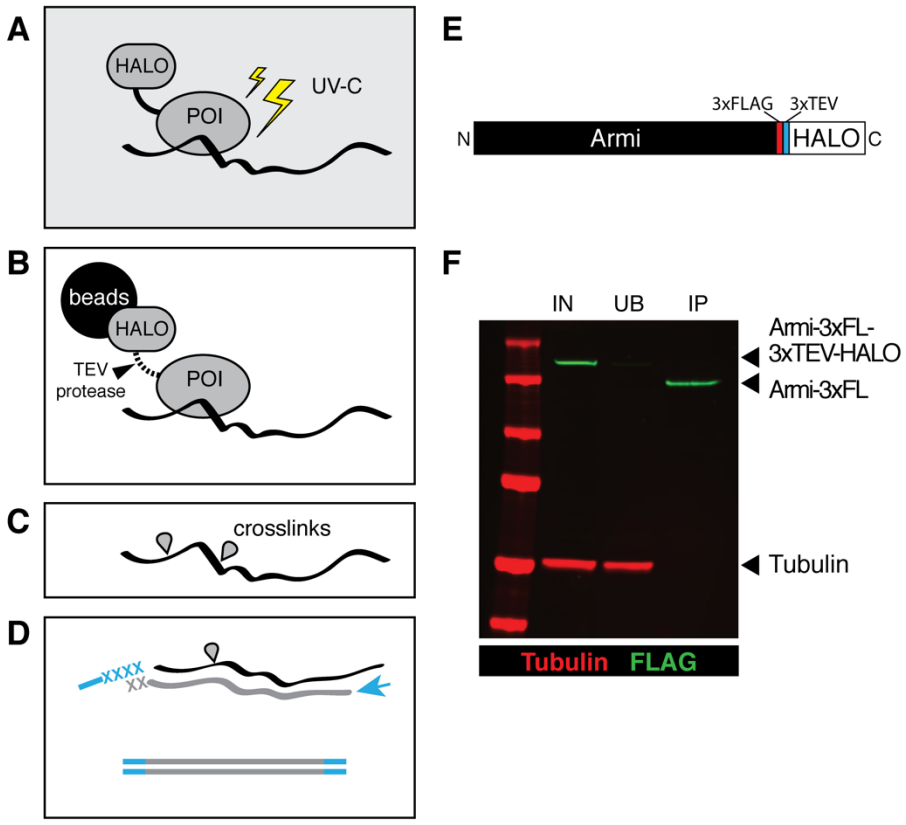


Figure 3.18 Establishment of CLIP-seq. A-D) Cartoon showing the principle of CLIP-seq experiments. **(A)** Cells expressing a HALO-tagged protein of interest (POI) are irradiated with UV-C light, inducing the formation of protein-RNA crosslinks. **(B)** Crosslinked protein-RNA complexes are isolated with a resin conjugated with the HALO ligand and the POI is released in solution via TEV cleavage. **(C)** Following Proteinase K digestion which leaves a small peptide at the crosslink sites, **(D)** library preparation is carried out using a template-switching mechanism. **E)** Schematic representation of the construct used for Armi CLIP-seq. The same construct has been used for CLIP-seq described in **Chapter IV** and **V**, with an inverted arrangement of the tag in case of an amino-terminal tagged POI. **F)** Representative western blot showing pulldown and elution of the Armi-HALO fusion. The size difference between IN and IP results from TEV cleavage of the HALO tag.

replicating until the end of the oligo (Chenchik, 1998). The obtained cDNA can then be amplified via PCR with Illumina primers. The SMARTer workflow is advantageous because it is well optimised for low input samples and because the template-switching mechanism is likely to overcome the obstacle of crosslinked peptides, which has been shown to stop some RTs. Overall, this HALO tag-based CLIP strategy significantly outperformed FLAG and StrepTagII, both in terms of low non-specific binding and efficient capture of the bait protein (a representative western blot is shown in **Figure 3.18F**). Hence, all CLIP-seq experiments described here and in **Chapter IV** and **V** have been carried out following this procedure.

When compared to a mock IP from cells not expressing a HALO fusion protein, Armi CLIP-seq from OSCs revealed strong enrichment of all types of somatic piRNA precursors, the uni-strand clusters *20A* and *flam* (red in **Figure 3.19A**) and coding genes known to give rise to piRNAs, such as *tj* (blue in **Figure 3.19A**) (Robine et al., 2009; Saito et al., 2009). Armi CLIP-seq signal is distributed along the entire length of the *flam* locus, which spans several hundreds of kbs (**Figure 3.19B**). Strikingly, Armi CLIP-seq signal over *tj* and other genic piRNA source loci is instead preferentially enriched on their 3'UTR, which is the region used for piRNA production (**Figure 3.19C**). This might suggest that the 3'UTR of genic piRNA source transcripts is cleaved off via an unknown mechanism and preferentially used for piRNA production (Li et al., 2003; Saito et al., 2009). In all cases, Armi CLIP-seq enriches for transcripts which give rise to piRNAs that are lost upon Armi knockdown (**Figure 3.19B-C**), thus suggesting that Armi binds to transcripts that are funnelled into Zuc-mediated, phased cleavages and possibly contributes to their specification as piRNA precursors. Inspection of transposon-mapping reads present in Armi CLIP-seq showed a preferential enrichment of antisense transposon sequences but not sense (**Figure 3.19D-E**), in line with the latter being generally repressed by the piRNA pathway. More specifically, the Armi CLIP-seq enrichment of each transposon family correlates with their copy number within *flam* (red in **Figure 3.19D**, with dot size proportional to their abundance within *flam*), thus indicating that this signal originates from Armi binding to this somatic piRNA cluster. No 1U bias was detected in Armi CLIP-seq, but this is likely a result of the library preparation procedure which involves a fragmentation step prior to RT and therefore does not retain information on the 5' end of the crosslinked RNAs. Additionally, this resembles what was reported for MOV10L1 CLIP-seq, where the 5'ends of the crosslinked RNAs did not show a 1U bias and were sensitive to nuclease treatment (Vourekas et al., 2015).

III.6 Armi binding to piRNA precursors depends on Piwi

The current model of the piRNA production process suggests that Piwi binds to the 5' end of a precursor transcript and directs Zuc to cleave at the first accessible uridine (Gainetdinov et al.). This cleavage event simultaneously specifies the 3' and the 5' end of two adjacent piRNAs, leading to the characteristic phasing signature (Han et al., 2015; Mohn et al., 2015). Based on the data discussed so far, Armi appears to participate to this process by bringing piRNA precursors to the mitochondria, where it localises in close vicinity to Zuc. To understand precisely when Armi RNA binding occurs and whether this process depends on other factors,

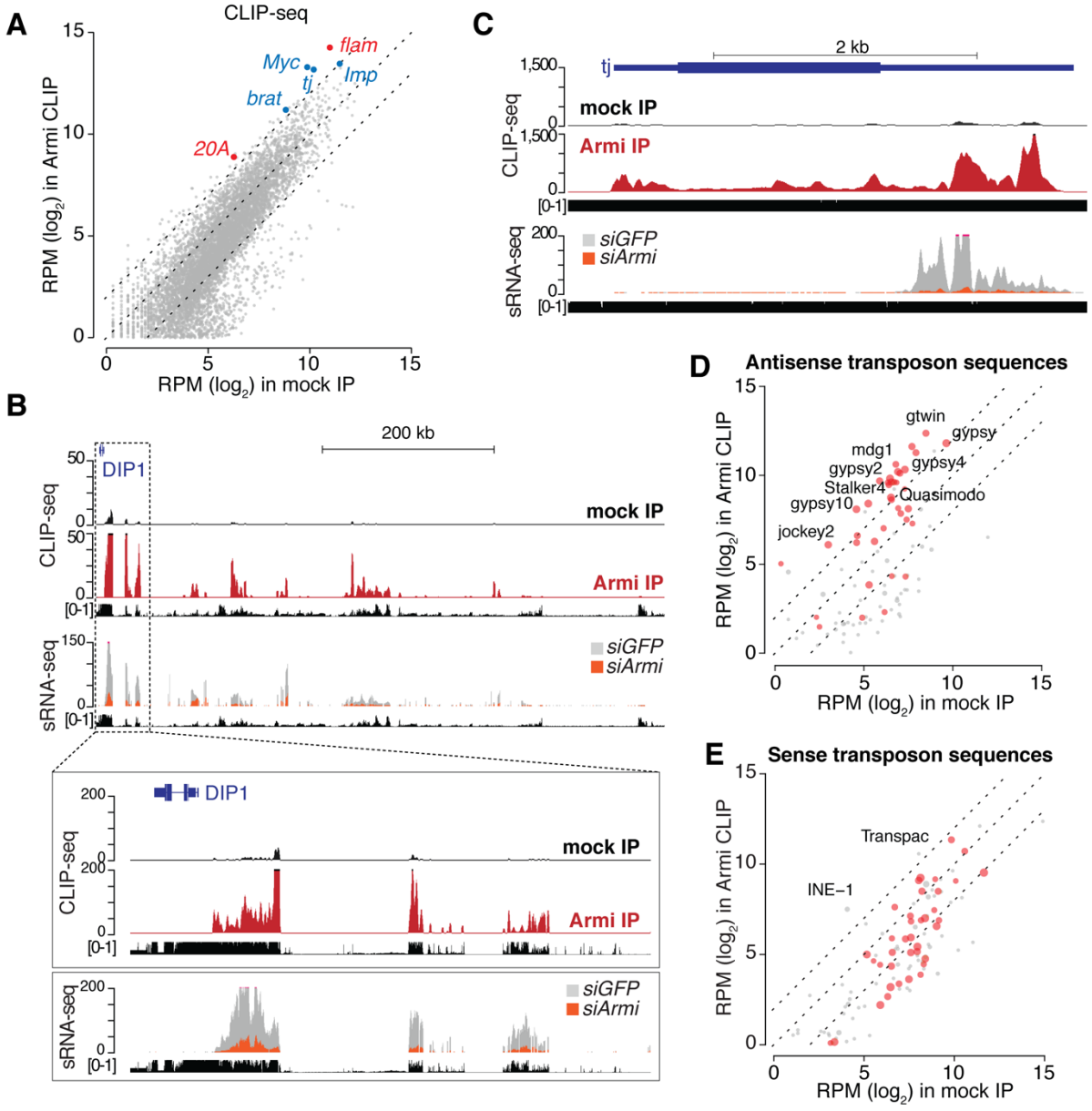


Figure 3.19 Armi binds to piRNA precursors. (A) Scatter plot showing expression levels (RPM) of genes in Armi CLIP-seq (n=4) against a mock IP (n=3). piRNA clusters expressed in OSCs are highlighted in red, selected protein coding genes producing piRNAs in blue. **(B)** UCSC genome browser shot displaying Armi CLIP-seq and small RNA-seq reads uniquely mapping to the piRNA cluster *flam* (upper panel) and a zoomed-in view of the first ~50 kb (bottom panel). Shown are normalised RPM. The mappability tracks for 50 bp and 25 bp read length, respectively, are shown below. **(C)** Same as in **(B)** but showing the protein-coding gene *tj*. **(D-E)** Scatter plots showing expression levels (RPM) of antisense and sense transposon sequences in Armi CLIP-seq against a mock IP. Transposon sequences present in *flam* are highlighted in red with dot size proportional to their abundance within *flam* according to dm6 Repeat Masker annotations.

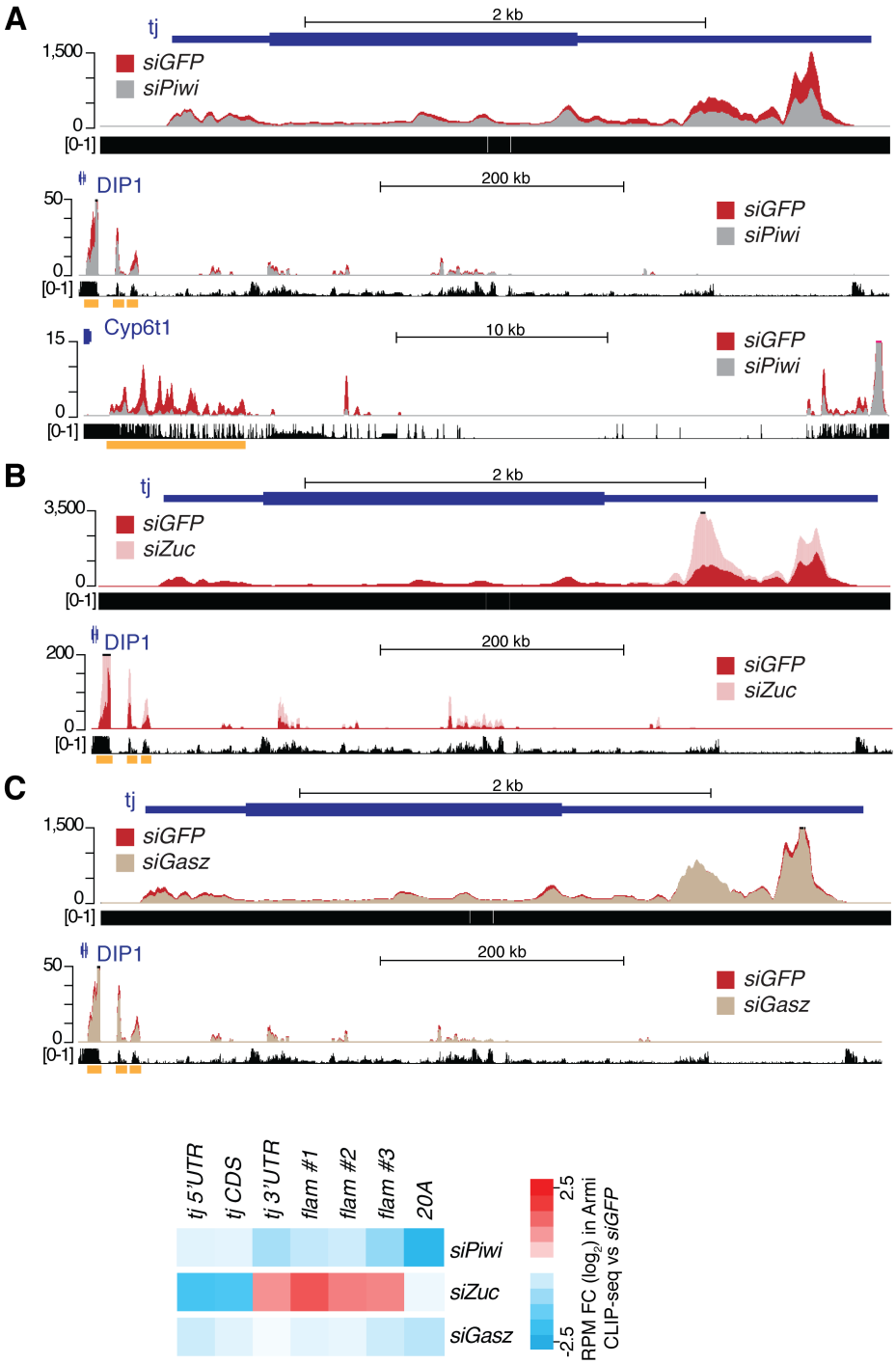


Figure 3.20 Armi binding to piRNA precursors is impaired upon *piwi* knockdown. (A-C) Genome browser shot displaying Armi CLIP-seq profiles from OSCs upon indicated knockdown. Shown are normalised RPM. (n=3 for *siZuc* and relative *siGFP* control; n=4 for *siPiwi* and relative *siGFP* control; n=3 for *siGasz* and relative *siGFP* control). Below each profile are the mappability tracks for 50 bp read length. **(D)** RPM Log_2FC in Armi CLIPseq upon indicated knockdown against their relative *siGFP* for selected regions of *tj, flam* and *20A* (orange boxes in panels **A-C**).

I performed Armi CLIP-seq from cells depleted of *piwi*, *zuc* and *gasz* and examined the changes in the interaction with piRNA precursors.

Loss of *piwi* reduced, but did not completely abolish, Armi binding to *tj, flam* and *20A* (**Figure 3.20A** and quantification in **3.20D**), indicating that Piwi recognition of the 5'ends of precursor transcripts might precede and facilitate the association with Armi. Despite strong reduction in Piwi protein levels, residual Piwi might still perform its function and therefore explain why Armi CLIP-seq signal is not entirely lost (**Figure 3.21**). Depletion of *zuc* showed the opposite effect, with substantially increased Armi binding to piRNA precursors (**Figure 3.20B** and quantification in **3.20D**). This is in line with a model in which, in the absence of Zuc, Armi accumulates on mitochondria bound to RNAs that cannot be processed. Interestingly, the increase of Armi footprint on *tj* mRNA was not even, but instead restricted to the 3'UTR. Finally, *gasz* knockdown did not impact the extent of Armi RNA binding activity, as shown by the unchanged CLIP-seq signal (**Figure 3.20C** and quantification in **3.20D**). Taken together, these data and those available in the literature (Yamashiro et al., 2019) support a model whereby stable Armi binding to piRNA precursors also requires prior recognition of their 5'end by Piwi, thereby assembling a trimeric complex of Armi-Piwi-pre-piRNA that is competent for exiting the Yb bodies and translocating to mitochondria, where Zuc-mediated piRNA processing will take place.

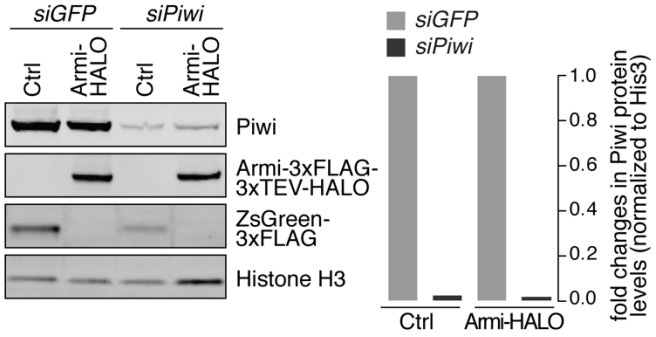


Figure 3.21 Validation of *piwi* knockdown for CLIP-seq. Western blot with relative quantification showing expression levels of Piwi in *siGFP* and *siPiwi* samples used for CLIP-seq in **Figure 3.20**.

Aiming to dissect the dependency of the Armi-Piwi-pre-piRNA interaction on other factors, I immunoprecipitated 3xFLAG-tagged Armi from OSCs depleted of various piRNA biogenesis factors and probed for the presence of endogenous Piwi (**Figure 3.22A** and quantification in **3.22B**). The same experiment was also carried out upon lysate treatment with RNase I, to address the role of RNA in the formation of this complex. In a control knockdown (*siGFP*), a low but reproducible coimmunoprecipitation of Piwi with Armi-3xFLAG was detected and it was insensitive to RNase treatment. This is in line with the fact that most Piwi is nuclear and only a small fraction is engaged in the biogenesis process. Depletion of *Zuc* increased the fraction of Piwi copurified with Armi-3xFLAG and this interaction was independent from RNA, indicating that, once stably anchored on mitochondria, the association between Armi and Piwi does not require RNA any more. Finally, Piwi and Armi still interacted in the absence of either *gasz* or *daed* but this was abolished by RNase A treatment, suggesting that, while in the Yb bodies, Armi and Piwi are only held together via RNA, until they reach mitochondria.

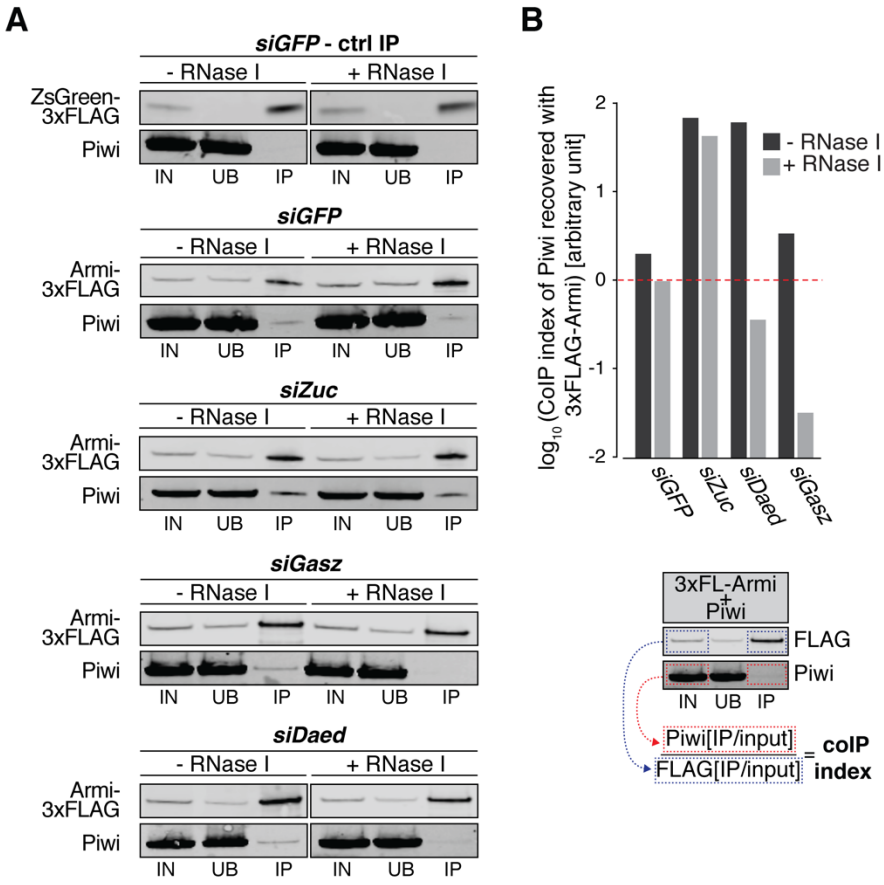


Figure 3.22 Dependencies of the Armi-Piwi complex. A) Western blots of ZsGreen- or Armi-3xFLAG co-immunoprecipitation from lysates of OSCs upon indicated knockdown and with/without RNase treatment. **B)** CoIP index showing enrichment of Piwi in Armi-3xFLAG pulldown upon indicated knockdowns and with/without RNase treatment.

CONCLUSION

Work described in this chapter adds a novel player to the network of proteins involved in piRNA production: the mitochondrial protein CG10880/Daedalus (Daed), which is unique to *Drosophilids* (**Figure 2.24**). We find that Daed is essential for Zuc-mediated, phased piRNA production and works in close association with Gasz (Handler et al., 2013). Together, Daed and Gasz constitute an "anchoring platform" on the mitochondrial surface that recruits and holds the RNA helicase Armi in close proximity to Zuc (model in **Figure 3.23**). We find that the recruitment of Armi to mitochondria is a crucial event, as Armi can bind to piRNA precursors in the nuage/Yb bodies and then deliver them to the processing centres. We dissect the molecular events underlying Armi binding to piRNA precursors and demonstrate that this occurs prior to its translocation to mitochondria and requires Piwi. Our data suggest that an Armi-Piwi-pre-piRNA complex is assembled in Yb bodies, licensed and then anchored on the outer mitochondrial membrane via Gasz and Daed during the cycles of Zuc cleavages (model in **Figure 3.23**). Taken together, these findings provide an advancement to our understanding of the molecular mechanisms underlying piRNA production.

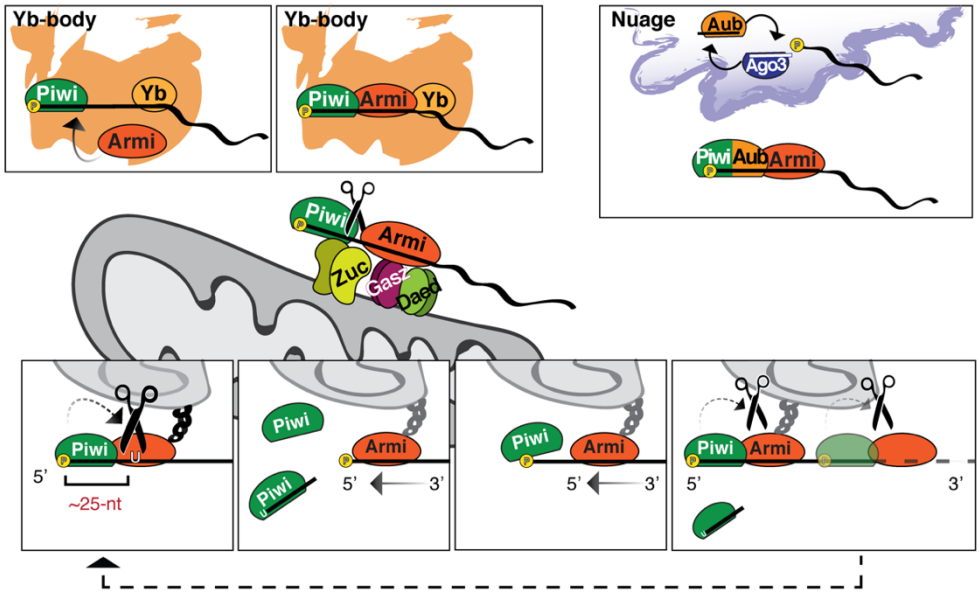


Figure 3.23 **A model for piRNA biogenesis.** In Yb-bodies, Piwi binds to the 5' phosphate of a piRNA precursor, and this in turn recruits Armi [top]. Piwi, Armi, and the pre-piRNA translocate to mitochondria where a Daed/Gasz complex stabilizes Armi while Zuc dimers cleave the transcript [middle]. Phased piRNA production then requires cycles of Piwi binding positioning Zuc to cleave at the first available uridine. While held in place by Daed and Gasz, Armi translocates along the transcript to provide the next segment of the RNA for piRNA biogenesis, and again Piwi binding to the newly-generated 5'-P restarts the process [bottom].

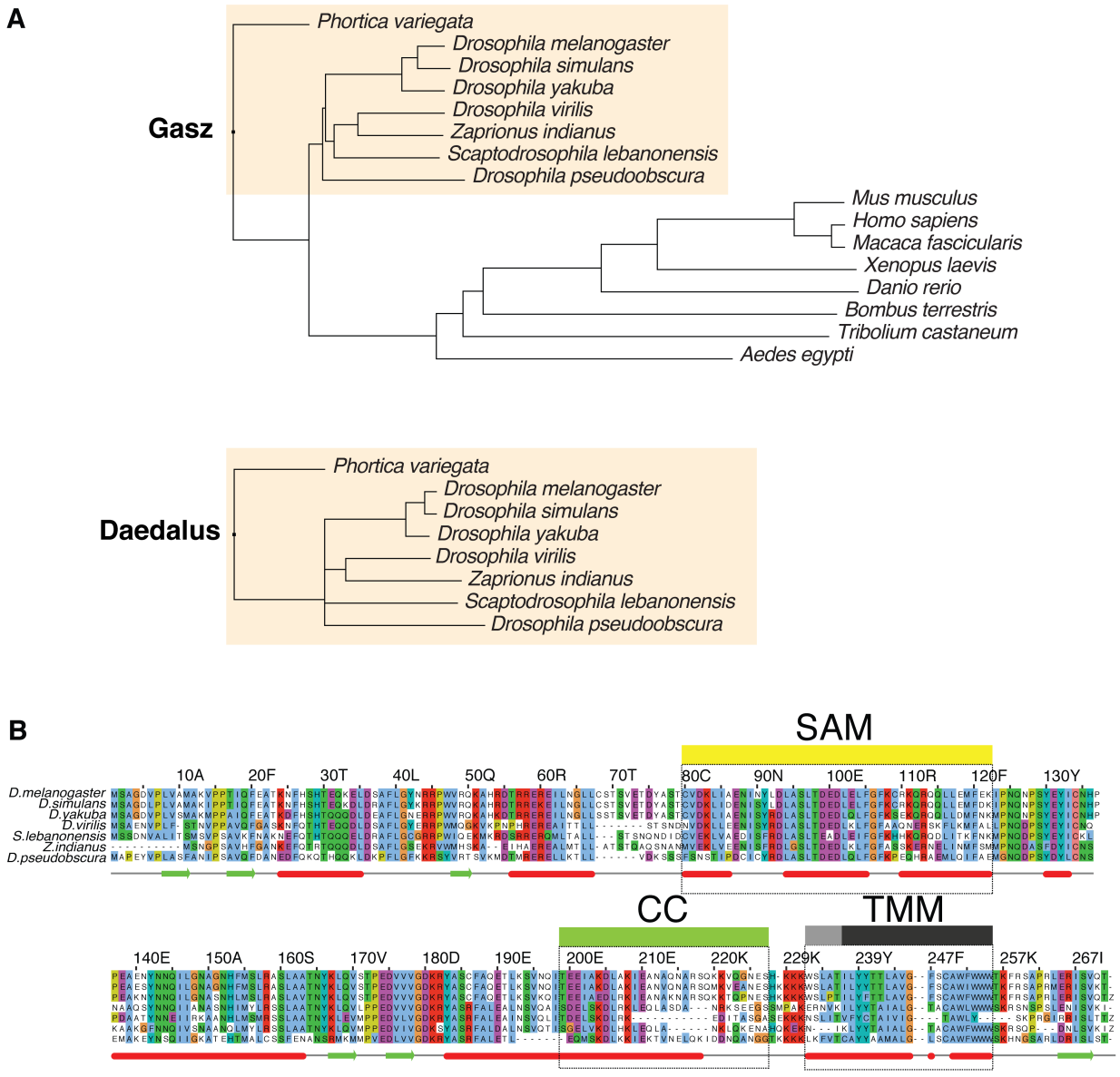


Figure 3.24 **Daed is unique to *Drosophilids*.** **(A)** Phylogenetic tree showing conservation of Gasz and Daed across species. Shaded in orange are *Drosophilidae* species. **(B)** Alignment of Daed amino acidic sequence from indicated species.

CHAPTER IV

Requirement of Nuclear Pore Complex subunits for transposon silencing

Author contributions

This chapter will be submitted for publication in a similar form. I performed all experiments with help from Alessandro Passera, Serena MacMillan, Victoria Lawless and B.C.. I analysed all sequencing data while the proteomics data analysis was carried out by the CRUK-CI Bioinformatics core. I conceived the project with inputs from B.C. and G.J.H.. I, B.C. and G.J.H. designed all experiments, discussed and interpreted the results. I wrote the manuscript draft with inputs from all authors.

BACKGROUND

The uni-strand piRNA cluster *flamenco* (*flam*) is the main source of piRNAs in somatic follicle cells (Li et al., 2009; Malone et al., 2009). It is a master regulator of *gypsy* family retroelements and mutations affecting *flam* expression severely compromise oogenesis (Desset et al., 2008; Lecher et al., 1997; Mevel-Ninio et al., 2007; Pelisson et al., 1994; Prud'homme et al., 1995; Sarot et al., 2004). *Flam* is a canonical RNA pol II transcript and as such undergoes capping and splicing and is exported by Nxf1/Nxt1 heterodimers (Dennis et al., 2016; Dennis et al., 2013; Goriaux et al., 2014; Tutucci and Stutz, 2011; Zanni et al., 2013). Nonetheless, *flam* must somehow be distinguished from other cellular transcripts and licensed for piRNA production. Although a sequence at the 5' end of the transcript has been shown to trigger piRNA production (Homolka et al., 2015; Ishizu et al., 2015), a complete picture of the transcription, export and specification of this unique piRNA cluster is still lacking. The only other uni-strand cluster expressed in somatic cells, *20A*, does not seem to share the same trigger sequence and is substantially smaller and expressed at lower levels, possibly implying that *flam* relies on a specialised molecular machinery.

Somatic cells feature discrete peri-nuclear granules of the DEAD box RNA helicase Yb, namely Yb bodies, which enable piRNA production (Murota et al., 2014; Saito et al., 2010). Like germline-specific nuage components, Yb is necessary for piRNA precursor specification, although the molecular mechanism remains unclear (Hirakata et al., 2019; Murota et al., 2014; Saito et al., 2010). Therefore, Yb bodies might be the somatic counterpart of nuage. Mutations that disrupt germline dual-strand cluster expression cause a disassembly of nuage (Czech et al., 2013; Klattenhoff et al., 2009; Pane et al., 2011; Zhang et al., 2012), indicating that expression and export of piRNA precursors is directly linked to the assembly of these non-membranous, peri-nuclear compartments. Given that *flam* is the major piRNA precursor expressed in somatic follicle cells, we investigated whether the formation of Yb bodies depends on *flam* and what other factor(s) might facilitate the nuclear export of transcripts arising from this unusually large, heterochromatic, uni-strand piRNA cluster.

RESULTS

IV.1. Yb bodies are the sites of *flamenco* export

In OSCs, previously described and used as an *in vitro* model of the somatic piRNA pathway, *flam* RNA localises to peri-nuclear granules, namely *flam* bodies, which are adjacent to Yb bodies (**Figure 4.1A**) (Murota et al., 2014). Somatic piRNA production is almost exclusively centred on *flam* and takes place in discrete peri-nuclear Yb bodies, thus we reasoned that the assembly of Yb bodies might depend on *flam* transcription and export. It has been shown that loss of Yb abolishes the formation of cytosolic *flam* bodies (Murota et al., 2014), therefore we asked whether loss of *flam* could have the reciprocal effect on Yb aggregation. I thus exploited two existing and previously reported mutant alleles (Brennecke et al., 2007; Malone et al., 2009; Mevel-Ninio et al., 2007) in which the *flam* locus is disrupted by P-element insertions near its 5' end (dashed blue lines in **Figure 4.1B**; $flam^{BG}$ refers to the allele $P\{GT1\}lncRNA:flam^{BG02658}$; $flam^{KG}$ refers to the allele $P\{SUPor\text{-}P\}lncRNA:flam^{KG00476}$). The $flam^{BG}$ allele carries a P-element insertion upstream of the annotated TSS, potentially disrupting the binding site of cubitus interruptus (ci), a transcription factor reported to drive the expression of *flam* (Goriaux et al., 2014). The $flam^{KG}$ allele instead carries a P-element insertion ~1.5 kb downstream of the annotated TSS and is expected to disrupt *flam* transcription·((Brennecke et al., 2007; Malone et al., 2009) and data described in this chapter). The ovaries are severely malformed in $flam^{BG/BG}$ homozygous mutants and completely absent in $flam^{KG/KG}$. For this reason, I crossed the two alleles and analysed their trans-heterozygous offspring, hereafter referred to as $flam^{BG/KG}$, whose phenotype, although still severe, is milder than that of $flam^{KG/KG}$ homozygous flies.

Aiming to determine whether *flam* is at all transcribed in the mutants or if they only fail to process the cluster transcript into piRNAs, I prepared RNA-seq libraries from $flam^{BG}$ heterozygous and homozygous ($flam^{BG/+}$ and $flam^{BG/BG}$), $flam^{KG}$ heterozygous ($flam^{KG/+}$) as well as trans-heterozygous $flam^{BG/KG}$ ovaries. As shown in the genome browser tracks in **Figure 4.1B**, *flam* transcription is globally unchanged in $flam^{BG/+}$ with respect to w^{1118} control flies. $Flam^{BG/+}$ ovaries also showed a normal size and morphology while $flam^{KG/+}$ were typically smaller, hence the former was used as reference for homozygous and trans-heterozygous mutant datasets. Both $flam^{BG/BG}$ and $flam^{BG/KG}$ mutants showed a striking reduction in the abundance of *flam* transcripts, with $flam^{BG/KG}$ showing the strongest decrease.

RNA-seq from *flam*$^{BG/KG}$ shows a few peaks upstream of the KG insertion site, perhaps corresponding to weak transcription initiation events from the *flam*^{KG} allele. Based on RNA-seq from *flam*$^{BG/BG}$, this allelic combination appears to still retain low level of RNA production all along the cluster, perhaps explaining the less severe impact on Yb bodies compared to the

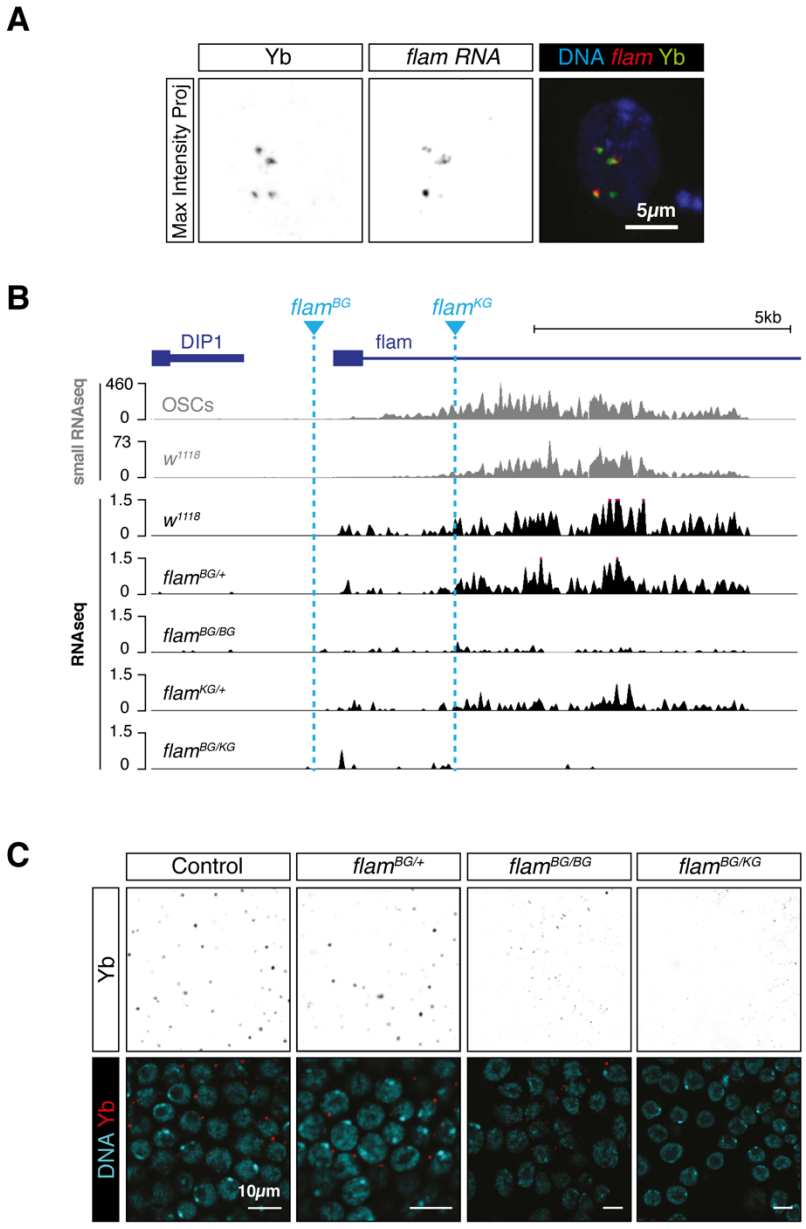
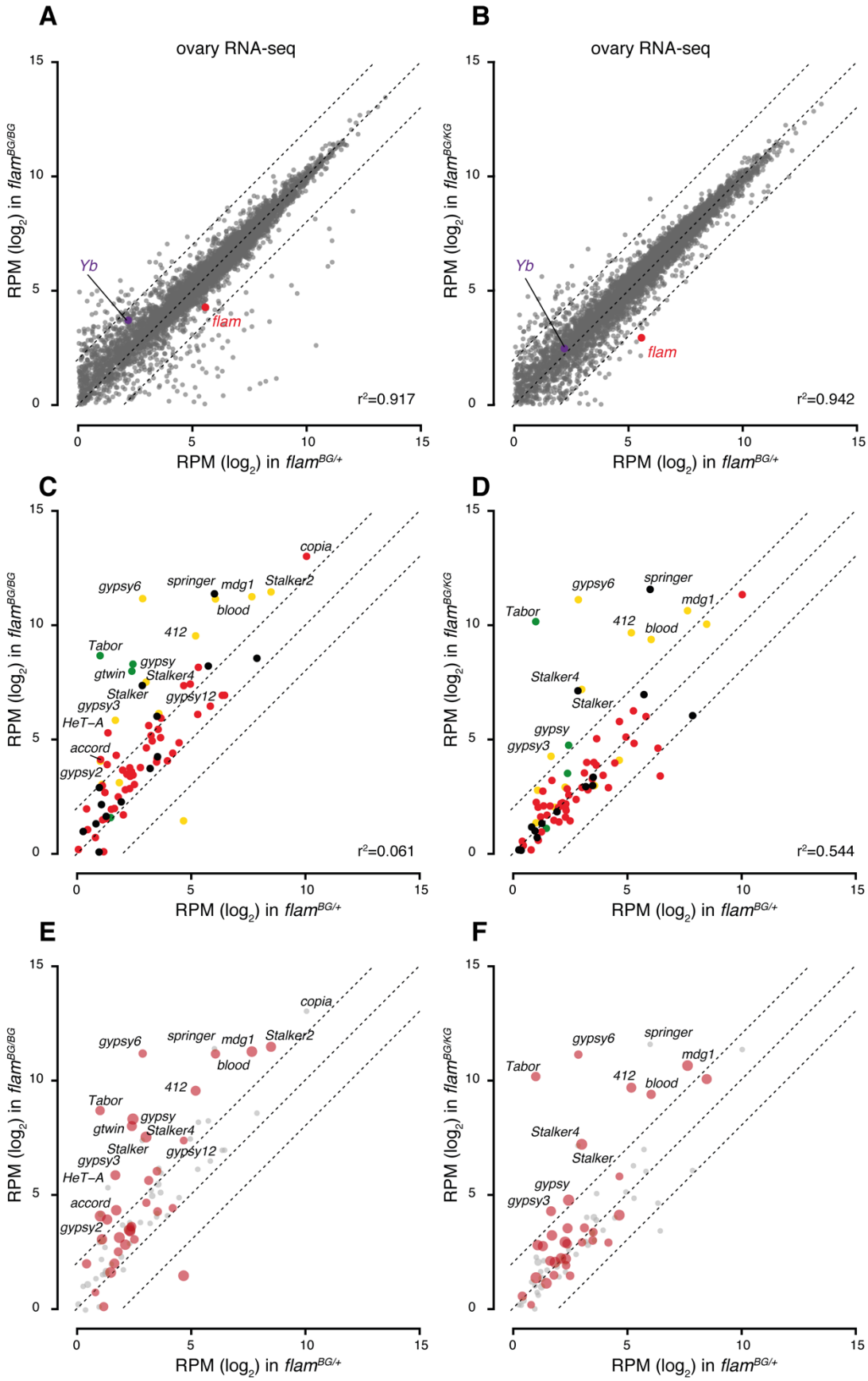


Figure 4.1. ***flam*** **mutants do not assemble Yb bodies. A)** Confocal images of Yb and *flam* RNA in OSCs. Scale bar, 5µm. **B)** Coverage plots of small RNA-seq (in grey) and RNA-seq (in black) reads from indicated genotypes uniquely mapped to the uni-strand cluster *flam*. Dashed blue lines indicate the location of the insertion of *flam*^{BG} and *flam*^{KG} alleles. **C)** Confocal images of Yb in ovaries. Scale bar, 10µm.

trans-heterozygous (see also **Figure 4.1C**). Having confirmed that *flam* RNA is severely reduced in the mutants, I asked whether this has consequences on the assembly of Yb bodies. Strikingly, while $flam^{BG/+}$ shows the usual peri-nuclear Yb foci, the signal is completely dispersed in $flam^{BG/BG}$ and $flam^{BG/KG}$ ovaries (**Figure 4.1C**). More specifically, Yb bodies become progressively smaller and disorganised in the $flam^{BG/BG}$ mutant and are completely lost in $flam^{BG/KG}$. This might be explained by the residual transcriptional activity in $flam^{BG/BG}$ seen in **Figure 4.2B**, which is instead completely absent in $flam^{BG/KG}$. Taken together, these data indicate that the assembly of Yb bodies is directly linked to the presence of *flam* transcripts in the cell. Since Yb is exclusively present on the outer side of the nuclear envelope, this indirectly links Yb bodies to the export of *flam.*

Next, I analysed the global transcriptome changes occurring in *flam* homozygous mutants compared to $flam^{BG/+}$. Both homozygous mutants showed a more pronounced gene de-regulation compared to typical piRNA pathway mutants, possibly due to wide-spread tissue damage and propagation of viral-like gypsy particles. 153 out of 7,414 genes (that were above the expression threshold of 1 RPM) were de-regulated more than 4-fold in $flam^{BG/BG}$ (r^2=0.917) and 86 out of 7,413 in $flam^{BG/KG}$ (r^2=0.942) (**Figure 4.2A-B**). Notably, the expression levels of *Yb* were not reduced in either homozygous mutant, indicating that impaired formation of Yb bodies is not caused by its decreased abundance. As expected, both *flam* mutant allelic combinations showed a dramatic de-repression of soma-specific and intermediate TEs (**Figure 4.2C-D**), with 26 out of 60 transposon families (that were above the expression threshold of 1 RPM) being de-repressed more than 4-fold in in $flam^{BG/BG}$ (r^2=0.061) and 10 in $flam^{BG/KG}$ (r^2=0.544). Furthermore, the most highly up-regulates TEs are encoded within *flam* (**Figure 4.2E-F**), indicating that most of the TE de-regulation observed is the consequence of impaired *flam* transcription and/or piRNA production.

Figure 4.2 ***flam*** **mutants RNA-seq analysis. A-B)** Scatter plot showing expression levels (RPM) of genes in ovary RNA-seq from indicated genotypes (n=3). **C-D)** Scatter plots showing expression levels (RPM) of sense transposon sequences in ovary RNAseq from indicated genotypes. Red dots correspond to ovarian germline-specific transposons, yellow dots to intermediate transposons and green dots to soma-specific ones. **E-F)** Same plot as in **C-D** but with transposon sequences present in *flam* highlighted in red. The dot size is proportional to their abundance within *flam* according to dm6 Repeat Masker annotations.


A
ovary RNA-seq
RPM (log2) in flamBG/BG
RPM (log2) in flamBG/+
Yb
flam
r2=0.917
B
ovary RNA-seq
RPM (log2) in flamBG/KG
RPM (log2) in flamBG/+
Yb
flam
r2=0.942
C
RPM (log2) in flamBG/BG
RPM (log2) in flamBG/+
copia
gypsy6
springer
mdg1
Stalker2
blood
412
Tabor
gypsy
gtwin
Stalker4
Stalker
gypsy12
gypsy3
HeT–A
accord
gypsy2
r2=0.061
D
RPM (log2) in flamBG/KG
RPM (log2) in flamBG/+
gypsy6
springer
mdg1
Tabor
412
blood
Stalker4
Stalker
gypsy
gypsy3
r2=0.544
E
RPM (log2) in flamBG/BG
RPM (log2) in flamBG/+
copia
gypsy6
springer
mdg1
Stalker2
blood
412
Tabor
gypsy
gtwin
Stalker4
Stalker
gypsy12
gypsy3
HeT–A
accord
gypsy2
F
RPM (log2) in flamBG/KG
RPM (log2) in flamBG/+
gypsy6
springer
mdg1
Tabor
412
blood
Stalker4
Stalker
gypsy
gypsy3

OSCs express *flam* and assemble functional Yb bodies, therefore I decided to use this system to investigate whether ablation of *flam* export impacts the formation of Yb bodies. Previous work has demonstrated that *flam* is exported via the canonical mRNA nuclear export route, which requires Nxt1 and Nxf1 heterodimers (Dennis et al., 2016; Dennis et al., 2013; Goriaux et al., 2014; Zanni et al., 2013). To confirm this finding, I depleted *nxf1* and *nxt1*, along with two other members of the NXF family that were reported as candidate piRNA pathway factors (*nxf2* and *nxf3*) (Czech et al., 2013). As expected, both *nxf1* and *nxt1* knockdown completely abolished the formation of Yb bodies, whereas that of *nxf2* and *nxf3* had no effect (**Figure 4.3**).This is in line with recent data, including **Chapter V** of this thesis, that implicate Nxf2 and Nxf3 in TGS and dual-strand piRNA cluster export, respectively (Batki et al., 2019; ElMaghraby et al., 2019; Fabry et al., 2019; Kneuss et al., 2019; Murano et al., 2019; Zhao et al., 2019). Notably, depletion of *nxf1* and *nxt1* had a dramatic effect on cell viability and a very small number of cells survived until the collection timepoint (96 hrs). In fact, the cells shown in **Figure 4.3** for *siNxf1* were imaged 48 hrs after the first siRNA nucleofection, as they did not survive for longer periods. This is expected since both factors are required for bulk mRNA export and their loss results in pleiotropic effects that extend beyond transposon control (Herold et al., 2001; Herold et al., 2000; Herold et al., 2003). Depletion of piRNA biogenesis factors acting downstream of Yb, such as Armi or Zuc, does not affect Yb bodies formation (Ishizu et al., 2015; Olivieri et al., 2012). Altogether, these data show that the disassembly of Yb bodies can be used as a proxy for impaired *flam* export and processing.

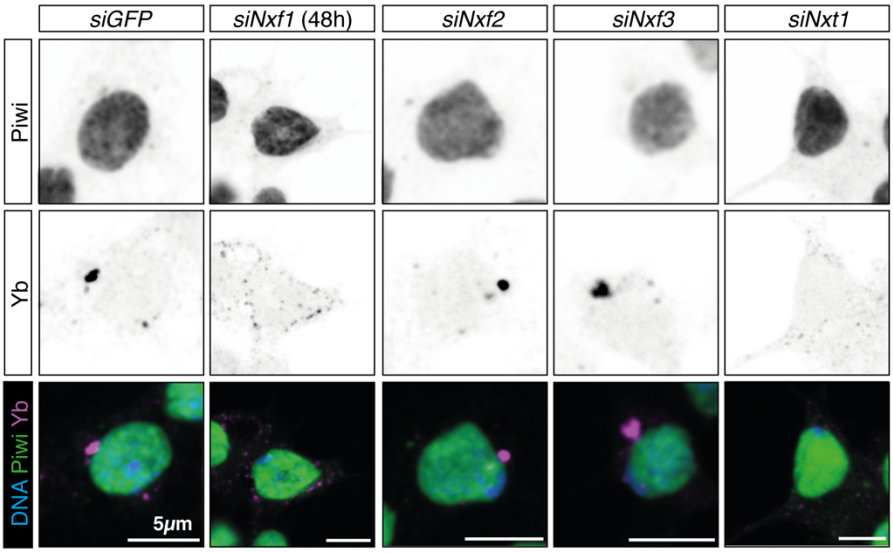
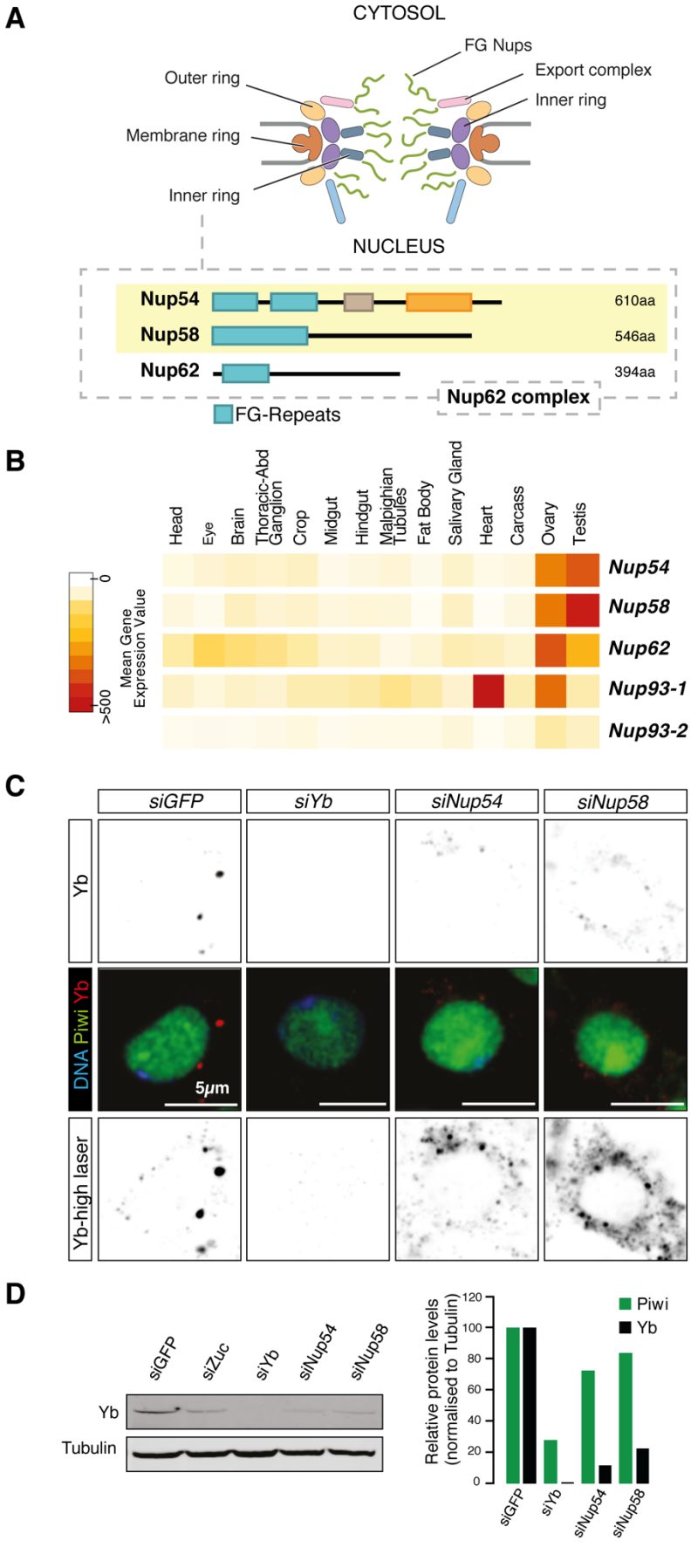


Figure 4.3 Yb bodies are the sites of *flam* export. Confocal images of Piwi and Yb in OSCs upon indicated knockdown. Unless otherwise specified, knockdowns cells are analysed 96h after the first siRNA nucleofection. Scale bar, 5µm.

IV.2. Nuclear Pore Complex subunits are required for transposon silencing

Flam is transcribed by RNA pol II and, as such, it has to be distinguished or detected among all other cellular RNAs and specifically funnelled to the piRNA production machinery. Data presented so far and those available in the literature assign a key role to Yb bodies in the cytosol (Dennis et al., 2016; Hirakata et al., 2019; Ishizu et al., 2015; Murota et al., 2014; Pandey et al., 2017; Qi et al., 2011; Saito et al., 2010). Nonetheless, a complete picture of the molecular events underlying *flam* specification for piRNA production is still missing. I therefore asked whether additional factor(s) facilitate *flam* export across the nuclear envelope.

A genetic screen for genes involved in the somatic piRNA pathway has identified several components of the Nuclear Pore Complex (NPC) among the strongest hits (Handler et al., 2013). The NPC is a large transmembrane selective channel that regulates the nucleo-cytoplasmic transport of macromolecules but is permeable to the diffusion of ions and other small molecules (Beck and Hurt, 2017). It is a cylindrical structure with 8-fold symmetry constituted by multiple copies of approximately 30 Nucleoporins (Nups). These Nups are gathered into sub-complexes forming higher-order structures called spokes, which in turn assemble larger modules within the pore (Kim et al., 2018). Each NPC can be divided into different sub-regions (summarized in the cartoon in **Figure 4.4A**), such as the inner and outer rings, the nuclear basket and the cytoplasmic export complexes (Beck and Hurt, 2017; Kim et al., 2018). The inner lumen of the pore acts as a barrier with selective permeability, thanks to a highly conserved class of "FG Nucleoporins" (FG Nups) (Beck and Hurt, 2017). The FG Nups carry multiple repeats of Phenylalanine and Glycine (FG), which form intrinsically disordered regions mostly lining the inner channel of the NPC but also present at its periphery. The FG repeats inside the lumen form a cohesive mesh that determines the permeability of the pore via interaction with nuclear transport receptors, such as Nxf1, various importins and others (Hulsmann et al., 2012; Patel et al., 2007). On the other hand, peripheral FG Nups on the nuclear basket act as repulsive filaments thanks to a non-cohesive assembly (Patel et al., 2007). FG Nups generally have shorter residence times in the NPC and are sometimes considered to be transient or shuttling components of the pore (Terry and Wente, 2009). Interestingly, two FG repeat Nucleoporins, Nup54 and Nup58, were among the top candidates in the aforementioned somatic piRNA pathway screen (Handler et al., 2013).


A
CYTOSOL
FG Nups
Export complex
Inner ring
Outer ring
Membrane ring
Inner ring
NUCLEUS
Nup54
610aa
Nup58
546aa
Nup62
394aa
Nup62 complex
FG-Repeats
B
Head
Eye
Brain
Thoracic-Abd Ganglion
Crop
Midgut
Hindgut
Malpighian Tubules
Fat Body
Salivary Gland
Heart
Carcass
Ovary
Testis
Nup54
Nup58
Nup62
Nup93-1
Nup93-2
Mean Gene Expression Value
0
>500
C
siGFP
siYb
siNup54
siNup58
Yb
DNA Piwi Yb
5μm
Yb-high laser
D
siGFP
siZuc
siYb
siNup54
siNup58
Yb
Tubulin
Relative protein levels (normalised to Tubulin)
120
100
80
60
40
20
0
Piwi
Yb
siGFP
siYb
siNup54
siNup58

Nup54 and Nup58 are FG Nups that assemble in the highly conserved Nup62•58•54 (Nup62) complex, whose yeast counterpart is the Nsp1•Nup49•Nup57 complex (or Nic96 complex) (Chug et al., 2015; Finlay et al., 1991; Grandi et al., 1993; Grandi et al., 1995; Guan et al., 1995; Schlaich et al., 1997). This complex protrudes towards the inner lumen of the pore and mediates the passage of macromolecules via its FG-repeats mesh. It is anchored to the main NPC scaffold via Nup93 (or Nic96 in yeast), which does not have FG repeats and also emerged as candidate in the piRNA pathway screen (Handler et al., 2013). The Nup62 complex constitutes a significant fraction of the total FG repeat mass of the pore and its components are all essential for viability (Chug et al., 2015). Intriguingly, knockdown of Nup54 and Nup58 in the *Drosophila* follicle cells did not significantly impact ovarian morphology (Handler et al., 2013). This represents a surprising finding, as Nup54 and 58 are core components of the NPC, and might indicate that these two proteins carry out specialised functions in the somatic cells of the ovary. In addition to Nup54 and Nup58, three other components of the NPC scored in the above-mentioned screen, although to lower degrees: Nup93-1, Nup214 and Gle1 (Handler et al., 2013). Interestingly, all three components of the Nup62 complex and Nup93-1 show higher expression in the fly ovary (**Figure 4.4B**) (Chintapalli et al., 2007), which is often predictive of an involvement in the piRNA pathway, although in this case it might also result from *de novo* NPC assembly and loading into the growing oocyte (Hampoelz et al., 2019). Nup 93-2, another fly homolog of yeast Nic96, is instead expressed at lower levels and does not show a bias for the reproductive tissues (**Figure 4.4B**). Given the specificity of Nup54 and Nup58 to somatic transposon control and their involvement in the selective permeability of the NPC, I decided to investigate whether they are connected to *flam* export and processing.

I first performed knockdowns of *nup54* and *nup58* in OSCs and, in stark contrast with what I had observed upon depletion of *nxf1* and *nxt1*, their depletion did not lead to dramatic cell death. Strikingly, immunofluorescence of Yb revealed that both knockdowns led to complete dispersal of Yb bodies (**Figure 4.4C**). However, Yb itself was still present in the cell, as shown by the same immunofluorescence captured with increased laser power (**Figure 4.4C**, bottom

Figure 4.4 Nup54 and Nup58 are required for Yb bodies formation. A) Cartoon representing the NPC and Nup62 components domain structure (adapted from Kim et al., 2018). **B)** Expression levels of *Nup54, 58, 62, 93-1* and *93-2* in different tissues of the adult fly. **C)** Confocal images of Yb in OSCs upon indicated knockdown. Scale bar, 5µm. **D)** Western blot with relative quantification showing expression levels of Piwi and Yb upon different knockdowns. (n=1)

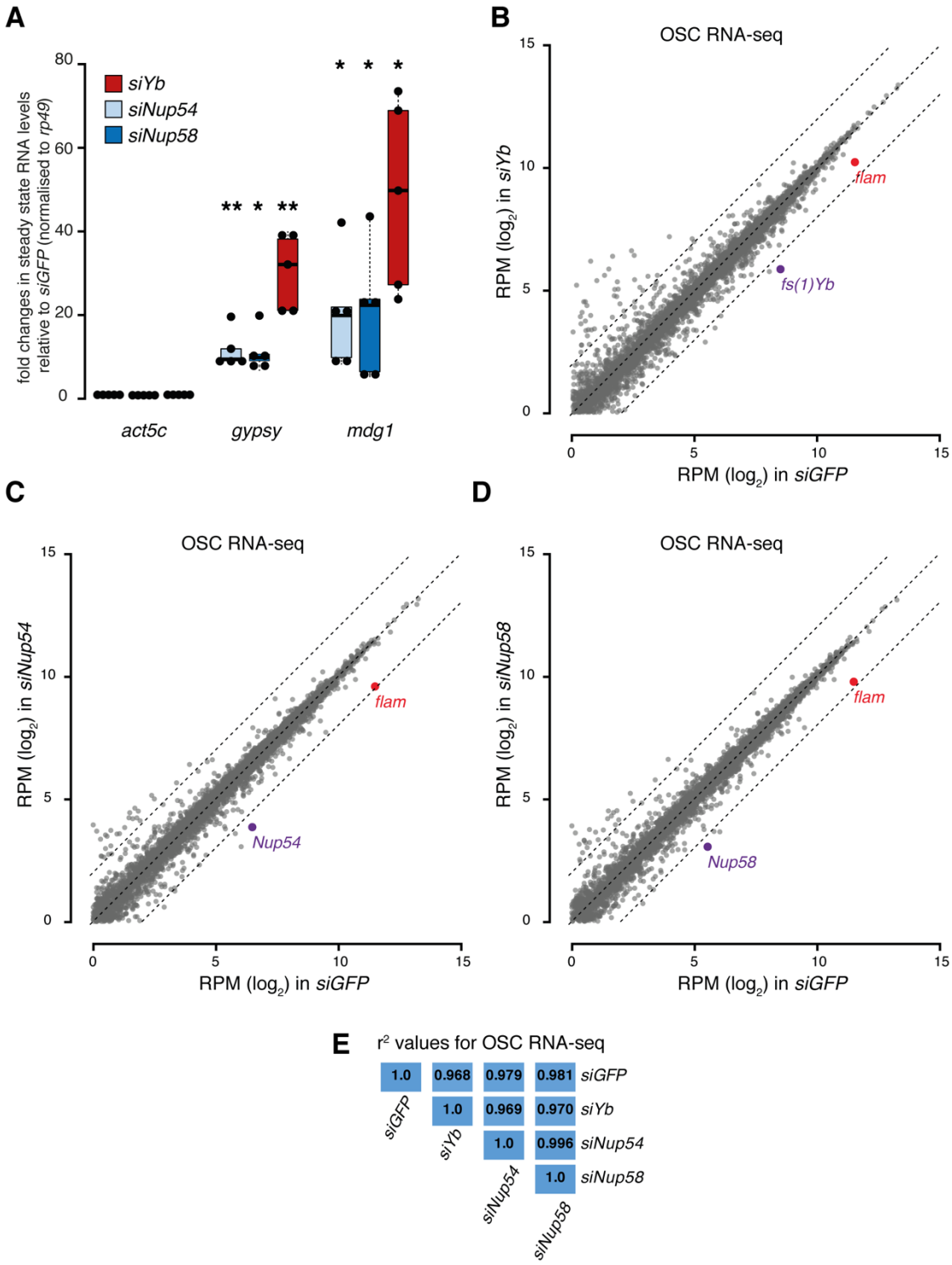


Figure 4.5 **Nup54 and Nup58 are involved in transposon control. A)** Fold changes in the steady-state RNA levels of somatic transposons in OSCs upon indicated knockdowns. Values are relative to *GFP* control knockdown and normalised to *rp49*. * = P value < 0.05; ** = P < 0.001 (unpaired t-test) (n=5). **B-D)** Scatter plot showing expression levels (RPM) of genes in OSC RNAseq from indicated knockdowns (n=4). **E)** r^2 values for OSC RNA-seq upon indicated knockdowns, calculated based on gene RPMs.

panel) and via western blot (**Figure 4.4D** and relative quantification). In contrast to knockdown of *yb*, Piwi protein levels were not dramatically reduced by *siNup54* and *siNup58*, indicating that piRNA production as a whole is not compromised. Furthermore, knockdown of *nup54* and *nup58* in OSCs led to de-repression of somatic transposons, although to a lesser extent than that caused by y*b* loss (**Figure 5.5A**).

Intrigued by these results, I performed RNA-seq to understand the transcriptome-wide changes caused by *yb*, *nup54* and *nup58* depletion (**Figure 4.5B-D**). Surprisingly, depletion of the two FG Nups did not substantially affect global gene expression (**Figure 4.5C-D**), with only 32 out of 7,826 genes (that were above the expression threshold of 1 RPM) showing de-regulation more than 4-fold in *siNup54* (r^2=0.979) and 31 out of 7,831 in *siNup58* (r^2=0.981). The changes in gene expression between *siNup54* and *siNup58* were nearly identical (r^2=0.996) and very similar to those caused by *siYb* (r^2=0.969 and 0.970, respectively) (**Figure 4.5E**). These results imply that, despite being core components of the NPC, they only have modest effects on global gene expression. In striking contrast, all three knockdowns led to specific up-regulation of somatic and intermediate TEs (yellow and green dots in **Figure 4.6 A-C**), with the majority of them being under the control of *flam* (Brennecke et al., 2007). The transposon families that showed de-repression upon loss of the either Nup were nearly identical to those up-regulated in *siYb*, although the overall changes were milder as observed via qPCR (**Figure 4.5A**). Finally, inspection of somatic piRNA source loci revealed that *flam* showed reduced RNA levels in all three knockdowns (**Figure 4.6D**), whereas *20A* or *tj* did not (nor did other genic piRNA precursors) (**Figure 4.6E-F**). This unexpected finding might underscore the existence of a link between the transcription of *flam* and its export to Yb bodies. The transcriptome-wide changes observed upon Nup54 and Nup58 loss in OSCs closely resemble a typical phenotype of impaired piRNA-mediated transposon control and are remarkably different from what would be expected by depletion of core NPC components or bulk mRNA export factors (Herold et al., 2003). These data, together with the results obtained in *flam* mutants, suggests that Yb bodies can only form when *flam* is transcribed and exported to the cytosol and that this step is likely failing in the absence of Nup54 and Nup58.

Aiming to dissect which portion of Nup54 and Nup58 is required for transposon control, I designed deletion mutants lacking either the FG repeats (ΔFG) or the carboxy-terminus (ΔC) (cartoon in **Figure 4.7A**). The carboxy-terminal region of Nup54 carries a putative ssDNA-binding region followed by a Nup54-family domain (brown and orange boxes in **Figure 4.7A**,

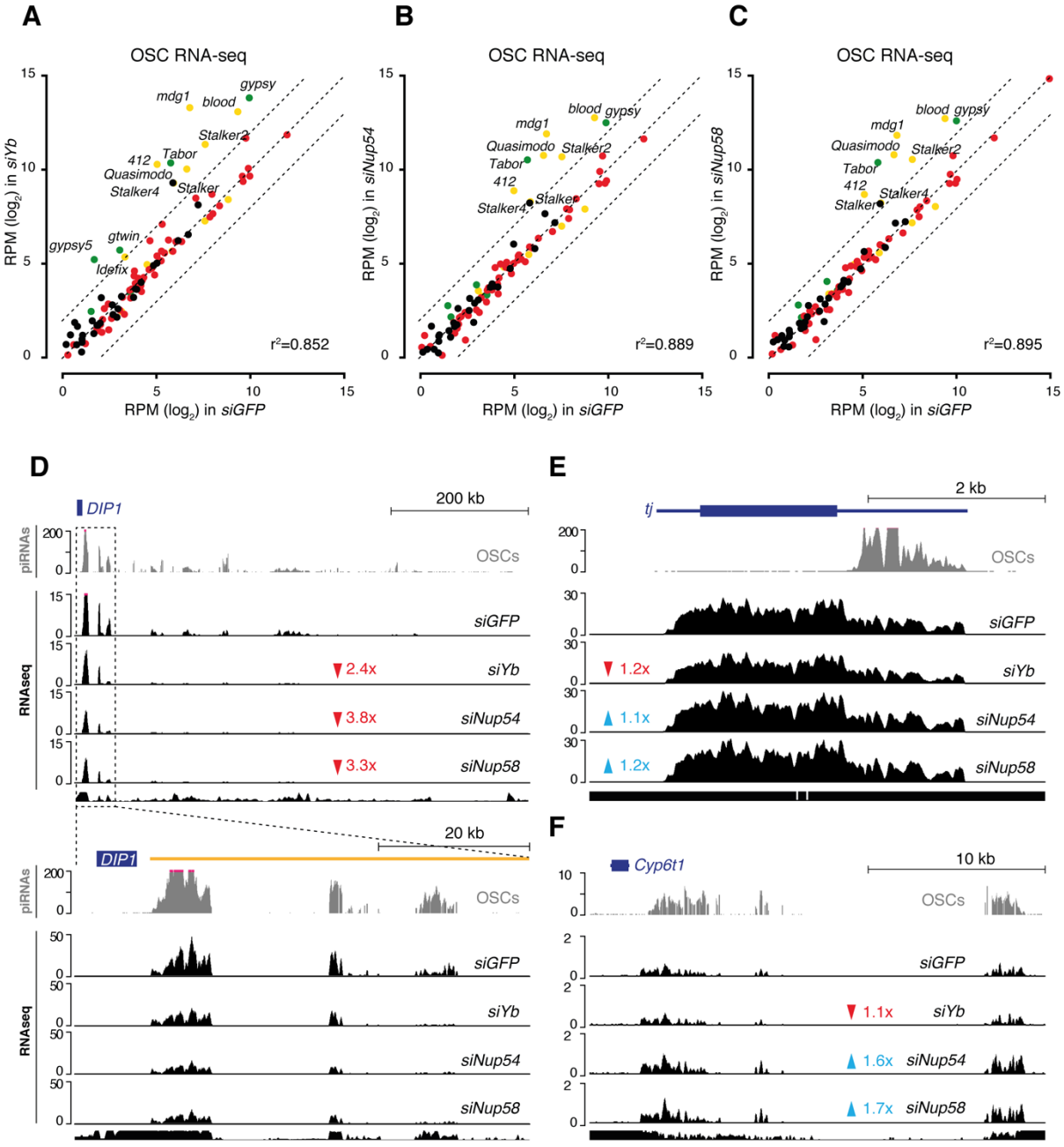


Figure 4.6 **Nup54 and Nup58 are involved in transposon control. A-C)** Scatter plots showing expression levels (RPM) of sense transposon sequences in OSC RNA-seq from indicated knockdowns. Red dots correspond to germline-specific transposons, yellow dots to intermediate transposons and green dots to soma-specific ones. **D-F)** Coverage plots of small RNA-seq (in grey) and RNA-seq (in black) reads from indicated knockdowns uniquely mapped to the uni-strand clusters *flam* (**D**; zoom on the first 50 kb in the bottom panel) and *20A* **(G)** and the protein-coding gene *tj* **(F)**. The mappability for an average 50 bp read length is shown at the bottom of each panel.

respectively), whereas that of Nup58 lacks readily identifiable protein domains (identified via NCBI Conserved Domain Search). All constructs are resistant to the siRNA used to knockdown either Nup and carry an HA tag. Unexpectedly, both Nups lacking the FG repeats could rescue transposon de-repression, whereas the ΔC constructs did not (**Figure 4.7B**). While the FG repeats contribute to the selective permeability of the NPC, the carboxy terminus of Nup54 and Nup58 anchors them to the scaffold of the NPC via interaction with Nup93 (Chug et al., 2015; Grandi et al., 1993; Grandi et al., 1995). In line with that, co-immunoprecipitation experiments with the same deletion constructs showed that only Nup54$^{\Delta C}$ failed to interact with full-length Nup58, and vice versa Nup58$^{\Delta C}$ did not coimmunoprecipitate with full length Nup54 (**Figure 4.7C**). On the other hand, both ΔFG constructs successfully coimmunoprecipitated with the other NPC components (**Figure 4.7C**), indicating that their ability to connect with the partner Nups and the NPC scaffold, and not the FG repeats *per se*, is required for transposon control. Intriguingly, Nup54 and Nup58 lacking the carboxy-terminal region exacerbated the transposon upregulation, possibly indicating a dominant negative effect (**Figure 4.7B**).

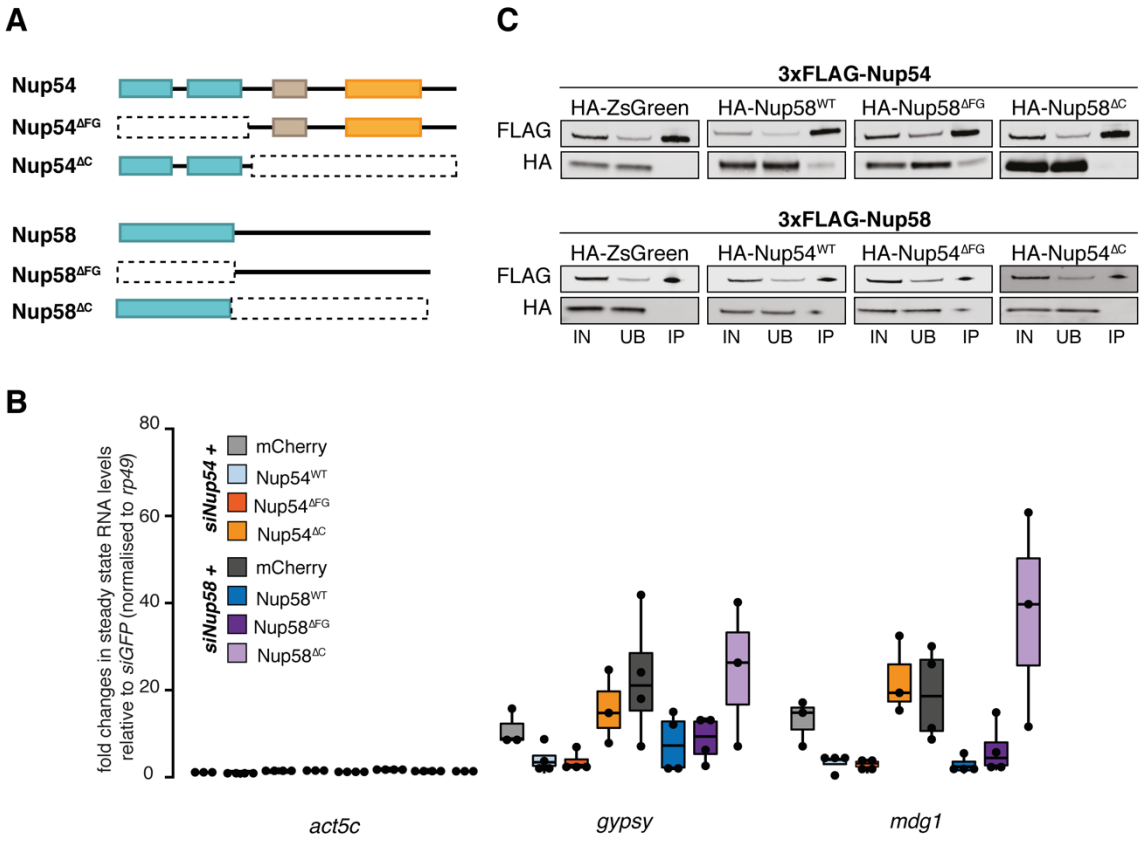


Figure 4.7 The FG repeats of Nup54 and Nup58 are dispensable for transposon control. A) Cartoon representing Nup54 and Nup58 domain mutant structure. **B)** Fold changes in the steady-state RNA levels of somatic transposons in OSCs nucleofected with siRNAs and various rescue constructs. Values are relative to *GFP* control knockdown and normalised to *rp49*. **C-D)** Western blots of FLAG-tag co-immunoprecipitations from lysates of S2 cells transfected with the indicated constructs.

IV.3. Yb is in close proximity to the NPC

Yb is an RNA helicase of the DEAD box family, whose members are involved in a multitude of cellular RNA metabolism steps, including the release of mRNPs on the cytosolic side of the NPC (Beck and Hurt, 2017; Tutucci and Stutz, 2011). Given that the formation of Yb bodies depends on *flam* export, I asked whether Yb could be an integral component of the NPC. I first performed a proximity biotinylation experiment with Yb-BASU, as described previously (**Chapter III** and detailed protocol in **Materials and Methods II.12**). Yb PL-MS enriched for the piRNA pathway factors Piwi and Armi, known interactors of Yb, as well as SoYb and others (**Figure 4.8A**). Notably, mitochondrial piRNA pathway factors such as Mino, Gasz and Daed were not significantly enriched in Yb PL-MS (green dots in **Figure 4.8A**), in line with a model whereby Armi and Piwi bound to pre-piRNAs translocate to mitochondria whilst Yb does not (compare with Armi PL-MS in **Figure 3.15A**). No NPC components were detected as significantly enriched, suggesting that Yb is not stably associated with the NPC. To further assess whether any transient interaction between Yb and the NPC exists, I also performed PL-MS experiments for both Nup54 and Nup58. The BioID proximity labelling approach has been initially developed for NPC components and is therefore most apt to identify their stable and transient interactors (Kim et al., 2014; Kim et al., 2016). I tested both an amino- and a carboxy-terminal fusion of BASU to the Nups, and found the former to be most stable and active. This is in line with the requirement of their carboxy-terminus for assembly of Nup54 and Nup58 in the NPC (Chug et al., 2015; Grandi et al., 1993; Grandi et al., 1995) (**Figure 4.7C**). Both Nup54 and Nup58 PL-MS enriched for the bait together with several other Nups (blue dots in **Figure 4.8B-C**) and canonical export factors. Among the most highly enriched proteins were all components of the Nup62 complex (Nup54, Nup58 and Nup62) and the linker Nup93-1 connecting the complex to the NPC scaffold, altogether validating the approach. Strikingly, both PL-MS experiments showed enrichment of Yb (orange dot in **Figure 4.8B-C**), but not of other piRNA pathway factors. This suggests that Yb alone or the Yb bodies are in close physical proximity to the NPC components, Nup54 and Nup58. However, this result does not reflect a direct interaction nor indicates that Yb is an integral component of the NPC. In fact, Yb PL-MS did not identify any Nup nor did immunoprecipitation of 3xFLAG-Yb copurify HA-tagged Nup54 and Nup58 and vice-versa (**Figure 4.8D**). Aside from other Nups and Yb, my PL-MS also provided a list of candidates that were strongly enriched in both Nup54 and Nup58 PL-MS experiments and whose potential involvement in the piRNA pathway will be further investigated.

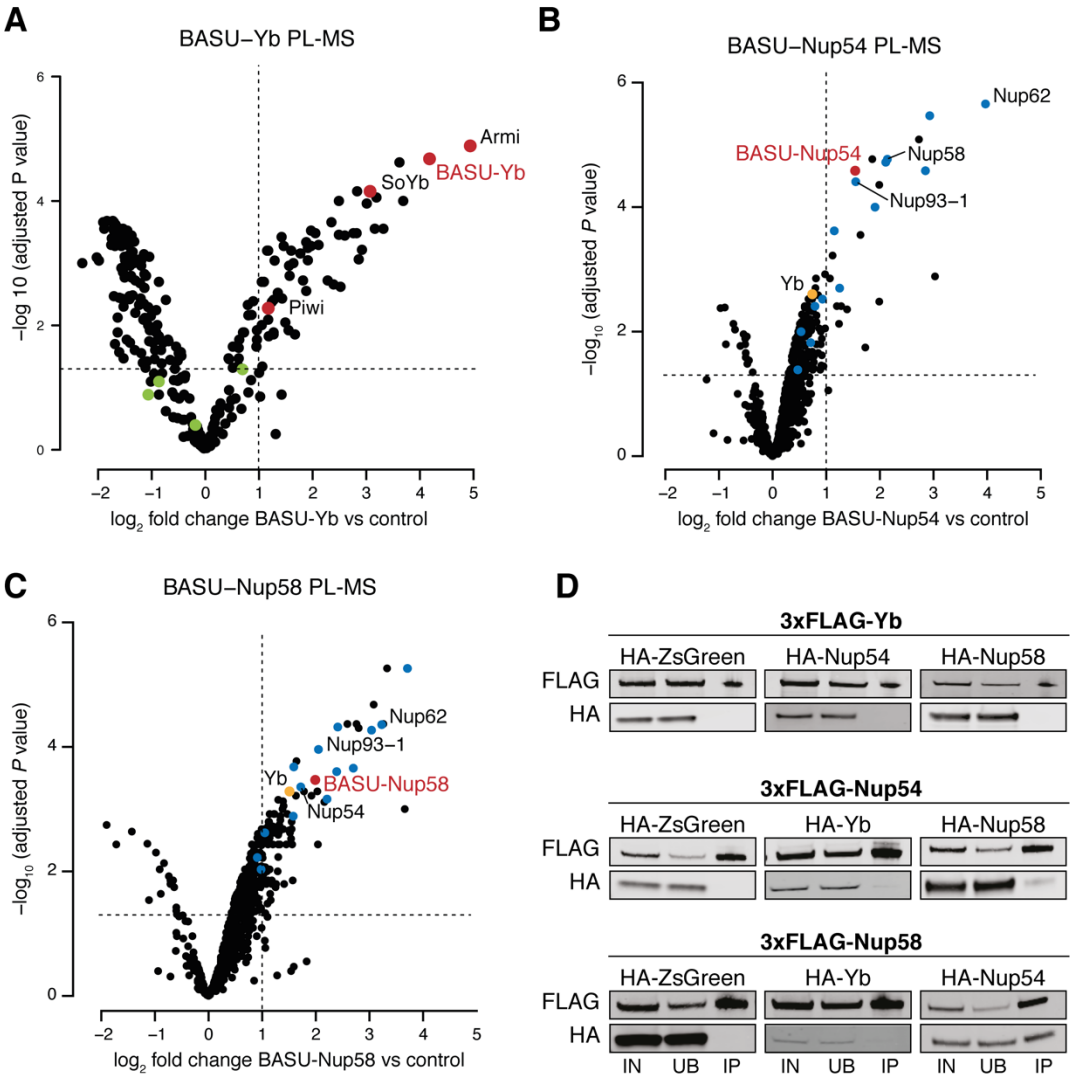


Figure 4.8 **Yb bodies are proximal to NPCs. A- C)** Volcano plots showing enrichment and corresponding significance of biotinylated proteins identified via PL-MS from OSCs expressing the indicated constructs against control (n=3). Highlighted in red are piRNA pathway factors, in green mitochondrial piRNA pathway factors, in blue the Nucleoporins and in orange Yb. See also supplementary **Tables 6-8. D)** Western blots of FLAG-tag co-immunoprecipitations from lysates of S2 cells transfected with the indicated constructs.

IV.4. Nup54 and 58 are required for piRNA production specifically from *flam*

Nup54 and Nup58 appear to be required for transposon control in the somatic compartment of the ovary. To understand whether piRNA production is affected upon their loss, I performed small RNA sequencing from OSCs depleted of *yb*, *nup54* and *nup58*. The population of transposon-mapping small RNAs in OSCs is predominantly antisense to TEs, in line with *flam* being the predominant source (**Figure 4.9A**, *siGFP* panel). Depletion of *yb* leads to a dramatic

collapse of antisense piRNAs and that of *nup54* and *nup58* causes a similar, although milder, effect, while 21nt-long siRNAs are mostly unchanged (**Figure 4.9A**). Residual antisense piRNAs observed in either Nup knockdown can also explain why *siNup54* and *siNup58* do not greatly reduce Piwi protein levels, when compared to *siYb* (**Figure 4.4C-D**). The levels of *flam*-mapping piRNAs were completely abolished upon loss of *yb*, in line with its key role in specifying *flam* for piRNA production. Knockdown of *nup54* and *nup58* also greatly reduced *flam* piRNA production, although to a lesser extent than that of *yb*. Strikingly, while loss of *yb* similarly impacted piRNA production from all source loci, including *tj* and *20A*, the defect caused by loss of the two Nups was unique to *flam* piRNAs. As shown in **Figure 4.9C** and **D**, the amount of piRNAs originating from *tj* and *20A* was essentially unchanged. Inspection of the global levels of genome-mapped piRNAs revealed substantial differences between knockdown of *yb* and that of *nup54* and *nup58*. Knockdown of *yb* had a vast effect on all piRNA populations (**Figure 4.10A**; r^2=0.546). Whilst processing of *flam, 20A, tj* and the other canonical piRNA processing substrates was reduced, there was a concomitant increased processing of other transcripts, which are normally not substrates of piRNA biogenesis. In the absence of Yb, piRNA processing does not seem to occur on selected substrates but instead shows promiscuous consumption of any available transcript. These results further point to Yb as the 'licensing factor' recognising selected cellular RNAs for processing into piRNAs. On the other hand, knockdown of *nup54* and *nup58* only had mild effects on genome-mapped piRNA levels (**Figure 4.10B-C**; r^2=0.930 and 0.931, respectively). Strikingly, *flam* was one of the very few loci showing decreased piRNA levels, further strengthening the notion that the main consequence of *siNup54* and *siNup58* is loss of *flam* piRNAs. If these knockdowns were causing a general impairment of RNA export, they would impact cell viability or affect piRNA production for all precursors equally. It is interesting to note that the ~3-fold reduction of *flam* piRNAs upon *siNup54* and *siNup58* is consistent with the similar reduction of *flam* transcripts detected by RNA-seq (**Figure 4.6D**). In contrast, the 30-fold reduction of *flam* piRNAs caused by *yb* depletion is much more pronounced than the decrease observed by RNA-seq (**Figure 4.6D**). If Yb is absent, any *flam* transcript that is successfully exported cannot be recognised and specified for piRNA production. On the other hand, any residual *flam* transcript that is exported in the Nups knockdown can still be processed thanks to the presence of Yb , which can perform its function even if not accumulated in Yb bodies (Hirakata et al., 2019; Ishizu et al., 2019). Altogether, this argues for a specific involvement of Nup54 and Nup58 in the export route of *flam* piRNA precursors, whereas Yb is responsible for licensing *flam* as

well as all other precursors for piRNA production.

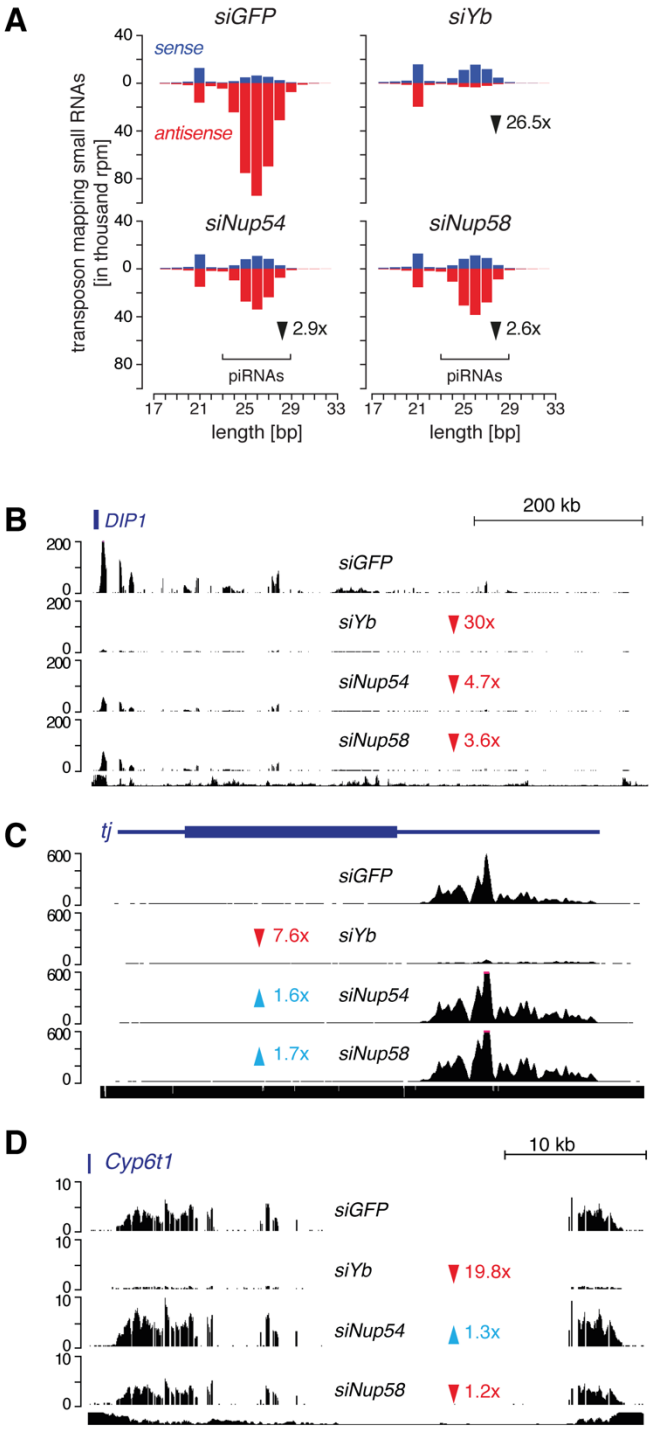


Figure 4.9 Depletion of *nup54* and *nup58* abolishes piRNA production from *flam*. A) Size distribution of transposon-mapping small RNAs from OSCs upon indicated knockdown. Sense reads are shown in blue, antisense in red. **B-D)** Coverage plots of small RNA reads uniquely mapped to the uni-strand clusters *flam* **(B)** and *20A* **(D)** and the coding gene *tj* **(C)**. Shown are normalised RPM. The mappability for an average 25 bp read length is shown at the bottom of each panel.

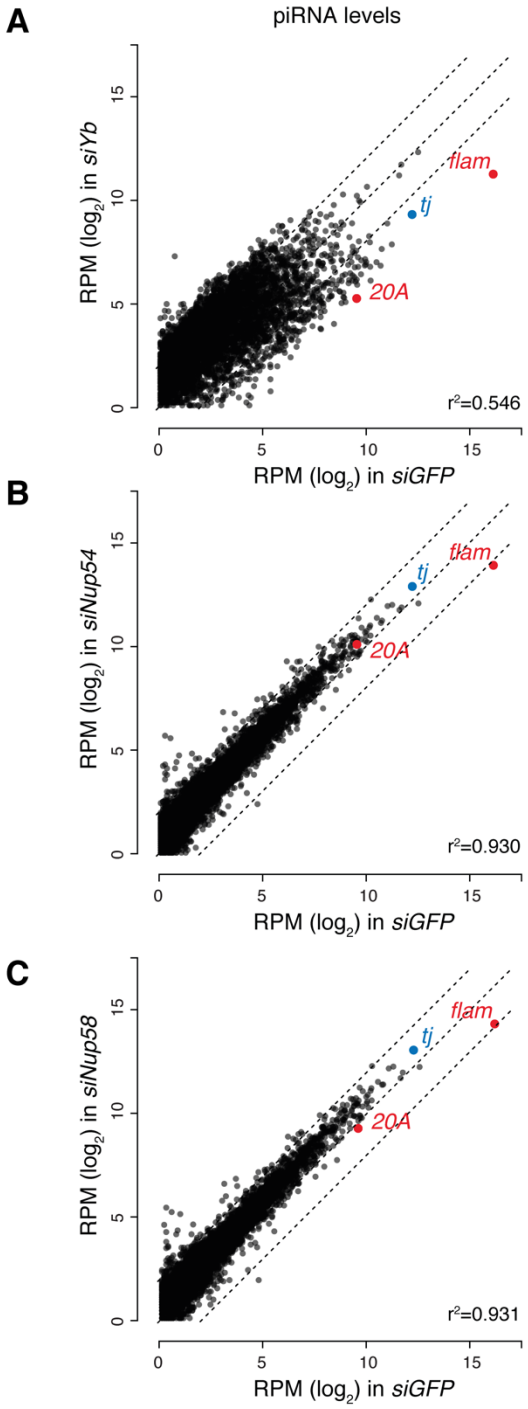


Figure 4.10 Nup54 and Nup58 are specifically required for *flam* piRNA production. A- C) Scatter plots showing expression levels (normalised RPM) of genome-mapped small RNAs in OSCs upon indicated knockdown.

IV.5. Identification of protein interactors of Yb

Aiming to understand the mechanism behind the specificity of Nup54 and Nup58 for *flam* export and specification for piRNA production, I tested three hypotheses. First, the specificity is dictated by protein-protein interactions, with Yb being linked to *flam* and the NPC via one or several adaptor protein(s). Second, Nup54 and Nup58 directly bind to *flam* RNA in the nucleus and promote its export through the NPC. Third, the NPC has undergone a specialisation to couple *flam* transcription and export. In this scenario, Nup54 and Nup58 would be expected to be asymmetrically distributed on the nuclear envelope. Alternatively, the *flam* DNA locus might be associated to the inner side of the NPC via components of the Nup62 complex, with this physical proximity promoting preferential export and processing of newly transcribed *flam.*

To test the first hypothesis, I used IP-MS to analyse the protein partners of full-length Yb and Yb forms lacking specific domains. The Helicase C domain (HelC) has recently been shown to be essential only for *flam* piRNA production and not for that of genic piRNAs (Hirakata et al., 2019). This mutant therefore phenocopies *nup54* and *nup58* knockdown results and might provide useful insights on candidate proteins specific to *flam* export. A potential "exportin" specific to *flam* would be associated with Yb^{WT} but not with $Yb^{\Delta HelC}$. Deletion of the extended Tudor (eTud) domain instead abolishes the interaction between Yb and Armi and causes a complete loss of specificity in piRNA production (Hirakata 2019). I thus expressed each 3xFLAG-tagged construct (Yb^{WT}, $Yb^{\Delta HelC}$ and $Yb^{\Delta eTud}$) in OSCs where the endogenous Yb had been depleted using an siRNA targeting the 3'UTR. Since Yb can also interact with other Yb molecules, depletion of the endogenous protein prevented the mutants from accumulating in pre-existing Yb bodies. In line with what has been previously reported (Hirakata et al., 2019), neither domain mutant was able to form Yb bodies (**Figure 4.11A**, detected with antibody against Yb and against the FLAG tag). Instead, $Yb^{\Delta HelC}$ and $Yb^{\Delta eTud}$ were dispersed in the cytosol and showed substantially lower levels of nuclear Piwi compared to Yb^{WT}. Furthermore, when I assayed the ability to rescue transposon upregulation, $Yb^{\Delta HelC}$ showed an intermediate behaviour between Yb^{WT} and $Yb^{\Delta eTud}$, with the latter completely failing to restore transposon repression, as expected (Hirakata et al., 2019) (**Figure 4.11B**). Having confirmed that the mutants behave as expected, I then carried out an immunoprecipitation against the FLAG tag, using a 3xFLAG-mCherry construct as control, and subjected the samples to Mass Spectrometry.

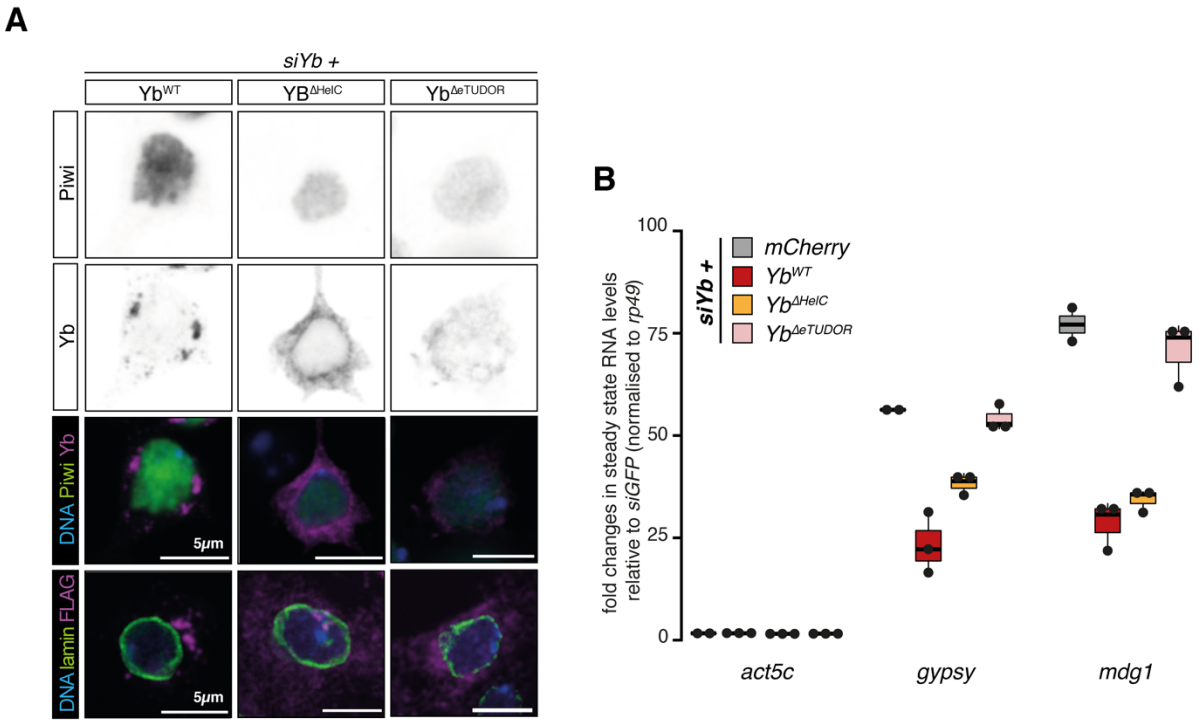


Figure 4.11 **Analysis of Yb domain mutants. A)** Confocal images of Piwi, lamin and 3xFLAG-tagged Yb expression constructs (detected with anti-Yb and anti-FLAG) in OSCs upon endogenous Yb knockdown. Scale bar, 5µm. **B)** Fold changes in the steady-state RNA levels of somatic transposons in OSCs nucleofected with *siYb* and indicated 3xFLAG rescue constructs. Values are relative to *GFP* control knockdown and normalised to *rp49*. (n=3 except for mCherry where n=2).

When I first compared the recovered proteins in Yb^{WT} to the mCherry control, Armi was the most significantly enriched protein and Piwi was also found in the immunoprecipitate (**Figure 4.12A-C**), similarly to Yb PL-MS (**Figure 4.8A**). Both Piwi and Armi interactions were lost in $Yb^{\Delta eTud}$ IP-MS and appeared reduced in $Yb^{\Delta HelC}$. When compared to Yb^{WT}, deletion of the HelC domain caused a less severe change in the interacting proteins, with Armi being the only one exclusively associated to the wild-type bait (**Figure 4.12D**). Conversely, deletion of the eTud domain had a much broader impact on Yb binding partners, with several proteins lost or gained (**Figure 4.12E**). Finally, when the mutants were compared to each other, a specific subset of interacting proteins was detected for either Yb mutant (**Figurec4.12F**). No other piRNA pathway factor, except for Armi and Piwi, nor screen hit was significantly associated with Yb^{WT} and lost in the mutants. Nevertheless, this experiment provided a list of candidate Yb-interacting proteins, whose involvement in transposon regulation and Yb bodies formation will be tested in future experiments. Among promising candidates are CG9684, a Tudor-domain

containing protein, Top2, a topoisomerase reported to interact with heterochromatic regions and with RNA, and Rbp1-like, an mRNA binding protein involved in RNA splicing and RNA metabolism. In particular CG9684 is associated with Yb^{WT} but not with the mutants and its mouse homolog has been linked to the piRNA pathway. Furthermore, the Tudor domain is frequently found in proteins involved in piRNA production and can potentially interact with the sDMA modification of some PIWI proteins (Handler et al., 2011; Vagin et al., 2009a; Vagin et al., 2009b)

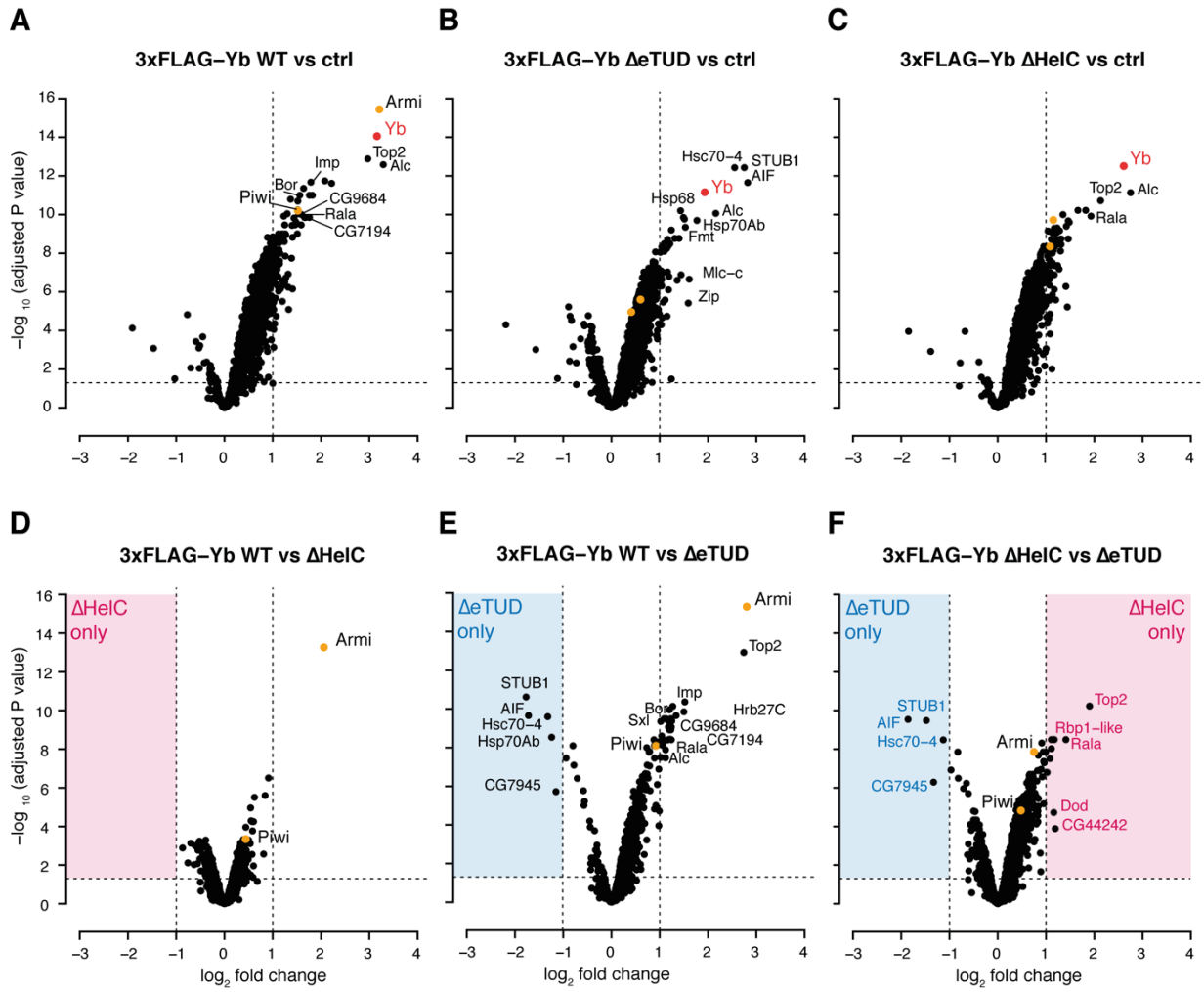
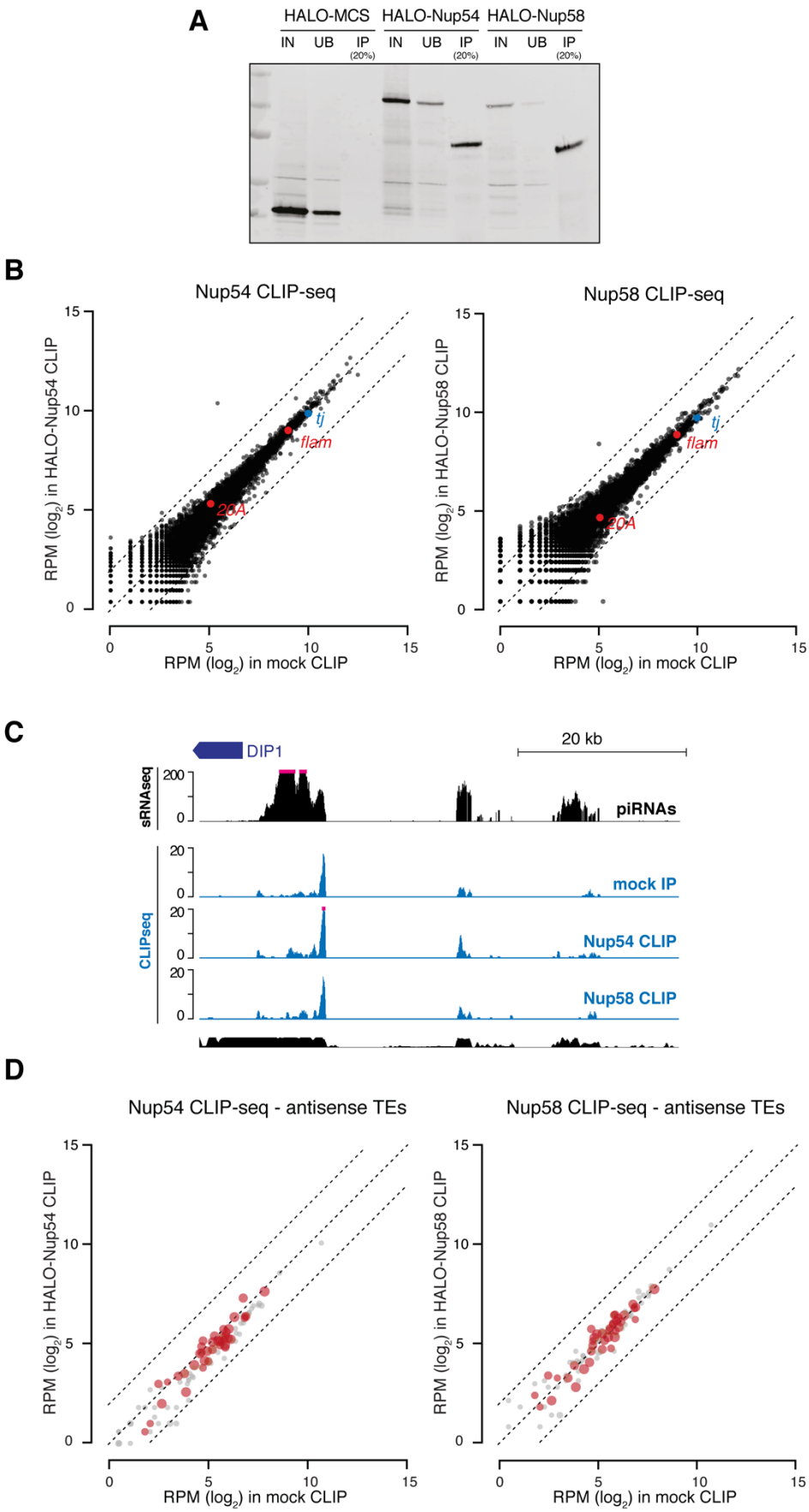


Figure 4.12 Identification of protein partners of Yb wild-type and mutants. A-F) Volcano plots showing enrichment and corresponding significance of biotinylated proteins identified via IP-MS from OSCs expressing the indicated constructs. **A-C)** each Yb construct (n=3) is compared to the mCherry control (n=2); **D-F)** Yb domain mutants are compared to full length Yb protein or to each other. See also supplementary **Tables 9-11.**

IV.6. Nup54 and Nup58 do not directly bind to *flam* RNA

To determine whether Nup54 and Nup58 bind directly to *flam* RNA to facilitate its passage through the NPC, I performed CLIP-seq of HALO-tagged Nup54 and Nup58 expressed in OSCs (**Figure 4.13A**). As a negative control, I used a HALO-3xFLAG empty vector. However, neither Nup54 nor Nup58 CLIP-seq showed significant enrichment of any RNA with respect to the control, except for their own mRNA which is often crosslinked to the bait during CLIP-seq experiments (**Figure 4.13B**). The uni-strand piRNA clusters *20A* and *flam* as well as genic piRNA precursors, exemplified by *tj*, were equally present in the control and the Nups CLIP-seq, as can also be seen from the coverage plot over the first 50 kb of *flam* (**Figure 4.13C**). Inspection of transposon-mapped reads also confirmed that no antisense transposon sequences were enriched in Nup54 and Nup58 CLIP-seq, further confirming that no direct Nup-*flam* interaction occurs. We cannot completely rule out that this negative result is due to a technical issue (i.e. the aminoacidic composition of the Nups is less prone to crosslinking with RNA) and a way to resolve this caveat would be to perform CLIP-seq with a different crosslinking method, such as PFA. However, this result is in line with the current model of RNA export, whereby Nxf1 directly binds to the RNA and mediates its translocation through the NPC via interaction with the FG repeats. Recent data in human and yeast also propose that Nxf1 transiently associates with the NPC and acts as a mobile nucleoporin (Ben-Yishay et al., 2019; Derrer et al., 2019). Furthermore, Yb has also been previously shown to bind to *flam.* Taken together, these data may indicate a model whereby Nxf1 binds to *flam* in the nucleus and likely brings it to the NPC, while Yb associates with *flam* RNA on the cytosolic side of the NPC and directs it to the biogenesis machinery. However, this still does not explain why the knockdown of *nup54* and *nup58* affects piRNA production only from *flam*, leaving all other piRNA precursors unaltered.

Figure 4.13 Nup54 an Nup58 do not directly bind to RNA. A) Representative western blot showing pulldown of the HALO-Nup fusions. **B)** Scatter plot showing expression levels (RPM) of genes in Nup54 or Nup58 CLIP-seq (n=3) against a mock experiment (n=3). piRNA clusters expressed in OSCs are highlighted in red, protein coding genes producing piRNAs in blue. **C)** UCSC genome browser shot displaying Nup54 and Nup58 CLIP-seq and small RNA-seq reads uniquely mapping to the first 50 kb of the piRNA cluster *flam*. The mappability track for 50 bp read length is shown below. **D)** Scatter plots showing expression levels (RPM) of antisense transposon sequences in Nup54 or Nup58 CLIP-seq against a mock experiment. Transposon sequences present in *flam* are highlighted in red with dot size proportional to their abundance within *flam* according to dm6 Repeat Masker annotations.


A
HALO-MCS
HALO-Nup54
HALO-Nup58
IN
UB
IP (20%)
B
Nup54 CLIP-seq
Nup58 CLIP-seq
RPM (log2) in HALO-Nup54 CLIP
RPM (log2) in HALO-Nup58 CLIP
RPM (log2) in mock CLIP
tj
flam
20A
C
DIP1
20 kb
sRNAseq
piRNAs
CLIPseq
mock IP
Nup54 CLIP
Nup58 CLIP
D
Nup54 CLIP-seq - antisense TEs
Nup58 CLIP-seq - antisense TEs

IV.7. Coupled *flam* transcription and export to specialised NPCs?

I finally explored the possibility of a specialised nuclear pore that would couple transcription and export of *flam*. We envisioned that, as an unusually long transcript whose production would take more than 5 hrs (with the canonical speed of RNA pol II being ~25 nt/sec) (Fukaya et al., 2017), the export of *flam* might initiate even before RNA pol II has reached the end of the genomic locus. In this scenario, Nup54 and Nup58 might be asymmetrically distributed in the nuclear envelope and primarily engage in *flam* export. Alternatively, or in addition to that, the DNA locus of *flam* might be preferentially localised in proximity to NPCs near Yb bodies while Nup54 and Nup58 might regulate its expression. In support of this hypothesis, several recent reports have found some NPC components to be directly involved in gene regulatory networks via direct binding to chromatin (Breuer and Ohkura, 2015; Gozalo et al., 2020; Iglesias et al., 2020; Kalverda and Fornerod, 2010).

To test this model, I first sought to determine the subcellular distribution of Nup54 and Nup58 in OSCs. Despite trying various tags and permeabilization methods, I was unable to detect a strong and reproducible signal for any expression construct, likely due to low expression levels and/or to inaccessibility of the tag within the NPC. Therefore, I decided to exploit proximity biotinylation, which successfully identified the interacting partners of Nup54 and Nup58 (**Figure 4.8B-C**), as a proxy for the distribution of both proteins within the cell. I expressed the two Nups amino-terminally fused to the TurboID biotin ligase, which has faster kinetics than BASU (Branon et al., 2018), then treated the cells with biotin for 1 hour and proceeded with a standard immunofluorescence protocol. Even though the HA tag did not show any discrete signal above background (represented by untransfected cells in **Figure 4.14**), the streptavidin signal was readily detectable and specific to the samples expressing a TurboID fusion protein (**Figure 4.14**). Strikingly, the Streptavidin signal generated by TurboID-Nup54 and TurboID-Nup58 was not an even ring surrounding the nucleus, but instead showed one or more aggregates adjacent to the nuclear envelope. This staining pattern was reproducible and led us to hypothesize that those granules might be adjacent or overlapping Yb bodies. Therefore, I expressed the TurboID fusion proteins and co-stained for Streptavidin and Yb. In contrast to our expectations, this revealed different scenarios, shown in **Figure 4.15A**. In some instances, Yb bodies and Nup54/Nup58 aggregates perfectly overlapped (**Figure 4.15A**, upper panel), but this was not always the case. The main signal for Nup54 and Nup58 was often distinct from the strongest Yb staining, but in most cases a dim signal for Yb was always detected in association with the Nup granules (arrowheads in **Figure 4.15A**). In addition to that, I noticed that a very

similar staining pattern could be observed for TurboID-Nxf1 (**Figure 4.15B**), and here the signal also did not always overlap with Yb.

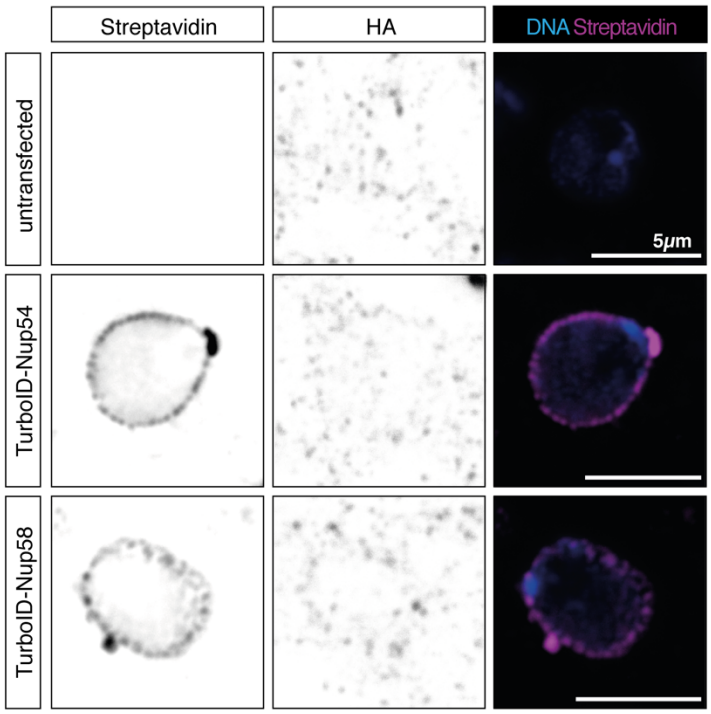


Figure 4.14 Subcellular localisation of TurboID-Nup54 and Nup58 in OSCs. Confocal images of HA and Streptavidin in OSCs transfected with HA-TurboID-tagged Nups. Scale bar, 5µm.

Since using proximity biotinylation as a proxy for the Nup localisation might generate artefacts due to construct overexpression and/or local accumulation of biotinylated proteins, I searched for an alternative approach to detect the endogenous FG-Nups. While no commercial antibodies that recognise fly Nup54 and Nup58 are available, I opted for a mouse monoclonal antibody raised against FG repeat-containing peptides from several Nups (mAb414, labelled as "FG Nups" in figure panels). Although not specific to Nup54 and Nup58, this approach has the advantage of detecting endogenous proteins, thus overcoming concerns about overexpression, and has been previously used against *Drosophila* Nups (Hampoelz et al., 2019). Furthermore, since the Nup62 complex forms a large fraction of the total FG repeat mass within the pore, we assumed that most of the observed signal originates from those. Staining for the FG Nups closely resembled what had been observed with the TurboID constructs, though with less prominent granules (**Figure 4.15C**), in line with them being enhanced by the accumulation of biotinylated proteins over time. As observed with TurboID, there was no consistent co-localisation between FG Nup granules and Yb bodies. For increased resolution of the FG Nup

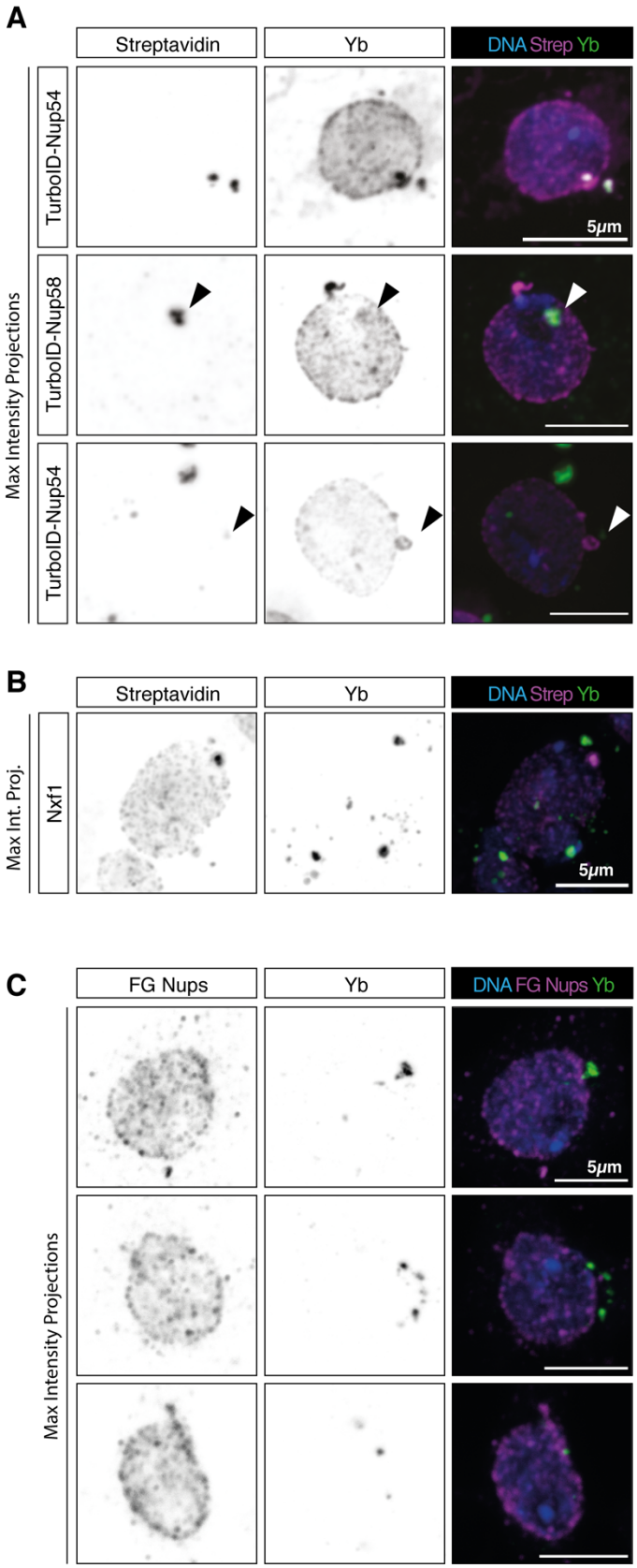


Figure 4.15 Subcellular localisation of Nups and Yb in OSCs. A-B) Confocal images of Yb and Streptavidin in OSCs transfected with HA-TurboID-tagged constructs. **C)** Confocal images of Yb and FG Nups in OSCs. Scale bars, 5μm.

aggregates in relation to Yb bodies, I also exploited STED microscopy, which has been successfully applied to NPCs (Thevathasan et al., 2019). Some FG Nup granules often appeared detached from the nuclear envelope and associated with Yb (arrowheads in **Figure 4.16**), although the main Yb bodies were not always in the vicinity of these granules, as described previously (**Figure 4.15**).

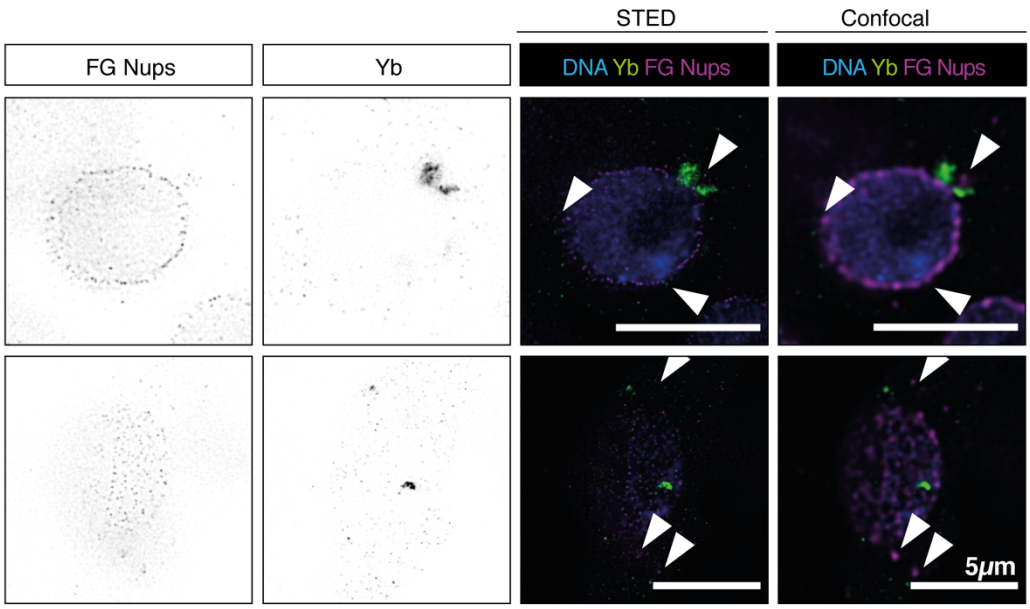


Figure 4.16 Subcellular localisation of Nups and Yb in OSCs. STED microscopy of Yb and FG Nups in OSCs. Respective confocal images are shown on the right for comparison.

A similar staining pattern for the FG Nups was also observed in the follicle cells of the ovary (arrowheads in in **Figure 4.17**). Interestingly, the FG Nups aggregates were more prominent in *flam* homozygous mutants (*flam*$^{BG/BG}$ in **Figure 4.17**). This might suggest that *flam*-dependent assembly of Yb bodies potentially influences NPC assembly as well, although we cannot exclude indirect effect of aberrant ovarian development in *flam*$^{BG/BG}$. In the *Drosophila* ovary, especially in nurse cells, NPCs are assembled from cytosolic Nup aggregates that fuse into *annulatae lamellae* and are deposited into the developing egg (Hampoelz et al., 2019). We therefore speculate that Yb might associate with growing FG Nup aggregates that have not yet been integrated into mature NPCs on the nuclear envelope. If this is the case, the observed cytosolic co-localisation between Yb and the Nups may represent an early stage of NPC biogenesis and contributes to bringing Yb to the NPCs where *flam* is exported. Since no direct interaction between Yb and either Nup54 or Nup58 was detected, this process might be mediated by other Nups present in the granules or by a yet unidentified adaptor protein.

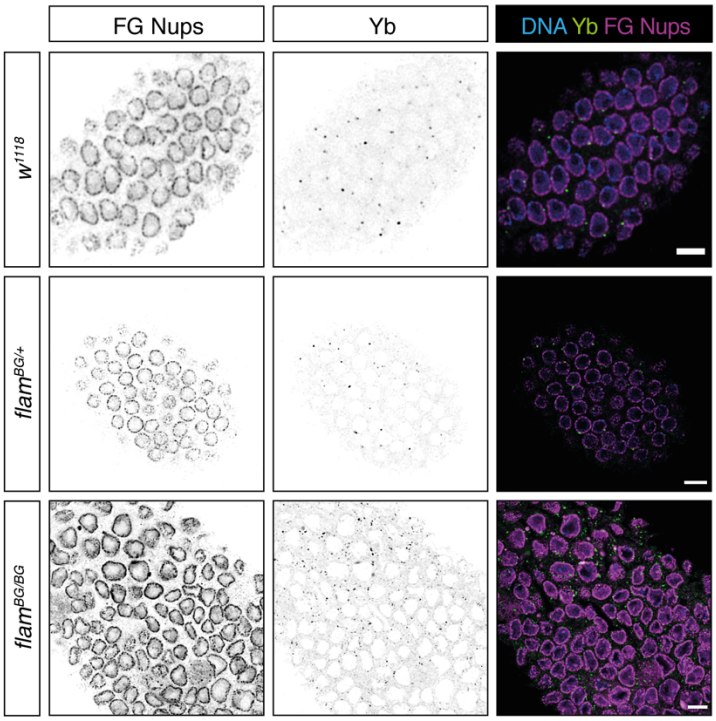


Figure 4.17 Subcellular localisation of Nups and Yb in follicle cells. Confocal images of Yb and FG Nups in ovaries of indicated genotypes. Scale bar, 5µm.

Finally, I set out to determine the physical localisation of the *flam* locus via DNA FISH in OSCs. If Nup54 and Nup58 directly couple transcription and export of *flam*, its DNA locus could be localised adjacent to the cytosolic Yb bodies. Because *flam* is a very repetitive region with poor mappability, I designed a set of probes against a unique region spanning *DIP1* and the first exon of *flam* (**Figure 4.18A**). Staining with these probes identified one or two foci for each cell, in most cases localised to the periphery of the nucleus and proximal to the nuclear envelope (**Figure 4.18B**), which is often the case for heterochromatic regions. As expected, DNA FISH foci were insensitive to RNase A treatment (always included in the staining protocol) but disappeared upon treatment with DNase I (**Figure 4.18B**). I then stained for *flam* DNA, Yb and the FG nups but did not identify any direct correlation between the nuclear position of *flam* and that of Yb bodies and FG Nup aggregates (**Figure 4.18C**). This is in line with previous reports failing to detect any correlation between the position of the genomic *flam* locus in the nucleus and that of Yb bodies in the cytosol (Murota et al., 2014) and is consistent with a model whereby *flam* RNA is transported to Yb bodies after its export from the NPC. However, this does not exclude a direct contact between the NPC and the *flam* genomic locus.

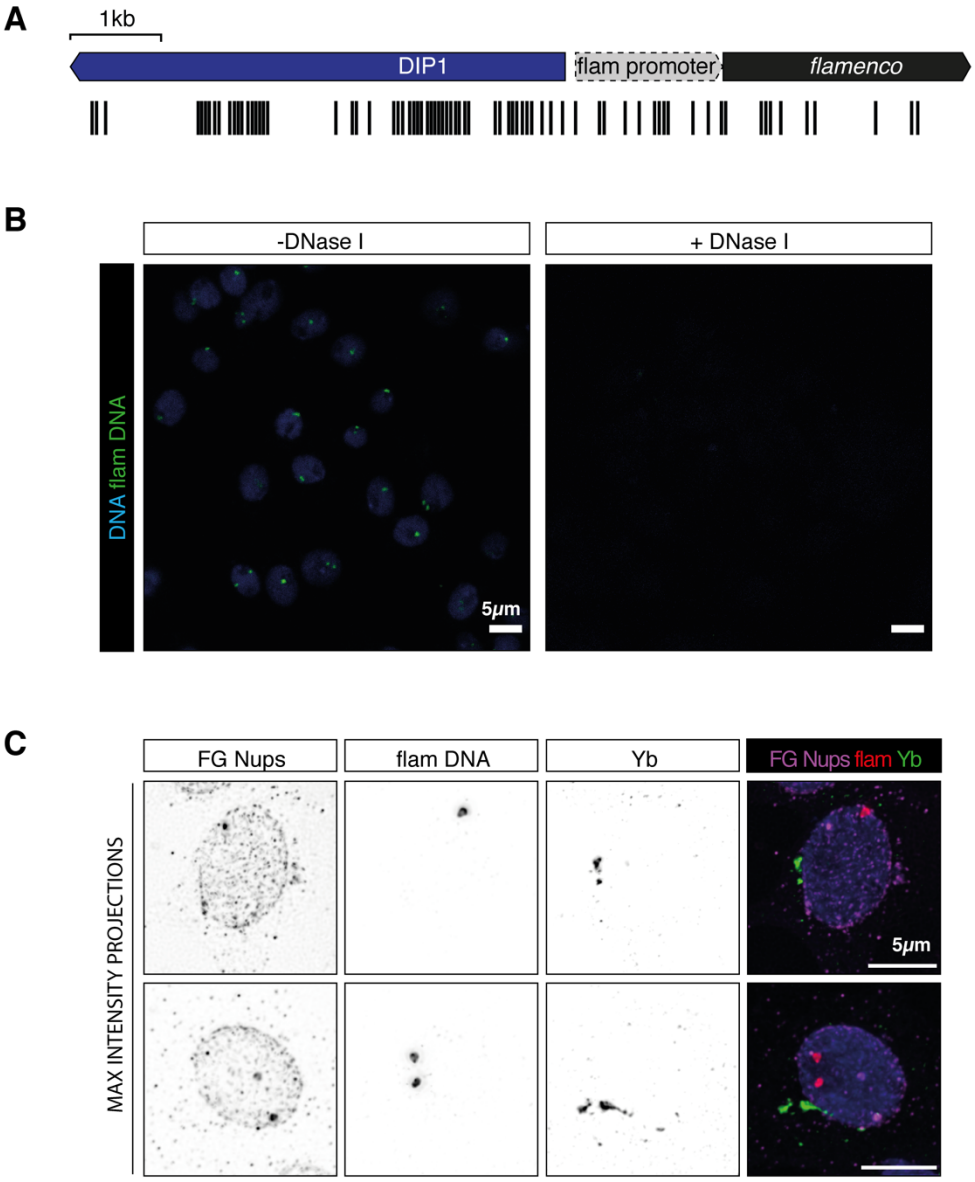


Figure 4.18 ***flam* locus DNA FISH. A)** Genomic distribution of the primary probes used to detect the genomic *flam* locus. **B)** Confocal images of *flam* DNA FISH in OSCs treated or not with DNase I. Scale bar, 5µm. **C)** Confocal images of Yb, FG Nups and *flam* DNA in OSCs. Scale bar, 5µm.

CONCLUSION

Work described in this chapter uncovers an unexpected involvement of Nuclear Pore Complex subunits in piRNA-mediated transposon control. We find that Nup54 and Nup58, previously identified as strong soma-specific piRNA pathway screen hits (Handler et al., 2013), participate to the production of piRNAs from the uni-strand cluster *flam*. Unlike Yb, which appears to dictate the recognition of piRNA biogenesis substrates on the cytosolic side of the NPC, loss of *Nup54* and *Nup58* only affects *flam* export and processing into piRNAs.

There are several possible scenarios to explain their remarkable specificity and, although the molecular mechanism remains elusive, we can exclude some hypotheses and speculate on others. For clarity of the representation, Nxf1/Nxt1 heterodimers bound to *flam* have been omitted from the models in **Figure 4.19**. First, Yb is not an integral component of the NPC, as we failed to identify any direct and stoichiometric interaction between Yb and the Nups (**Figure 4.19A**). However, Yb is at least transiently coming into contact with NPC components, as shown by PL-MS experiments and by its peri-nuclear localisation. Whether an adaptor protein stabilises this transient Yb-NPC interaction and promotes *flam* export remains an open question (**Figure 4.19B**). The specificity is also not dictated by a direct Nup-*flam* RNA interaction, thus excluding such a model as well (**Figure 4.19C**). The effect of *yb, nup54* and *nup58* knockdown on *flam* transcript levels strongly argues for a co-transcriptional export model in which the NPC might have specialised together with this uni-strand cluster. If any such coupling of transcription and export exists, then it is not due to mere physical proximity of *flam* DNA and Yb bodies, separated by a "specialised" NPC (**Figure 4.19D**). However, what I can hypothesize based on our data and the published literature (Breuer and Ohkura, 2015; Gozalo et al., 2020; Iglesias et al., 2020) is that Nup54, Nup58 or Nup93-1, also a piRNA pathway screen hit and reported to interact with chromatin, directly bind to the *flam* locus and assemble a machinery that is capable to promote *flam* RNA export while it is still being transcribed (**Figure 4.19E**). In this scenario, disruption of *flam* export would also affect the transcriptional output from this unusually long cluster. Once exported, *flam* RNA would be immediately bound by Yb, transiently associated to the cytosolic side of the NPC, and stored in Yb bodies until further processing takes place (**Figure 4.19E**). Finally, Yb might participate to the assembly of specialised NPCs starting from cytosolic FG Nup aggregates, as reported (Hampoelz et al., 2019), although the lack of direct interaction between Yb and the Nups presumes the existence of additional players (**Figure 4.19F**).

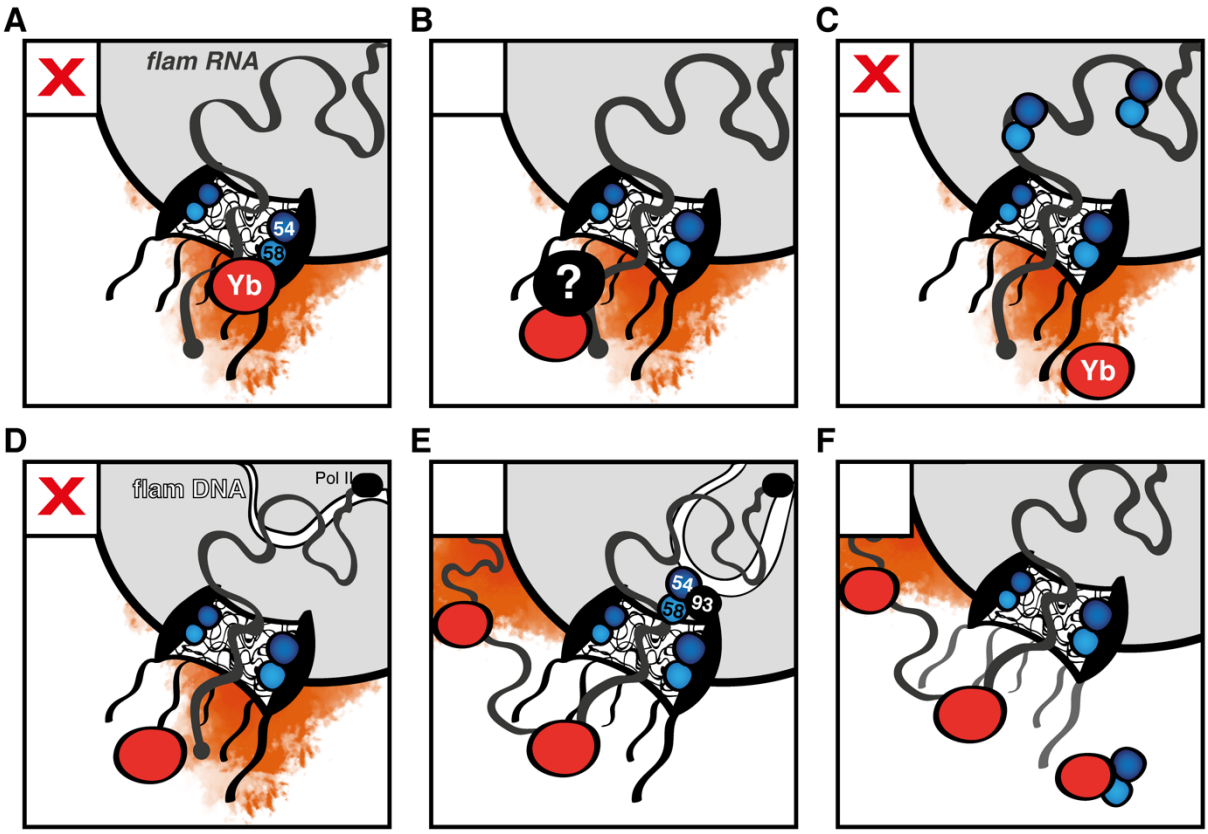


Figure 4.19 Possible models for *flam* export. A) Yb is an integral component of the NPC and directly associates with Nup54 and Nup58. **B)** Yb transiently associates with the NPC and promotes *flam* export thanks to an adaptor protein. **C)** Nup54 and Nup58 directly bind to *flam* RNA to facilitate its export through the NPC. **D)** The *flam* DNA locus is directly opposite Yb bodies, on the inner side of the nuclear envelope. **E)** Components of the Nup62 complex directly bind to *flam* DNA locus to couple its transcription and export. Once exported, *flam* RNA is then directed to Yb bodies. **F)** Yb participates to the assembly of Nup54 and Nup58 into specialised NPCs that are capable of exporting *flam*.

CHAPTER V

piRNA-guided transposon silencing co-opts nuclear export factors

Author contributions

This chapter has been published in similar form as Fabry, M.H.*, Ciabrelli, F**.*, Munafò, M.***, Eastwood, E.L., Kneuss, E., Falciatori, I., Falconio, F.A., Hannon, G.J., and Czech, B. (2019). piRNA-guided co-transcriptional silencing co-opts nuclear export factors. ELife *8*, e47999.

*** equal contribution**

The project was conceived by B.C. and G.J.H., with inputs from M.H.F., F.C., myself and E.L.E.. I identified the PICTS complex via IP-MS. M.H.F. performed RNA-seq and ChIP-seq from OSCs and performed all computational analyses of sequencing data, except for CLIP- and RIP-seq. M.H.F. performed rescue experiments, DNA tethering assays and ChIP-seq from DNA and RNA tethering samples. M.H.F. and myself performed coIP and IF experiments, respectively, to characterise the PICTS complex in S2 cells. F.C. performed RNA-seq and ChIP-seq from ovaries, characterised the genetic dependencies within PICTS and performed qPCR from tethering experiments. E.L.E. generated *nxf2* mutant flies and carried out their characterisation. I performed small RNA-seq, CLIP-seq, RIP-seq and RNA tethering experiments and analysed CLIP-seq and RIP-seq data. M.H.F. performed ChIP-seq and RNA FISH from RNA tethering experiments. E.K. generated GFP-Nxf2 knock-in flies. F.A.F. generated expression constructs. I.F. conducted early work on Panx and Nxf2 interaction. B.C. and G.J.H. supervised the project and wrote the first draft of the manuscript. All authors edited and proofread the final manuscript.

BACKGROUND

Small RNA-guided epigenetic silencing is a very conserved mechanisms which can be found in animals and plants and can be used to regulate gene expression and dampen transposon activity (Holoch and Moazed, 2015). In *Drosophila*, this consists in piRNA-guided deposition of repressive histone marks along transposon bodies, primarily H3K9me3, followed by compaction into heterochromatin (Le Thomas et al., 2013; Rozhkov et al., 2013; Sienski et al., 2012; Wang and Elgin, 2011). piRNA-dependent epigenetic silencing targets active elements and is initiated during transcription, hence it is commonly referred to as "co-transcriptional gene silencing" (TGS). Piwi in complex with an antisense piRNA is thought to engage with complementary, nascent transposon RNAs to then recruit downstream effectors (Donertas et al., 2013; Muerdter et al., 2013; Ohtani et al., 2013; Sienski et al., 2015; Sienski et al., 2012; Yu et al., 2015). Previous studies have identified Panx as a critical effector of TGS, as its recruitment to a nascent RNA is necessary and sufficient to induce epigenetic silencing of the corresponding locus (Sienski et al., 2015; Yu et al., 2015). Panx does not have any domains of known function and no direct homolog has yet been identified in vertebrates. In *Drosophila*, the deposition of repressive H3K9me3 marks at transposon loci is catalysed by dSETDB1/Eggless, a conserved histone methyltransferase acting downstream of Piwi and Panx. The histone demethylase dLsd1/Su(var)3-3 has instead been implicated in removing H3K4me2/3 from target loci to facilitate the establishment of a repressive state (Iwasaki et al., 2016; Rangan et al., 2011; Sienski et al., 2015; Wang and Elgin, 2011; Yu et al., 2015). Notably, both Egg and dLsd1 are ubiquitously expressed and are general chromatin silencing factors whose activity must be directed towards the desired targets. Current knowledge places Panx at a critical node of the TGS response, bridging Piwi-mediated target recognition to the downstream general chromatin silencing machinery (Sienski et al., 2015; Yu et al., 2015). Nonetheless, how Panx recruits these histone-modifying enzymes and what other factors are involved in the process is an outstanding question.

RESULTS

V.1 Identification of the PICTS complex

Aiming to deepen our understanding of piRNA-guided TGS, we sought to identify what proteins associate with Panx *in vivo*. I therefore established a protocol to perform immunoprecipitation (IP) from ovary lysates of a GFP-tagged Panx transgene expressed under its endogenous regulatory elements (Handler et al., 2013). As a negative control, I performed the same IP from a fly line not expressing GFP (**Figure 5.1A**) and the isolated protein complexes were subjected to quantitative Mass Spectrometry (IP-MS) (for a detailed protocol see **Materials and Methods II.13.1 and II.13.3**) (Papachristou et al., 2018). Three proteins showed a striking enrichment and significance in GFP-Panx IP over the control, these were Panx itself, Nxf2 and Nxt1 (**Figure 5.1B**). Interestingly, both Nxf2 and Nxt1 are related to the family of RNA nuclear export factors and were strong hits in genetic screens for genes involved in transposon control (Czech et al., 2013; Handler et al., 2013). While Nxt1 is ubiquitously present in the adult fly, with the highest expression levels in the female gonads, Nxf2 is almost exclusively expressed in the ovary (modENCODE et al., 2010).

Nxf2 belongs to the evolutionarily conserved Nuclear Export Factor (NXF) family, which in *Drosophila* comprises four members: Nxf1/sbr, Nxf2, Nxf3 and Nxf4 (**Figure 5.1C**). Nxf1 is essential for bulk mRNA export from the nucleus, similarly to its yeast homolog Mex67p (Stutz and Izaurralde, 2003), whereas the other three appear dispensable in S2 cells (Herold et al., 2001; Herold et al., 2000). Proteins of the NXF family share a similar domain organisation (**Figure 5.1C**). Starting from the amino-terminus, they possess an RNA-binding domain (RBD), followed by Leucine-Rich Repeats (LRR), a region structurally similar to Nuclear Transport Factor 2 (NTF2) and a Ubiquitin-Associated domain (UBA). While the amino-terminal region of NXF proteins generally is responsible for cargo binding, the NTF2 and UBA domains are involved in contacting the FG repeats of the NPC (Braun et al., 2001; Fribourg et al., 2001; Grant et al., 2002; Levesque et al., 2001). Nxf2 possesses all of these domains and, compared to Nxf1, has an extended amino-terminus (NTR) and one additional set of LRR upstream of the RBD. Nxt1, also known as p15, has instead been described to hetero-dimerise with the NTF2 fold of NXF family proteins to assemble one functional and structural unit (Herold et al., 2001). The role of the Nxf1/Nxt1 heterodimer in mRNA export is highly conserved in metazoans, whereas yeast Mex67p hetero-dimerizes with Mtr2p, whose sequence appears unrelated to Nxt1 despite performing a similar function (Stutz and Izaurralde, 2003).

Finally, *Drosophila* Nxt1 has been previously reported to interact with Nxf2, likely via its NTF2 fold (Herold et al., 2001; Herold et al., 2000).

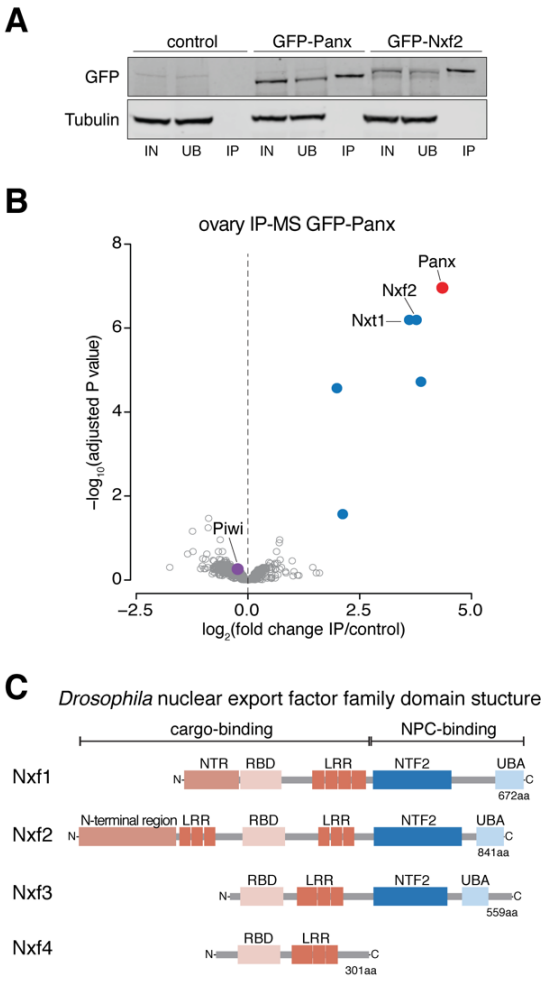


Figure 5.1 Panx interacts with nuclear export factors. A) Western blot showing a representative replicate of the immunoprecipitation of GFP-Panx and GFP-Nxf2 from ovary lysates that was submitted for Mass Spectrometry. IN=input; UB=unbound; IP=immunoprecipitate. **B)** Volcano plot showing enrichment values and corresponding significance levels for proteins co-purified with GFP-Panx from ovary lysates (n=4 for GFP-Panx and n=3 for control ovaries). Proteins with fold change > 2 and adjusted P value < 0.05 are highlighted in blue. The bait protein is labelled in red and Piwi in purple. See also supplementary **Table 12. C)** Cartoon displaying the *Drosophila* NXF family domain structure. NTR, amino-terminal region; LRR, leucine rich repeats; RBD, RNA-binding domain; NTF2, NTF2-like domain; UBA, Ubiquitin associated domain.

Since little was known about the function of Nxf2, we used CRISPR/Cas9 to generate knock-in flies expressing a GFP-Nxf2 fusion protein from the endogenous *nxf2* locus. GFP-Nxf2 predominantly localises to the nuclei of germline and somatic cells of the ovary (**Figure 5.2A**, top panel), closely resembling the localisation of GFP-Panx (**Figure 5.2A**, bottom panel). IP-

MS of GFP-Nxf2 knock-in flies confirmed that Nxf2, Nxt1 and Panx form a nuclear complex *in vivo* (**Figure 5.2B**). Given the established role of Panx in TGS, we hypothesized that this complex is constantly at war against invading transposons and therefore named it PICTS (Panx-induced co-transcriptional silencing) complex, after the warrior tribes of northern Scotland that valiantly fought the Roman intruders.

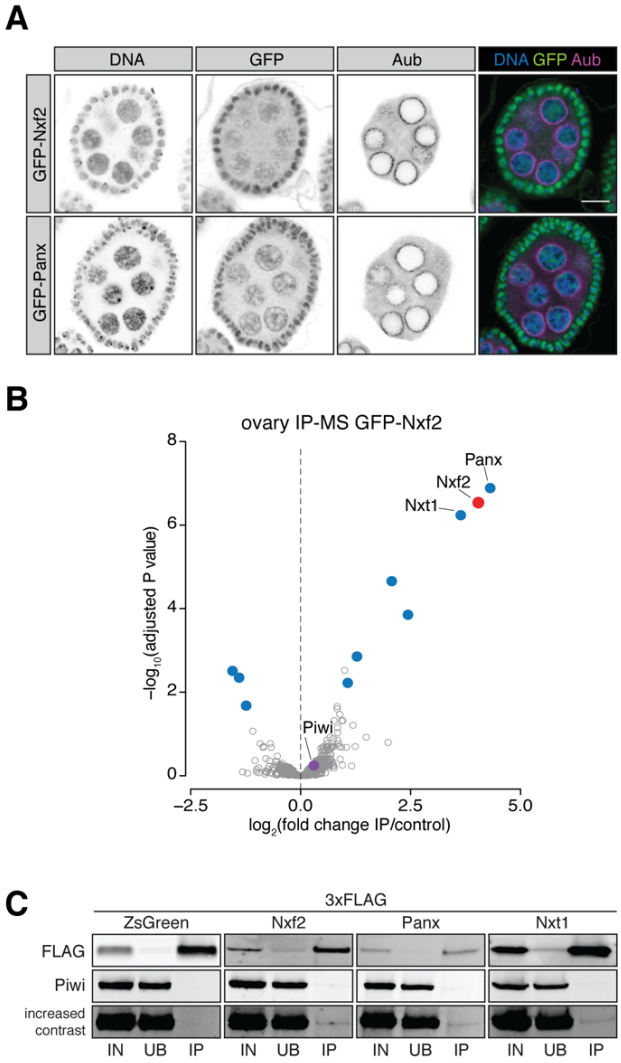


Figure 5.2. Nxf2 is a nuclear protein that interacts with Panx. A) Confocal images of GFP-Nxf2, GFP-Panx and Aub in ovaries. Scale bar, 10µm. **B)** Volcano plot showing enrichment values and corresponding significance levels for proteins co-purified with GFP-Nxf2 from ovary lysates (n=4 for GFP-Nxf2 and n=3 for control ovaries). Proteins with fold change > 2 and adjusted P value < 0.05 are highlighted in blue. The bait protein is labelled in red and Piwi in purple. See also supplementary **Table 13. C)** Western blot of FLAG-tagged immunoprecipitations from OSC lysates transfected with indicated constructs. The bottom panel shows Piwi signal with increased contrast.

Unexpectedly, and in contrast with previous findings (Sienski et al., 2015; Yu et al., 2015), GFP-Panx IP-MS did not show an enrichment of Piwi, nor did GFP-Nxf2 IP-MS (**Fig 5.1B, 5.2B**). These results suggest that Piwi is not an integral component of the PICTS complex but leaves the questions of whether any direct interaction occurs and how PICTS is recruited to active transposons. I therefore expressed FLAG-tagged Panx, Nxf2, Nxt1 and a ZsGreen control in OSCs, performed an anti-FLAG IP and probed for the presence of endogenous Piwi in the immunoprecipitate via western blot. These data revealed a weak association between Piwi and the components of PICTS, but not with the negative control (**Figure 5.2C**). If Piwi only recruits PICTS to nascent transposon mRNAs which are mostly absent in an unperturbed cellular context, it follows that only a tiny fraction of Piwi will be directly binding to PICTS. Hence, such rare and transient binding events might be below the detection limit of Mass Spectrometry but still detectable at low levels via western blot.

Finally, IP-MS experiments did not identify any significant enrichment of chromatin silencing factors such as Egg or dLsd1, again suggesting that their interaction with PICTS might be labile and below detection limit. However, we noticed that one other protein was significantly enriched in both IP-MS for GFP-Panx and GFP-Nxf2: cut-up (ctp), a dynein light chain protein that is reported to interact with Egg (Stabell et al., 2006). Ongoing work in the lab is investigating whether ctp is acting as a link between PICTS and the chromatin silencing machinery.

V.2 Nxf2 is a piRNA pathway factor that functions in Transcriptional Gene Silencing

Nxt1 is known to contribute to bulk mRNA export and its depletion causes dramatic gene mis-regulation (Herold et al., 2003). As described in **Chapter IV**, its depletion in OSCs severely impaired cell viability, thus precluding further functional studies. On the other hand, Nxf2 is dispensable for mRNA export (Herold et al., 2001; Herold et al., 2003) and its depletion in the germline leads to TE upregulation without affecting the overall ovarian morphology (Czech et al., 2013). Therefore, we focused on investigating the role of Nxf2 in the piRNA pathway and generated null mutants with CRISPR-Cas9, using two gRNAs directed towards the beginning of the coding sequence. We recovered two null alleles, $nxf2^{F10*}$ and $nxf2^{\Delta 1*}$, harbouring premature stop codons that disrupt the coding sequence of Nxf2 from amino acid 10 onwards (**Figure 5.3A**). The deletion was confirmed via genotyping PCR with primers spanning the gRNA target region, followed by sequencing of the PCR product. Given the strong association between Nxf2 and Panx and the established role of the latter in TGS, throughout the following experiments we compared *nxf2* null alleles to *panx* null alleles, namely $panx^{M1}$ and $panx^{M4}$ (Yu

et al., 2015). For all subsequent analyses, the two mutant alleles of each gene have been crossed and the resulting trans-heterozygous F1 flies (e.g. *nxf2*$^{F10*/\Delta1}$ and *panx*$^{M1/M4}$, hereafter referred to as homozygous mutants) have been compared to their heterozygous siblings, which are a mixed population carrying either of the two alleles over a balancer chromosome.

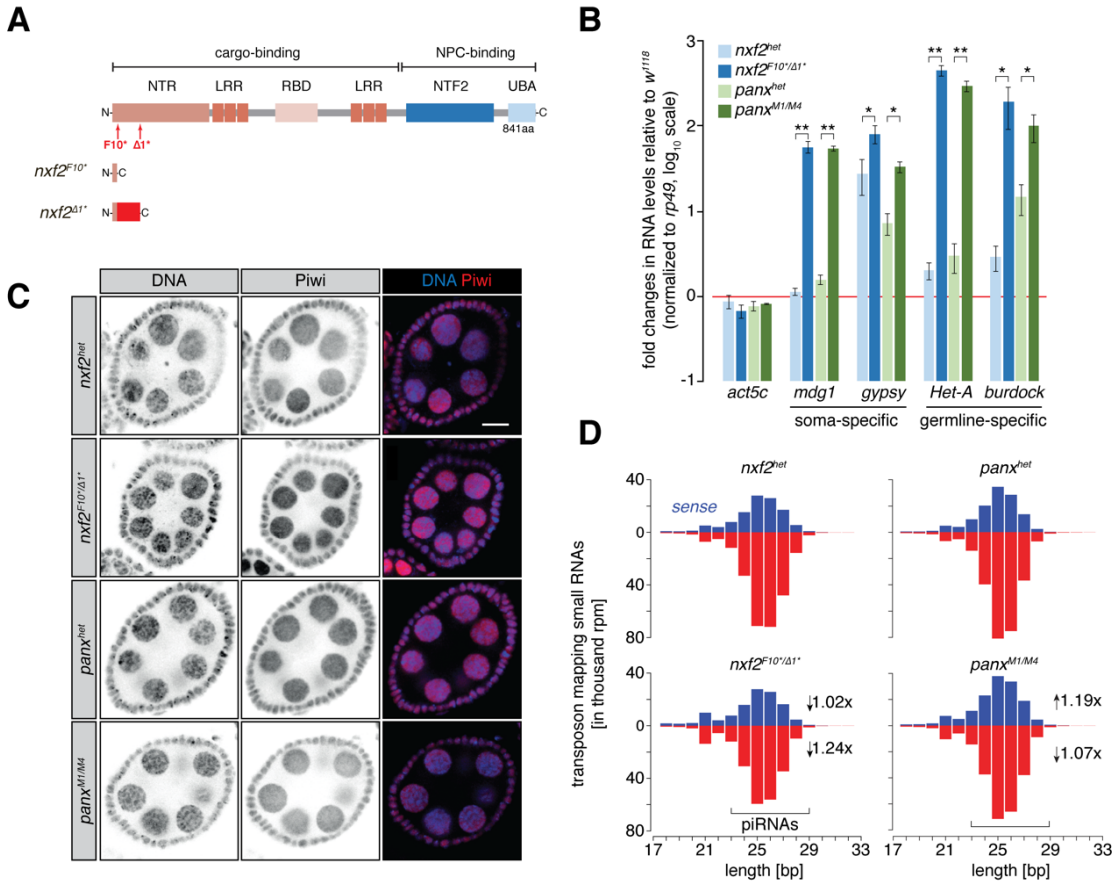


Figure 5.3. Nxf2 is a piRNA pathway factor that functions in TGS. A) Cartoon displaying the domain structure of Nxf2 and of the two mutant alleles. Highlighted in red is the portion of the CDS that goes out of frame. **B)** Fold changes in steady-state RNA levels of the indicated soma- and germline-specific transposons from ovaries. Values are relative to *w*^{1118} flies and normalized to *rp49*. * = P value < 0.05; ** = P < 0.001 (unpaired t-test). Error bars indicate standard deviation (n=3). **C)** Confocal images of Piwi in ovaries. Scale bar, 10µm. **D)** Size distribution of transposon-mapping small RNAs from ovaries of indicated genotypes. Sense reads are shown in blue, antisense in red.

Nxf2 homozygous flies are female sterile, a hallmark of all mutants in which the piRNA pathway is impaired. qPCR analysis of RNA isolated from *nxf2* mutant ovaries showed strong de-repression of soma- (*mdg1* and *gypsy*) and germline-specific (*HeT-A* and *burdock*) transposable elements in the homozygous mutants (**Figure 5.3B**) with respect to control flies (w^{1118}). The extent of transposon de-repression was comparable to that observed in *panx* homozygous mutants. This indicates that Nxf2 participates to transposon repression in both compartments of the ovary. To rule out a possible involvement of Nxf2 in the production of piRNAs, we performed immunofluorescence staining for Piwi in ovaries from *nxf2* and *panx* mutant flies (**Figure 5.3C**). As expected from a factor involved in TGS, and therefore acting downstream of Piwi target engagement, Piwi levels in nuclei were unaffected by loss of either component of PICTS. I also sequenced small RNAs isolated from ovaries of *panx* and *nxf2* mutant flies (**Figure 5.3D**). This showed that *panx* and *nxf2* homozygous mutants have essentially identical small RNA levels as the respective heterozygous control, thus proving that Nxf2 functions downstream of piRNA production. Taken together, these data point towards an involvement of Nxf2 in TGS, since its loss causes transposons upregulation in both tissues of the ovary, and rule out a role in germline piRNA cluster biology or piRNA biogenesis.

RNA-seq analysis from mutant flies and from OSCs depleted of *piwi, panx* and *nxf2* demonstrated that loss of PICTS components does not globally alter the expression levels of protein-coding genes but almost exclusively affects transposon levels (Fabry et al., 2019). Since TGS ultimately results in the deposition of the repressive histone mark H3K9me2/3 on transposon loci, we profiled its distribution genome-wide via Chromatin Immuno-Precipitation and sequencing (ChIP-seq). In the absence of Panx and Nxf2, H3K9me3 was completely lost from transposons, leading to their transcriptional reactivation (Fabry et al., 2019). H3K9me3 was specifically lost only from loci targeted by Piwi, whereas its genome-wide distribution was essentially unchanged. This further indicates that neither Nxf2 nor Panx are involved in global establishment of heterochromatin but only act on transposon loci targeted by Piwi. In all aforementioned experiments, the effects caused by loss of *panx* and *nxf2* showed a high degree of similarity, further strengthening the notion that they act together as part of the PICTS complex.

V.3 Characterisation of the interactions within PICTS

Components of multi-protein complexes often depend on each other for their respective localisation and stability. We noticed that transient transfection of Panx constructs in S2 cells, which lack an active piRNA pathway, or OSCs, which express relatively low levels of Panx and Nxf2, consistently resulted in only moderate expression of the transgene (an example of which can be seen in **Figure 5.9B**). This was in stark contrast with proteins of similar molecular weight and expressed from the same vector backbone, and thus led us to hypothesize that Panx might depend on Nxf2 for its stability. To address this question, I carried out a western blot on ovary lysates from *panx* and *nxf2* mutants and probed for endogenous Panx (**Figure 5.4A**). As expected, *panx* heterozygous and homozygous mutants show diminished and null levels of Panx, respectively, when compared to w^{1118} controls. Interestingly, Panx levels are also dramatically affected upon loss of Nxf2, with *nxf2* homozygous mutants only retaining approximately 15% of Panx protein (**Figure 5.4A**). An even stronger effect was observed when the GFP-Nxf2 allele was expressed in a $panx^{M1}$ mutant background, with the GFP signal completely absent in $panx^{M1/M1}$ (Fabry et al., 2019). Hence, Panx and Nxf2 are interdependent for their protein stability. This prompted us to investigate if Panx and Nxf2 depend on each other also for their subcellular localisation and which domains mediate the interaction.

To this end, we generated HA- and 3xFLAG-tagged constructs of Panx and Nxf2, either full-length or carrying domain deletions or point mutations, and expressed them in various combination in S2 cells. To characterise the domains required for Panx and Nxf2 interaction, we carried out IP of the FLAG-tagged bait and probed via western blot for the presence of the HA-tagged prey in the immunoprecipitate (Fabry et al., 2019) . To test for their respective subcellular localisation, I performed an immunofluorescence (IF) from the same samples against HA and FLAG tags and against Lamin. In all following IF panels, the Lamin signal has been used to draw a red outline that depicts the nuclear envelope.

We first generated HA-tagged constructs of Panx full-length and split into two, by removing either the amino- or the carboxy-terminal portion (FL, ΔN and ΔC, respectively; schematic in **Figure 5.4B**) and co-expressed them with 3xFLAG-tagged Nxf2. As expected, full-length Panx and Nxf2 strongly co-immunoprecipitated and co-localised to the nucleus even in the absence of other piRNA pathway factors (**Figure 5.4C-D**, "Panx-FL" lane). Interestingly, deletion of the carboxy-terminal region of Panx still allowed its nuclear localisation but abolished the interaction with Nxf2, which appeared redistributed to the cytoplasm ("$Panx^{\Delta C}$"

lane). Similarly, 3xFLAG-Nxf2 expressed in the complete absence of Panx remained largely cytoplasmic ("ZsGreen" lane). On the other hand, Panx lacking the amino-terminal half was still able to interact with Nxf2 but failed to enter the nucleus ("Panx$^{\Delta N}$" lane).

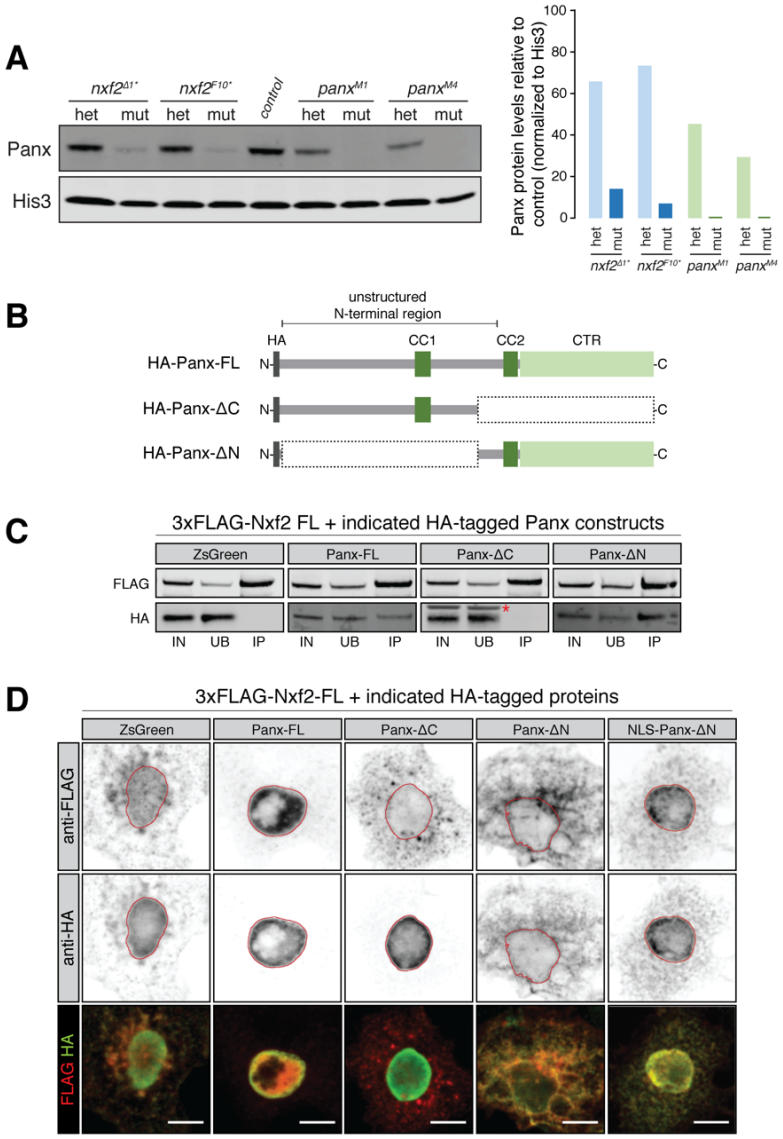


Figure 5.4. Protein stability and localization of Nxf2 and Panx is reciprocally co-dependent. **A)** Western blot showing Panx protein levels in ovaries of indicated genotypes with relative quantification, normalized to His3 signal. **B)** Schematic of HA-tagged Panx full-length and mutant constructs. CC= Coiled Coil; CTR= Carboxy-terminal region. **C)** Western blot analyses of FLAG-tag co-immunoprecipitation from lysates of S2 cells transfected with the indicated constructs. Asterisk indicates unspecific band from anti-HA antibody. **D)** Confocal images of the indicated 3xFLAG-Nxf2 (top, red in the merge) and HA-Panx (bottom, green in the merge) constructs transfected in S2 cells. Scale bar, 5µm.

As a consequence, Nxf2 was also aberrantly localised in the cytoplasm. Forced nuclear localisation of Panx$^{\Delta N}$ via addition of a Nuclear Localisation Signal (NLS) induced nuclear re-localisation of Nxf2 as well ("NLS-Panx$^{\Delta N}$" lane). Altogether, these data demonstrate that Nxf2 is dependent on Panx to achieve nuclear localisation and that their interaction occurs via the carboxy-terminus of Panx. The amino-terminus of Panx instead harbours the information for nuclear import.

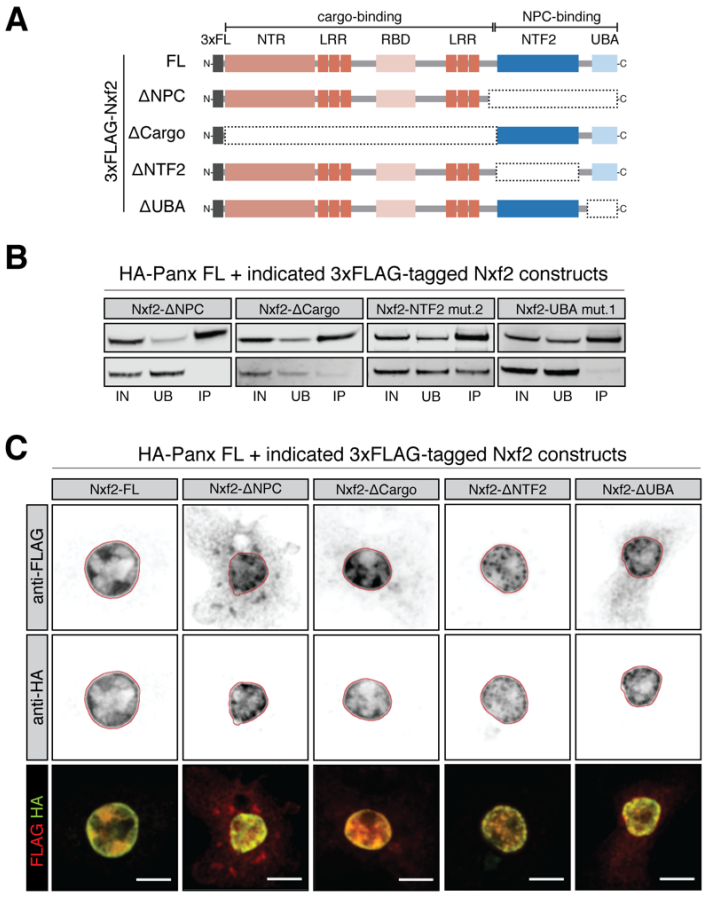


Figure 5.5 Requirements for PICTS assembly. A) Schematic of 3xFLAG-tagged Nxf2 full-length and domain mutant constructs. **B)** Western blot analyses of FLAG-tag co-immunoprecipitation from lysates of S2 cells transfected with the indicated constructs. **C)** Confocal images of the indicated 3xFLAG-Nxf2 (top, red in the merge) and HA-Panx (bottom, green in the merge) constructs transfected in S2 cells. Scale bar, 5µm.

Next, we generated 3xFLAG-Nxf2 mutants lacking the NPC-binding, the cargo-binding, the NTF2 or the UBA domain (ΔNPC, Δcargo, ΔNTF2 and ΔUBA, schematic in **Figure 5.5A**) and co-expressed them with full-length, HA-tagged Panx. Panx nuclear signal was unaffected in the presence of any Nxf2 domain mutants (**Figure 5.5C**), in line with the fact that Nxf2 depends

on Panx for its nuclear localisation but not vice-versa. While full-length Nxf2 was exclusively nuclear and strongly co-immunoprecipitated with Panx (**Figure 5.5B-C**), Nxf2$^{\Delta NPC}$ failed to interact with Panx and showed increased cytoplasmic signal. The reciprocal deletion of the cargo-binding region instead retained the interaction with Panx and was mostly restricted to the nucleus. To narrow down which portion of the NPC-binding domain interacts with Panx we further analysed Nxf2 lacking the NTF2 or the UBA domain individually. Among these, only Nxf2$^{\Delta UBA}$ showed higher cytoplasmic signal similar to Nxf2$^{\Delta NPC}$ (**Figure 5.5B**) and failed to copurify with Panx (Fabry et al., 2019). Taken together, this implicates the UBA domain within the carboxy-terminal NPC-binding region of Nxf2 in the interaction with Panx.

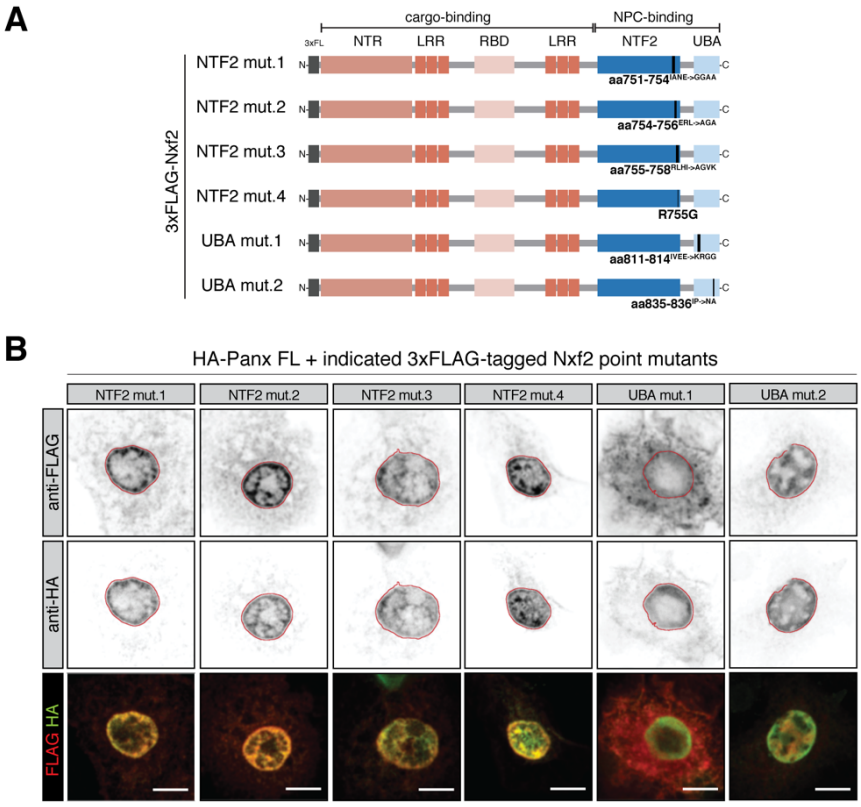


Figure 5.6 **Requirements for PICTS assembly. A)** Schematic of 3xFLAG-tagged Nxf2 point mutant constructs. **B)** Confocal images of the indicated 3xFLAG-Nxf2 (top, red in the merge) and HA-Panx (bottom, green in the merge) constructs transfected in S2 cells. Scale bar, 5µm.

Finally, aiming to investigate whether any conserved residues in the NPC-binding region are required to assemble PICTS, we generated 3xFLAG-tagged Nxf2 constructs carrying point mutations in 2-4 conserved amino acids (schematic and nomenclature in **Figure 5.6A**) in the NTF2 and UBA domain. Among these, UBA mut.1 had the strongest effect in de-localising Nxf2 (**Figure 5.6B**), further implicating the UBA domain in the interaction with Panx. The point mutations in the NTF2 domain targeted residues known to be essential for the interaction with Nxt1, which is expressed in S2 cells, and therefore indirectly probed the involvement of Nxt1 in silencing. None of the NTF2 point mutations significantly changed the localisation of Nxf2 (**Figure 5.6B**). We therefore concluded that, although Nxt1 is part of PICTS and involved in TGS, it is dispensable for the direct interaction between Panx and Nxf2.

V.4 Nxf2 and Nxt1 tethering to RNA induces silencing

The distinctive feature of piRNA-guided transposon silencing is that it happens via recognition of nascent RNA, as opposed to a specific DNA sequence. Previous work has demonstrated that artificial recruitment of Panx to a nascent reporter RNA triggers epigenetic silencing of the corresponding locus via deposition of H3K9me3, thus recapitulating what happens *in vivo* to transposon loci (Sienski et al., 2015; Yu et al., 2015). This assay is commonly referred to as "RNA tethering" and is based on the λN-BoxB system (Baron-Benhamou et al., 2004; Keryer-Bibens et al., 2008), represented in **Figure 5.7**. The reporter construct (namely "BoxB sensor") contains the Ubiquitin promoter from *D. simulans* driving the expression of a fluorescent protein, in this case ZsGreen. The 3'UTR of the reporter RNA carries 9 BoxB sites, which fold into 19-nt long hairpins that are recognised by the bacteriophage lambda N protein (λN). The second component of the assay is a construct expressing the 22 amino-terminal amino acids of λN-protein, sufficient to bind BoxB, fused to the protein of interest (POI). Interaction between the BoxB hairpins and the λN tag enables recruitment of the POI onto the reporter RNA, thereby bypassing the requirement for piRNA target sites. The resulting effects on transcription, translation and epigenetic states of the reporter can be assayed via qPCR, western blot and ChIP-seq, respectively (**Figure 5.7**). The use of a *D. simulans* promoter allows us to perform qPCR or ChIP-seq assays specific to the reporter construct. The 5'UTR of the reporter RNA carries a spliced intron, which enables detection of the nascent, un-spliced RNA via qPCR.

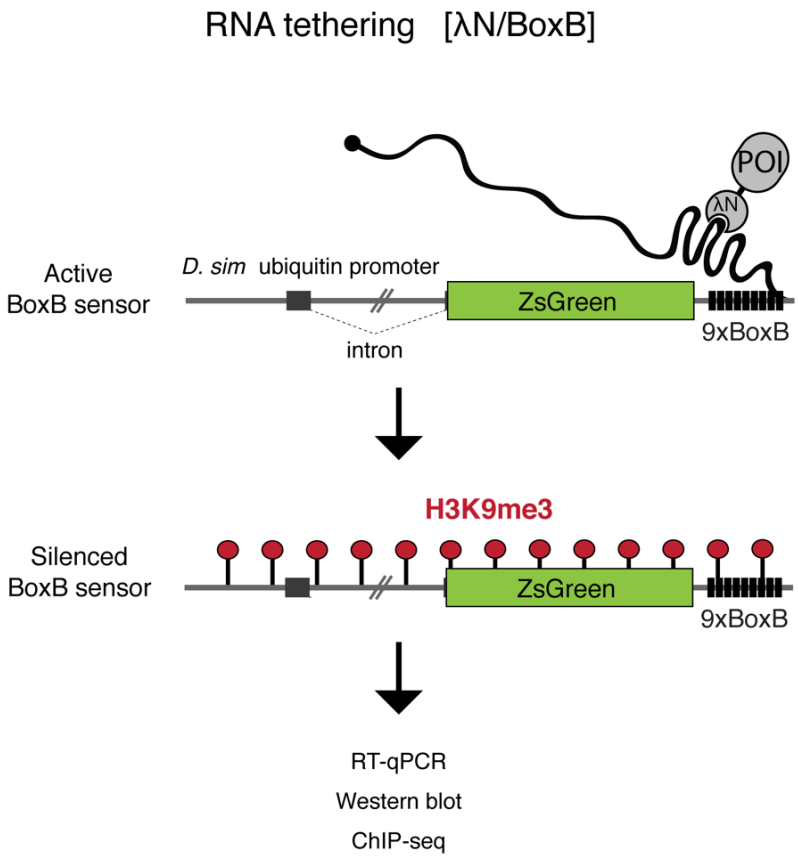


Figure 5.7 The RNA tethering assay. Schematic of the RNA λN/BoxB tethering system used to probe for epigenetic silencing. A λN fusion to a protein of interest (POI) is recruited to the nascent sensor RNA via interaction with BoxB hairpins. If the POI is capable of inducing silencing, the reporter becomes covered by H3K9me3 and is transcriptionally inactivated. The effects on reporter transcription, translation and epigenetic states can be measured by qPCR, western blot and ChIP.

With the aim of testing whether Nxf2 and Nxt1 share the same silencing capability of Panx, I set out to establish an RNA tethering system in OSCs. The first version I tested is depicted in **Figure 5.8A** and includes the previously described HA-ZsGreen reporter in combination with λN-3xFLAG-POI constructs followed by a T2A self-cleaving peptide upstream of myc-tagged mCherry. This system was designed to use flow cytometry as a readout for reporter activity, in addition to western blots. The cells that have received the tethering constructs can be identified based on mCherry expression, and the relative intensity of ZsGreen fluorescence is measured only in the mCherry positive population.

I first generated an OSC line with a stable integration of the BoxB sensor. This was achieved via co-transfection of the sensor and a helper plasmid carrying a Puromycin resistance gene, followed by selection with Puromycin-containing media and isolation of ZsGreen positive

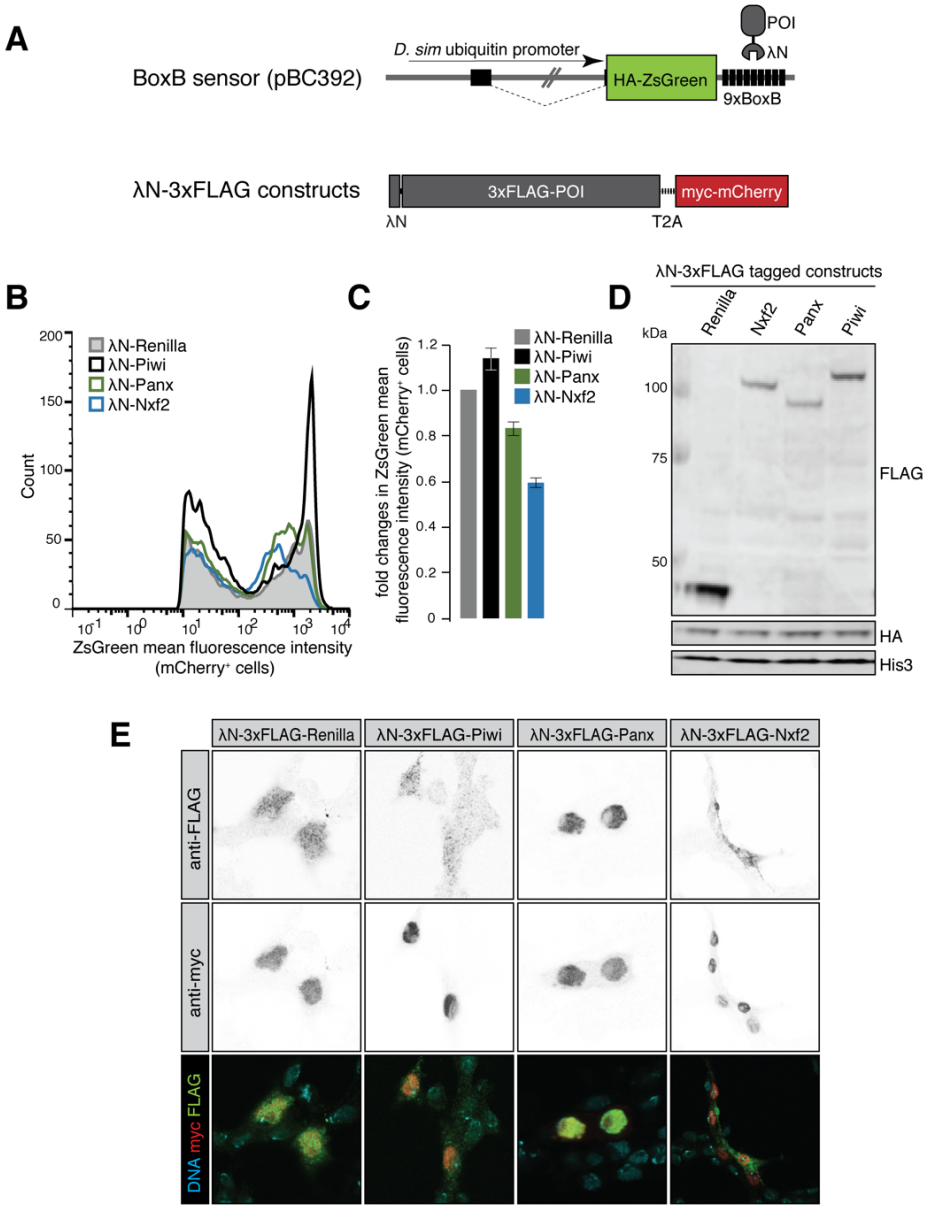


Figure 5.8 **Establishment of the RNA tethering system. A)** Cartoon of the BoxB sensor and λN constructs used for RNA tethering. **B)** Representative replicate showing flow cytometry analysis of ZsGreen mean fluorescence intensity (MFI) in the mCherry positive cell population. **C)** Fold-changes in ZsGreen MFI, as in (B). Error bars show standard deviation (n=3). **D)** Representative western blot showing expression levels of HA-tagged ZsGreen sensor and λN-3xFLAG-tagged proteins. **E)** Confocal images showing localization of λN-3xFLAG-tagged proteins and myc-tagged mCherry in OSCs.

clones. I then nucleofected λN-3xFLAG-tagged Piwi, Panx, Nxf2 or a Renilla negative control, harvested the cells four days later and analysed reporter expression levels. Flow cytometry analysis of the mean fluorescence intensity of ZsGreen showed only a mild but reproducible decrease upon tethering of Panx and Nxf2 but not Piwi (one representative replicate is shown in **Figure 5.8B** and triplicate quantification in **Figure 5.8C**). However, the observed reduction was too subtle to be detected via western blot (HA panel in **Figure 5.8D**), despite good expression levels of each λN construct (FLAG panel in **Figure 5.8D**). Previous work has shown a substantially higher silencing capability of Panx in similar tethering experiments (Sienski et al., 2015; Yu et al., 2015), suggesting that something is hampering the sensitivity of this assay. I therefore asked whether the λN-3xFLAG-tagged constructs were correctly localised to the nucleus and performed an immunofluorescence anti-FLAG. This revealed that Piwi and Nxf2 were aberrantly localised to the cytosol, whereas only Panx seemed to retain its correct nuclear localisation (**Figure 5.8E**). This might have been caused by poor efficiency of the T2A self-cleavage, leading to an aberrantly extended carboxy-terminus that impairs the interaction between Panx and Nxf2 and the overall functionality of tethered proteins.

Hence, I moved on to a second version of the RNA tethering system, depicted in **Figure 5.9A**. I generated an OSC line stably expressing another BoxB sensor which carries a Nano-Luciferase (NLuc) in addition to HA-ZsGreen, thus providing an alternative and highly sensitive readout for reporter expression levels (Hall et al., 2012). The λN-3xFLAG-tagged constructs did not carry any fusion at their carboxy-terminus that could impact their expression and localisation. Nucleofection of λN-Panx, Nxf2 and Nxt1 caused robust reporter silencing after 96 hrs that was readily detectable via western blot (one representative replicate is shown in **Figure 5.9B** and relative quantification of replicates in **Figure 5.9C**). Notably, reporter silencing caused by Nxf2 and Nxt1 tethering was stronger than that of Panx, with Nxf2 having the strongest effect. In contrast, Piwi tethering to the reporter did not induce silencing, as reported previously (Sienski et al., 2015; Yu et al., 2015). This is line with the model that Piwi can instruct silencing only when in complex with a piRNA, and not when artificially recruited to a transcript. This seems to support the notion that only the trimeric interaction Piwi:piRNA:target can lead to a conformational change in Piwi that enables recruitment of downstream effectors. Of note, instead of leading to silencing, tethering of λN-Piwi caused a modest but reproducible increase in reporter expression.

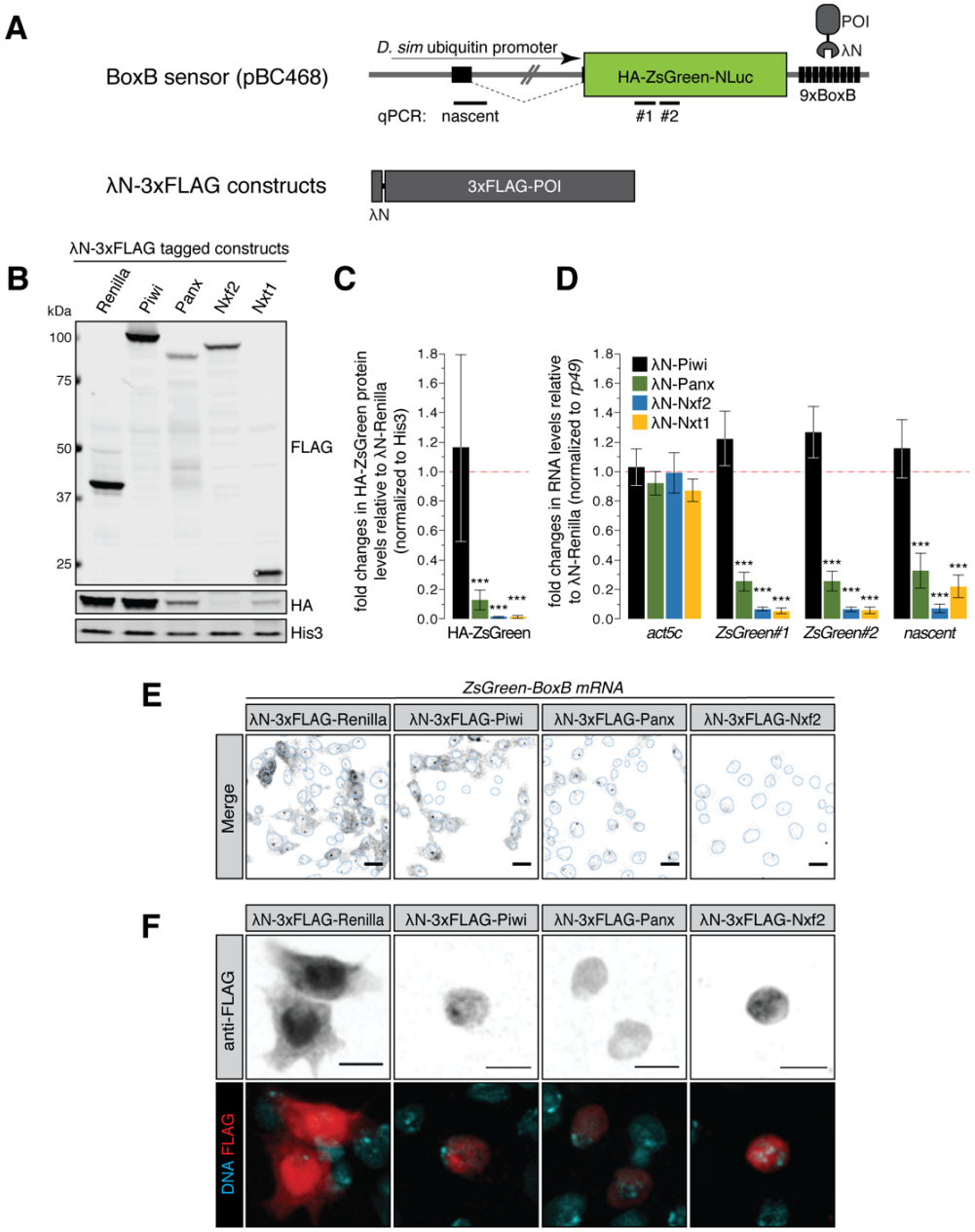


Figure 5.9 Recruitment of PICTS components to nascent RNA results in epigenetic silencing. **A)** Schematic representation of the BoxB sensor and λN construct used for RNA tethering. **B)** Representative western blot showing expression levels of HA-tagged ZsGreen sensor and of λN-3xFLAG-tagged proteins. **C)** Fold changes in protein levels of HA-ZsGreen in lysates from OSCs transfected with the indicated λN constructs (relative to a λN-Renilla and normalized to His3). *** = P value < 0.0001 (unpaired t-test). Error bars indicate standard deviation (n=4). **D)** Fold changes in steady-state RNA levels of the sensor and *act5c* (relative to λN-Renilla and normalized to *rp49*). *** = P value < 0.0001 (unpaired t-test). Error bars indicate standard deviation (n=4). **E)** Confocal images showing RNA FISH for *ZsGreen-BoxB* reporter transcripts. The blue nuclear outline has been drawn based on Lamin signal. Scale bar, 5µm. **F)** Confocal images showing localization of λN-3xFLAG-tagged proteins in OSCs. Scale bar, 5µm.

To test whether reporter silencing happens at the transcriptional level, we isolated RNA from the same experiments and performed qPCR for the reporter transcript using three different primer pairs (shown in **Figure 5.9A**). One primer pair spans the un-spliced junction between the 5'UTR exon and the downstream intron and therefore only detects nascent RNA (labelled as "nascent"). Tethering of each PICTS component, but not of Piwi, caused a robust decrease in the RNA levels of the reporter, indicating that these factors act at the (co-)transcriptional level (**Figure 5.9D**). The transcriptional repression was further demonstrated via RNA FISH for the reporter transcript, which showed a substantial decrease in cytosolic and nucleoplasmic staining as well as diminished signal from the transcription foci upon Panx and Nxf2 recruitment (**Figure 5.9E**). ChIP-seq from these samples confirmed that the repression is indeed the result of the reporter being decorated by H3K9me3 marks (Fabry et al., 2019). Finally, all λN-3xFLAG tagged constructs showed correct nuclear localisation (**Figure 5.9F**), further confirming that the RNA tethering assay is recapitulating what happens *in vivo* on transposon loci. Altogether, these results demonstrate that each component of the PICTS complex has an intrinsic capability to trigger co-transcriptional silencing of a locus, by instructing the deposition of repressive histone marks.

Additional work in the lab has characterised the silencing capability of PICTS upon recruitment to DNA. Interestingly, in this assay Panx and Nxf2 elicit epigenetic silencing whereas Nxt1 does not (Fabry et al., 2019). Further dissection of the domains required for silencing revealed that the amino-terminal portion of Panx alone is necessary and sufficient to induce silencing (Fabry et al., 2019). This indicates that the Panx amino-terminal region constitutes the silencing engine within PICTS.

V.5 Testing the interaction of Nxf2 with RNA

All members of the *Drosophila* NXF family carry an RNA-binding domain (RBD in **Figure 5.1C**), which for Nxf1 has been implicated in binding to spliced mRNAs that are ready to be exported to the cytoplasm (Liker et al., 2000; Tutucci and Stutz, 2011). Given that piRNA-guided TGS targets nascent RNAs and that Nxf2 recruitment to RNA causes an even stronger silencing than Panx itself (**Figure 5.9D**), we hypothesized that Nxf2 might be the RNA-binding unit of PICTS that stabilises its interaction with the target. To test this hypothesis, I adapted the protocol described in **Chapter III** to perform CLIP-seq of a HALO-3xTEV-3xFLAG-Nxf2 fusion expressed in OSCs (a detailed protocol is provided in **Materials and Methods II.10**). As shown in **Figure 5.10A**, this experiment resulted in successful enrichment of 3xFLAG-Nxf2

in the eluate, although to a lower extent compared to Armi (**Figure 3.18**). However, sequencing of Nxf2-bound RNA did not detect enrichment of mRNAs, somatic piRNA clusters or sense transposon transcripts, when compared to a mock IP (**Figure 5.10B**, purple dots correspond to transposons). I reasoned that over-expressed Nxf2 might be unstable if not in complex with Panx and thus incapable of binding its targets. Therefore, I decided to perform a similar *in vivo* experiment (namely RIP-seq) using the previously described GFP-Nxf2 fly line (a detailed protocol is provided in **Materials and Methods II.11**). Since the GFP tag is knocked into the endogenous *nxf2* locus, the protein is expressed at normal levels and forms a functional complex with Panx and Nxt1 (**Figure 5.2B**). I also decided to perform chemical crosslinking of ovarian lysates with PFA prior to RIP-seq, given that the UV crosslinking used for CLIP-seq in OSCs is often less effective on intact tissue. As a positive control, I carried out the identical experiment from GFP-Panx expressing flies, since Panx has been previously shown to bind to sense transposon fragments (Sienski et al., 2015). Aiming to avoid the confounding effect of background transposon expression in different strains, I prepared libraries from each total input lysate and used them as controls. Finally, since PFA crosslinking might increase background signal, both IP and input RNA samples were depleted of rRNA prior to library preparation. Western blot analysis confirmed enrichment of the respective GFP-tagged bait protein in the IP from crosslinked ovary lysates (**Figure 5.10C**). Nonetheless, GFP-Nxf2 RIP-seq did not show any enrichment of mRNAs nor sense transposon transcripts, recapitulating the results obtained by CLIP-seq (**Figure 5.10D**). On the other hand, GFP-Panx RIP-seq successfully enriched for a number of sense transposon transcripts, thus validating my experimental procedure and confirming what was observed by Sienski and colleagues (**Figure 5.10E**). Since the publication of our work, three other groups reported the identification of Nxf2 and Nxt1 as Panx interactors (Batki et al., 2019; Murano et al., 2019; Zhao et al., 2019). Of note, their data indicate that Nxf2 does have the ability to bind RNA, but this interaction is likely very labile and undetectable in my experiments and will be further discussed in **Chapter VI**.

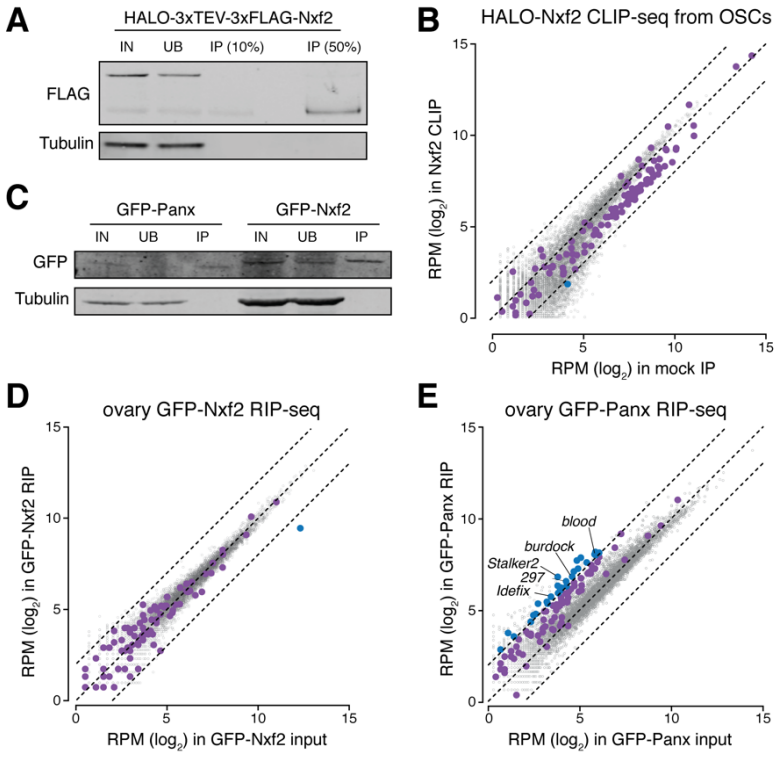


Figure 5.10 Probing the interaction of Nxf2 with RNA. A) Western blot showing immunoprecipitation of HALO-3xTEV-3xFLAG-Nxf2 from OSCs. **B)** Scatter plot showing expression levels (RPM) of genes (in grey) and sense transposons (in purple) in Nxf2 CLIP-seq (n=3) against a mock immunoprecipitation (n=3). **C)** Western blot showing immunoprecipitation of GFP-tagged Panx and Nxf2 from ovary lysates used for RIP-seq. **D)** Scatter plots showing expression levels (RPM) of genes (in grey) and sense transposons (in purple) in GFP-Nxf2 and GFP-Panx RIP-seq (n=3) against the respective input controls.

CONCLUSION

Work presented in this chapter contributed to the identification of a trimeric complex required for TGS of transposons in *Drosophila*: the PICTS complex, which comprises Panx, a protein with no known domains, and two nuclear export factors, Nxf2 and Nxt1 (**Figure 5.11**). While Nxt1 associates with all the NXF-family proteins and is implicated in various processes, we show that Nxf2 has evolved specifically for epigenetic silencing of transposable elements. Nxf2, together with Panx, is required for piRNA-guided transposon control and heterochromatin deposition in the fly ovary. Nxf2 and Panx are interdependent for the respective protein stability, subcellular localisation and silencing function. Each component of PICTS is *per se* capable of inducing silencing when recruited to RNA, likely through recruitment of Panx. Both Panx and Nxf2 can bind to target RNA, although Panx seems to do so in a more stable fashion, which further strengthens the requirement of active transcription to enable Piwi target engagement and downstream recruitment of PICTS (**Figure 5.11**).

Taken together, our findings highlight that the evolution of transposon defence mechanisms involved exaptation of nuclear export factors. The NXF family has undergone an expansion in metazoans, with two copies in *C. elegans*, four in *D. melanogaster* and five in human (Tan et al., 2000; Tan et al., 2005; Yang et al., 2001). Since they do not seem to perform redundant functions, it is plausible that they have diversified their functions to certain tissues or substrates (ElMaghraby et al., 2019; Kneuss et al., 2019).

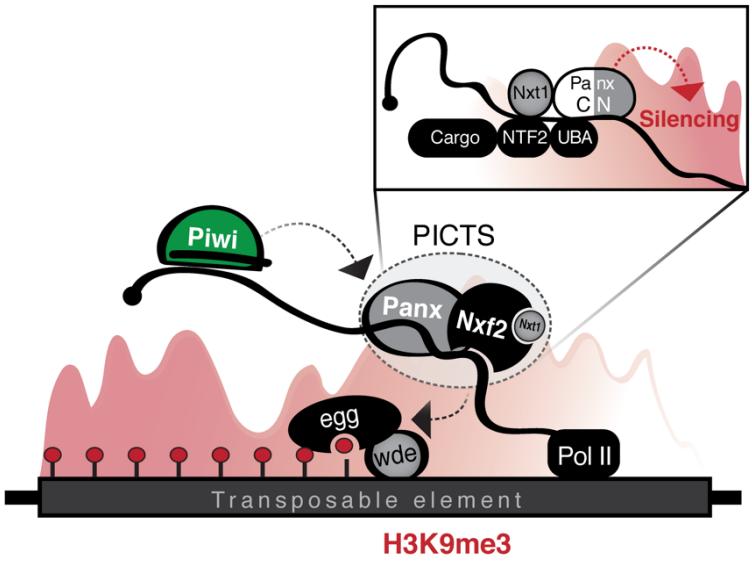


Figure 5.11 Model for piRNA-guided co-transcriptional silencing. Piwi bound to a piRNA recognises nascent transposon RNAs and recruits the PICTS complex, composed of Panx, Nxf2 and Nxt1. This in turn induces epigenetic silencing of the locus via H3K9me3 deposition, catalysed by Egg and its co-factor Wde.

CHAPTER VI

DISCUSSION

VI.1 Towards an understanding of the molecular mechanisms of piRNA biogenesis

The biogenesis of piRNAs requires a highly specialised machinery that recognises the correct substrates and processes them into strings of consecutive piRNAs. With the work presented in **Chapter III**, I have expanded the repertoire of factors that participate to piRNA production and provided mechanistic insights into how the RNA substrates are presented to the biogenesis machinery.

CG10880/Daedalus (Daed) is a protein anchored to the outer surface of mitochondria that is expressed predominantly in the *Drosophila* ovary and had been previously identified as required for germline transposon control (Czech et al., 2013). Dead has a similar domain structure to Gasz, another mitochondrially-anchored protein involved in piRNA production (Handler et al., 2013; Munafò et al., 2019; Yamashiro et al., 2019; Zhang et al., 2016). Gasz and Daed assemble into homo- and hetero-polymeric complexes on the surface of mitochondria and are important to recruit the RNA helicase Armi and hold it close to the nuclease Zuc, thus enabling piRNA production. In the absence of either Gasz or Daed, piRNA production collapses since Armi-bound piRNA precursors cannot be stably brought to mitochondria. Interestingly, Daed is unique to *Drosophilids* (**Figure 3.24**) whereas Gasz is conserved and its mouse homolog has been implicated in piRNA production and mitochondrial fusion in male germ cells (Zhang et al., 2016). It is not clear why *Drosophilids* possess two proteins to carry out similar functions, but it is apparent that they are not redundant. I hypothesize that this stems from some of the differences in piRNA production between *Drosophila* and other animals, such as the absence of prominent 3' trimming following Zuc cleavages.

Recruitment of Armi to the mitochondrial surface is essential to drive piRNA production (Ge et al., 2019; Munafò et al., 2019; Yamashiro et al., 2019) and, among known piRNA biogenesis

factors, Armi is unique in the ability to "shuttle" between nuage/Yb bodies and mitochondria. This capability is key, since Armi transfers selected piRNA precursors from the sites of licensing (nuage/Yb bodies) to the sites of processing (mitochondria), thus bridging the two major events and subcellular compartments of piRNA biogenesis. Armi binding to RNA has emerged as a tightly controlled process, likely to prevent random binding. In the absence of such control mechanisms, Armi promotes piRNA production from spurious substrates (Ge et al., 2019; Ishizu et al., 2019). Conversely, recruitment of Armi onto reporter transcripts determines their processing into *bona fide* piRNAs (Pandey et al., 2017; Rogers et al., 2017). This implies that Armi alone is necessary and sufficient to specify any cellular transcripts for Zuc-mediated processing, whereas substrate recognition requires its association with other factors. The binding of Armi to RNA is regulated by Yb and Piwi, as well as by Armi's own ATP-dependent helicase activity (Ge et al., 2019; Ishizu et al., 2019; Munafò et al., 2019; Pandey et al., 2017; Yamashiro et al., 2019). In somatic cells, Yb, via a yet not fully clarified mechanism, selects specific mRNAs and cluster transcripts. In its absence, piRNA production does not completely collapse but instead occurs on abundant cellular RNAs, which are then aberrantly bound by Armi (Ishizu et al., 2019). The 5'-P of Yb-selected substrates is loaded by Piwi to form a pre-piRISC that can be bound by Armi (**Figure 3.23**, upper panel). Although the precise hierarchy of each molecular event has not been clarified yet, various lines of evidence, including **Chapter III**, suggest that, if Piwi cannot bind to the 5'-P of a precursor, this will not be efficiently recognised by Armi (Munafò et al., 2019; Yamashiro et al., 2019).

Overall, what emerges from these studies is that only a complex consisting of Piwi, Armi and a pre-piRNA is fully competent to leave Yb bodies and translocate to mitochondria. This trimeric complex only binds to the correct substrate if Yb is present and Armi ATPase activity is intact, thus evidencing a multi-layered regulation that prevents spurious processing. Since Armi and Piwi are expressed in both follicle and nurse cells, this likely represents a unified mechanism applicable to all compartments of the fly ovary. However, the mechanisms upstream certainly differ, as somatic cells rely on Yb whereas germ cells have the ping-pong cycle. It is interesting to note that Yb is not proximal to mitochondrial biogenesis factors (**Figures 3.10-11-15** and **4.8A**), further indicating that only the Piwi-Armi-pre-piRNA complex leaves Yb bodies, whilst the role of Yb terminates once the right substrates have been licensed.

Once on mitochondria, the Piwi-Armi-pre-piRNA complex is anchored via Gasz and Daed in the vicinity of Zuc. Although other cytosolic piRNA biogenesis factors were found to be nearby Zuc, Armi appeared to be the one in closest proximity to the cleavage-competent dimer (**Figure**

3.23). This might indicate that Armi ensures the processivity of Zuc cleavages by presenting the RNA substrates. In strong support of this hypothesis, a recent study in *Bombyx mori* reports that overexpression of Armi wild type, but not an ATPase mutant, strongly promotes Zuc-driven piRNA production *in vitro* (Izumi et al., 2020). In this context, the ATP-dependent helicase activity of Armi might provide an explanation for the regular spacing of Zuc cleavages. Overall, these data support the model depicted in **Figure 3.23**, where Piwi and Armi first bind to a precursor RNA in the nuage/Yb bodies, the 5'-P of the RNA precursor is loaded by Piwi MID domain and this allows Armi stable binding to the transcript (**Figure 3.23**). The trimeric Piwi-Armi-pre-piRNA complex is competent to depart from nuage/Yb bodies and translocate to mitochondria, perhaps thanks to a conformational switch in Armi or additional factors like the uncharacterised SoYb, that are enriched in our PL-MS for mitochondrial factors. Interestingly, the interaction between Piwi and Armi becomes RNase-sensitive upon loss of *daed* and *gasz*, but is not in unperturbed cells (**Figure 3.22**). This would suggest that the interaction of Piwi and Armi is primarily mediated by RNA in Yb bodies, whereas Daed and Gasz provide a more stable anchoring platform on mitochondria. A typical piRNA production cycle would occur as follows (**Figure 3.23**): Piwi bound to the 5'-P of the precursor directs Zuc to cleave at the first accessible uridine (Gainetdinov et al., 2018) or consensus motif, at least in *B. mori* (Izumi et al., 2020). Once the first piRNA is released and readily incorporated into Piwi, mitochondrially-anchored Armi unwinds or more likely translocates along the transcript to allow Piwi sequential binding to a new 5'-P released by the previous Zuc cleavage. This re-initiates the cycle, as the incoming Piwi footprint determines the next Zuc cleavage site. It will be of interest for future to studies to investigate whether specific chaperones or other factors enable this "cyclic" binding of Piwi to free 5'-P.

How widely applicable is this model of piRNA biogenesis? To date, studies in mouse point towards a relatively similar mechanism to produce piRNAs during spermatogenesis. The key players are conserved, such as the nuclease MITOPLD (homolog of Zuc), the RNA helicase MOV10L1 (homolog of Armi) and the mitochondrial ankyrin repeat protein GASZ. Structural studies on MITOPLD show a cleavage bias nearly identical to that observed in flies (Ipsaro et al., 2012) whereas MOV10L1 has been shown to bind piRNA precursors and promote their processing (Vourekas et al., 2015). Furthermore, the footprint of PIWI proteins dictates piRNA size in mouse as well (Gainetdinov et al., 2018). Taken together, this strongly supports a unifying model of mitochondrial piRNA biogenesis whereby PIWI proteins bind 5'P RNA precursors and Armi/MOV10L1 ensures the processivity of Zuc/MITOPLD cleavages. In contrast with flies, MITOPLD generates slightly longer phased piRNA intermediates that will

then be trimmed to their mature size by the exonuclease PNLDC1 (Han et al., 2015; Mohn et al., 2015). Notably, mouse GASZ interacts with mitofusins to promote mitochondrial fusion and deletion of its mitochondrial targeting sequence causes aberrant mitochondrial fission and altered metabolism (Zhang et al., 2016). MOV10L1 localises in proximity to mitochondria in a structure known as inter-mitochondrial cement (ICM) and physically associates with PIWI proteins (Zheng et al., 2010). Therefore, it is possible that a similar "shuttling" of MOV10L1 between nuage and ICM occurs in mice as well. Work in *B.morii* has also uncovered a very similar picture of the mitochondrial piRNA production route and has assigned prominent roles to BmZuc and BmArmi (Nishida et al. 2018; Izumi et al. 2020). The fact that a phasing signature has been detected in most animals and that Zuc is well conserved (Gainetdinov et al., 2018; Ozata et al., 2019), makes it tempting to speculate that the core mitochondrial piRNA biogenesis process is functionally similar in most animals. Future studies will define the extent to which the mitochondrial biogenesis machinery is conserved. An exception to this seems again to be *C.elegans*, where 21-U RNA biogenesis has so far not been linked to mitochondria. 21-U RNA precursors are very short, implying that an RNA translocase activity like that of Armi might be dispensable, and instead rely on a dedicated processing machinery evolved from the small nuclear RNA pathway (Cordeiro Rodrigues et a., 2019). Nonetheless, 21-U RNA production requires the generation of a 5'P followed by binding of the PIWI protein PRG-1 and resection of the 3'ends by the exonuclease PARN-1, closely resembling what observed in other species (Ozata et al., 2019).

VI.2 *flam* export relies on Yb and specific Nuclear Pore Complex subunits

The substrates of somatic piRNA biogenesis carry canonical features of RNA pol II-transcribed genes but can nonetheless be specifically recognised among all other mRNAs. The uni-strand cluster *flam* gives rise to the largest fraction of somatic piRNAs and the work described in **Chapter IV** provides novel evidence on the molecular mechanisms underlying its export and specification as a piRNA precursor.

Recent work, including **Chapter III** and **IV**, point to Yb as the key 'licensing factor' that selects newly exported RNAs for piRNA production (Dennis et al., 2016; Hirakata et al., 2019; Ishizu et al., 2019; Murota et al., 2014; Sokolova et al., 2019). What features of these RNAs promote binding of Yb is not entirely clear, although a 'trigger sequence' has been identified for some (Homolka et al., 2015; Ishizu et al., 2015). We find that the NPC transiently comes into contact

with Yb, the latter likely scanning all transcripts and binding to a subset, among which is *flam* that is then transported to Yb bodies (Ishizu et al., 2015; Murota et al., 2014). Because the Yb bodies are liquid-like, phase-separated compartments (Hirakata et al., 2019), we can hypothesize that Yb molecules bound to newly exported *flam* have a high propensity to aggregate into pre-existing Yb bodies, thus providing a sort of 'directionality' to the export of *flam*. Compartmentalisation of cellular events into spatially separated, non-membranous organelles is a strategy often used to optimise biochemical reactions or bring together proteins that participate to the same pathway. Yb bodies, which are unique to somatic follicle cells and depend on *flam* export, might in this case provide a temporary storage site for piRNA precursor molecules, prior to their licensing for mitochondrial processing. Work from the Siomi lab has suggested that somatic precursors are cleaved into shorter pre-piRNAs prior to their translocation to mitochondria (Murota et al., 2014; Saito et al., 2010; Yamashiro et al., 2019). Although direct evidence of this is still lacking, we can speculate that the Yb bodies provide a temporary storage place for *flam* pre-piRNAs, where they are shielded from undesired RNA decay pathways. Non-membranous organelles have been linked to other small RNA pathways, indicating that some properties of these liquid-like droplets might be particularly advantageous. One notable and widespread example are P-bodies, RNA and protein granules which seem to act as mRNA storage sites and include proteins involved in the miRNA pathway, RNA decay and translational repression (Standart and Weil, 2018). Small RNA pathways in *C.elegans* rely on various types of germline liquid-like organelles that are spatially and temporally regulated, namely the P-granules, Z-granules and Mutator foci (Wan et al., 2018). Although all three types of granules are closely associated, they do not overlap and are always arranged in a specific order, which may help to organise and coordinate the different steps of RNA biogenesis pathways. In line with that, cytosolic foci of Yb and *flam* do not perfectly overlap, possibly suggesting a similar spatial compartmentalisation of *flam* RNA processing steps.

Aiming to dissect what other player(s) determine the export of *flam* prior to Yb binding on the outer side of the nuclear envelope, we uncovered an unexpected requirement of Nup54 and Nup58. Both Nups emerged as strong candidates for somatic transposon control (Handler et al., 2013) and we now discover that they are specifically involved in piRNA production from *flam*. The phenotype observed upon loss of *nup54* and *nup58* is remarkably specific: despite being core components of the NPC channel (Chug et al., 2015; Kim et al., 2018; Solmaz et al., 2011), their knockdown does not impair cell viability (**Chapter IV**) or disrupts ovarian morphology (Handler et al., 2013). Furthermore, depletion of Nup54 and Nup58 only causes modest changes in the expression of the vast majority of mRNAs, whereas transposons are severely mis-

regulated (**Figure 4.6**). The fact that both Nups belong to the NPC and that their loss causes a disassembly of Yb bodies indirectly proves their involvement in *flam* export. Nonetheless, we also observe an intriguing reduction in the overall transcript levels of *flam*, which might imply a mechanism that unifies transcription and export. A likewise reduction of *flam* transcripts also occurs upon *yb* knockdown. Since, to our knowledge, Yb is only present on the cytosolic side of the NPC, the only likely explanation is that impaired *flam* export negatively affects the upstream transcriptional process. It has been shown that canonical Nxf1/Nxt1 heterodimers mediate *flam* translocation across the NPC (Dennis et al., 2016). Notably, Nxf1 and Nxt1 typically recognise splicing events marked by the Exon Junction Complex, which within *flam* have been reported in the 5' region. We can therefore speculate that, once the *flam* 5' end is marked by Nxf1/Nxt1, the transcript can be funnelled into the NPC and another mechanism ensures continued transcription while the rest of the RNA is exported. As Nxf1 has also been shown to permanently reside in the NPC (Ben-Yishay et al., 2019; Derrer et al., 2019), this potentially explains why the *flam* 5' end appears anchored to the nuclear envelope, while the remainder of the transcript is associated with Yb bodies (Sokolova et al., 2019).

The export of piRNA precursors is an essential task that, at least in *Drosophila* germ cells, has evolved a specialised machinery. It has been recently shown that a gonad-specific member of the NXF family, Nxf3, has been repurposed to mediate export of non-canonical dual-strand cluster transcripts (ElMaghraby et al., 2019; Kneuss et al., 2019). It is interesting to note that in *nxf3* null mutants the production of piRNAs from those clusters is not completely abolished (ElMaghraby et al., 2019; Kneuss et al., 2019), thus implying that alternative export routes can still escort some RNAs out of the nucleus. Similarly, knockdown of *nup54* and *nup58* does not completely eliminate piRNA production from *flam* (**Figure 4.10**), although we cannot exclude that this results from residual Nup54 and Nup58 levels due to incomplete knockdown. To overcome this potential limitation, we have attempted to generate our own CRISPR/Cas9 null mutants for these two Nups, however we did not recover any mutant alleles and other available alleles are not viable in homozygosis. The ubiquitous presence of Nup54 and Nup58 in the NPC makes the dissection of any specific function more challenging and we must therefore rely on conditional mutants, knockdowns or hypomorphic alleles, where available.

Interestingly, several Nup genes are highly expressed in the gonads and, most importantly, show signs of rapid adaptive evolution that results in hybrid incompatibilities. Nup160 from *Drosophila simulans* is incompatible with one or more factors on the *D. melanogaster* X chromosome, as hybrids carrying *D. simulans* Nup160 and a *D. melanogaster* X chromosome

have been shown to be lethal in both sexes (Tang and Presgraves, 2009). *D. melanogaster* females carrying autosomal deficiencies (Df) were crossed with *D. simulans* males and, among the offspring, males that inherited the deficiency were analysed. These males express only the *D. simulans* gene(s) within the deficiency but possess an X chromosome from *D. melanogaster*, hence their death implies that the X is incompatible with an autosomal gene from *D. simulans*. A similar hybrid incompatibility has also been reported for Nup96 (Presgraves et al., 2003), suggesting that selection driven by an evolutionary conflict involving pathogens or transposable element might explain the fast adaptation of Nup genes (Tang and Presgraves, 2009). Notably, signatures of rapid divergence between *D. melanogaster* and *D. simulans* have been identified also for Nup75, Nup107, Nup133 and Nup153, all part of the same NPC subcomplex as Nup96 (Presgraves and Stephan, 2007). It is tempting to speculate that the hybrid incompatibility is, at least partially, due to the inability of Nups from *D. simulans* to recognise *flam* in *D. melanogaster*. However, it has to be noted that *flam* mutants are not lethal (Prud'homme et al., 1995), therefore this might be part of a more complex phenotype. Nonetheless, these findings further underscore the fast adaptation of some Nup genes which, in some instances, might have diverged to originate novel, non-canonical functions. Another recent study on the involvement of Nup54 in *Drosophila* sex differentiation did not identify a striking global signature of rapid adaptation (Haussmann, 2019). However, this manuscript identified a stretch of amino acids of increasing length in different species and four hotspots for substitution in the *nup54* promoter. Hence, it will be interesting to investigate the conservation of *flam* across other *Drosophilids* and whether *D. simulans* Nups are able to export *flam* in *D. melanogaster.*

Finally, a growing body of evidence in yeast and in flies points towards the idea that NPCs are not passive channels waiting for their cargoes, but that some Nups are relatively mobile and can directly bind to and regulate genes. This can involve preferential association between Nups and actively transcribed regions or re-localisation of genes to the NPC in response to activation (Kalverda and Fornerod, 2010; Pascual-Garcia et al., 2017). In *Drosophila,* this has been mostly shown for 'dynamic' Nups, such as Nup62, Nup98 and Nup153, which have the ability come on and off the NPC during interphase. On the other hand, 'stable' Nups are integral components of the NPC throughout the cell cycle but they also appear to bind to chromatin. Interestingly, *Drosophila* Nup107 and Nup93, both considered to be stable, bind to mutually exclusive regions of the genome, each marked by a distinctive epigenetic signature (Gozalo et al., 2020). While Nup107 preferentially interacts with actively transcribed regions, Nup93 appears to specifically target silenced, Polycomb-bound regions (Gozalo et al., 2020). This underscores that a specialisation of NPCs in different areas of the nucleus can indeed occur, despite them

all sharing the same subunits. In the context of the piRNA pathway, this argues for a model whereby NPC components bind to the *flam* locus, preferentially located at the nuclear periphery, to promote its expression. A conceptually similar mechanism coupling transcription and export has been found in yeast, where the SAGA complex physically connects with the NPC via adaptor proteins, such as Mex67, a protein with a domain structure similar to Nxf1 (Kohler and Hurt, 2007). This physical proximity is thought to help creating a more favourable environment for RNA processing, quality control and export of specific loci. In this scenario, failed export might directly affect transcription, possibly explaining the reduced levels of *flam* upon knockdown of *yb, nup54* and *nup58.* It is apparent that many non-canonical and specialised (epi)genetic programmes operate in germ cells (ElMaghraby et al., 2019; Hampoelz et al., 2019; Hopes T., 2020; Inoue et al., 2017; Kneuss et al., 2019; Xia et al., 2020). Such extreme specialisation is intimately linked to the necessity of preparing the development of the new organism and protecting the offspring's genome from harm, which includes preventing transposon mobilisation. The existence of germline-specific mechanisms can also occur within constitutive protein complexes, as reported for *Drosophila* ribosomal proteins via paralog switching (Hopes T., 2020). Intriguingly, specialised NPCs within the same cell have been identified in the ciliate *Tetrahymena termophila,* where small RNAs drive genomic rearrangements at the onset of reproduction. Here, the macro- and micro-nucleus are equipped with their own distinct subset of nucleoporins, and therefore display differential permeability (Malone et al., 2008). Taken together, our results suggest that components of the NPC might have acquired specialised functions to carry out transposon control in the *Drosophila* ovary.

The *flam* locus seems present only in a subset of *Drosophila* species (J.v.L., B.C. and G.J.H., unpublished observation), thus it is possible that this NPC-based mechanism has evolved in parallel with the cluster itself. Nonetheless, dedicated export machineries exist for other non-coding transcripts, such as Exportin-t for tRNAs, Exportin-5 for pre-miRNAs and CRM1 for snRNAs and rRNA precursors (Kohler and Hurt, 2007) and fly dual-strand piRNA cluster transcripts (ElMaghraby et al., 2019; Kneuss et al., 2019). These exportins recognise either structural features of their substrate RNAs, such as the 2-nucleotide 3' overhang generated by Drosha cleavage on pre-miRNAs, or adaptor proteins, such as Nxf3. Therefore, eukaryotic cells appear to possess many parallel, non-overlapping export routes that efficiently coordinate the many complex RNA processing pathways occurring at any given time.

VI.3 Co-option of cellular functions for epigenetic silencing of TEs

piRNA-guided transposon silencing occurs on nascent TE transcripts and involves epigenetic repression of the locus. A critical mediator of the silencing response is Panx (Sienski et al., 2015; Yu et al., 2015), which we now discover is part of the PICTS complex, together with Nxf2 and Nxt1 (described in **Chapter V**). In the absence of PICTS, Piwi is unable to induce silencing of active transposons. Artificial recruitment of PICTS to a locus, however, can bypass Piwi and is sufficient to initiate the silencing cascade.

The emerging model of piRNA-guided TGS implies that only the interaction Piwi-piRNA-target is competent to elicit silencing (**Figure 5.11**). In line with that, we and others find that forced recruitment of Piwi to a reporter locus fails to initiate silencing (**Figure 5.9**) (Batki et al., 2019; Fabry et al., 2019; Murano et al., 2019; Sienski et al., 2015; Yu et al., 2015; Zhao et al., 2019). A generally accredited explanation is that, upon target binding, Piwi undergoes a conformational change, and this in turn enables recruitment of PICTS. Notably, this provides an additional layer of regulation to the piRNA pathway and prevents uncontrolled silencing events of unwanted genomic loci. We and others failed to detect enrichment of Piwi in our IP-MS for PICTS components, and only observe a weak association via coIP (**Figure 5.1-2**) (Batki et al., 2019; Fabry et al., 2019). This can be explained because only a tiny fraction of nuclear Piwi directly engages with PICTS in unperturbed cells, where TEs are mostly inactive and robustly wired into piRNA-mediated silencing. Together with what I discussed above (**VI.1**), it is evident that each step of the piRNA pathway is finely regulated to prevent spurious silencing or processing of inadequate substrates. We often find that the most important junctures of the pathway are indeed gated by the interplay of various factors (e.g. Piwi-piRNA-target at the onset of TGS or Piwi-Armi-pre-piRNA during precursor licensing), providing robustness and specificity to the system and avoiding interference with other cellular activities.

Additional evidence supporting this precise regulation stems from the interdependency between Panx and Nxf2 for their protein stability and subcellular localisation (**Figure 5.4**) (Batki et al., 2019; Fabry et al., 2019). In the absence of Nxf2, Panx is rapidly destabilised due to the presence of a so-called “degron” sequence in its carboxy-terminal part (Batki et al., 2019), meaning that it can only exert its function where Nxf2 is expressed. Conversely, in cells depleted of Panx, Nxf2 does not localise to the nucleus. Of note, reports that other factors, namely Nxf1, Mael and Arx, associate to PICTS (Murano et al., 2019; Zhao et al., 2019), are

inconsistent with our results and call for a better dissection of the hierarchy and functional relevance of these interactions during the TGS silencing cascade.

Our experiments find that both Nxf2 and Nxt1 are required for effective transposon control, however, the actual silencing engine within PICTS resides in the amino-terminus of Panx. This therefore raises the question of what is the function of Nxf2 and Nxt1 within PICTS. Although we failed to identify a stable association between Nxf2 and transposon mRNAs (**Figure 5.10**), other work indicates that Nxf2 at least contributes to the RNA-binding activity within PICTS. The amino-terminal RRM of Nxf2 can bind to RNA *in vitro* (Batki et al., 2019), similarly to that of Nxf1 (Liker et al., 2000), and CLIP-seq for Nxf2 enriches for sense transposon transcripts (Murano et al., 2019; Zhao et al., 2019). It is worth highlighting that those CLIP-seq experiments were performed either upon depletion of downstream factors (e.g. *HP1A* or *Mael*, thus increasing the substrate availability for Nxf2 binding) or from a stable cell line with concomitant depletion of endogenous *nxf2*. Taken together, these data strengthen the notion that PICTS binding to RNA is indeed a transient event, which can only be detected in specific experimental setups. However, some open questions remain, such as what is the fate of the PICTS-engaged transcript once the locus has been recognised for silencing and why is the RNA binding activity of PICTS so critical. A potential explanation comes from yeast TGS, where Swi6, an HP1 homolog, has a strong affinity for RNA and is thought to help stabilising the transcript on the chromatin until it is degraded by the exosome (Martienssen and Moazed, 2015). This at the same time stabilises RITS recruitment onto a locus and prevents export of an unwanted transcript. Another avenue is that Nxf2 somehow shields the nascent transcript from the canonical RNA export machinery, again preventing TE mRNAs from being translated.

It is worth pointing out that the link between PICTS and the downstream epigenetic machinery remains elusive. In line with other TGS pathways, Panx is thought to recruit an H3K9 methyltransferase (Egg) but this interaction is either very transient or occurs via bridge proteins that escaped detection. PICTS itself does not seem to directly bind to histone marks, an activity that is instead present within yeast RITS (Duempelmann et al., 2020). One of the components of RITS is the chromodomain protein Chp1, which can directly recognise H3K9me3. As consequence, loci that become decorated by H3K9me3 in response to Ago1 target engagement can recruit more RITS via Chp1, thus establishing a self-reinforcing loop that stabilises the silencing. It is likely that several copies of PICTS are recruited onto each active TE locus, thanks to the broad array of Piwi-loaded phased piRNAs, but this process has not yet been dissected in detail.

It is most fascinating that nuclear export factors have been co-opted and repurposed for piRNA-driven heterochromatin formation in germ cells. Exaptation of 'housekeeping gene' paralogs appears to be a recurrent theme in the piRNA pathway, with several other examples being discovered over the last decade (Andersen et al., 2017; Chen et al., 2016; ElMaghraby et al., 2019; Kneuss et al., 2019; Vermaak et al., 2005; Vermaak and Malik, 2009). The NXF protein family is highly conserved and, while the yeast genome encodes for only one member (Mex67p), *Drosophila* has four different ones (Herold et al., 2001). The ubiquitous Nxf1 is critical for bulk mRNA export, whereas both Nxf2 and Nxf3 show gonad-specific expression patterns and are required for transposon silencing. While Nxf2 functions in co-transcriptional gene silencing, Nxf3 promotes the export of dual-strand cluster transcripts, bypassing the Nxf1-mediated route (ElMaghraby et al., 2019; Kneuss et al., 2019). Nxf4 is expressed only in testis and its role is yet to be established. The striking functional diversification of NXF family members in *Drosophila* suggests a more general pattern, where the numerous NXF variants present in metazoans have acquired tissue-specific roles (Tan et al., 2000; Tan et al., 2005; Yang et al., 2001). Mice express testis-specific NXF variants and *nxf2* null mutants are sterile (Pan et al., 2009), although no direct link to the piRNA pathway has yet been found. Overall, we can speculate that the NXF protein family has undergone a compelling evolutionary radiation that has led to diversified functions, also thanks to the host-transposon conflict.

VI.4 Future perspectives

Since the discovery of piRNAs nearly 15 years ago, substantial work has provided insights into the mechanisms underlying their production and silencing modes. Early work identified the involvement of conserved, germline-specific proteins and outlined major differences with other RNAi pathways (Aravin et al., 2006; Brennecke et al., 2007). Later on, several genetic screens and follow-up studies shed light onto the existence of highly specialised molecular machineries that have evolved for transposon control, often by co-opting paralogs of 'housekeeping genes' (Andersen et al., 2017; Czech et al., 2013; Handler et al., 2013; Mohn et al., 2015; Muerdter et al., 2013). A substantial challenge that current and future studies face is to understand how this specialised machinery connects with the general cellular pathways. Although factors like Panx or the RDC complex carry out highly specialised roles, they ultimately interface with and feed into some conserved general pathways. For instance, once Panx has specified a locus for silencing, the chromatin environment is remodelled by Egg, a conserved methyl-transferase

whose roles extend beyond transposon control (Rangan et al., 2011; Sienski et al., 2015; Stabell et al., 2006; Yu et al., 2015). Likewise, the dual-strand cluster export machinery involves ubiquitous factors such as Nxt1 and UAP56, which accumulate in Rhi-positive foci. To disentangle the transposon-specific role of these general factors is fundamentally difficult, as their loss has wide-spread effects on cellular functions and null mutants are often not viable. Recent studies on Su(var)2-10, Ovary absent, Nup54 and Nup58 as well as ongoing work in the lab are starting to bridge this gap (Ninova et al., 2019a; Yang et al., 2019), although a precise dissection of these molecular mechanisms remains challenging. Ultimately, understanding this will greatly advance our knowledge of how the evolution of novel molecular machineries is integrated within the general cellular functions.

Another major challenge that lies ahead is that of dissecting the dynamics of each molecular event of the piRNA pathway. Regarding piRNA biogenesis, the current model postulates cycles of PIWI binding to pre-piRNAs, Zuc cleavages and Armi unwinding. Nonetheless, the evidence supporting this model is mostly indirect, therefore calling for additional experimental proof. One avenue to tackle this question would be that of reconstituting piRNA production *in vitro* and precisely dissecting the contribution of each factor. In stark contrast with the elegant *in vitro* studies that have elucidated the mechanisms of RNAi (Bernstein et al., 2001; Denli et al., 2004; Gregory et al., 2004; Park et al., 2011), *in vitro* assays with recombinant Zuc failed to recapitulate the signatures of piRNA production observed *in vivo* (Ipsaro et al., 2012; Nishimasu et al., 2012). This further suggests that future studies aimed at reconstituting fly piRNA production *in vitro* will have to involve a wider number of factors, similarly to what has begun to be possible in *B. mori* extracts (Izumi et al., 2016; Izumi et al., 2020). It is indeed clear that the specificity of Zuc cleavages does not lie within a single factor but likely arises from the interplay of many.

Another outstanding question emerging from my work is the involvement of mitochondria in the piRNA pathway. The localisation of the piRNA biogenesis machinery on these organelles seems a conserved feature and has been found in several animals. Nonetheless, the reason behind this preferential localisation remains a mystery. Some of the factors involved in piRNA production also have an impact on the correct assembly of mitochondria in flies and mice (Huang et al., 2011; Munafò et al., 2019; Olivieri et al., 2010; Zhang et al., 2016), which further underscores the link between mitochondria and piRNA biogenesis, but also poses the question of whether this extends beyond transposon control. Is the functionality of mitochondria important for piRNA biogenesis? Or, conversely, does the presence of the piRNA production

machinery also have an impact on the functionality of germline mitochondria? Because germ cells are a highly specialised cell type, as discussed above, it would not be surprising if also their mitochondria had undergone some specialisation. One additional intriguing feature is that Zuc belongs to the conserved PLD family of phosphodiesterases but, unlike other members of the family, has lost the ability to hydrolyse lipids and functions instead as a nuclease (Ipsaro et al., 2012; Nishimasu et al., 2012). Zuc itself is remarkably conserved (Ipsaro et al., 2012; Nishida et al., 2018; Nishimasu et al., 2012) and similar enzymes are also found in prokaryotes. An example is the bacterial Nup that cleaves nucleic acids *in vitro* (Selvy et al., 2011), thus raising the interesting possibility that an ancestral prokaryotic enzyme present on mitochondria has been adopted for transposon control.

In summary, my thesis work has shed new light onto the molecular mechanisms that govern piRNA-mediated transposon control in *Drosophila.* Given the enormous progress made over the last decade and the emergence of new technology that will allow us to answer more precise questions, I envision that an even deeper understanding of this pathway is within reach.

CHAPTER VII

BIBLIOGRAPHY

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., *et al.* (2000). The genome sequence of Drosophila melanogaster. Science *287*, 2185-2195.

Almeida, M.V., Andrade-Navarro, M.A., and Ketting, R.F. (2019). Function and Evolution of Nematode RNAi Pathways. Noncoding RNA 5.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Andersen, P.R., Tirian, L., Vunjak, M., and Brennecke, J. (2017). A heterochromatin-dependent transcription machinery drives piRNA expression. Nature *549*, 54-59.

Ang, Y.S., and Yung, L.Y. (2016). Rational design of hybridization chain reaction monomers for robust signal amplification. Chem Commun (Camb) *52*, 4219-4222.

Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M.J., Kuramochi-Miyagawa, S., Nakano, T., *et al.* (2006). A novel class of small RNAs bind to MILI protein in mouse testes. Nature *442*, 203-207.

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

Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K., and Hannon, G.J. (2007). Developmentally regulated piRNA clusters implicate MILI in transposon control. Science *316*, 744-747.

Bagga, S., Bracht, J., Hunter, S., Massirer, K., Holtz, J., Eachus, R., and Pasquinelli, A.E. (2005). Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. Cell *122*, 553-563.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120-124.

Baron-Benhamou, J., Gehring, N.H., Kulozik, A.E., and Hentze, M.W. (2004). Using the lambdaN peptide to tether proteins to RNAs. Methods Mol Biol *257*, 135-154.

Bastock, R., and St Johnston, D. (2008). Drosophila oogenesis. Curr Biol *18*, R1082-1087.

Batki, J., Schnabl, J., Wang, J., Handler, D., Andreev, V.I., Stieger, C.E., Novatchkova, M., Lampersberger, L., Kauneckaite, K., Xie, W., *et al.* (2019). The nascent RNA binding complex SFiNX licenses piRNA-guided heterochromatin formation. Nat Struct Mol Biol *26*, 720-731.

Baulcombe, D.C. (1996). RNA as a target and an initiator of post-transcriptional gene silencing in transgenic plants. Plant Mol Biol *32*, 79-88.

Beck, M., and Hurt, E. (2017). The nuclear pore complex: understanding its function through structural insight. Nat Rev Mol Cell Biol *18*, 73-89.

Beliveau, B.J., Kishi, J.Y., Nir, G., Sasaki, H.M., Saka, S.K., Nguyen, S.C., Wu, C.T., and Yin, P. (2018). OligoMiner provides a rapid, flexible environment for the design of genome-scale oligonucleotide in situ hybridization probes. Proc Natl Acad Sci U S A *115*, E2183-E2192.

Ben-Yishay, R., Mor, A., Shraga, A., Ashkenazy-Titelman, A., Kinor, N., Schwed-Gross, A., Jacob, A., Kozer, N., Kumar, P., Garini, Y., *et al.* (2019). Imaging within single NPCs reveals NXF1's role in mRNA export on the cytoplasmic side of the pore. J Cell Biol *218*, 2962-2981.

Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., and Golani, I. (2001). Controlling the false discovery rate in behavior genetics research. Behav Brain Res *125*, 279-284.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363-366.

Blumenstiel, J.P. (2019). Birth, School, Work, Death, and Resurrection: The Life Stages and Dynamics of Transposable Element Proliferation. Genes (Basel) *10*.

Boutros, M., and Ahringer, J. (2008). The art and design of genetic screens: RNA interference. Nat Rev Genet *9*, 554-566.

Branon, T.C., Bosch, J.A., Sanchez, A.D., Udeshi, N.D., Svinkina, T., Carr, S.A., Feldman, J.L., Perrimon, N., and Ting, A.Y. (2018). Efficient proximity labeling in living cells and organisms with TurboID. Nat Biotechnol *36*, 880-887.

Braun, I.C., Herold, A., Rode, M., Conti, E., and Izaurralde, E. (2001). Overexpression of TAP/p15 heterodimers bypasses nuclear retention and stimulates nuclear mRNA export. J Biol Chem *276*, 20536-20543.

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

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

Breuer, M., and Ohkura, H. (2015). A negative loop within the nuclear pore complex controls global chromatin organization. Genes Dev *29*, 1789-1794.

Caplen, N.J., Parrish, S., Imani, F., Fire, A., and Morgan, R.A. (2001). Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems. Proc Natl Acad Sci U S A *98*, 9742-9747.

Chang, T.H., Mattei, E., Gainetdinov, I., Colpan, C., Weng, Z., and Zamore, P.D. (2019). Maelstrom Represses Canonical Polymerase II Transcription within Bi-directional piRNA Clusters in Drosophila melanogaster. Mol Cell *73*, 291-303 e296.

Chen, Y.A., Stuwe, E., Luo, Y., Ninova, M., Le Thomas, A., Rozhavskaya, E., Li, S., Vempati, S., Laver, J.D., Patel, D.J.*, et al.* (2016). Cutoff Suppresses RNA Polymerase II Termination to Ensure Expression of piRNA Precursors. Mol Cell *63*, 97-109.

Chen, E.S., Zhang, K., Nicolas, E., Cam, H.P., Zofall, M., and Grewal, S.I. (2008). Cell cycle control of centromeric repeat transcription and heterochromatin assembly. Nature 451, 734-737.

Chenchik, A., Zhu, Y., Diatchenko, L., Li., R., Hill, J. & Siebert, P. (1998). Generation and use of high-quality cDNA from small amounts of total RNA by SMART PCR. RT-PCR Methods for Gene Cloning and Analysis, 305–319.

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.

Choi, H.M., Beck, V.A., and Pierce, N.A. (2014). Multiplexed in situ hybridization using hybridization chain reaction. Zebrafish *11*, 488-489.

Chug, H., Trakhanov, S., Hulsmann, B.B., Pleiner, T., and Gorlich, D. (2015). Crystal structure of the metazoan Nup62*Nup58*Nup54 nucleoporin complex. Science *350*, 106-110.

Cordeiro Rodrigues, R.J., de Jesus Domingues, A.M., Hellmann, S., Dietz, S., de Albuquerque, B.F.M., Renz, C., Ulrich, H.D., Sarkies, P., Butter, F., and Ketting, R.F. (2019). PETISCO is a novel protein complex required for 21U RNA biogenesis and embryonic viability. Genes Dev 33, 857-870.

Cosby, R.L., Chang, N.C., and Feschotte, C. (2019). Host-transposon interactions: conflict, cooperation, and cooption. Genes Dev *33*, 1098-1116.

Cox, D.N., Chao, A., and Lin, H. (2000). piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. Development *127*, 503-514.

Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J.A., Sachidanandam, R.*, et al.* (2008). An endogenous small interfering RNA pathway in Drosophila. Nature *453*, 798-802.

Czech, B., Munafo, M., Ciabrelli, F., Eastwood, E.L., Fabry, M.H., Kneuss, E., and Hannon, G.J. (2018). piRNA-Guided Genome Defense: From Biogenesis to Silencing. Annu Rev Genet *52*, 131-157.

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*, 749-761.

Czech, B., Zhou, R., Erlich, Y., Brennecke, J., Binari, R., Villalta, C., Gordon, A., Perrimon, N., and Hannon, G.J. (2009). Hierarchical rules for Argonaute loading in Drosophila. Mol Cell *36*, 445-456.

Darricarrere, N., Liu, N., Watanabe, T., and Lin, H. (2013). Function of Piwi, a nuclear Piwi/Argonaute protein, is independent of its slicer activity. Proc Natl Acad Sci U S A *110*, 1297-1302.

Daugherty, M.D., and Malik, H.S. (2012). Rules of engagement: molecular insights from host-virus arms races. Annu Rev Genet *46*, 677-700.

Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F., and Hannon, G.J. (2004). Processing of primary microRNAs by the Microprocessor complex. Nature *432*, 231-235.

Dennis, C., Brasset, E., Sarkar, A., and Vaury, C. (2016). Export of piRNA precursors by EJC triggers assembly of cytoplasmic Yb-body in Drosophila. Nat Commun *7*, 13739.

Dennis, C., Brasset, E., and Vaury, C. (2019). flam piRNA precursors channel from the nucleus to the cytoplasm in a temporally regulated manner along Drosophila oogenesis. Mob DNA *10*, 28.

Dennis, C., Zanni, V., Brasset, E., Eymery, A., Zhang, L., Mteirek, R., Jensen, S., Rong, Y.S., and Vaury, C. (2013). "Dot COM", a nuclear transit center for the primary piRNA pathway in Drosophila. PLoS One *8*, e72752.

Derrer, C.P., Mancini, R., Vallotton, P., Huet, S., Weis, K., and Dultz, E. (2019). The RNA export factor Mex67 functions as a mobile nucleoporin. J Cell Biol *218*, 3967-3976.

Desset, S., Buchon, N., Meignin, C., Coiffet, M., and Vaury, C. (2008). In Drosophila melanogaster the COM locus directs the somatic silencing of two retrotransposons through both Piwi-dependent and -independent pathways. PLoS One *3*, e1526.

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

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

Donertas, D., Sienski, G., and Brennecke, J. (2013). Drosophila Gtsf1 is an essential component of the Piwi-mediated transcriptional silencing complex. Genes Dev *27*, 1693-1705.

Doolittle, W.F.S., C. (1980). Selfish genes, the phenotype paradigm and genome evolution. Nature *284*, 601-603.

Drinnenberg, I.A., Weinberg, D.E., Xie, K.T., Mower, J.P., Wolfe, K.H., Fink, G.R., and Bartel, D.P. (2009). RNAi in budding yeast. Science *326*, 544-550.

Duempelmann, L., Skribbe, M., and Buhler, M. (2020). Small RNAs in the Transgenerational Inheritance of Epigenetic Information. Trends Genet 36, 203-214.

Ecker, J.R., and Davis, R.W. (1986). Inhibition of gene expression in plant cells by expression of antisense RNA. Proc Natl Acad Sci U S A *83*, 5372-5376.

Eichhorn, S.W., Guo, H., McGeary, S.E., Rodriguez-Mias, R.A., Shin, C., Baek, D., Hsu, S.H., Ghoshal, K., Villen, J., and Bartel, D.P. (2014). mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. Mol Cell *56*, 104-115.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001a). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature *411*, 494-498.

Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W., and Tuschl, T. (2001b). Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. EMBO J *20*, 6877-6888.

ElMaghraby, M.F., Andersen, P.R., Puhringer, F., Hohmann, U., Meixner, K., Lendl, T., Tirian, L., and Brennecke, J. (2019). A Heterochromatin-Specific RNA Export Pathway Facilitates piRNA Production. Cell *178*, 964-979 e920.

Encell, L.P., Friedman Ohana, R., Zimmerman, K., Otto, P., Vidugiris, G., Wood, M.G., Los, G.V., McDougall, M.G., Zimprich, C., Karassina, N*., et al.* (2012). Development of a dehalogenase-based protein fusion tag capable of rapid, selective and covalent attachment to customizable ligands. Curr Chem Genomics *6*, 55-71.

England, C.G., Luo, H., and Cai, W. (2015). HaloTag technology: a versatile platform for biomedical applications. Bioconjug Chem *26*, 975-986.

Fabry, M.H., Ciabrelli, F., Munafo, M., Eastwood, E.L., Kneuss, E., Falciatori, I., Falconio, F.A., Hannon, G.J., and Czech, B. (2019). piRNA-guided co-transcriptional silencing coopts nuclear export factors. Elife *8*.

Fedoroff, N.V. (2012). Presidential address. Transposable elements, epigenetics, and genome evolution. Science *338*, 758-767.

Filipowicz, W., Bhattacharyya, S.N., and Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat Rev Genet *9*, 102-114.

Findley, S.D., Tamanaha, M., Clegg, N.J., and Ruohola-Baker, H. (2003). Maelstrom, a Drosophila spindle-class gene, encodes a protein that colocalizes with Vasa and RDE1/AGO1 homolog, Aubergine, in nuage. Development *130*, 859-871.

Finlay, D.R., Meier, E., Bradley, P., Horecka, J., and Forbes, D.J. (1991). A complex of nuclear pore proteins required for pore function. J Cell Biol *114*, 169-183.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-811.

Forstemann, K., Horwich, M.D., Wee, L., Tomari, Y., and Zamore, P.D. (2007). Drosophila microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. Cell *130*, 287-297.

Fribourg, S., Braun, I.C., Izaurralde, E., and Conti, E. (2001). Structural basis for the recognition of a nucleoporin FG repeat by the NTF2-like domain of the TAP/p15 mRNA nuclear export factor. Mol Cell *8*, 645-656.

Fukaya, T., Lim, B., and Levine, M. (2017). Rapid Rates of Pol II Elongation in the Drosophila Embryo. Curr Biol *27*, 1387-1391.

Gainetdinov, I., Colpan, C., Arif, A., Cecchini, K., and Zamore, P.D. (2018). A Single Mechanism of Biogenesis, Initiated and Directed by PIWI Proteins, Explains piRNA Production in Most Animals. Mol Cell *71*, 775-790 e775.

Gatto, L., and Lilley, K.S. (2012). MSnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. Bioinformatics *28*, 288-289.

Ge, D.T., Wang, W., Tipping, C., Gainetdinov, I., Weng, Z., and Zamore, P.D. (2019). The RNA-Binding ATPase, Armitage, Couples piRNA Amplification in Nuage to Phased piRNA Production on Mitochondria. Mol Cell *74*, 982-995 e986.

Genzor, P., and Bortvin, A. (2015). A Unique HMG-box Domain of Mouse Maelstrom Binds Structured RNA but Not Double Stranded DNA. PLoS One *10*, e0120268.

Ghildiyal, M., Seitz, H., Horwich, M.D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E.L., Zapp, M.L., Weng, Z.*, et al.* (2008). Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science *320*, 1077-1081.

Gilbert, C., and Feschotte, C. (2018). Horizontal acquisition of transposable elements and viral sequences: patterns and consequences. Curr Opin Genet Dev *49*, 15-24.

Goriaux, C., Desset, S., Renaud, Y., Vaury, C., and Brasset, E. (2014). Transcriptional properties and splicing of the flamenco piRNA cluster. EMBO Rep *15*, 411-418.

Gottlin, E.B., Rudolph, A.E., Zhao, Y., Matthews, H.R., and Dixon, J.E. (1998). Catalytic mechanism of the phospholipase D superfamily proceeds via a covalent phosphohistidine intermediate. Proc Natl Acad Sci U S A *95*, 9202-9207.

Gozalo, A., Duke, A., Lan, Y., Pascual-Garcia, P., Talamas, J.A., Nguyen, S.C., Shah, P.P., Jain, R., Joyce, E.F., and Capelson, M. (2020). Core Components of the Nuclear Pore Bind Distinct States of Chromatin and Contribute to Polycomb Repression. Mol Cell *77*, 67-81 e67.

Grandi, P., Doye, V., and Hurt, E.C. (1993). Purification of NSP1 reveals complex formation with 'GLFG' nucleoporins and a novel nuclear pore protein NIC96. EMBO J *12*, 3061-3071.

Grandi, P., Schlaich, N., Tekotte, H., and Hurt, E.C. (1995). Functional interaction of Nic96p with a core nucleoporin complex consisting of Nsp1p, Nup49p and a novel protein Nup57p. EMBO J *14*, 76-87.

Grant, R.P., Hurt, E., Neuhaus, D., and Stewart, M. (2002). Structure of the C-terminal FG-nucleoporin binding domain of Tap/NXF1. Nat Struct Biol *9*, 247-251.

Gregory, R.I., Yan, K.P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. (2004). The Microprocessor complex mediates the genesis of microRNAs. Nature *432*, 235-240.

Grewal, S.I., and Jia, S. (2007). Heterochromatin revisited. Nat Rev Genet *8*, 35-46.

Gu, J., Wang, M., Yang, Y., Qiu, D., Zhang, Y., Ma, J., Zhou, Y., Hannon, G.J., and Yu, Y. (2018). GoldCLIP: Gel-omitted Ligation-dependent CLIP. Genomics Proteomics Bioinformatics *16*, 136-143.

Guan, T., Muller, S., Klier, G., Pante, N., Blevitt, J.M., Haner, M., Paschal, B., Aebi, U., and Gerace, L. (1995). Structural analysis of the p62 complex, an assembly of O-linked glycoproteins that localizes near the central gated channel of the nuclear pore complex. Mol Biol Cell *6*, 1591-1603.

Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. Science *315*, 1587-1590.

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

Hall, M.P., Unch, J., Binkowski, B.F., Valley, M.P., Butler, B.L., Wood, M.G., Otto, P., Zimmerman, K., Vidugiris, G., Machleidt, T.*, et al.* (2012). Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. ACS Chem Biol *7*, 1848-1857.

Hampoelz, B., Schwarz, A., Ronchi, P., Bragulat-Teixidor, H., Tischer, C., Gaspar, I., Ephrussi, A., Schwab, Y., and Beck, M. (2019). Nuclear Pores Assemble from Nucleoporin Condensates During Oogenesis. Cell *179*, 671-686 e617.

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

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

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

Hannon, G.J., and Rossi, J.J. (2004). Unlocking the potential of the human genome with RNA interference. Nature *431*, 371-378.

Harris, A.N., and Macdonald, P.M. (2001). Aubergine encodes a Drosophila polar granule component required for pole cell formation and related to eIF2C. Development *128*, 2823-2832.

Haussmann, I., Nallasivan, K., Scocchia, D., Civetta, A., and Soller, M. (2019). Channel nuclear pore protein 54 directs sexual differentiation and neuronal wiring required for female reproductive behaviors in Drosophila. biorxiv *DOI: 10.1101/806380*.

Hay, B., Jan, L.Y., and Jan, Y.N. (1990). Localization of vasa, a component of Drosophila polar granules, in maternal-effect mutants that alter embryonic anteroposterior polarity. Development *109*, 425-433.

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

Heitz, E. (1928). Das Heterochromatin der Moose. I Jahrb Wiss Bot *69*, 762–818.

Hell, S.W., and Wichmann, J. (1994). Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. Opt Lett *19*, 780-782.

Herold, A., Klymenko, T., and Izaurralde, E. (2001). NXF1/p15 heterodimers are essential for mRNA nuclear export in Drosophila. RNA *7*, 1768-1780.

Herold, A., Suyama, M., Rodrigues, J.P., Braun, I.C., Kutay, U., Carmo-Fonseca, M., Bork, P., and Izaurralde, E. (2000). TAP (NXF1) belongs to a multigene family of putative RNA export factors with a conserved modular architecture. Mol Cell Biol *20*, 8996-9008.

Herold, A., Teixeira, L., and Izaurralde, E. (2003). Genome-wide analysis of nuclear mRNA export pathways in Drosophila. EMBO J *22*, 2472-2483.

Hirakata, S., Ishizu, H., Fujita, A., Tomoe, Y., and Siomi, M.C. (2019). Requirements for multivalent Yb body assembly in transposon silencing in Drosophila. EMBO Rep *20*, e47708.

Holoch, D., and Moazed, D. (2015). RNA-mediated epigenetic regulation of gene expression. Nat Rev Genet *16*, 71-84.

Homolka, D., Pandey, R.R., Goriaux, C., Brasset, E., Vaury, C., Sachidanandam, R., Fauvarque, M.O., and Pillai, R.S. (2015). PIWI Slicing and RNA Elements in Precursors Instruct Directional Primary piRNA Biogenesis. Cell Rep *12*, 418-428.

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

Hopes T., A.M., Norris K., McCarthy C.G.P., O'Connell M.J., Fontana J., Aspden J.L. (2020). Specialisation of ribosomes in gonads through paralog-switching. biorxiv *DOI: 10.1101/2020.01.20.913020v1*.

Hoskins, R.A., Carlson, J.W., Wan, K.H., Park, S., Mendez, I., Galle, S.E., Booth, B.W., Pfeiffer, B.D., George, R.A., Svirskas, R.*, et al.* (2015). The Release 6 reference sequence of the Drosophila melanogaster genome. Genome Res *25*, 445-458.

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*, RESEARCH0085.

Houwing, S., Kamminga, L.M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Filippov, D.V., Blaser, H., Raz, E., Moens, C.B.*, et al.* (2007). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. Cell *129*, 69-82.

Huang, B., Babcock, H., and Zhuang, X. (2010). Breaking the diffraction barrier: super-resolution imaging of cells. Cell *143*, 1047-1058.

Huang, X.A., Yin, H., Sweeney, S., Raha, D., Snyder, M., and Lin, H. (2013). A major epigenetic programming mechanism guided by piRNAs. Dev Cell *24*, 502-516.

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

Hulsmann, B.B., Labokha, A.A., and Gorlich, D. (2012). The permeability of reconstituted nuclear pores provides direct evidence for the selective phase model. Cell *150*, 738-751.

Huppertz, I., Attig, J., D'Ambrogio, A., Easton, L.E., Sibley, C.R., Sugimoto, Y., Tajnik, M., Konig, J., and Ule, J. (2014). iCLIP: protein-RNA interactions at nucleotide resolution. Methods *65*, 274-287.

Hutvagner, G., and Zamore, P.D. (2002). A microRNA in a multiple-turnover RNAi enzyme complex. Science *297*, 2056-2060.

Iglesias, N., Paulo, J.A., Tatarakis, A., Wang, X., Edwards, A.L., Bhanu, N.V., Garcia, B.A., Haas, W., Gygi, S.P., and Moazed, D. (2020). Native Chromatin Proteomics Reveals a Role for Specific Nucleoporins in Heterochromatin Organization and Maintenance. Mol Cell *77*, 51-66 e58.

Inoue, A., Jiang, L., Lu, F., Suzuki, T., and Zhang, Y. (2017). Maternal H3K27me3 controls DNA methylation-independent imprinting. Nature *547*, 419-424.

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.

Ishizu, H., Iwasaki, Y.W., Hirakata, S., Ozaki, H., Iwasaki, W., Siomi, H., and Siomi, M.C. (2015). Somatic Primary piRNA Biogenesis Driven by cis-Acting RNA Elements and trans-Acting Yb. Cell Rep *12*, 429-440.

Ishizu, H., Kinoshita, T., Hirakata, S., Komatsuzaki, C., and Siomi, M.C. (2019). Distinct and Collaborative Functions of Yb and Armitage in Transposon-Targeting piRNA Biogenesis. Cell Rep *27*, 1822-1835 e1828.

Iwasaki, Y.W., Murano, K., Ishizu, H., Shibuya, A., Iyoda, Y., Siomi, M.C., Siomi, H., and Saito, K. (2016). Piwi Modulates Chromatin Accessibility by Regulating Multiple Factors Including Histone H1 to Repress Transposons. Mol Cell *63*, 408-419.

Izumi, N., Shoji, K., Sakaguchi, Y., Honda, S., Kirino, Y., Suzuki, T., Katsuma, S., and Tomari, Y. (2016). Identification and Functional Analysis of the Pre-piRNA 3' Trimmer in Silkworms. Cell *164*, 962-973.

Izumi, N., Shoji, K., Suzuki, Y., Katsuma, S., and Tomari, Y. (2020). Zucchini consensus motifs determine the mechanism of pre-piRNA production. Nature.

J. Yuyang Lu, L.C., Tong Li, Ting Wang, Yafei Yin, Ge Zhan, Ke Zhang, Michelle Percharde, Liang Wang, Qi Peng, Pixi Yan, Hui Zhang, Xue Han, Xianju Bi, Wen Shao, Yantao Hong, Zhongyang Wu, Peizhe Wang, Wenzhi Li, Jing Zhang, Zai Chang, Yingping Hou, Pilong Li,

Miguel Ramalho-Santos, Jie Na, Wei Xie, Yujie Sun, Xiaohua Shen (2019). L1 and B1 repeats blueprint the spatial organization of chromatin. biorxiv, doi: https://doi.org/10.1101/802173.

Jacobs, F.M., Greenberg, D., Nguyen, N., Haeussler, M., Ewing, A.D., Katzman, S., Paten, B., Salama, S.R., and Haussler, D. (2014). An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. Nature *516*, 242-245.

Jih, G., Iglesias, N., Currie, M.A., Bhanu, N.V., Paulo, J.A., Gygi, S.P., Garcia, B.A., and Moazed, D. (2017). Unique roles for histone H3K9me states in RNAi and heritable silencing of transcription. Nature *547*, 463-467.

Jinek, M., and Doudna, J.A. (2009). A three-dimensional view of the molecular machinery of RNA interference. Nature *457*, 405-412.

Kalverda, B., and Fornerod, M. (2010). Characterization of genome-nucleoporin interactions in Drosophila links chromatin insulators to the nuclear pore complex. Cell Cycle *9*, 4812-4817.

Kawamura, Y., Saito, K., Kin, T., Ono, Y., Asai, K., Sunohara, T., Okada, T.N., Siomi, M.C., and Siomi, H. (2008). Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. Nature *453*, 793-797.

Kelleher, E.S., Edelman, N.B., and Barbash, D.A. (2012). Drosophila interspecific hybrids phenocopy piRNA-pathway mutants. PLoS Biol *10*, e1001428.

Kennerdell, J.R., Yamaguchi, S., and Carthew, R.W. (2002). RNAi is activated during Drosophila oocyte maturation in a manner dependent on aubergine and spindle-E. Genes Dev *16*, 1884-1889.

Keryer-Bibens, C., Barreau, C., and Osborne, H.B. (2008). Tethering of proteins to RNAs by bacteriophage proteins. Biol Cell *100*, 125-138.

Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes Dev *15*, 2654-2659.

Khvorova, A., Reynolds, A., and Jayasena, S.D. (2003). Functional siRNAs and miRNAs exhibit strand bias. Cell *115*, 209-216.

Kim, A., Terzian, C., Santamaria, P., Pelisson, A., Purd'homme, N., and Bucheton, A. (1994). Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of Drosophila melanogaster. Proc Natl Acad Sci U S A *91*, 1285-1289.

Kim, D.I., Birendra, K.C., Zhu, W., Motamedchaboki, K., Doye, V., and Roux, K.J. (2014). Probing nuclear pore complex architecture with proximity-dependent biotinylation. Proc Natl Acad Sci U S A *111*, E2453-2461.

Kim, D.I., Jensen, S.C., Noble, K.A., Kc, B., Roux, K.H., Motamedchaboki, K., and Roux, K.J. (2016). An improved smaller biotin ligase for BioID proximity labeling. Mol Biol Cell *27*, 1188-1196.

Kim, S.J., Fernandez-Martinez, J., Nudelman, I., Shi, Y., Zhang, W., Raveh, B., Herricks, T., Slaughter, B.D., Hogan, J.A., Upla, P.*, et al.* (2018). Integrative structure and functional anatomy of a nuclear pore complex. Nature *555*, 475-482.

Kim, V.N. (2005). MicroRNA biogenesis: coordinated cropping and dicing. Nat Rev Mol Cell Biol *6*, 376-385.

King, F.J., Szakmary, A., Cox, D.N., and Lin, H. (2001). Yb modulates the divisions of both germline and somatic stem cells through piwi- and hh-mediated mechanisms in the Drosophila ovary. Mol Cell *7*, 497-508.

Kirino, Y., Kim, N., de Planell-Saguer, M., Khandros, E., Chiorean, S., Klein, P.S., Rigoutsos, I., Jongens, T.A., and Mourelatos, Z. (2009). Arginine methylation of Piwi proteins catalysed by dPRMT5 is required for Ago3 and Aub stability. Nat Cell Biol *11*, 652-658.

Kishi, J.Y., Lapan, S.W., Beliveau, B.J., West, E.R., Zhu, A., Sasaki, H.M., Saka, S.K., Wang, Y., Cepko, C.L., and Yin, P. (2019). SABER amplifies FISH: enhanced multiplexed imaging of RNA and DNA in cells and tissues. Nat Methods *16*, 533-544.

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.

Klenov, M.S., Lavrov, S.A., Korbut, A.P., Stolyarenko, A.D., Yakushev, E.Y., Reuter, M., Pillai, R.S., and Gvozdev, V.A. (2014). Impact of nuclear Piwi elimination on chromatin state in Drosophila melanogaster ovaries. Nucleic Acids Res *42*, 6208-6218.

Kneuss, E., Munafo, M., Eastwood, E.L., Deumer, U.S., Preall, J.B., Hannon, G.J., and Czech, B. (2019). Specialization of the Drosophila nuclear export family protein Nxf3 for piRNA precursor export. Genes Dev *33*, 1208-1220.

Kohler, A., and Hurt, E. (2007). Exporting RNA from the nucleus to the cytoplasm. Nat Rev Mol Cell Biol *8*, 761-773.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Kruse, K.D., N.; Enriquez-Gasca, R.; Gaume, X.; Torres-Padilla, M. E.; Vaquerizas, J. M. (2019). Transposable elements drive reorganisation of 3D chromatin during early embryogenesis. biorxiv, DOI: 10.1101/523712.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116-120.

Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S., and Narlikar, G.J. (2017). Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. Nature *547*, 236-240.

Lasko, P.F., and Ashburner, M. (1990). Posterior localization of vasa protein correlates with, but is not sufficient for, pole cell development. Genes Dev *4*, 905-921.

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

Le Thomas, A., Stuwe, E., Li, S., Du, J., Marinov, G., Rozhkov, N., Chen, Y.C.A., Luo, Y., Sachidanandam, R., Toth, K.F*., et al.* (2014). Transgenerationally Inherited piRNAs Trigger piRNA Biogenesis by Changing the Chromatin of piRNA Clusters and Inducing Precursor Processing. Genes Dev *28*, 1667-1680.

Lecher, P., Bucheton, A., and Pelisson, A. (1997). Expression of the Drosophila retrovirus gypsy as ultrastructurally detectable particles in the ovaries of flies carrying a permissive flamenco allele. J Gen Virol *78 (Pt 9)*, 2379-2388.

Lee, F.C.Y., and Ule, J. (2018). Advances in CLIP Technologies for Studies of Protein-RNA Interactions. Mol Cell *69*, 354-369.

Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., and Carthew, R.W. (2004). Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. Cell *117*, 69-81.

Levesque, L., Guzik, B., Guan, T., Coyle, J., Black, B.E., Rekosh, D., Hammarskjold, M.L., and Paschal, B.M. (2001). RNA export mediated by tap involves NXT1-dependent interactions with the nuclear pore complex. J Biol Chem *276*, 44953-44962.

Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. Cell *115*, 787-798.

Lewis, S.H., Quarles, K.A., Yang, Y., Tanguy, M., Frezal, L., Smith, S.A., Sharma, P.P., Cordaux, R., Gilbert, C., Giraud, I., et al. (2018). Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. Nat Ecol Evol 2, 174-181.

Li, X.Z., Roy, C.K., Dong, X., Bolcun-Filas, E., Wang, J., Han, B.W., Xu, J., Moore, M.J., Schimenti, J.C., Weng, Z., et al. (2013). An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. Mol Cell 50, 67-81.

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.* (2009a). Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. Cell *137*, 509-521.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009b). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Li, M.A., Alls, J.D., Avancini, R.M., Koo, K., and Godt, D. (2003). The large Maf factor Traffic Jam controls gonad morphogenesis in Drosophila. Nat Cell Biol *5*, 994-1000.

Li, X.Z., Roy, C.K., Dong, X., Bolcun-Filas, E., Wang, J., Han, B.W., Xu, J., Moore, M.J., Schimenti, J.C., Weng, Z*., et al.* (2013). An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. Mol Cell *50*, 67-81.

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

Liker, E., Fernandez, E., Izaurralde, E., and Conti, E. (2000). The structure of the mRNA export factor TAP reveals a cis arrangement of a non-canonical RNP domain and an LRR domain. EMBO J *19*, 5587-5598.

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

Lingel, A., Simon, B., Izaurralde, E., and Sattler, M. (2003). Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain. Nature *426*, 465-469.

Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science *305*, 1437-1441.

Liu, J., Rivas, F.V., Wohlschlegel, J., Yates, J.R., 3rd, Parker, R., and Hannon, G.J. (2005). A role for the P-body component GW182 in microRNA function. Nat Cell Biol *7*, 1261-1266.

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

Liu, Q., Rand, T.A., Kalidas, S., Du, F., Kim, H.E., Smith, D.P., and Wang, X. (2003). R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. Science *301*, 1921-1925.

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.

Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E., and Kutay, U. (2004). Nuclear export of microRNA precursors. Science *303*, 95-98.

Ma, J.B., Ye, K., and Patel, D.J. (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. Nature *429*, 318-322.

Ma, J.B., Yuan, Y.R., Meister, G., Pei, Y., Tuschl, T., and Patel, D.J. (2005). Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. Nature *434*, 666-670.

Macrae, I.J., Zhou, K., Li, F., Repic, A., Brooks, A.N., Cande, W.Z., Adams, P.D., and Doudna, J.A. (2006). Structural basis for double-stranded RNA processing by Dicer. Science *311*, 195-198.

Malone, C.D., Falkowska, K.A., Li, A.Y., Galanti, S.E., Kanuru, R.C., LaMont, E.G., Mazzarella, K.C., Micev, A.J., Osman, M.M., Piotrowski, N.K., et al. (2008). Nucleus-specific importin alpha proteins and nucleoporins regulate protein import and nuclear division in the binucleate Tetrahymena thermophila. Eukaryot Cell 7, 1487-1499.

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.

Martienssen, R., and Moazed, D. (2015). RNAi and heterochromatin assembly. Cold Spring Harb Perspect Biol 7, a019323.

Matranga, C., Tomari, Y., Shin, C., Bartel, D.P., and Zamore, P.D. (2005). Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. Cell *123*, 607-620.

Matsumoto, N., Sato, K., Nishimasu, H., Namba, Y., Miyakubi, K., Dohmae, N., Ishitani, R., Siomi, H., Siomi, M.C., and Nureki, O. (2015). Crystal Structure and Activity of the Endoribonuclease Domain of the piRNA Pathway Factor Maelstrom. Cell Rep *11*, 366-375.

McGinn, J., and Czech, B. (2014). Small RNA library construction for high-throughput sequencing. Methods Mol Biol *1093*, 195-208.

Meister, G. (2013). Argonaute proteins: functional insights and emerging roles. Nat Rev Genet *14*, 447-459.

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. (2004). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol Cell *15*, 185-197.

Mevel-Ninio, M., Pelisson, A., Kinder, J., Campos, A.R., and Bucheton, A. (2007). The flamenco locus controls the gypsy and ZAM retroviruses and is required for Drosophila oogenesis. Genetics *175*, 1615-1624.

mod, E.C., Roy, S., Ernst, J., Kharchenko, P.V., Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C.A., Ma, L.*, et al.* (2010). Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science *330*, 1787-1797.

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

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

Molaro, A., Falciatori, I., Hodges, E., Aravin, A.A., Marran, K., Rafii, S., McCombie, W.R., Smith, A.D., and Hannon, G.J. (2014). Two waves of de novo methylation during mouse germ cell development. Genes Dev *28*, 1544-1549.

Mouse Genome Sequencing, C., Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M.*, et al.* (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.

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

Muerdter, F., Olovnikov, I., Molaro, A., Rozhkov, N.V., Czech, B., Gordon, A., Hannon, G.J., and Aravin, A.A. (2012). Production of artificial piRNAs in flies and mice. RNA *18*, 42-52.

Munafo, M., Manelli, V., Falconio, F.A., Sawle, A., Kneuss, E., Eastwood, E.L., Seah, J.W.E., Czech, B., and Hannon, G.J. (2019). Daedalus and Gasz recruit Armitage to mitochondria, bringing piRNA precursors to the biogenesis machinery. Genes Dev *33*, 844-856.

Murano, K., Iwasaki, Y.W., Ishizu, H., Mashiko, A., Shibuya, A., Kondo, S., Adachi, S., Suzuki, S., Saito, K., Natsume, T.*, et al.* (2019). Nuclear RNA export factor variant initiates piRNA-guided co-transcriptional silencing. EMBO J *38*, e102870.

Murota, Y., Ishizu, H., Nakagawa, S., Iwasaki, Y.W., Shibata, S., Kamatani, M.K., Saito, K., Okano, H., Siomi, H., and Siomi, M.C. (2014). Yb integrates piRNA intermediates and processing factors into perinuclear bodies to enhance piRISC assembly. Cell Rep *8*, 103-113.

Nagao, A., Mituyama, T., Huang, H., Chen, D., Siomi, M.C., and Siomi, H. (2010). Biogenesis pathways of piRNAs loaded onto AGO3 in the Drosophila testis. RNA *16*, 2503-2515.

Nakanishi, K., Weinberg, D.E., Bartel, D.P., and Patel, D.J. (2012). Structure of yeast Argonaute with guide RNA. Nature *486*, 368-374.

Napoli, C., Lemieux, C., and Jorgensen, R. (1990). Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. Plant Cell *2*, 279-289.

Ni, J.Q., Liu, L.P., Binari, R., Hardy, R., Shim, H.S., Cavallaro, A., Booker, M., Pfeiffer, B.D., Markstein, M., Wang, H.*, et al.* (2009). A Drosophila resource of transgenic RNAi lines for neurogenetics. Genetics *182*, 1089-1100.

Ni, J.Q., Markstein, M., Binari, R., Pfeiffer, B., Liu, L.P., Villalta, C., Booker, M., Perkins, L., and Perrimon, N. (2008). Vector and parameters for targeted transgenic RNA interference in Drosophila melanogaster. Nat Methods *5*, 49-51.

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 U S A *103*, 16325-16330.

Ninova, M., Chen, Y.A., Godneeva, B., Rogers, A.K., Luo, Y., Fejes Toth, K., and Aravin, A.A. (2019a). Su(var)2-10 and the SUMO Pathway Link piRNA-Guided Target Recognition to Chromatin Silencing. Mol Cell.

Ninova, M., Godneeva, B., Chen, Y.A., Luo, Y., Prakash, S.J., Jankovics, F., Erdelyi, M., Aravin, A.A., and Fejes Toth, K. (2019b). The SUMO Ligase Su(var)2-10 Controls Hetero- and Euchromatic Gene Expression via Establishing H3K9 Trimethylation and Negative Feedback Regulation. Mol Cell.

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

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

Noto, T., and Mochizuki, K. (2017). Whats, hows and whys of programmed DNA elimination in Tetrahymena. Open Biol 7.

Ohtani, H., Iwasaki, Y.W., Shibuya, A., Siomi, H., Siomi, M.C., and Saito, K. (2013). DmGTSF1 is necessary for Piwi-piRISC-mediated transcriptional transposon silencing in the Drosophila ovary. Genes Dev *27*, 1656-1661.

Okamura, K., Chung, W.J., Ruby, J.G., Guo, H., Bartel, D.P., and Lai, E.C. (2008). The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. Nature *453*, 803-806.

Okamura, K., Liu, N., and Lai, E.C. (2009). Distinct mechanisms for microRNA strand selection by Drosophila Argonautes. Mol Cell *36*, 431-444.

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.

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.

Orgel, L.E.C., F. H. C. (1980). Selfish DNA: the ultimate parasite. Nature *284*, 604-607.

Osumi, K., Sato, K., Murano, K., Siomi, H., and Siomi, M.C. (2019). Essential roles of Windei and nuclear monoubiquitination of Eggless/SETDB1 in transposon silencing. EMBO Rep *20*, e48296.

Ozata, D.M., Gainetdinov, I., Zoch, A., O'Carroll, D., and Zamore, P.D. (2019). PIWI-interacting RNAs: small RNAs with big functions. Nat Rev Genet 20, 89-108.

Ozata, D.M., Yu, T., Mou, H., Gainetdinov, I., Colpan, C., Cecchini, K., Kaymaz, Y., Wu, P.H., Fan, K., Kucukural, A., et al. (2020). Evolutionarily conserved pachytene piRNA loci are highly divergent among modern humans. Nat Ecol Evol 4, 156-168.

Pan, J., Eckardt, S., Leu, N.A., Buffone, M.G., Zhou, J., Gerton, G.L., McLaughlin, K.J., and Wang, P.J. (2009). Inactivation of Nxf2 causes defects in male meiosis and age-dependent depletion of spermatogonia. Dev Biol *330*, 167-174.

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

Pane, A., Jiang, P., Zhao, D.Y., Singh, M., and Schupbach, T. (2011). The Cutoff protein regulates piRNA cluster expression and piRNA production in the Drosophila germline. EMBO J *30*, 4601-4615.

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

Papachristou, E.K., Kishore, K., Holding, A.N., Harvey, K., Roumeliotis, T.I., Chilamakuri, C.S.R., Omarjee, S., Chia, K.M., Swarbrick, A., Lim, E.*, et al.* (2018). A quantitative mass spectrometry-based approach to monitor the dynamics of endogenous chromatin-associated protein complexes. Nat Commun *9*, 2311.

Pardue, M.L., and DeBaryshe, P.G. (2008). Drosophila telomeres: A variation on the telomerase theme. Fly (Austin) *2*, 101-110.

Parhad, S.S., Tu, S., Weng, Z., and Theurkauf, W.E. (2017). Adaptive Evolution Leads to Cross-Species Incompatibility in the piRNA Transposon Silencing Machinery. Dev Cell *43*, 60-70 e65.

Park, J.E., Heo, I., Tian, Y., Simanshu, D.K., Chang, H., Jee, D., Patel, D.J., and Kim, V.N. (2011). Dicer recognizes the 5' end of RNA for efficient and accurate processing. Nature *475*, 201-205.

Parker, J.S., Roe, S.M., and Barford, D. (2005). Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. Nature *434*, 663-666.

Pascual-Garcia, P., Debo, B., Aleman, J.R., Talamas, J.A., Lan, Y., Nguyen, N.H., Won, K.J., and Capelson, M. (2017). Metazoan Nuclear Pores Provide a Scaffold for Poised Genes and Mediate Induced Enhancer-Promoter Contacts. Mol Cell *66*, 63-76 e66.

Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnan, B., Muller, P.*, et al.* (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature *408*, 86-89.

Patel, S.S., Belmont, B.J., Sante, J.M., and Rexach, M.F. (2007). Natively unfolded nucleoporins gate protein diffusion across the nuclear pore complex. Cell *129*, 83-96.

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

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

Payer, L.M., and Burns, K.H. (2019). Transposable elements in human genetic disease. Nat Rev Genet *20*, 760-772.

Pelisson, A., Song, S.U., Prud'homme, N., Smith, P.A., Bucheton, A., and Corces, V.G. (1994). Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the Drosophila flamenco gene. EMBO J *13*, 4401-4411.

Perkins, L.A., Holderbaum, L., Tao, R., Hu, Y., Sopko, R., McCall, K., Yang-Zhou, D., Flockhart, I., Binari, R., Shim, H.S.*, et al.* (2015). The Transgenic RNAi Project at Harvard Medical School: Resources and Validation. Genetics *201*, 843-852.

Port, F., Chen, H.M., Lee, T., and Bullock, S.L. (2014). Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. Proc Natl Acad Sci U S A *111*, E2967-2976.

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.

Presgraves, D.C., Balagopalan, L., Abmayr, S.M., and Orr, H.A. (2003). Adaptive evolution drives divergence of a hybrid inviability gene between two species of Drosophila. Nature *423*, 715-719.

Presgraves, D.C., and Stephan, W. (2007). Pervasive adaptive evolution among interactors of the Drosophila hybrid inviability gene, Nup96. Mol Biol Evol 24, 306-314.

Prud'homme, N., Gans, M., Masson, M., Terzian, C., and Bucheton, A. (1995). Flamenco, a gene controlling the gypsy retrovirus of Drosophila melanogaster. Genetics *139*, 697-711.

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.

Ramanathan, M., Majzoub, K., Rao, D.S., Neela, P.H., Zarnegar, B.J., Mondal, S., Roth, J.G., Gai, H., Kovalski, J.R., Siprashvili, Z.*, et al.* (2018). RNA-protein interaction detection in living cells. Nat Methods *15*, 207-212.

Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res *44*, W160-165.

Rangan, P., Malone, C.D., Navarro, C., Newbold, S.P., Hayes, P.S., Sachidanandam, R., Hannon, G.J., and Lehmann, R. (2011). piRNA production requires heterochromatin formation in Drosophila. Curr Biol *21*, 1373-1379.

Rehwinkel, J., Behm-Ansmant, I., Gatfield, D., and Izaurralde, E. (2005). A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. RNA *11*, 1640-1647.

Robine, N., Lau, N.C., Balla, S., Jin, Z., Okamura, K., Kuramochi-Miyagawa, S., Blower, M.D., and Lai, E.C. (2009). A broadly conserved pathway generates 3'UTR-directed primary piRNAs. Curr Biol *19*, 2066-2076.

Rodriguez-Terrones, D., and Torres-Padilla, M.E. (2018). Nimble and Ready to Mingle: Transposon Outbursts of Early Development. Trends Genet *34*, 806-820.

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

Romano, N., and Macino, G. (1992). Quelling: transient inactivation of gene expression in Neurospora crassa by transformation with homologous sequences. Mol Microbiol *6*, 3343-3353.

Ross, R.J., Weiner, M.M., and Lin, H. (2014). PIWI proteins and PIWI-interacting RNAs in the soma. Nature 505, 353-359.

Roth, S., and Lynch, J.A. (2009). Symmetry breaking during Drosophila oogenesis. Cold Spring Harb Perspect Biol *1*, a001891.

Roush, S., and Slack, F.J. (2008). The let-7 family of microRNAs. Trends Cell Biol *18*, 505-516.

Roux, K.J., Kim, D.I., Raida, M., and Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. J Cell Biol *196*, 801-810.

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. (2014). RNAi and overexpression of genes in ovarian somatic cells. Methods Mol Biol *1093*, 25-33.

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-Groschene, G., Bucheton, A., and 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.

Sato, K., Iwasaki, Y.W., Shibuya, A., Carninci, P., Tsuchizawa, Y., Ishizu, H., Siomi, M.C., and Siomi, H. (2015). Krimper Enforces an Antisense Bias on piRNA Pools by Binding AGO3 in the Drosophila Germline. Mol Cell *59*, 553-563.

Schlaich, N.L., Haner, M., Lustig, A., Aebi, U., and Hurt, E.C. (1997). In vitro reconstitution of a heterotrimeric nucleoporin complex consisting of recombinant Nsp1p, Nup49p, and Nup57p. Mol Biol Cell *8*, 33-46.

Schopp, I.M., Amaya Ramirez, C.C., Debeljak, J., Kreibich, E., Skribbe, M., Wild, K., and Bethune, J. (2017). Split-BioID a conditional proteomics approach to monitor the composition of spatiotemporally defined protein complexes. Nat Commun *8*, 15690.

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

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

Shiromoto, Y., Kuramochi-Miyagawa, S., Daiba, A., Chuma, S., Katanaya, A., Katsumata, A., Nishimura, K., Ohtaka, M., Nakanishi, M., Nakamura, T.*, et al.* (2013). GPAT2, a

mitochondrial outer membrane protein, in piRNA biogenesis in germline stem cells. RNA *19*, 803-810.

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

Sienski, G., Donertas, 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.

Sokolova, O.A., Ilyin, A.A., Poltavets, A.S., Nenasheva, V.V., Mikhaleva, E.A., Shevelyov, Y.Y., and Klenov, M.S. (2019). Yb body assembly on the flamenco piRNA precursor transcripts reduces genic piRNA production. Mol Biol Cell *30*, 1544-1554.

Solmaz, S.R., Chauhan, R., Blobel, G., and Melcak, I. (2011). Molecular architecture of the transport channel of the nuclear pore complex. Cell *147*, 590-602.

Song, J.J., Liu, J., Tolia, N.H., Schneiderman, J., Smith, S.K., Martienssen, R.A., Hannon, G.J., and Joshua-Tor, L. (2003). The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. Nat Struct Biol *10*, 1026-1032.

Stabell, M., Bjorkmo, M., Aalen, R.B., and Lambertsson, A. (2006). The Drosophila SET domain encoding gene dEset is essential for proper development. Hereditas *143*, 177-188.

Standart, N., and Weil, D. (2018). P-Bodies: Cytosolic Droplets for Coordinated mRNA Storage. Trends Genet 34, 612-626.

Stein, C.B., Genzor, P., Mitra, S., Elchert, A.R., Ipsaro, J.J., Benner, L., Sobti, S., Su, Y., Hammell, M., Joshua-Tor, L.*, et al.* (2019). Decoding the 5' nucleotide bias of PIWI-interacting RNAs. Nat Commun *10*, 828.

Strom, A.R., Emelyanov, A.V., Mir, M., Fyodorov, D.V., Darzacq, X., and Karpen, G.H. (2017). Phase separation drives heterochromatin domain formation. Nature *547*, 241-245.

Stuckey, J.A., and Dixon, J.E. (1999). Crystal structure of a phospholipase D family member. Nat Struct Biol *6*, 278-284.

Sydor, A.M., Czymmek, K.J., Puchner, E.M., and Mennella, V. (2015). Super-Resolution Microscopy: From Single Molecules to Supramolecular Assemblies. Trends Cell Biol *25*, 730-748.

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.

Tan, W., Zolotukhin, A.S., Bear, J., Patenaude, D.J., and Felber, B.K. (2000). The mRNA export in Caenorhabditis elegans is mediated by Ce-NXF-1, an ortholog of human TAP/NXF and Saccharomyces cerevisiae Mex67p. RNA *6*, 1762-1772.

Tan, W., Zolotukhin, A.S., Tretyakova, I., Bear, J., Lindtner, S., Smulevitch, S.V., and Felber, B.K. (2005). Identification and characterization of the mouse nuclear export factor (Nxf) family members. Nucleic Acids Res *33*, 3855-3865.

Tang, S., and Presgraves, D.C. (2009). Evolution of the Drosophila nuclear pore complex results in multiple hybrid incompatibilities. Science *323*, 779-782.

Terry, L.J., and Wente, S.R. (2009). Flexible gates: dynamic topologies and functions for FG nucleoporins in nucleocytoplasmic transport. Eukaryot Cell *8*, 1814-1827.

Thevathasan, J.V., Kahnwald, M., Cieslinski, K., Hoess, P., Peneti, S.K., Reitberger, M., Heid, D., Kasuba, K.C., Hoerner, S.J., Li, Y.*, et al.* (2019). Nuclear pores as versatile reference standards for quantitative superresolution microscopy. Nat Methods *16*, 1045-1053.

Tomari, Y., Du, T., and Zamore, P.D. (2007). Sorting of Drosophila small silencing RNAs. Cell *130*, 299-308.

Trojer, P., and Reinberg, D. (2007). Facultative heterochromatin: is there a distinctive molecular signature? Mol Cell *28*, 1-13.

Tutucci, E., and Stutz, F. (2011). Keeping mRNPs in check during assembly and nuclear export. Nat Rev Mol Cell Biol *12*, 377-384.

Vagin, V.V., Hannon, G.J., and Aravin, A.A. (2009a). Arginine methylation as a molecular signature of the Piwi small RNA pathway. Cell Cycle *8*, 4003-4004.

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

Vagin, V.V., Wohlschlegel, J., Qu, J., Jonsson, Z., Huang, X., Chuma, S., Girard, A., Sachidanandam, R., Hannon, G.J., and Aravin, A.A. (2009b). Proteomic analysis of murine Piwi proteins reveals a role for arginine methylation in specifying interaction with Tudor family members. Genes Dev *23*, 1749-1762.

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

Van Valen, L. (1973). A new evolutionary law. Evolutionary Theory *1*, 1-30.

Vermaak, D., Henikoff, S., and Malik, H.S. (2005). Positive selection drives the evolution of rhino, a member of the heterochromatin protein 1 family in Drosophila. PLoS Genet *1*, 96-108.

Vermaak, D., and Malik, H.S. (2009). Multiple roles for heterochromatin protein 1 genes in Drosophila. Annu Rev Genet *43*, 467-492.

Vicidomini, G., Bianchini, P., and Diaspro, A. (2018). STED super-resolved microscopy. Nat Methods *15*, 173-182.

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

Wan, G., Fields, B.D., Spracklin, G., Shukla, A., Phillips, C.M., and Kennedy, S. (2018). Spatiotemporal regulation of liquid-like condensates in epigenetic inheritance. Nature 557, 679-683.

Wang, G., and Reinke, V. (2008). A C. elegans Piwi, PRG-1, regulates 21U-RNAs during spermatogenesis. Curr Biol 18, 861-867.

Wang, J., Jia, S.T., and Jia, S. (2016). New Insights into the Regulation of Heterochromatin. Trends Genet *32*, 284-294.

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 U S A *108*, 21164-21169.

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

Wang, X.H., Aliyari, R., Li, W.X., Li, H.W., Kim, K., Carthew, R., Atkinson, P., and Ding, S.W. (2006). RNA interference directs innate immunity against viruses in adult Drosophila. Science *312*, 452-454.

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

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

Weick, E.M., and Miska, E.A. (2014). piRNAs: from biogenesis to function. Development 141, 3458-3471.

Xia, B., Yan, Y., Baron, M., Wagner, F., Barkley, D., Chiodin, M., Kim, S.Y., Keefe, D.L., Alukal, J.P., Boeke, J.D.*, et al.* (2020). Widespread Transcriptional Scanning in the Testis Modulates Gene Evolution Rates. Cell *180*, 248-262 e221.

Yamashiro, H., Negishi, M., Kinoshita, T., Ishizu, H., Ohtani, H., and Siomi, M.C. (2019). Armitage determines Piwi-piRISC processing from precursor formation and quality control to inter-organelle translocation. EMBO Rep, e48769.

Yan, K.S., Yan, S., Farooq, A., Han, A., Zeng, L., and Zhou, M.M. (2003). Structure and conserved RNA binding of the PAZ domain. Nature *426*, 468-474.

Yang, F., Quan, Z., Huang, H., He, M., Liu, X., Cai, T., and Xi, R. (2019). Ovaries absent links dLsd1 to HP1a for local H3K4 demethylation required for heterochromatic gene silencing. Elife *8*.

Yang, J., Bogerd, H.P., Wang, P.J., Page, D.C., and Cullen, B.R. (2001). Two closely related human nuclear export factors utilize entirely distinct export pathways. Mol Cell *8*, 397-406.

Yashiro, R., Murota, Y., Nishida, K.M., Yamashiro, H., Fujii, K., Ogai, A., Yamanaka, S., Negishi, L., Siomi, H., and Siomi, M.C. (2018). Piwi Nuclear Localization and Its Regulatory Mechanism in Drosophila Ovarian Somatic Cells. Cell Rep *23*, 3647-3657.

Yigit, E., Batista, P.J., Bei, Y., Pang, K.M., Chen, C.C., Tolia, N.H., Joshua-Tor, L., Mitani, S., Simard, M.J., and Mello, C.C. (2006). Analysis of the C. elegans Argonaute family reveals that distinct Argonautes act sequentially during RNAi. Cell *127*, 747-757.

Youngman, E.M., and Claycomb, J.M. (2014). From early lessons to new frontiers: the worm as a treasure trove of small RNA biology. Front Genet *5*, 416.

Yu, B., Lin, Y.A., Parhad, S.S., Jin, Z., Ma, J., Theurkauf, W.E., Zhang, Z.Z., and Huang, Y. (2018). Structural insights into Rhino-Deadlock complex for germline piRNA cluster specification. EMBO Rep *19*.

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

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.

Zanni, V., Eymery, A., Coiffet, M., Zytnicki, M., Luyten, I., Quesneville, H., Vaury, C., and Jensen, S. (2013). Distribution, evolution, and diversity of retrotransposons at the flamenco locus reflect the regulatory properties of piRNA clusters. Proc Natl Acad Sci U S A *110*, 19842-19847.

Zhang, F., Wang, J., Xu, J., Zhang, Z., Koppetsch, B.S., Schultz, N., Vreven, 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.

Zhang, G., Tu, S., Yu, T., Zhang, X.O., Parhad, S.S., Weng, Z., and Theurkauf, W.E. (2018a). Co-dependent Assembly of Drosophila piRNA Precursor Complexes and piRNA Cluster Heterochromatin. Cell Rep *24*, 3413-3422 e3414.

Zhang, H., Kolb, F.A., Jaskiewicz, L., Westhof, E., and Filipowicz, W. (2004). Single processing center models for human Dicer and bacterial RNase III. Cell *118*, 57-68.

Zhang, J., Wang, Q., Wang, M., Jiang, M., Wang, Y., Sun, Y., Wang, J., Xie, T., Tang, C., Tang, N.*, et al.* (2016). GASZ and mitofusin-mediated mitochondrial functions are crucial for spermatogenesis. EMBO Rep *17*, 220-234.

Zhang, Y., Li, T., Preissl, S., Amaral, M.L., Grinstein, J.D., Farah, E.N., Destici, E., Qiu, Y., Hu, R., Lee, A.Y.*, et al.* (2019). Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. Nat Genet *51*, 1380-1388.

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

Zhang, Z., Koppetsch, B.S., Wang, J., Tipping, C., Weng, Z., Theurkauf, W.E., and Zamore, P.D. (2014a). Antisense piRNA amplification, but not piRNA production or nuage assembly, requires the Tudor-domain protein Qin. EMBO J *33*, 536-539.

Zhang, Z., Wang, J., Schultz, N., Zhang, F., Parhad, S.S., Tu, S., Vreven, T., Zamore, P.D., Weng, Z., and Theurkauf, W.E. (2014b). The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. Cell *157*, 1353-1363.

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

Zhao, K., Cheng, S., Miao, N., Xu, P., Lu, X., Zhang, Y., Wang, M., Ouyang, X., Yuan, X., Liu, W.*, et al.* (2019). A Pandas complex adapted for piRNA-guided transcriptional silencing and heterochromatin formation. Nat Cell Biol *21*, 1261-1272.

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

Zhou, R., Czech, B., Brennecke, J., Sachidanandam, R., Wohlschlegel, J.A., Perrimon, N., and Hannon, G.J. (2009). Processing of Drosophila endo-siRNAs depends on a specific Loquacious isoform. RNA *15*, 1886-1895.

CHAPTER VIII

APPENDICES

VIII.1 Fiji custom scripts for image processing

The following script is a representative example of those used to process immunofluorescence images. The same dynamic range has been applied to individual channels across all samples. For each channel, a monochromatic inverted image has been exported in RGB format and a custom LUT has been applied for the merge. No other filters have been applied to the channels.

```
run("Duplicate...", "duplicate");
img=getTitle()

run("Split Channels");

selectWindow("C2-"+img);
run("Duplicate...", " ");
run("Invert LUT");
setMinAndMax(0, 2500);
run("RGB Color");
saveAs("Tiff", "/Users/munafo01/Desktop/out/"+img+"_yb");
close();

selectWindow("C2-"+img);
run("Magenta");
setMinAndMax(0, 2500);

selectWindow("C3-"+img);
run("Duplicate...", " ");
run("Invert LUT");
setMinAndMax(0, 4095);
run("RGB Color");
saveAs("Tiff", "/Users/munafo01/Desktop/out/"+img+"_piwi");
close();

selectWindow("C3-"+img);
```

```
run("Green");
setMinAndMax(0, 4095);

selectWindow("C1-"+img);
run("Blue");
setMinAndMax(0, 4095);

waitForUser("Merge channels, then hit OK")
run("RGB Color");
waitForUser("check RGB, then hit OK")
run("Scale Bar...", "width=5 height=20 font=14 color=White
background=None location=[Lower Right] bold hide");
saveAs("Tiff", "/Users/munafo01/Desktop/out/"+img);
```

The following scripts has been used to extract the nuclear outline from the anti-lamin channel, related to **Figures 5.4-6**.

```
selectWindow("C3-"+img);
setMinAndMax(110, 2100);
run("Gaussian Blur...", "sigma=3");
run("Make Binary");
run("Fill Holes");
run("Outline");
run("Dilate");
run("Dilate");
//apply custom red LUT
run("Edit LUT...");
run("RGB Color");
saveAs("Tiff",
"/Users/munafo01/Desktop/S2_CoIFs_panxnxf2_sep2018/"+img+"_lam
in");
close();
```

VIII.2 Custom scripts for count of transposon reads

The following custom scripts have been used to extract counts for each transposon consensus sequence from RNA-seq and CLIP-seq libraries. For RNA-seq libraries, files flagged as “0” correspond to sense transposon reads, whereas those flagged as “16” correspond to antisense. For stranded CLIP-seq libraries “16” corresponds to sense and “0” to antisense.

```
for i in *.te.Aligned.sortedByCoord.out.bam; do (
        filename=${i%%.te.Aligned.sortedByCoord.out.bam}
        samtools view -f 0x10 -b $i > $filename.0.bam;
        samtools index $filename.0.bam;
        samtools idxstats $filename.0.bam | cut -f 1,3 >
$filename.0.te.chrom_reads.txt;
        ) &
```

```
for i in *.te.Aligned.sortedByCoord.out.bam; do (
        filename=${i%%.te.Aligned.sortedByCoord.out.bam}
        samtools view -F 0x10 -b $i > $filename.16.bam;
        samtools index $filename.16.bam;
        samtools idxstats $filename.16.bam | cut -f 1,3 >
$filename.16.te.chrom_reads.txt;
        ) &
```

VIII.3 R custom scripts for gene differential expression analysis

The following scripts has been used to plot RPM counts for genome-mapped reads and to extract up- and down-regulated genes.

```
library(ggplot2)
library(ggrepel)

dm6_reads_ctrl <-
read.table("*.Aligned.sortedByCoord.out.bam.count.htseq",
as.is =TRUE)
dm6_reads_kd <-
read.table("*.Aligned.sortedByCoord.out.bam.count.htseq",
as.is = TRUE)

dm6_reads <- cbind(dm6_reads_ctrl, dm6_reads_kd[,2])
dm6_reads <- dm6_reads[1:17622,]

dm6_reads[,2] <- dm6_reads[,2]*(1000000/sum(dm6_reads[,2]))
dm6_reads[,3] <- dm6_reads[,3]*(1000000/sum(dm6_reads[,3]))

dm6_reads_rpm <- dm6_reads
rownames(dm6_reads_rpm) <- dm6_reads[,1]
colnames(dm6_reads_rpm) <- c("gene",  "ctrl", "kd")

#Removes all rows with at least one value < 1
dm6_reads_rpm <- dm6_reads_rpm[!rowSums(dm6_reads_rpm < 1),]

#Calculate FC for kd vs ctrl
fc <- as.data.frame(dm6_reads_rpm[,3]/dm6_reads_rpm[,2])
rownames(fc) <- dm6_reads_rpm[,1]
rownames(fc) <- dm6_reads_rpm[,1]
colnames(fc) <- c("kd_ctrl")
write.csv(fc, file="./*_dm6.csv", row.names = TRUE)

dm6_reads_rpm_log2 <- log2(dm6_reads_rpm[2:3])
dm6_reads_rpm_log2 <- data.frame(dm6_reads_rpm_log2,"gene")
dm6_reads_rpm_log2 <-
cbind(rownames(dm6_reads_rpm_log2),dm6_reads_rpm_log2)

colnames(dm6_reads_rpm_log2) <- c( "gene", "ctrl", "kd",
"class")

plot<-
     ggplot(dm6_reads_rpm_log2, aes(x=ctrl, y=kd)) +
```

```
      geom_point(data=dm6_reads_rpm_log2[1:7401,], aes(x=ctrl,
     y=kd), colour="black", pch=16, alpha =0.6 ) +
      geom_point(data=dm6_reads_rpm_log2[grep("\\bflam\\b",
     dm6_reads_rpm_log2[,1]),], aes(x=ctrl, y=kd),
     colour="red", size=2, pch=16) +
     geom_text_repel(data=dm6_reads_rpm_log2[grep("\\bflam\\b"
     , dm6_reads_rpm_log2[,1]),], aes(label=gene,hjust=0,
     vjust=0), colour="red", size=3, pch=1) +
      xlim(-0.1, 15)+ylim(-0.1, 15)+
     xlab("RPM (log2) in siGFP") +
     ylab("RPM (log2) in siX") +

     stat_function(fun=function(x)x, linetype="dotted",
colour="black") +
     stat_function(fun=function(x)x+2, linetype="dotted",
colour="black") +
     stat_function(fun=function(x)x-2, linetype="dotted",
colour="black") +
     theme_bw()

     plot + theme(aspect.ratio=1, legend.position="none")

#Calculate correlation
X <- dm6_reads_rpm_log2[,2]
Y <- dm6_reads_rpm_log2[,3]
cor(X,Y)^2
```

VIII.4 R custom scripts for transposon differential expression analysis

The following scripts has been used to plot RPM counts for transposon-mapped reads.

```
library(ggplot2)
library(ggrepel)

te_class <- read.table("~/te_list_gl_vs_soma_merged.txt",
as.is = TRUE)

te_reads <-
read.table("./ctrl_vs_kd_comb4_SENSE.te.chrom_reads.txt",
as.is = TRUE, header = FALSE)

te_reads <- te_reads[1:122,]
rownames(te_reads) <- te_class[,1]

#Normalize to the appropriate scaling factor, calculated for
each sample as 1 000 000/ (sum of genome-mapped reads)
te_reads[,2] <- te_reads[,2]*(0.014468113)
te_reads[,3] <- te_reads[,3]*(0.016149348)
te_reads <- te_reads[!rowSums(te_reads < 1),]

#Add "class" column
library(stringr)
test <- str_split_fixed(rownames(te_reads), "_", 2)
te_reads[,1] <- test[,1]
te_reads <- cbind(te_reads, test[,2])

#Calculate FC kd vs ctrl
te_reads[,5]<-(te_reads[,3]/te_reads[,2]>4)
colnames(te_reads) <- c("te","ctrl", "kd","class","fc4_te")
te_reads[,2:3] <- log2(te_reads[,2:3])

plot<-
     ggplot(te_reads, aes(x=ctrl, y=kd)) +
     geom_point(data=te_reads, aes(x=ctrl, y=kd),
colour="grey", pch=16, alpha =0.6 ) +

      geom_point(data=te_reads[grep("intermediate",
     te_reads[,4]),], aes(x=ctrl, y=kd), colour="gold",
     size=3, pch=20) +
     geom_point(data=te_reads[grep("soma", te_reads[,4]),],
     aes(x=ctrl, y=kd), colour="green4", size=3, pch=20) +
```

```
     geom_point(data=te_reads[grep("germline",
     te_reads[,4]),], aes(x=ctrl, y=kd), colour="red", size=3,
     pch=20) +
     geom_point(data=te_reads[grep("unknown", te_reads[,4]),],
     aes(x=ctrl, y=kd), colour="black", size=3, pch=20) +

     geom_text_repel(data=te_reads[grep("TRUE",
     te_reads[,5]),], aes(label=te,hjust=0, vjust=0),
     colour="black", size=3, pch=1) +


     xlab("RPM (log2) in sigfp_comb") +
     ylab("RPM (log2) in sinup54_comb")+
     xlim(-0.1, 15)+ylim(-0.1, 15)+

     stat_function(fun=function(x)x, linetype="dotted",
colour="black") +
     stat_function(fun=function(x)x+2, linetype="dotted",
colour="black") +
     stat_function(fun=function(x)x-2, linetype="dotted",
colour="black") +
     theme_bw()

     plot + theme(aspect.ratio=1, legend.position="none")

#Calculate correlation
X <- te_reads[,2]
Y <- te_reads[,3]
cor(X,Y)^2
```

VIII.5 List of antibodies

The following antibodies have been used for Western Blots (with relative dilution factor)

• Piwi	Rabbit	Brennecke et al., 2007	1:5,000
• Yb	Mouse	Saito et al., 2010	1:1,000
• Tubulin	Rabbit	ab18251	1:5,000
• Histone H3	Mouse	ab10799	1:1,000
• HA tag	Rabbit	Cell Signalling #C29F4	1:1,000
• FLAG tag	Mouse	Sigma	1:2,500
• GFP	Chicken	ab13970	1:5,000
• Atp5a	Mouse	ab 14748	1: 1,000
• Streptavidin	IRDye® 800CW	LiCor # 926-32230	1: 4,000
• Panx	Mouse	Sienski et al., 2015	1: 20

The following antibodies have been used for immunofluorescence in ovaries and cells (with relative dilution factor)

• Piwi	Rabbit	Brennecke et al., 2007	1:500
• Aub	Rabbit	Brennecke et al., 2007	1:500
• Ago3	Mouse	Senti et al., 2015 [7B4-C2]	1:500
• Yb	Mouse	Murota et al., 2014	1:500
• Armi	Mouse	Saito et al., 2010	1:500
• Vasa	Rat	DSHB #760351	1:500
• Lamin	Mouse	DSHB #ADL67.10	1:200
• Atp5a	Mouse	ab14748	1:500
• GFP	Chicken	ab13970	1:1,000
• HA tag	Rabbit	ab9110	1:500
• FLAG tag	Mouse	Sigma M2 #F1804	1:500
• FG Nups	Mouse	Biolegend mAb414	1:200
• Streptavidin	Alexa Fluor® 555	Thermo Fisher Scientific S-21381	1:500

VIII.6 List of oligos

The following oligos have been used for RT-qPCR

	FW	REV
rp49	GTCGGATCGATATGCTAAGCTG	CAGATACTGTCCCTTGAAGCG
act5c	GCATCCACGAGACCACCTACAAC	CGGTGATCTCCTTCTGCATACGG
gypsy	AGAAAGTCGCCGTCTACCCTGTA	GTGTGACATTGAGCAGCGTTTCC
mdg1	TATACGAACACTCCACCACCCCA	GGCTTTTCGGATTGGGAGTTGGA
HeT-A	CGCGCGGAACCCATCTTCAGA	CGCCGCAGTCGTTTGGTGAGT
TAHRE	CCCTCTCACACAGCGATCATAGC	GCTTATGTTTGTGGCCTGGCTTG
ZsGreen sensor#1	CTACTTCAAGAACTCCTGCCCCG	GGTACATGCAGTTCTCCTCCACG
ZsGreen sensor #2	CCCCGTGATGAAGAAGATGACCG	CGTCCTTCAGCAGCAGGTACATG
ZsGreen sensor Nascent	GCAGCAGCAAGTACAAGCAAAAAG	TGGCCGAACAAAGACCTTGAAATG

The following siRNAs have been used for OSC knockdowns

	sense	antisense
siRNA-GFP	AGCUGGAGUACAACUACAACA	UUGUAGUUGUACUCCAGCUUG
siRNA-piwi	CGGUCAUGCUGCAGACGAACU	UUCGUCUGCAGCAUGACCGGG
siRNA-zuc	CGAACUUGAUGCACAACAAAU	UUGUUGUGCAUCAAGUUCGUG
siRNA-armi	CGUCGGAUGUGUAGUGGAAUA	UUCCACUACACAUCCGACGUG
siRNA-yb	GGAGCAAGAGAUAAAUUUAAG	UAAAUUUAUCUCUUGCUCCUU
siRNA-daed	AGAAGAAAGUGCAGAGCAACG	UUGCUCUGCACUUUCUUCUGG
siRNA-gasz	GCAGCUACGAUGGAUUCUACU	UAGAAUCCAUCGUAGCUGCUA
siRNA-nup54	CAACAUGCUUGUUCACAUAGA	UAUGUGAACAAGCAUGUUGGG
siRNA-nup58	CGAGAUCACGAACUUGAAAUG	UUUCAAGUUCGUGAUCUCGUG
siRNA-yb[3'UTR]	GCAUUCUCUUGCGAAAGAAAC	UUCUUUCGCAAGAGAAUGCUG
siRNA-nxf1	GCAGUUUAUCAGCGAAGUACG	UACUUCGCUGAUAAACUGCUG
siRNA-nxf2	GGUACUUCACGGAAAUAAACU	UUUAUUUCCGUGAAGUACCAG
siRNA-nxf3	GGUGUUCACUUCGAGUCUACG	UAGACUCGAAGUGAACACCUG
siRNA-nxt1	CCGUCAAGUUCGCAGAUCAGC	UGAUCUGCGAACUUGACGGAG

The following gRNAs have been used to generate KO alleles:

gRNAs against *CG10880* locus:

gGACGCACACAATGTCCGCG GCCTCAAACTGAATGGTCGG

gRNAs against *gasz* locus:

GACTCTGATAGCAGCTACGA GCGTTGAAGTCAATGAACAA

The following probes have been used for *ZsGreen* and *flam* RNA FISH

oMF0512_ZsGreen_hcr_probe1

GCCCTTACTCCCAATTCCaaaaaACTTGTGGCCGTCCACGCAGCCCTCCATGCGGTACTTCATGGTCATCTCCTT

oMF0513_ZsGreen_hcr_probe2

GCCCTTACTCCCAATTCCaaaaaAGTTCTTGAAGTAGTCGACGATGTCCTGGGGGGTACTCGGTGAACACGCGGTT

oMF0514_ZsGreen_hcr_probe3

GCCCTTACTCCCAATTCCaaaaaGTACATGCAGTTCTCCTCCACGCTCACGGTGATGTCGGCGTTGCAGATGCAC

oMF0515_ZsGreen_hcr_probe4

GCCCTTACTCCCAATTCCaaaaaGGCACGGGGATGATCTTCTCGCAGGAGGGCTCCCAGTTGTCGGTCATCTTCT

oMF0516_ZsGreen_hcr_probe5

GCCCTTACTCCCAATTCCaaaaaTGCCACTTCTGGTTCTTGGCGTCGCTGCGGTCCTCGCGGGTCAGCTTGTGCT

oMF0517_ZsGreen_hcr_probe6

GCCCTTACTCCCAATTCCaaaaaGCCAGTCGGGCATCTTGCGGGGCACGGACTTGGCCTTGTACACGGTGTCGAA

oMF0518_ZsGreen_hcr_probe7

GCCCTTACTCCCAATTCCaaaaaGCAGCGCAAGCGGCCACCGTCCTTCAGCAGCAGGTACATGCTCACGTCGCCC

oMF0519_ZsGreen_hcr_probe8

GCCCTTACTCCCAATTCCaaaaaGTTGATGGCCTGCTTGCCCTTGAAGGGGTAGCCGATGCCCTCGCCGGTGATC

RF_EK13 Flam_7_5'

GCCCTTACTCCCAATTCCaaaaaCTAGCTTGCCCTCTGGACCAAACTGGATCTATTTCCTGGACACAGGACCAAA

RF_EK14 Flam_8_5'

GCCCTTACTCCCAATTCCaaaaaGCGCGTTCTCACAACTCGTATTTAGTTTTCGCAATCTACCCGTTTGCCTAAC

RF_EK15 Flam_9_5'

GCCCTTACTCCCAATTCCaaaaaGAGGAAAATGGGATTTTCCTGTAATGGTTTGGTCATACAGCTTGTAGGGAAA

RF_EK16 Flam_10_5'

GCCCTTACTCCCAATTCCaaaaaAGGGAAGAGTTCTATCCGAGAGAACTCGGGCCAATTTATTACCAATTTTTAC

RF_EK18 Flam_18_5'

GCCCTTACTCCCAATTCCaaaaaTCACAATGATATCTAGGACCCTACCAACTCTGGCAAGCTTCCAGCATGACTA

RF_EK20 Flam_20_5'

GCCCTTACTCCCAATTCCaaaaaTAAAGGGTGCCGCCTTTCAGATTTTGATCGATCACAAACCCCTCGTTTGGCT

RF_EK21 Flam_21_5'

GCCCTTACTCCCAATTCCaaaaaTTGAGTTGCCCTAATTTCAAACCAACCATTTCGAGTCACTACTGACGCTAGC

The following probes have been used for *flam* DNA FISH

oligopaint_01 TTGATACCCAATCTGCCACTGAAATCCTCTTCTGTGGtttCAACTTAAC
oligopaint_02 CCGTGTATTTTGCGACACACTGTGGGTTTGGTTTTtttCAACTTAAC
oligopaint_03 TGCAGTTCAAAATTATTCGTTGTTGAAAACGGCTCCGGtttCAACTTAAC
oligopaint_04 GGCCGTGAAAGTGTGCAACTCTCTGTTCGGCtttCAACTTAAC
oligopaint_05 GGGCGATTTTCTTTGATCTGCCGTTGGCGTGtttCAACTTAAC
oligopaint_06 GCATGCAGCATCTGGGAATCATGGTGGACAAtttCAACTTAAC
oligopaint_07 GTCGCCAGATGACCCGTAGTCCACGTAGTTGtttCAACTTAAC
oligopaint_08 GCCAGCCCACCTTACTACCTTACTATGTACAGTGTCAGtttCAACTTAAC
oligopaint_09 AAGAATGGTCGCCGGGTGCATGGTCTCCCAGtttCAACTTAAC
oligopaint_10 GAGATGCACCCTTCGGCCAATGCTGCCCAACtttCAACTTAAC
oligopaint_11 TTCCCACTCCGCGAACAGCTTGTAGATGGCAtttCAACTTAAC
oligopaint_12 TGCACCGGATGATGATCTGCCCATGTTGAATCtttCAACTTAAC
oligopaint_13 CCTCGTTGATATCCATTGGCTCCGAACCCATCTtttCAACTTAAC
oligopaint_14 GCGAGAAGGCTTGGCGCGATTTTATTATTGCAAAAtttCAACTTAAC
oligopaint_15 CTTGGCTGTCATTTTCGAGGTACCCTTGCCCtttCAACTTAAC
oligopaint_16 GGATTCACGGCCGTTGTCACTGTCAACTCGAtttCAACTTAAC
oligopaint_17 TTGCTGTCGATGGTGAAATCGCTGATGGTCAtttCAACTTAAC

oligopaint_18 CCCCAAGAACGCTCTGATGGCCCTCAACGAGtttCAACTTAAC
oligopaint_19 CGGTTCTGTCGCAACTTCTTGTCCCGAGTGCtttCAACTTAAC
oligopaint_20 GCCCATAGCGAAGTTTATCCTTGTCTGAATATGCCTTTGGtttCAACTTAAC
oligopaint_21 CCGAAACGTAGAAGCAAATCGGCTAGGTTCACCtttCAACTTAAC
oligopaint_22 AATCGGAAGATGGTGCTGTCCTCGGCATCTAtttCAACTTAAC
oligopaint_23 CAAGCGTTGCCAGGACCTAGTGGACACCATGtttCAACTTAAC
oligopaint_24 GCAGACCGGAAATTACTTACAGTTGTGTTGCGAGTtttCAACTTAAC
oligopaint_25 GAAACCTAGTGCTTTGTGTGAGCGTGTGAAAATTCTGGtttCAACTTAAC
oligopaint_26 GAAGTCGTCTGGTAAGCGAGGAAACCTTTGGTGtttCAACTTAAC
oligopaint_27 CTGGCACTGGTACTGGCACTGGTATTGTCCTtttCAACTTAAC
oligopaint_28 GCCGATCTCTGTCAACGATGAACCGTCCGTCtttCAACTTAAC
oligopaint_29 TGATAACTGGCATGGGCAGTTTGGAGGCAATtttCAACTTAAC
oligopaint_30 CCAACTCGTCCGACCAGCAGAATAATCCGGAtttCAACTTAAC
oligopaint_31 AAAATGCGTCCTGTGGCGATTGAACAGCCACtttCAACTTAAC
oligopaint_32 AGCAGTCTTCTCCGGTGGCTTTATATCCGACATCtttCAACTTAAC
oligopaint_33 ATTGGCGGCCTGCAACGGCATTTGGATCAAAtttCAACTTAAC
oligopaint_34 ACTCTTTCGCAAACGCAGAATCCCAATGCCAtttCAACTTAAC
oligopaint_35 GATTAGTTAGCGAAGGCAACGGGTTCCGCCTtttCAACTTAAC
oligopaint_36 CAGTCCGCTGCAAGTGATTTTGTCTAGGGTAGCtttCAACTTAAC
oligopaint_37 AAAAGCGACTCGCTGGCCAACGGAAACCATCtttCAACTTAAC
oligopaint_38 TTTTAAGTAGACCAGCCGACCACCCTGACGGtttCAACTTAAC
oligopaint_39 GGAACTAGCCATGGCCATTCCTGCGACCCAAtttCAACTTAAC
oligopaint_40 AAACGGAGATCCACAGTCGCTGGGCAGTTGGtttCAACTTAAC
oligopaint_41 GCCGAGGACTCCGGGACCAGGTAGTCTATGCtttCAACTTAAC
oligopaint_42 CGATGACCGAAGGATCCCCATAACTATTGTTTCATGCTAAtttCAACTTAAC
oligopaint_43 TCGTTTCGGGCACTTTCGGTGGAGTGAATGGtttCAACTTAAC
oligopaint_44 ACGCTGGACTTGCGGTTATCGCTAACTTTTATGGtttCAACTTAAC
oligopaint_45 CAGTGCTGGCCATATCTAGTTGATCGAAAGCACGtttCAACTTAAC
oligopaint_46 ACATTCGTAACCCACCCATTTGCATTGCACTCATTtttCAACTTAAC
oligopaint_47 GGGGTGCATTGTGAATGCTGGTGGAATTCGGtttCAACTTAAC
oligopaint_48 TGCAATAGCCACTAAAAGCGGGTTAATCGAAAGTGCtttCAACTTAAC
oligopaint_49 AATTGTGAATGAACACTTTCACTTTGTGCGCTGCAAATTtttCAACTTAAC
oligopaint_50 TGACGTCACGAAGATATGTATTCCGAGTGAAAATGCATCAtttCAACTTAAC
oligopaint_51 TGCCTCCAAAGACAAATGGCTTGTTGTTTCCAGCtttCAACTTAAC
oligopaint_52 TGCAGTACGATTCTTTCTGTGACCTGAGCAGTGAAAAtttCAACTTAAC
oligopaint_53 CTGGCCATAGTAATATAAGCAAAGTGATGGAAAGACGGCGtttCAACTTAAC
oligopaint_54 GCAAGTCATGAAAAGTCATGCGCTTATGGTTCGAGTTCTAtttCAACTTAAC
oligopaint_55 TGCCTGTCTGCAGCGTGGGCAAAATCTATAAtttCAACTTAAC
oligopaint_56 CCGTCGCACTCCCACTATCTAAGTAACAGGTATCTTACCtttCAACTTAAC
oligopaint_57 GGTGTTCACACTGTGTTCACAAGAGTTTTGCGTTTCCtttCAACTTAAC
oligopaint_58 GCCTATGCAGCGCAGTTTGACACTGCTTAAATATTCGTAtttCAACTTAAC
oligopaint_59 AGTCCCTTAAGGTTCGCGAGGTTTCAAAGGTCAtttCAACTTAAC
oligopaint_60 GGATCACAAGACGAGTCGAGCTGTCCCATATTCTTTTtttCAACTTAAC
oligopaint_61 GGGAACTCAATACACACATATACCAACCCACTTGCAGAtttCAACTTAAC
oligopaint_62 GCTCTTGTACCGCAGTGAAGAAAACTTTTGCGACAGTATTtttCAACTTAAC
oligopaint_63 TAATTGTTTAGTCGCGTTTCCACATATGCACATACGCACtttCAACTTAAC
oligopaint_64 AACTGAGAGAAATCAAAAGAAGACTCACGAAAGGAACCCAtttCAACTTAAC

oligopaint_65	GCCGGGCGCGTGCGCATTTAAAACAATTCTCtttCAACTTAAC
oligopaint_66	ACACTTACCGCTTGAAAATGTACAAAATCTTGGGTGACGtttCAACTTAAC
oligopaint_67	ATGGGACATTGGAGATTTGTACATGGATGCTAGTGGGtttCAACTTAAC
oligopaint_68	GCGAACTTGACACCCAACCAGTCAGCGAAATTTTAAtttCAACTTAAC
oligopaint_69	ATGTGTGGAATTTGGTTGGGTCACTTTTCTCGATGTGtttCAACTTAAC
oligopaint_70	GGTTTGCTGAGCCGAAAGAATACTGACCTCTGGTAAAAtttCAACTTAAC
oligopaint_71	GCTTATCGGATCATTAGGCTAGTTACCGTTAAGCGGTtttCAACTTAAC
oligopaint_72	CCATTGCATAGATAGCAAAACACGAAACCCCAACTTGTtttCAACTTAAC
oligopaint_73	TTGGACTGGGTGTTAATTATTTCTATGCCGGTTTGCTGCtttCAACTTAAC
oligopaint_74	CCGTTCGCTTGAAAGCTAGGAAGCTAATTGATCAATGACAtttCAACTTAAC

VIII.7 – Supplementary tables

The following tables report proteins significantly enriched in the PL-MS or IP-MS experiments from Chapter III, IV and V.

Table 1 – Relative to Figure 3.10D

Proteins showing $Log_2FC>1.5$ and adjusted P value < 0.05 in BASU-Daed PL-MS. Only proteins with more than one unique peptide are reported.

Accessions	GeneSymbol	Unique_peptides	log2FC	adj.P.Val
CG7082	papi	5	4.75	0.000022
CG1458	Cisd2	3	4.73	0.00001
CG9424	bocks	3	4.38	0.000022
CG5581	Ote	12	4.37	0.00001
CG5014	Vap33	6	4.16	0.000022
CG44154	koi	2	4.03	0.00016
CG10880	CG10880	34	3.96	0.00001
CG3186	eEF5	3	3.79	0.00015
CG4147	Hsc70-3	20	3.67	0.000016
CG43770	Sxl	4	3.51	0.00012
CG10504	Ilk	4	3.47	0.00012
CG5508	mino	8	3.35	0.000097
CG15081	Phb2	7	3.31	0.0002
CG3249	spoon	2	3.31	0.00024
CG14894	CG14894	2	3.3	0.00085
CG5676	CG5676	2	3.21	0.00012
CG46280	CG46280	3	3.09	0.000047
CG6701	CG6701	5	3.08	0.0011
CG6718	iPLA2-VIA	2	3.04	0.00035
CG8522	SREBP	2	3.03	0.0017
CG33129	CG33129	2	2.95	0.00021
CG14981	mge	3	2.94	0.00054
CG8547	CG8547	2	2.88	0.0023
CG5271	RpS27A	7	2.86	0.000022
CG10198	Nup98-96	15	2.75	0.000047
CG6479	CG6479	4	2.55	0.00016
CG12047	mud	3	2.47	0.00075
CG3948	zetaCOP	2	2.46	0.00037
CG6214	MRP	3	2.45	0.00045
CG7838	BubR1	5	2.3	0.00017
CG7961	alphaCOP	2	2.29	0.00089
CG17952	LBR	6	2.28	0.00016
CG11513	armi	49	2.23	0.000047
CG4183	Hsp26	2	2.18	0.0014
CG4264	Hsc70-4	18	2.13	0.00012
CG31755	SoYb	9	2.13	0.00017
CG2028	CkIalpha	3	2.09	0.00031
CG2183	Gasz	3	2.07	0.00016
CG11844	vig2	3	1.95	0.00038
CG32675	Tango5	3	1.94	0.001
CG9710	nudC	14	1.9	0.00016
CG32016	4E-T	5	1.88	0.0038
CG5436	Hsp68	2	1.83	0.0065
CG8258	CCT8	5	1.82	0.0017

CG6122	piwi	8	1.79	0.0013
CG13388	Akap200	2	1.77	0.0025
CG4170	vig	4	1.77	0.012
CG6838	ArfGAP3	2	1.74	0.017
CG3820	Nup214	2	1.73	0.0028
CG8367	cg	6	1.72	0.0013
CG18445	oys	2	1.69	0.0017
CG11154	ATPsynbeta	4	1.63	0.0008
CG31363	Jupiter	2	1.58	0.00097
CG14788	Ns3	2	1.56	0.0071
CG7439	AGO2	7	1.53	0.00047

Table 2 – Relative to Figure 3.11

Proteins showing $Log_2FC>1.5$ and adjusted P value < 0.05 in BASU-Gasz PL-MS. Only proteins with more than one unique peptide are reported.

Accessions	GeneSymbol	Unique_peptides	log2FC	P.Value
CG31755	SoYb	3	4.87	0.04
CG5271	RpS27A	3	4.68	0.0000021
CG8948	Graf	2	3.82	0.0011
CG16973	msn	3	3.56	0.00034
CG10811	eIF4G1	2	3.52	0.00063
CG10198	Nup98-96	2	3.48	0.000095
CG6756	Tom70	3	3.42	0.00007
CG11710	CG11710	3	3.31	0.00026
CG3249	spoon	3	3.18	0.0003
CG7878	CG7878	4	3.08	0.022
CG46280	CG46280	5	3.05	0.0000052
GASZ+BASU	BASU-Gasz	18	2.99	0.00000074
CG11513	armi	8	2.94	0.00036
CG4170	vig	2	2.87	0.000072
CG42551	larp	2	2.87	0.0015
CG11844	vig2	3	2.76	0.00026
CG1242	Hsp83	4	2.72	0.0013
CG14894	CG14894	3	2.6	0.000069
CG1633	Jafrac1	3	2.56	0.00083
CG10283	CG10283	2	2.56	0.0015
CG4264	Hsc70-4	4	2.54	0.001
CG8367	cg	2	2.47	0.00025
CG18174	Rpn11	2	2.45	0.0026
CG5205	obe	3	2.41	0.00082
CG11856	Nup358	4	2.33	0.00012
CG6122	piwi	10	2.29	0.00033
CG30069	CG30069	2	2.27	0.00074
CG30084	Zasp52	3	2.24	0.00093
CG1898	HBS1	2	2.23	0.0042
CG7838	BubR1	4	2.22	0.00039
CG1458	Cisd2	2	2.17	0.00043
CG30404	Tango11	2	2.17	0.0011
CG43726	qin	2	2.13	0.0035
CG6493	Dcr-2	4	2.06	0.00059
CG7139	CG7139	2	2.04	0.00046
CG5208	Patr-1	2	2.04	0.001
CG7439	AGO2	15	2.01	0.00072
CG9710	nudC	7	1.98	0.0038
CG6946	glo	2	1.95	0.00037
CG9684	CG9684	3	1.9	0.011

CG32555	RhoGAPp190	3	1.86	0.0019
CG5581	Ote	3	1.8	0.00088
CG32031	Argk	2	1.76	0.0038
CG11660	RIOK1	3	1.65	0.0098
CG8280	eEF1alpha1	8	1.62	0.0011
CG9765	tacc	3	1.6	0.0044
CG5726	CG5726	2	1.6	0.0046
CG32315	dlt	2	1.6	0.0094
CG1913	alphaTub84B	2	1.54	0.0041
CG16944	sesB	4	1.52	0.006
CG1404	Ran	3	1.51	0.017

Table 3 – Relative to Figure 3.15A

Proteins showing $Log_2FC>1.5$ and adjusted P value < 0.05 in Armi-BASU PL-MS. Only proteins with more than one unique peptide are reported. Due to space limitations, only the top 50 proteins are reported.

Accessions	GeneSymbol	Unique_peptides	log2FC	P.Value
CG5166	Atx2	2	7.5	0.00000045
CG8963	CG8963	2	6.54	0.000012
CG7838	BubR1	7	6.07	6.5E-08
CG31755	SoYb	6	5.97	1.6E-08
CG5508	mino	3	5.9	0.0000024
CG4916	me31B	2	5.85	0.00000002
CG8280	eEF1alpha1	8	5.79	0.00000028
CG11183	DCP1	4	5.74	0.00000034
CG18174	Rpn11	2	5.74	0.00001
CG4878	eIF3b	2	5.69	0.00000067
CG8258	CCT8	3	5.56	4.3E-08
ARMI+BASU	Armi-BASU	51	5.52	0.000031
CG1633	Jafrac1	4	5.51	0.00000015
CG14894	CG14894	6	5.49	0.00000012
CG14648	lost	4	5.42	0.0000021
CG1691	Imp	10	5.4	0.00000036
CG5205	obe	4	5.4	0.000059
CG11844	vig2	3	5.35	0.0000017
CG10811	eIF4G1	13	5.32	0.0000003
CG5581	Ote	5	5.32	0.000032
CG3249	spoon	3	5.28	1.7E-08
CG5940	CycA	2	5.23	3.5E-08
CG9684	CG9684	2	5.19	2.1E-08
CG1404	Ran	3	5.17	0.0000011
CG1341	Rpt1	9	5.15	3.8E-08
CG7082	papi	3	5.14	0.00000035
CG9805	eIF3a	6	5.13	0.00000021
CG8975	RnrS	2	5.1	0.00000019
CG10423	RpS27	2	5.04	3.2E-08
CG42551	larp	3	4.93	0.00000015
CG11139	p47	2	4.87	0.0000021
CG9710	nudC	4	4.86	0.0071
CG6838	ArfGAP3	2	4.79	0.00001
CG5686	chico	3	4.77	0.00000012
CG1242	Hsp83	8	4.75	0.00032
CG4735	shu	6	4.73	0.000001
CG3845	NAT1	6	4.72	0.000003
CG32016	4E-T	9	4.7	0.00000029

CG11660	RIOK1	3	4.69	7.4E-08
CG31363	Jupiter	3	4.68	0.00000021
CG9281	CG9281	2	4.68	0.0000062
CG11856	Nup358	17	4.65	0.000044
CG5119	pAbp	12	4.63	5.7E-08
CG11238	l(3)04053	2	4.63	0.000081
CG11061	GM130	2	4.61	0.0000019
CG9888	Fib	3	4.6	0.00000027
CG4237	ArfGAP1	3	4.57	6.2E-08
CG41099	CG41099	7	4.51	7.3E-08
CG17255	nocte	8	4.51	0.0000018
CG10638	CG10638	2	4.49	0.00000078

Table 4 – Relative to Figure 3.15C

Proteins showing $Log_2FC>1.5$ and adjusted P value < 0.05 in Zuc-BASU PL-MS. Only proteins with more than one unique peptide are reported.

Accessions	GeneSymbol	Unique_peptides	log2FC	P.Value
CG3249	spoon	3	4.92	0.00000032
CG30404	Tango11	3	4.69	0.000047
CG11513	armi	33	4.57	0.00000022
CG5508	mino	3	4.51	0.000019
CG2238	eEF2	7	4.37	0.013
CG17952	LBR	3	4.27	0.000025
CG3291	pcm	3	4.26	0.0013
CG9710	nudC	4	4.22	0.0000043
CG9424	bocks	2	4.06	0.034
ZUC+BASU	Zuc-BASU	11	3.8	0.00013
CG6944	Lam	2	3.75	0.0000044
CG3869	Marf	9	3.41	0.00000089
CG42551	larp	3	3.19	0.0000072
CG4237	ArfGAP1	2	3.07	0.000022
CG7838	BubR1	7	3.05	0.0000053
CG46280	CG46280	8	3.03	0.0000014
CG14894	CG14894	5	2.98	0.0000011
CG6756	Tom70	5	2.91	0.0000035
CG10198	Nup98-96	6	2.82	0.00078
CG18174	Rpn11	2	2.79	0.000063
CG7082	papi	3	2.69	0.0000075
CG10880	CG10880	3	2.63	0.000012
CG8258	CCT8	3	2.62	0.000058
CG31755	SoYb	6	2.55	0.0000017
CG5395	nmd	4	2.5	0.000019
CG8280	eEF1alpha1	8	2.41	0.00012
CG10364	msb1l	3	2.4	0.000086
CG9126	Stim	4	2.31	0.000011
CG11844	vig2	3	2.28	0.000074
CG9213	CG9213	2	2.25	0.024
CG6701	CG6701	5	2.23	0.00083
CG6214	MRP	6	2.22	0.000036
CG18361	dsh	2	2.22	0.00083
CG2183	Gasz	2	2.2	0.00066
CG1633	Jafrac1	4	2.16	0.000014
CG5271	RpS27A	4	2.16	0.0021
CG9684	CG9684	2	2.1	0.000028
CG3937	cher	10	2.02	0.034
CG32016	4E-T	7	2.01	0.007

CG2092	scra	2	1.99	0.015
CG8465	Ankle2	2	1.96	0.000049
CG14648	lost	2	1.94	0.00057
CG1913	alphaTub84B	2	1.92	0.0012
CG6479	CG6479	3	1.91	0.0022
CG12051	Act42A	4	1.87	0.022
CG42250	lqfR	3	1.85	0.0021
CG11856	Nup358	15	1.74	0.000093
CG8522	SREBP	4	1.71	0.0073
CG32031	Argk	7	1.7	0.000058
CG1945	faf	5	1.57	0.0035
CG14964	CG14964	2	1.57	0.0086
CG33113	Rtnl1	2	1.57	0.033
CG5205	obe	5	1.53	0.0023
CG5581	Ote	6	1.52	0.0028
CG31363	Jupiter	3	1.51	0.0003

Table 5 – Relative to Figure 3.16C

Proteins showing $Log_2FC>1.5$ and adjusted P value < 0.05 in Zuc Split BioID PL-MS. Only proteins with more than one unique peptide are reported.

Accessions	GeneSymbol	Unique_peptides	log2FC	P.Value
CG30404	Tango11	2	2.42	0.000016
ZUC+SPLITBIO	Zuc-SplitBioID	13	2.67	0.00007
CG11173	Snap29	1	2.94	0.00024
CG5581	Ote	8	2.04	0.00048

Table 6 – Relative to Figure 4.8A

Proteins showing $Log_2FC>1.5$ and adjusted P value < 0.05 in BASU-Yb PL-MS. Only proteins with more than one unique peptide are reported.

Accessions	GeneSymbol	Unique_peptides	log2FC	adj.P.Val
CG11513	armi	49	4.95	0.000013
CG2706	fs(1)Yb	75	4.19	0.000021
CG8367	cg	6	3.7	0.0001
CG9684	CG9684	19	3.63	0.000024
CG3186	eEF5	3	3.33	0.00028
CG11183	DCP1	5	3.2	0.000088
CG4620	unk	2	3.17	0.00028
CG31755	SoYb	9	3.08	0.00007
CG6311	Edc3	10	3.02	0.00011
CG32016	4E-T	5	2.93	0.00061
CG9132	CG9132	3	2.87	0.00087
CG9710	nudC	14	2.84	0.00007
CG11881	dgt6	3	2.83	0.00033
CG10504	Ilk	4	2.77	0.00033
CG5166	Atx2	3	2.61	0.00036
CG4170	vig	4	2.53	0.0024
CG1691	Imp	10	2.5	0.0001
CG5939	Prm	5	2.5	0.00035
CG14894	CG14894	2	2.4	0.0022
CG11414	CG11414	12	2.36	0.00022
CG14788	Ns3	2	2.29	0.0019

CG5208	Patr-1	4	2.12	0.00051
CG7838	BubR1	5	2.06	0.00033
CG10423	RpS27	2	1.98	0.0003
CG31363	Jupiter	2	1.98	0.00053
CG11844	vig2	3	1.92	0.00057
CG30084	Zasp52	3	1.92	0.0009
CG14648	lost	5	1.89	0.00046
CG3820	Nup214	2	1.89	0.0028
CG5886	CG5886	2	1.83	0.0019
CG43770	Sxl	4	1.78	0.0007
CG7878	CG7878	4	1.68	0.014
CG4264	Hsc70-4	18	1.65	0.001
CG8258	CCT8	5	1.61	0.00063
CG10811	eIF4G1	4	1.58	0.0016
CG7518	CG7518	6	1.57	0.0011
CG11154	ATPsynbeta	4	1.56	0.012
CG1913	alphaTub84B	2	1.5	0.009

Table 7 – Relative to Figure 4.8B

Proteins showing $Log_2FC>1.5$ and adjusted P value < 0.05 in BASU-Nup54 PL-MS. Only proteins with more than one unique peptide are reported.

Accessions	GeneSymbol	Unique_peptides	log2FC	adj.P.Val
CG6251	Nup62	11	3.97	0.0000022
CG14712	CG14712	17	3.03	0.0013
CG8771	Nup188	24	2.93	0.0000034
CG10198	Nup98-96	39	2.85	0.000026
CG33180	Ranbp16	4	2.73	0.0000082
CG7360	Nup58	17	2.14	0.000017
CG11943	Nup205	15	2.11	0.000019
CG2637	Fs(2)Ket	2	1.99	0.000044
CG6479	CG6479	5	1.99	0.0033
CG4453	Nup153	14	1.91	0.0001
CG17023	Dbp80	5	1.86	0.000017
CG34407	Not1	2	1.73	0.018
CG10318	NC2alpha	2	1.64	0.00028
CG11092	Nup93-1	12	1.55	0.000039
CG8831	Nup54	40	1.54	0.000026

Table 8 – Relative to Figure 4.8C

Proteins showing $Log_2FC>1.5$ and adjusted P value < 0.05 in BASU-Nup58 PL-MS. Only proteins with more than one unique peptide are reported.

Accessions	GeneSymbol	Unique_peptides	log2FC	adj.P.Val
CG8771	Nup188	24	3.71	0.0000055
CG14712	CG14712	17	3.66	0.001
CG17023	Dbp80	5	3.33	0.0000055
CG6479	CG6479	5	3.26	0.000043
CG6251	Nup62	11	3.23	0.000044
CG33180	Ranbp16	4	3.08	0.000021
CG10198	Nup98-96	39	3.04	0.000054
CG10318	NC2alpha	2	2.8	0.00005
CG2637	Fs(2)Ket	2	2.76	0.000043

CG4453	Nup153	14	2.7	0.00022
CG14788	Ns3	13	2.59	0.000043
CG11943	Nup205	15	2.41	0.000048
CG3820	Nup214	20	2.39	0.00025
CG2158	Nup50	19	2.21	0.00069
CG7583	CtBP	2	2.16	0.00077
CG11092	Nup93-1	12	2.05	0.00011
CG5794	puf	2	2.04	0.0037
CG13387	emb	2	2.03	0.00052
CG7360	Nup58	40	1.99	0.00034
CG17660	CG17660	3	1.92	0.00061
CG4464	RpS19a	4	1.78	0.00052
CG8831	Nup54	17	1.72	0.00044
CG2207	Df31	7	1.64	0.00017
CG13349	Rpn13	3	1.63	0.00061
CG7262	Nup93-2	5	1.59	0.00021
CG4579	Nup154	2	1.58	0.0013
CG34407	Not1	2	1.57	0.014
CG4622	CG4622	2	1.56	0.0037
CG2095	Sec8	2	1.55	0.0016
CG1708	cos	2	1.55	0.0021
CG7917	Nlp	4	1.54	0.003
CG6521	Stam	2	1.52	0.0024
CG2706	fs(1)Yb	20	1.51	0.00052

Table 9 – Relative to Figure 4.12A

Proteins showing $Log_2FC>1.5$ and adjusted P value < 0.05 in 3xFLAG-Yb WT IP-MS. Only proteins with more than one unique peptide are reported.

Accessions	GeneSymbol	Unique_peptides	log2FC	adj.P.Val
CG8057	alc	2	3.29	2.6E-13
CG11513	armi	56	3.21	3.6E-16
CG2706	fs(1)Yb	49	3.16	8.8E-15
CG10223	Top2	53	2.97	1.3E-13
CG1475	RpL13A	7	2.22	2.4E-12
CG6510	RpL18A	10	2.08	1.8E-12
CG4897	RpL7	12	1.82	1E-11
CG1691	Imp	18	1.79	2.1E-12
CG2099	RpL35A	4	1.76	1E-11
CG2849	Rala	4	1.76	1.4E-10
CG7194	CG7194	8	1.68	1.4E-10
CG8615	RpL18	4	1.66	1.4E-10
CG5502	RpL4	19	1.64	4.4E-12
CG7283	RpL10Ab	9	1.6	1E-10
CG9354	RpL34b	4	1.59	3.4E-10
CG6815	bor	19	1.56	1E-11
CG9684	CG9684	4	1.54	9E-11
CG6253	RpL14	7	1.52	2E-11
CG3314	RpL7A	11	1.52	5.7E-11
CG6122	piwi	26	1.52	6.3E-11
CG17420	RpL15	4	1.51	9.9E-10

Table 10 – Relative to Figure 4.12B

Proteins showing $Log_2FC>1.5$ and adjusted P value < 0.05 in 3xFLAG-Yb ΔeTud IP-MS. Only proteins with more than one unique peptide are reported.

Accessions	GeneSymbol	Unique_peptides	log2FC	adj.P.Val
CG7263	AIF	5	2.82	2.2E-12
CG5203	STUB1	4	2.75	3.7E-13
CG4264	Hsc70-4	28	2.55	3.7E-13
CG8057	alc	2	2.16	8.6E-11
CG2706	fs(1)Yb	49	1.93	7E-12
CG18743	Hsp70Ab	7	1.77	2E-10
CG3201	Mlc-c	7	1.61	0.00000022
CG15792	zip	52	1.59	0.0000038
CG10289	fmt	9	1.53	4.5E-10
CG5436	Hsp68	9	1.51	1.7E-10

Table 11 – Relative to Figure 4.12C

Proteins showing $Log_2FC>1.5$ and adjusted P value < 0.05 in 3xFLAG-Yb ΔHelC IP-MS. Only proteins with more than one unique peptide are reported.

Accessions	GeneSymbol	Unique_peptides	log2FC	adj.P.Val
CG8057	alc	2	2.75	7.3E-12
CG2706	fs(1)Yb	49	2.61	3.1E-13
CG10223	Top2	53	2.13	1.9E-11
CG2849	Rala	4	1.93	1.2E-10
CG1475	RpL13A	7	1.82	6E-11
CG6510	RpL18A	10	1.66	6E-11

Table 12 – Relative to Figure 5.1B

Proteins showing $Log_2FC>2$ and adjusted P value < 0.05 in GFP-Panx IP-MS. Only proteins with more than one unique peptide are reported.

Accessions	GeneSymbol	Unique_peptides	log2FC	adj.P.Val
CG9754	Panx	22	4.35	0.00000011
CG17108	CG17108	2	3.87	0.000019
CG4118	nxf2	33	3.77	0.00000064
CG12752	Nxt1	6	3.61	0.00000064
CG18067	CG18067	4	2.12	0.027
CG6998	ctp	2	1.99	0.000027

Table 13 – Relative to Figure 5.2B

Proteins showing $Log_2FC>2$ and adjusted P value < 0.05 in GFP-Nxf2 IP-MS. Only proteins with more than one unique peptide are reported.

Accessions	GeneSymbol	Unique_peptides	log2FC	adj.P.Val
CG9754	Panx	22	4.31	0.00000013
CG4118	nxf2	33	4.04	0.00000029

CG12752	Nxt1	6	3.64	0.00000058
CG6863	tok	3	2.44	0.00014
CG6998	ctp	2	2.07	0.000022

VIII.8 Source data

Chapter III

Raw data from proteomics and high-throughput sequencing experiments relative to this chapter are available on Proteomics Identifications (PRIDE) database (accession numbers: **PXD013417**, **PXD013405**, **PXD013404**, and **PXD013403**) and Gene Expression Omnibus (GEO) (accession number: **GSE129321**).

Chapter IV

Raw data from proteomics and high-throughput sequencing experiments relative to this chapter are available on PRIDE (accession numbers: **PXD019674, PXD019671, PXD019670**) and GEO (accession number: **GSE152297**).

Chapter V

Raw data from proteomics and high-throughput sequencing experiments relative to this chapter are available on PRIDE (accession number: **PXD011415**) and GEO (accession number: **GSE121661**).