piRNA-dependent transcriptional gene silencing during *Drosophila* oogenesis and embryogenesis

Martin Heinrich Ulrich Fabry

St John's College
Cancer Research UK - Cambridge Institute
University of Cambridge

This thesis is submitted for the degree of
Doctor of Philosophy

May 2020

Declaration

I hereby declare that the content of this thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specific reference to the work of others is made in the text. The content of this dissertation is original and has not been submitted before for any degree or other qualification.
This dissertation does not exceed the specified word limit of 60,000 words as defined by the Degree Committee for the Faculty of Clinical Medicine and Clinical Veterinary Medicine.

Martin Heinrich Ulrich Fabry, May 2020

Acknowledgments

I would like to thank and acknowledge all my colleagues that contributed to this work and supported me throughout my PhD:

First of all, I would like to thank my supervisor Greg Hannon for giving me the opportunity to work in his lab for the last five years and allow me to pursue sometimes crazy ideas independently. I very much valued your support, guidance and expertise throughout my studies.

I would like to especially thank Ben Czech for his support and supervision. His mentorship and encouragement throughout the years greatly contributed to the success of this work and was fundamental for my development as a scientist.

Further, I would like to thank all current and previous members of the Hannon lab for stimulating scientific discussions and the friendly environment. I would especially like to thank Marzia Munafò, Emma Kneuss and Evelyn Eastwood for their scientific and personal support and for making this PhD a great experience.

I would like to thank all piRNA colleges and especially Marzia Munafò and Ben Czech that contributed to the Nxf2 project for their collaboration. Further, I would like to thank Federica Falconio as well as Emily Lythgoe for their support on the Degron and *roo* projects.

I would like to thank Cancer Research UK for their funding and support during my PhD. The scientific support by the Core facilities at CRUK CI was crucial for this work. I would especially like to acknowledge the Genomics core facility for sequencing hundreds of libraries generated for this thesis, the Proteomics core facility for generating mass spectrometry data, the Light Microscopy core facility, and especially Fadwa Joud, for the amazing expertise and support, the Bioinformatics core facility for support with mass spectrometry data analysis and the RICS facility for their help with countless experiments. I would also like to thank St. John's college for their support.

Finally, I would like to thank especially my family and friends for their support throughout my PhD. Without you none of this work would have been possible.

Summary

piRNA-dependent transcriptional gene silencing during *Drosophila* oogenesis and embryogenesis

Martin Heinrich Ulrich Fabry

The PIWI-interacting RNA (piRNA) pathway is a small RNA based intracellular immune system protecting animal gonads from the deleterious effects of transposons, thus maintaining transgenerational genome integrity.

In *Drosophila melanogaster* ovaries, piRNA-Piwi complexes localise to the nucleus and scan nascent transcripts for transposon expression by using complementary antisense piRNAs as guides. Following target engagement, the gonad-specific protein Panoramix (Panx) is recruited and induces transcriptional gene silencing (TGS) by connecting to the general chromatin silencing machinery of the cell resulting in changes of the epigenetic chromatin state, thus shutting down transcription. However, whether Panx acts on its own or if other proteins are involved in piRNA-dependent TGS remains unknown.

During my PhD I studied the protein-protein interactions of Panx and co-discovered the Panx induced co-transcriptional silencing (PICTS) complex comprised of Panx, Nxf2 and Nxt1. The PICTS complex induces TGS at active transposon insertions in *Drosophila* ovaries. Furthermore, I studied the effects of epigenetic inheritance of piRNA-Piwi complexes and the PICTS complex during early *Drosophila* embryogenesis. Piwi showed no zygotic transcription in somatic cells but strong maternal deposition and localised not only to pole cells, the germ line precursors, but was also strongly enriched in somatic nuclei. Additionally, the PICTS complex was both maternally deposited and zygotically transcribed and co-localised with Piwi in somatic nuclei. Several transposons showed strong zygotic expression during early embryogenesis. However, transcriptional gene silencing occurred at individual transposon insertions and repressive chromatin marks accumulated around the genomic location of transposons targeted by maternally deposited piRNAs. Depletion

of maternally deposited Piwi resulted in deregulation of transposons and loss of repressive chromatin marks at associated genomic regions.

My PhD project uncovered an epigenetic transposon regulatory complex that showed expression not only in gonadal tissue but also in somatic cells during early embryogenesis and revealed a novel function of the piRNA pathway in transposon control by inducing epigenetic chromatin changes during early *Drosophila* development.

Table of Contents

Table of Figures

List of Tables

List of Code

List of abbreviations

act5c	Actin-5C
AEL	After egg laying
Ago	Argonaute
AID	Auxin-induced degradation
ANOVA	Analysis of variance
Armi	Armitage
Arx	Asterix
Aub	Aubergine
BAC	Bacterial artificial chromosome
Boot	Bootlegger
BWA	Burrow-Wheeler Aligner
C. elegans	*Caenorhabditis elegans*
CC	Coiled coil
cDNA	Complementary DNA
ChIP-Seq	Chromatin immunoprecipitation followed by sequencing
Co-IP	Co-immunoprecipitation
CRISPR	Clustered regulatory interspersed short palindromic repeats
CRUK CI	Cancer Research UK - Cambridge Institute
CRY2	Cytochrome C
Cuff	Cutoff
Daed	Daedalus
DAPI	4′,6-diamidino-2-phenylindole
Del	Deadlock
dsDNA	double-stranded DNA
Egg	Eggless
env	Envelope protein
FISH	Fluorescence *in situ* hybridisation
fRIP	Formaldehyde RIP
gag	Group-specific antigen-like protein

GFP	Green-fluorescent protein
GO	Gene ontology
H1	Histone H1
H3K27me3	Histone H3 lysine 27 trimethylation
H3K4me2	Histone H3 lysine 4 dimethylation
H3K9me3	Histone H3 lysine 9 trimethylation
HA	Hemagglutinin
HCR	Hybridisation Chain Reaction
HP1a	Heterochromatin protein 1a
HR	Homologous recombination
IAA	Indole-3-acetic acid
IF	Immunofluorescence
JT	JabbaTrap
Krimp	Krimper
LD	Lipid droplet
LEXY	Light-induced Nuclear Export system
LMA	Low melting point agarose
Loq	Loquacious
LRR	Leucine rich repeat
Lsd1	Lysine-specific histone demethylase 1
LSFM	Light sheet fluorescence microscopy
LTR	Long terminal repeat
Mael	Maelstrom
miRNA	micro RNA
Moon	Moonshiner
mRNA	Messenger RNA
MZT	Maternal to zygotic transition
Nbr	Nibbler
NC	Nuclear cycle
NHEJ	Non-homologous end joining
NLS	Nuclear localisation signal
nos	Nanos

NTF2	NTF2-like domain
NTR	Amino-terminal region
Nxf1	Nuclear Export Factor 1
Nxf2	Nuclear Export Factor 2
Nxf3	Nuclear Export Factor 3
Nxt1	Nuclear Transport Factor 2 Like Export Factor 1
ORF	Open reading frame
OSC	Ovarian somatic cells
osk	Oskar
PAGE	Polyacrylamide gel electrophoresis
Panx	Panoramix
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PICTS	Panx-induced Co-transcriptional silencing complex
piRNA	Piwi-interacting RNA
PIWI	P-element-induced wimpy testis
POI	Protein of interest
pol	Reverse transcriptase
Pol II	RNA polymerase II
polyA	Polyadenylation
PTGS	Post-transcriptional gene silencing
qRT-PCR	Quantitative reverse transcriptase PCR
RAG	Recombinase-activating gene
RBD	RNA-binding domain
RFP	Red-fluorescent protein
Rhi	Rhino
fRIP-Seq	Formaldehyde RNA immunoprecipitation followed by sequencing
rp49	Ribosomal protein 49
RPM	Reads per million
rRNA	Ribosomal RNA
RT	Room temperature
S. pombe	*Schizosaccharomyces pombe*

S2	Schneider 2 cells
siRNA	single-interfering RNA
Smg	Smaug
sna	snail
sry	Serendipity
SSC	Saline-sodium citrate
Su(var)	Suppressor of variegation
TAD	Topological associated domain
TBS	Tris-buffered saline
TE	Transposable Element
TF	Transcription factor
TGS	Co-transcriptional gene silencing
tub	alpha-Tubulin
twi	twist
UBA	Ubiquitin associated domain
ubi	Ubiquitin-63E
Vas	Vasa
Wde	Windei
WGS	Whole genome sequencing
ZGA	Zygotic genome activation
Zld	Zelda
Zuc	Zucchini

1 Introduction

Genome integrity is fundamental for the transmission of intact genetic information to future generations. Prevention of deleterious mutations that can affect the fitness of offspring is crucial. Transposable elements (TEs) are a serious threat to genomic stability. TEs are parasitic DNA sequences that can change their position or copy number in the genome by replicating and integrating in different locations, thus potentially disrupting gene structures or regulatory regions.

Germ cells, which gives rise to the next generation, have adapted several mechanisms to ensure genome integrity. The PIWI-interacting RNA (piRNA) pathway is crucial for regulating transposable elements in animal gonads. In *Drosophila melanogaster*, transcripts derived from genomic clusters are processed into 23-30nt long piRNAs. These small RNA species predominantly contain sequences capable of regulating transposon families. At the centre of this pathway are PIWI-clade Argonaute proteins, such as Piwi, which is loaded with piRNAs in the cytoplasm and subsequently translocates into the nucleus. piRNA-Piwi complexes scan the transcriptome for complementary targets present in the nucleus. Engagement of the complex with nascent transcripts leads to the recruitment of a piRNA pathway-specific downstream effector complex. This complex recruits the general silencing machinery via a currently unknown mechanism and leads to the deposition of repressive chromatin marks. The targeted genomic locus is then transformed into transcriptionally inactive heterochromatin, thus preventing further TE expression. The maintenance of the transcriptionally silenced genomic state is Piwi-dependent for several transposons with loss of piRNA pathway factors leading to re-expression of formally silenced TE insertions.

1.1 Genome Integrity

Animal genomes are under permanent evolutionary pressure to adapt to changing environments. Small changes in the DNA sequence may lead to improved survival or increased fertility, therefore making propagation of the affected alleles throughout populations more likely. However, major genomic changes and even small mutations can result in deleterious effects on cellular or developmental pathways through altered protein expression or function (Eyre-Walker and Keightley, 2007). A broad range of threats to genome integrity exists. Chemicals present in food or the environment as well as UV and high-energy radiation can have mutagenic effects and often cause random DNA damage (Langie et al., 2015; Singer and Kusmierek, 1982). Maintaining genome integrity is crucial to prevent accumulation of disadvantageous mutations and thus is imperative for the survival of a species. Failing to protect the genome from damaging mutations can lead to various genomic diseases including cancer (Stratton et al., 2009). Several mechanisms are in place to ensure genome integrity. Some physiological adaptations protect cells from radiation such as melanin production in the skin, thus preventing mutations in the first place. Other molecular pathways are activated once DNA damage occurs.

The DNA damage response (DDR) pathway is activated by a broad range of mutations, including single base lesions and DNA breakage (Jackson and Bartek, 2009). Repair is accomplished by different pathways such as non-homologous end joining (NHEJ) and homologous recombination (HR). While NHEJ is the preferred method of repair for somatic cells in G1 phase, HR is mainly activated during S and G2 phase of the cell cycle (Symington and Gautier, 2011). During NHEJ, double-strand brakes are re-ligated. This can lead to error-free or error-prone repair in a stochastic manner (Chang et al., 2017). Homologous recombination utilizes homologous DNA sequences usually from complementary chromosomes as a template to repair the damaged strand (Ciccia and Elledge, 2010).

These mechanisms have been widely exploited for genome engineering. For instance, the Clustered Regulatory Interspersed Short Palindromic Repeat (CRISPR)/Cas9 system introduces double-strand breaks at genomic target loci such

as protein-coding genes (Hsu et al., 2014). Cellular repair mechanisms then detect DNA damage and are activated. If not repaired error-free, this results in small mutations that can have an impact on reading frames for translation, often resulting in premature stop codons. Additionally, utilising HR enables the introduction of precise genetic information at targeted loci by providing homologous sequences to be used as a repair template during HR (Cubbon et al., 2018).

Genome protection of germ cells is especially crucial to prevent the inheritance of mutations in further generations. While the DDR pathway aims at repairing genomic alterations but can lead to error-prone repairs, preventing mutations in the first place is the method of choice for germ cells. In the nematode *Caenorhabditis elegans* (*C. elegans*) and *Drosophila melanogaster* for instance, germ cells are defined by maternal inheritance of factors during early embryogenesis (see chapter 1.7.2) ultimately preventing somatic differentiation and defining germ cell fate (Cinalli et al., 2008; Seydoux and Braun, 2006). Maternally inherited factors are involved in priming future germ cells for the protection from mutagens such as transposon expression early on in development and therefore strongly contributing to genome integrity (see chapter 1.6)

However, germ cells in other animals are defined in later stages of embryogenesis. For instance, in mice a specific subpopulation of embryonic cells is exposed to factors ultimately leading to germ cell differentiation (Hayashi et al., 2007). *C. elegans* and *Drosophila* represent a class of animals with a preformed germline, while the murine germline is induced later during embryogenesis. In contrast, other organisms such as flowering plants do not set aside germ cells during embryogenesis but rather maintain undifferentiated stem cells that experience a somatic-to-reproductive transition (Berger and Twell, 2011). The differences in definition of germ cells across organisms illustrate the need for multiple sophisticated mechanisms to protect genome integrity of germ cells and their progenitors during embryogenesis and beyond. The piRNA pathway represents a complex mechanism to control transposon propagation in most animal gonads and therefore ensures transgenerational genome stability (see also chapter 1.5).

1.2 Repetitive DNA Sequences

Repetitive sequences are highly abundant in eukaryotic genomes. Around 66-69% of the human genome consists of repeats (de Koning et al., 2011), while 85% of some plant genomes, such as that of maize, is occupied by repetitive sequences (Schnable et al., 2009). There are many types of repetitive elements that range from few base pairs in length, such as simple repeats/microsatellites, to several kilobases, such as transposons. Repetitive elements are often concentrated at the centromeres, pericentromeres, and telomeres and are essential factors of constitutive heterochromatin (Biscotti et al., 2015; Saksouk et al., 2015). However, several repetitive elements show a more diverse localisation within the genome. Transposable elements, for instance, are a class of dispersed repeats that can be found throughout the genome but show enrichment at heterochromatic regions. The function of repetitive elements within the genome is not completely understood. Although repeats are often located in heterochromatin, some show active transcription especially in particular developmental stages (Ugarkovic, 2005). However, evidence suggests a contribution of tandem repeats such as satellite DNA (satDNA) in establishment of heterochromatin at telomeres and centromeres, thus contributing greatly to genome stability (Biscotti et al., 2015; Seller et al., 2019; Yuan and O'Farrell, 2016).

1.2.1 Transposable Elements

Transposable elements (TEs) are mobile genomic elements and are abundant throughout the genomes of the animal kingdom. Due to their ability to integrate at various locations in the genome, TEs represent one class of highly repetitive DNA sequences. Organisms with small and complex genomes show lower abundance of transposons with some yeast genomes being comprised of <5% TEs (Wostemeyer and Kreibich, 2002). The genome of *Drosophila* is made up by about 20% transposon sequences (Barron et al., 2014), while about half of the human genome is transposon-derived (Lander et al., 2001). Organisms with large genomes such as plants can be comprised of TEs by more than 85% (Schnable et al., 2009). Interestingly, the diversity

of transposon families varies greatly between organisms. For instance, TEs in yeast are mostly comprised of the *Copia*-family, Ty1-5, whereas flies have more than 100 different families (Bleykasten-Grosshans et al., 2013; McCullers and Steiniger, 2017). While the majority of TEs in humans are inactive, it has been estimated that around 30% of *Drosophila* transposon insertions are active (McCullers and Steiniger, 2017). This makes *Drosophila* a particularly interesting model organism for studying transposon-related mechanisms.

TEs can be categorized into two distinct classes: Retrotransposons (class-I) and DNA-Transposons (class-II) (Finnegan, 1989; McClintock, 1950). Retrotransposons are transcribed and often code for multiple retrovirus-related proteins including group-specific antigen (*gag)*, reverse transcriptase (*pol)* and in some cases envelope proteins (*env)* (Finnegan, 2012). Retrotransposons are in many ways similar to retroviruses. Indeed, at least two *Drosophila* retrotransposons, *gypsy* and *ZAM*, have been shown to produce infectious particles (Kim et al., 1994; Leblanc et al., 2000) and it is believed that both retroviruses and retrotransposons share a common ancestor (Nefedova and Kim, 2017). The proteins encoded by retrotransposons allow the integration of TE-derived RNA transcripts by reverse transcription and the subsequent integration into the host genome by a 'copy and paste' mechanism. Although there seems to be some bias in guiding transposons to genomic regions for insertions, little is known about their specificity for DNA sequences or whether negative selection is responsible for allowing transposon insertions at specific loci while excluding them from others (Sultana et al., 2017). Retrotransposons can be further divided in LTR and non-LTR transposons. LTRs (long terminal repeats) contain sequences that are recognised by endogenous transcription factors, thereby driving TE transcription in a tissue- and developmental-controlled fashion (Bronner et al., 1995; Thompson et al., 2016). Additionally, LTRs control mRNA processing and mediate integration into the host genome (Benachenhou et al., 2013). Non-LTR transposons share a similar genomic structure with LTR transposons but lack flanking LTRs. Several non-LTR family members can autonomously integrate into the genome such as LINE1 elements in human and the *I-element* in *Drosophila* (Han, 2010). However, many non-LTR elements lack coding sequences for proteins necessary for

transposition, thus making them dependent on other factors supplied by the host cell or other active transposons (Goodier and Kazazian, 2008).

DNA transposons do not directly rely on their transcripts to change position within the genome but rather change location by a 'cut and paste' mechanism (Munoz-Lopez and Garcia-Perez, 2010). Terminal inverted repeats (TIRs) flank DNA transposons. These are recognised by a transposase in a sequence-dependent manner leading to excision of the transposon DNA from its donor position. The transposon DNA can then integrate at a new acceptor position. The gap at the donor position caused by the excision of the TE can be either repaired by NHEJ, removing the TE from its previous position but leaving a "scar", or by homologous recombination, thereby allowing the duplication of the transposon sequence by utilising the homologous chromosome still containing the TE as a template (see also chapter 1.1).

1.2.2 Transposons and genome evolution

Transposition events occur in all cells including the germ line. Activity of transposons in somatic cells can result in genetic heterogeneity and has been linked to tumorigenesis in humans (Rodriguez-Martin et al., 2020; Shukla et al., 2013) and neuronal diseases such as schizophrenia (Bundo et al., 2014) as well as neurodegeneration in *Drosophila* (Krug et al., 2017). However, only the transposition events that occur in germ cells can result in transgenerational propagation of new insertions. It is therefore not surprising that many transposon families are active in gonads. Recent studies in *Drosophila* revealed hijacking of endogenous pathways during oogenesis that lead to enrichment of transposon insertions specifically in oocytes (Wang et al., 2018). However, animals have developed sophisticated mechanisms, primarily expressed in germ cells, that control transposon expression by either removing their RNA intermediates or preventing transcription of the TE entirely by changing the epigenetic landscape (see also chapter 1.5).

While transposons represent a great threat to genome integrity, they are also key drivers of genome evolution. Integration of transposons in close proximity to genes can have a profound impact on expression patterns. Regulatory elements found in the

transposon sequence can be co-opted by the host genome (Batut et al., 2013), which may lead to improved adaptation to environmental situations. Additionally, several transposon-derived genes have developed critical functions in several organisms including humans (Sinzelle et al., 2009). For instance, in mammals, Rag1 and Rag2 are involved in rearrangement of DNA and are key players in the V(D)J recombination mechanism of the adaptive immune response. Both proteins form an enzyme that is able to excise DNA containing specific recombination signals from a donor site and subsequently integrate it into a DNA acceptor (Agrawal et al., 1998). It has later been shown that both Rag1 and Rag2 are partially derived from the Transib transposon family found in several organisms (Kapitonov and Jurka, 2005).

1.3 Transcriptional regulation

Differential gene expression in tissues and during development is crucial for the versatile functions performed by cells that are necessary to form an organism. Gene expression is controlled by complex mechanisms and pathways. For example, tissue-specific master transcription factors can drive transcriptional regulatory networks thereby directing cell fate decisions (Wilkinson et al., 2017). Genic transcript levels are also fine-tuned by the degradation of mRNAs following nuclear export in a process called post-transcriptional gene silencing (PTGS), which is one of the most fundamental mechanisms underlying the RNAi pathway (see chapter 1.4).

However, another way to regulate gene expression is by controlling the transcriptional output at gene loci directly in a process called transcriptional gene activation (TGA) or transcriptional gene silencing (TGS). This is achieved by modulating the accessibility of DNA by modifying biophysical properties of the highly organised chromatin structure. The fundamental unit of DNA organisation is the nucleosome. About 147bp of double-stranded DNA are wrapped around an octamer comprised of the core histones H2A, H2B, H3 and H4 (Luger and Hansen, 2005). Introduction of post-translational modifications (PTMs) at histones can interfere with binding of transcription factors or the transcription machinery. Euchromatic regions are

enriched in activating marks such as methylated histone 3 lysine 4 (H3K4), while heterochromatic regions show strong deposition of methyl groups at histone 3 lysine 9 (H3K9). Genes residing in euchromatin are generally associated with high levels of transcription, while transcriptional output at heterochromatic genes is predominantly low (Hyun et al., 2017; Richards and Elgin, 2002). H3K9 methylation is associated with transcriptional repression, however, it is still debated whether these marks are the cause or a consequence of repression.

In *Drosophila*, three H3K4 methyltransferases have been identified: dSet1, TRX and TRR. Different methylation states of H3K4 can be found at active genomic regions. Enhancers are enriched in H3K4me1 (Heintzman et al., 2007), while the 5' terminus of transcribed genes shows strong H3K4me2 signal (Kim and Buratowski, 2009). H3K4me3 shows the highest abundance at transcription start sites (TSS) of genes downstream of the nucleosome-depleted region (Soares et al., 2017). Interestingly, dSet1 is mostly involved in trimethylation of H3K4 and loss of dSet1 results in retention of RNA polymerase II (Pol II) (Ardehali et al., 2011). This suggests that H3K4me3 marks facilitate proper Pol II-dependent elongation and transcription. H3K4 methylation is actively removed by lysine-specific demethylase 1 (Lsd1/Su(var)3-3) (Shi et al., 2004). The chromatin-associated protein CoRest binds Lsd1 and guides the complex to active gene targets (Lee et al., 2005), thereby contributing to transcriptional gene silencing and heterochromatin formation in *Drosophila* (Rudolph et al., 2007). Lsd1 has been shown to be involved in piRNA-dependent transposon silencing (Czech et al., 2013) (see chapter 1.5.8) (Figure 1.1).

In *Drosophila*, three SET domain methyltransferases, Eggless/dSETDB1 (Egg), Su(var)3-9 and G9a, are present that have been implicated in heterochromatin formation by modifying the methylation state of H3K9. Little is known about the specificity or redundancy of these methyltransferases towards their target loci and their developmental regulation. However, Egg, the homolog of human SETDB1, has been associated with heterochromatin formation on the fourth chromosome and is responsible for H3K9me3 maintenance in adult tissues (Seum et al., 2007b; Tzeng et al., 2007). Egg is the only essential H3K9 methyltransferase in *Drosophila* and has further been shown to be involved in transposon control by deposition of H3K9me3 in

ovaries (Rangan et al., 2011; Sienski et al., 2015; Yu et al., 2015) (see chapter 1.5.8). Windei (Wde), homolog of human Activating transcription factor 7-interacting protein 1 (ATF-IP) (Wang et al., 2003), is a co-factor of Egg and is believed to recruit Egg to chromatin at target genes for silencing (Osumi et al., 2019) (Figure 1.1). Both Egg and Wde are essential for germ line development (Koch et al., 2009). Recent studies have revealed Egg's pivotal role in heterochromatin formation during the mid-blastula transition (MBT) during *Drosophila* embryogenesis (Seller et al., 2019). Egg is associated with mono-, di- and trimethylation of H3K9 (Seum et al., 2007b). The methyltransferase Su(var)3-9 functions predominantly during embryogenesis and has been linked to constitutive heterochromatin formation rather than transcriptional gene silencing (Allshire and Madhani, 2018; Elgin and Reuter, 2013). However, Su(var)3-9 has been implicated in transposon regulation (Czech et al., 2013; Handler et al., 2013). Recent studies revealed that Su(var)3-9 is not necessary for constitutive heterochromatin formation but can contribute, together with Egg, towards the re-establishment of heterochromatin in early nuclear cycles during *Drosophila* embryogenesis (Seller et al., 2019). G9a, the third SET domain methyltransferase in *Drosophila*, shows uniform low expression during development but is highly transcribed in gonads (Stabell et al., 2006). Learning and memory formation in *Drosophila* have been linked to epigenetic regulation by G9a (Kramer et al., 2011). While multiple roles for both Su(var)3-9 and G9a have been described, neither are essential for either development or fertility in *Drosophila* (Seum et al., 2007a; Tschiersch et al., 1994).

It is currently unclear how different methyltransferases are recruited to specific genomic sequences. Transposons are targeted by the piRNA pathway and Egg is believed to be recruited to sites of TE transcription following recognition of transcripts by Piwi-bound piRNAs via an unknown mechanism (see chapter 1.5.8). While H3K9me3 is highly enriched at TE insertions in *Drosophila* gonads, the mechanism leading to recruitment of methyltransferases to piRNA pathway-independent regions such as constitutive heterochromatin is unknown. Recent studies suggest that the E3 SUMO ligase Su(var)2-10 is involved in the recruitment of Egg to chromatin by SUMOylation of chromatin associated proteins and indeed, Su(var)2-10 interacts with

the piRNA pathway machinery, potentially providing a link between Piwi and Eggless recruitment (see chapter 1.5.8). However, the mechanism responsible for Su(var)2-10 recruitment to piRNA-independent loci is not yet resolved (Ninova et al., 2020a; Ninova et al., 2020b).

Following H3K9me3 deposition, homodimers of Heterochromatin Protein 1a (HP1a/Su(var)205) bind H3K9me3 marks via their chromodomain. This mechanism is highly conserved in organisms from yeast (Bannister et al., 2001) to mammals (Lachner et al., 2001). HP1a is believed to recruit several proteins involved in heterochromatin formation including transcriptional repressors, histone deacetylases and chromatin remodelers through its chromoshadow domain (Hall et al., 2002; Lechner et al., 2005; Yamada et al., 2005). Recent studies in human and *Drosophila* cells have shown that binding of HP1a results in the conversion of chromatin into phase separated compartments that prevent diffusion of proteins, thus providing an additional or alternative explanation how heterochromatin formation achieves transcriptional gene silencing beyond chromatin compaction (Larson et al., 2017; Strom et al., 2017) (Figure 1.1).

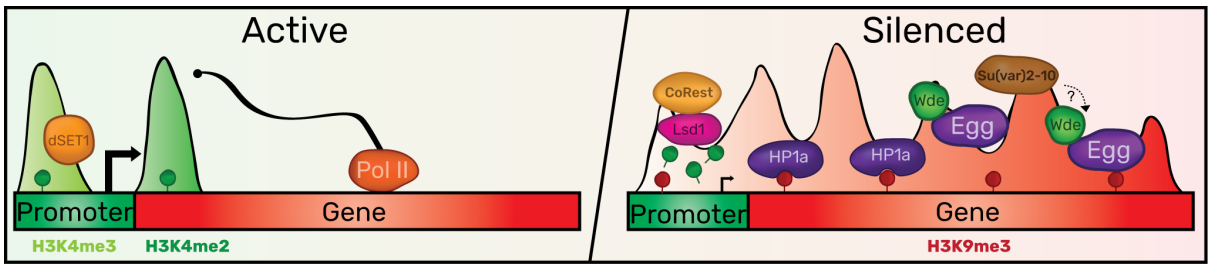

Figure 1.1: Epigenetic states of permissive and restrictive chromatin

Transcribed genes are enriched in H3K4me3 at their promoters and H3K4me2 at the 5' UTR. dSet1 is involved in depositing H3K4me3 at promoters. Silenced genes are characterised by repressive chromatin modifications such as H3K9me3. In *Drosophila*, the methyltransferase Eggless (Egg) in complex with Windei (Wde) is involved in deposition of H3K9me3 at silenced genes and constitutive heterochromatin. Su(var)2-10 might recruit Egg and Wde to chromatin. HP1a binds H3K9me3 marks with its chromo domain and changes the biophysical properties of the genomic locus, thus preventing transcription.

Several additional mechanisms have been described that regulate transcriptional activity at the chromatin level. Among the most prominent is the Polycomb group (PcG) of proteins, which are involved in transcriptional gene silencing, especially during development and differentiation (Beuchle et al., 2001). PcG proteins assemble into complexes with the best studied being PRC1 and PRC2. In *Drosophila*, cis-regulatory elements termed Polycomb gene responsive elements (PREs) are targeted by PcG proteins and lead to the deposition of repressive chromatin marks (Golbabapour et al., 2013). Characteristic targets are the Hox genes in *Drosophila* such as *ultrabithorax* (*Ubx*) (Chan et al., 1994). Enhancer of zeste, E(z), is a component of PRC2. The SET domain-containing methyltransferase catalyses methylation of histone-3-lysine-27 (H3K27me3), a histone mark characteristic for Polycomb-associated gene silencing (Czermin et al., 2002).

In plants and many animals, DNA methylation is used to regulate gene expression. Methylation of Cytosine at position 5 (5mC) accumulates at CpG islands and is associated with transcriptional repression in mammals (Ben-Hattar and Jiricny, 1988; Iguchi-Ariga and Schaffner, 1989; Watt and Molloy, 1988). However, methylation is not limited to CpG islands in plants. Asymmetric CHH and CHG (where H represents A, T or C) methylation has been show to accumulate at transposon insertions in plants and is maintained by the RNA-directed DNA methylation (RdDM) pathway (Zhang et al., 2018). CHH methylation islands are necessary to enforce boundaries between transposon-enriched heterochromatin and euchromatin domains (Li et al., 2015). Hypomethylation can result in reactivation of previously silenced transposons. DNA methylation is indispensable for proper development during mammalian early embryogenesis. Intriguingly, however, methylation states are erased and re-established between generations. The majority of DNA methylation is first lost during gametogenesis and second during early embryogenesis in mammals (Greenberg and Bourc'his, 2019). While DNA methylation is important in mammals, methylation states in *Drosophila* are low (Capuano et al., 2014) and are widely viewed to have no impact on gene regulation.

1.4 RNA interference pathways

Regulating mRNA levels by transcriptional control mechanisms is an effective strategy to promote differential expression in cells of various developmental stages or tissues. However, additional pathways have evolved to fine-tune expression even further. RNA interference (RNAi) was discovered by Andrew Fire and Craig Mello more than 20 years ago (Fire et al., 1998) for which they received the Nobel Prize in Physiology or Medicine in 2006. However, initial observations made in plants greatly contributed to the discovery (Baulcombe, 1996; van der Krol et al., 1990). Two decades of extensive research have uncovered many mechanistic principles of this small RNA-centred pathway.

RNAi is involved in the process of post-transcriptional gene silencing (PTGS). At the heart of the pathway are small RNAs, ~20-30nt in size, bound by Argonaute proteins that are able to target transcripts by base complementarity.

Argonaute proteins are found throughout various organisms from bacteria and archaea to eukaryotes (Olina et al., 2018). In *Drosophila* five Argonaute proteins are known that can be divided into two classes. The AGO-clade includes Ago1 and Ago2, involved in microRNA (miRNA) or small-interfering RNA (siRNA)-mediated silencing respectively. The PIWI-clade Argonaute proteins include Aubergine (Aub), Argonaute-3 (Ago3), and P-element induced wimpy testis (Piwi), which have evolved a function in a separate gonad-specific RNAi system, the piRNA pathway (see chapter 1.5). The number of Argonaute proteins varies greatly between organisms. Yeasts such as *S. pombe* code for only one Argonaute protein, while *C. elegans* has evolved a third class of Argonaute proteins, the clade of WAGO-family proteins (Hutvagner and Simard, 2008).

Argonaute proteins bind small RNAs originating from distinct sources. Ago1 mainly binds miRNAs (>97% in *Drosophila* S2 cells), while Ago2 binds endogenous siRNAs (endo-siRNAs) (Czech et al., 2008; Okamura et al., 2004). miRNAs are genetically encoded, often in clusters or can be derived from intronic regions (Olena and Patton, 2010). Transcripts forming hairpin-like structures are called pri-miRNAs and their expression is developmentally and tissue-specifically regulated (Wienholds

and Plasterk, 2005). Transcription is promoter- and Pol II-dependent and transcripts undergo splicing as well as polyadenylation (Bartel, 2004). pri-miRNAs are recognised by the microprocessor complex comprised of the nuclear proteins Drosha and its co-factor Pasha (Han et al., 2004; Kim and Kim, 2007). Pasha recognises the junctions of pri-miRNA transcripts (Wilson and Doudna, 2013). This is likely to be required to position Drosha correctly and leads to the Drosha-dependent cleavage of transcripts at the stem. The arising intermediate is called pre-miRNA and is bound by Exportin-5/Ranbp21 and RanGTP, resulting in export to the cytoplasm (Lund and Dahlberg, 2006) (Figure 1.2).

siRNAs can be of endogenous origin or be derived from exogenous sources. Genic endo-siRNAs derived from exonic or intronic regions are highly abundant in *Drosophila* (Czech et al., 2008; Ghildiyal et al., 2008; Okamura et al., 2008). Additionally, transposon-mapping RNA has been shown to be a major source of endo-siRNA precursors. However, siRNAs can be also derived from viral RNA as a response to infections (Czech et al., 2008; Galiana-Arnoux et al., 2006; Wang et al., 2006).

In the cytoplasm, dsRNA intermediates of miRNAs and siRNAs are further processed by Dicer enzymes (Bernstein et al., 2001). miRNA precursor cleavage is mediated by Dicer-1, while siRNA precursors are believed to be processed by Dicer-2. Processing of miRNA and siRNA precursors is further dependent on specific isoforms of the dsRNA-binding protein Loquacious, Loqs-PB or Loqs-PD respectively (Fukunaga et al., 2012; Hartig et al., 2009; Marques et al., 2010; Zhou et al., 2009). miRNA precursors are loaded into Ago1 by the RNA-induced silencing complex loading complex (RLC) comprised of Dicer-1, Ago1 and Loqs-PB. siRNAs are loaded into Ago2 by an RLC comprised of Dicer-2, Ago2 and R2D2 (Hartig and Forstemann, 2011; Marques et al., 2010).

Argonaute proteins are comprised of two lobes. One lobe contains an amino-terminal PAZ domain, while the second lobe is comprised of MID and PIWI domains. The PAZ domain anchors the 2-nucleotide 3' overhang of the presented dsRNAs (Ma et al., 2004). The 5' phosphate interacts with the MID domain, thus stabilising the binding between RNA and protein (Boland et al., 2011; Frank et al., 2010). The formed

complex is called the pre-RNA-induced silencing complex (RISC). Argonaute-bound dsRNA, containing a guide and a complementary passenger strand (referred to miR and miR* respectively for miRNA processing), is further modified by unwinding and ejecting of the passenger strand for miRNAs or Ago2-mediated cleavage for siRNAs, thereby giving rise to the functional RISC (Wilson and Doudna, 2013) (Figure 1.2). Crystal structures for the Argonaute protein from *Pyrococcus furiosus* revealed that the PIWI domain is similar to ribonuclease H (RNase H) and harbours a conserved active motif comprised of a catalytic tetrad (Song et al., 2004; Yamaguchi et al., 2020). In *Drosophila*, unlike humans, both Ago1 and Ago2 exhibit slicing activity (Miyoshi et al., 2005). However, experiments have shown that Ago1 is an ineffective nuclease *in vivo* (Forstemann et al., 2007).

Following assembly, RISC can interact with transcripts by complementary base pairing. Binding of RISC can have several consequences on post-transcriptional control. Binding may result in inhibition of translation, thus preventing the production of proteins coded by the targeted transcripts (Baek et al., 2008; Hendrickson et al., 2009). However, the predominant impact of the RNAi pathway is the reduction of target transcripts by degradation (Eichhorn et al., 2014; Guo et al., 2010) (Figure 1.2). Several studies have shown the recruitment of the PAN2-PAN3 and the CCR4-NOT complex that are involved in deadenylation of target mRNA and its subsequent degradation (Braun et al., 2011; Chekulaeva et al., 2011; Chen et al., 2009; Christie et al., 2013; Eisen et al., 2020; Tuschl et al., 1999).

The RNAi pathway allows cells to adjust their transcriptome by removing excess or harmful transcripts. This has crucial implications for differentiation of cells and has been shown to be utilised to drive development during *Drosophila* embryogenesis (Aboobaker et al., 2005).

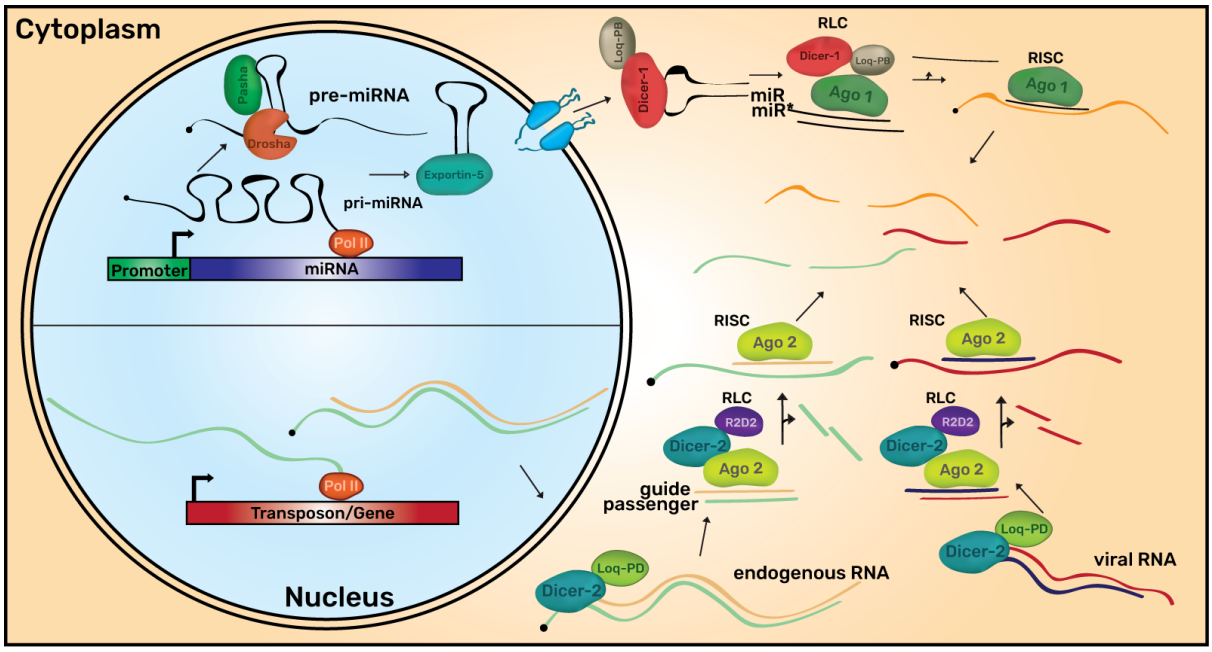

Figure 1.2: The RNAi pathway

Top: miRNA precursors (pri-miRNAs) are transcribed by Pol II and processed by Drosha/Pasha. The resulting short hairpin pre-miRNAs are exported by Exportin-5. In the cytoplasm, Dicer further cleaves pri-miRNAs, thus generating dsRNA precursors. The RISC loading complex (RLC) comprised of Dicer-1, Loqs-PB and Ago1 loads dsRNA precursors. The passenger strand (miR*) is unwound and ejected, thus forming a mature RISC. **Bottom**: siRNA precursors are derived from genic transcripts or transposons as well as viral RNA triggered by infection. dsRNA precursors are processed by Dicer-2 further facilitated by Loqs-PD. The RLC is comprised of Dicer-2, R2D2 and Ago2. The passenger strand is cleaved by Ago2 and removed, thus giving rise to mature RISC. Target engagement of RISC leads to degradation of transcripts.

1.5 The piRNA pathway

RNAi has been shown to affect several pathways including transcriptional regulation during animal development, viral defence, and transposon control using small RNAs loaded into Argonaute proteins (Czech et al., 2008; Galiana-Arnoux et al., 2006; Wang et al., 2006; Wienholds and Plasterk, 2005) thereby protecting cells from deleterious events such as mutations that could compromise genome integrity. Components of the RNAi pathways are expressed in both somatic and germ cells (Graveley et al., 2011).

However, animals have developed additional mechanisms to precisely control the expression of parasitic elements such as transposons in the germline in order to prevent accumulation of transposon-induced DNA damage, thus protecting not only the individual animal but also their offspring. The most studied pathway involved in TE control in germ cells is called the piRNA pathway. It consists of a complex network of factors guided by short non-coding RNAs called Piwi-interacting RNAs (piRNAs) with sizes ranging from ~23-30nt that are loaded into the key effector protein Piwi, a member of the PIWI-clade Argonaute family. Proteins involved in piRNA pathway-dependent transposon control are numerous and several factors act additionally in further cellular pathways (Czech et al., 2013; Handler et al., 2013; Muerdter et al., 2013). Unlike miRNAs and siRNAs used in the canonical RNAi pathway, piRNA precursors originate from single-stranded RNA transcripts encoded by genomic clusters (up to hundreds of kb in size) that are exported, processed and loaded in a fundamentally different way than Dicer-dependent miRNAs and siRNAs. piRNAs predominantly target transposon sequences in *Drosophila* and can guide the piRNA-Piwi complex to sites of nascent transcription as well as influence the processing of transposon transcripts in the nucleus (Teixeira et al., 2017). Piwi-dependent transposon defence relies on nuclear transcriptional gene silencing (TGS) and cytoplasmic post transcriptional gene silencing (PTGS) to control the production of TE transcripts as well as their destiny within the cell after transcription (reviewed in (Czech et al., 2018)).

1.5.1 piRNA clusters

piRNAs originate from genomic loci called piRNA clusters that consist of transposon fragments of various families. Most of the fragments have diverged from the functional consensus sequence of active TEs and lost the ability to transpose. Those clustered regions of “dead” transposons within the genome thereby resemble a graveyard of previously mobile genomic elements. Clusters can be classified into unistrand and dual-strand clusters. Unistrand clusters give rise to transcripts from only one genomic strand and appear dependent on the canonical transcription and export machineries, while dual-strand clusters can be transcribed from both strands and export is dependent on piRNA pathway-specific factors.

1.5.2 Unistrand clusters

Transcripts derived from unistrand clusters are transcribed by RNA polymerase II (Pol II) and rely on canonical RNA processing including capping, splicing, and polyadenylation (Goriaux et al., 2014; Mohn et al., 2014). The promoter regions of unistrand clusters are covered with H3K4me3 marks. However, the body of the cluster is embedded in H3K9me3, which is usually associated with regions of low transcriptional activity (Goriaux et al., 2014). Clusters 20A and *flamenco* (*flam*) are the main source of unistrand-derived piRNAs. While 20A is expressed in both germ cells and follicle cells, *flam* is predominantly active in follicle cells and is involved in suppressing *gypsy*-like transposons (Brennecke et al., 2007; Lau et al., 2009; Malone et al., 2009; Pelisson et al., 1994; Prud'homme et al., 1995). Research further suggests that unistrand clusters such as *flam* co-opt neighbouring gene promoters for their transcription and rely on the same transcription factors controlling these genes (Goriaux et al., 2014). The function of piRNAs depends on their complementary sequence. Therefore, piRNAs derived from unistrand transcripts are sensitive to the direction of transcription. Interestingly, most transposon insertions present in unistrand clusters give rise to piRNAs in antisense orientation relative to transposon mRNAs (Brennecke et al., 2007). piRNAs derived from these clusters are therefore able to

target sense transposon transcripts. To this date, it is not well understood how unistrand clusters evolved in a way to promote antisense insertions of transposon fragments rather than random orientation. Whether there is an active mechanism driving this process or passive selection is not known.

1.5.3 Dual-strand clusters

Dual-strand clusters are transcribed from both genomic strands and are the main source of piRNAs in germ cells. In contrast to unistrand clusters, dual-strand clusters are independent of promoters of nearby genes and use a fundamentally different approach for their transcription (Andersen et al., 2017). Dual-strand clusters are embedded in repressive H3K9me3 marks. The HP1a homolog Rhino (Rhi) is expressed in germ cells only and can selectively bind to dual-strand cluster-associated H3K9me3 through its chromo domain (Klattenhoff et al., 2009; Zhang et al., 2012; Zhang et al., 2014). However, H3K9me3 marks are not exclusively found at dual-strand clusters but rather mark large parts of the genome. The mechanism guiding Rhi specifically to clusters is unknown but it is widely established that Rhi-binding is specific to dual-strand clusters. Rhi is regarded as a reader protein, able to read the epigenetic histone code. However, whether additional epigenetic chromatin modifications besides H3K9me3 or entirely different mechanisms are affecting Rhino's specific recruitment to dual-strand clusters remains to be resolved. Interestingly, the HMG protein Maelstrom (Mael) has been proposed to supress transcription of genes in close proximity to dual-strand clusters, thus favouring Rhi-dependent cluster transcription (Chang et al., 2019) (Figure 1.3).

Rhi is recruited to dual-strand clusters and serves as a binding platform for other germline-specific piRNA factors. Deadlock (Del) binds to the chromoshadow domain of Rhi and recruits Cutoff (Cuff) (Pane et al., 2011). Together, Rhi, Del and Cuff form the dual-strand cluster-associated RDC complex (Mohn et al., 2014). Del additionally recruits Moonshiner (Moon), which assembles a protein complex consisting of the TATA-box binding protein-related factor 2 (TRF2) and additional components of the transcription activation complex (Andersen et al., 2017). This leads

to the recruitment of Pol II and the initiation of transcription at H3K9me3-covered dual-strand clusters. Arising transcripts are believed to be bound by Cuff, which shows close resemblance to decapping enzymes. While Cuff likely lost its decapping activity, it might protect the 5' end of transcripts from post-transcriptional processing and could prevent Pol-II termination, thus affecting the interaction with the canonical exporting machinery (Chen et al., 2016) (Figure 1.3).

1.5.4 piRNA precursor export

Canonical mRNA transcripts are primarily exported via the Nxf1-Nxt1 pathway. Emerging transcripts are modified with a 5' cap structure after about 20-30 nucleotides have been synthesised (Proudfoot et al., 2002). This is followed by splicing and recruitment of the transcription and export complex (TREX) to nascent transcripts (Herold et al., 2001; Kohler and Hurt, 2007). Following polyadenylation of the 3' terminus, Nxf1 and Nxt1 are recruited and the mature mRNA-protein complex exported from the nucleus. This pathway allows quality control of RNAs and prevents export of immature or aberrant transcripts.

Unistrand clusters such as 20A and *flam* seem to depend on canonical export (Dennis et al., 2016; Mohn et al., 2014) but it is currently unclear how those transcripts are licenced for subsequent piRNA production and guided towards processing centres, while other mRNAs are translated and do not give rise to substantial piRNA levels.

Dual-strand cluster transcripts, however, are capped but are neither spliced nor polyadenylated, thereby lacking common features required for canonical mRNA export (Andersen et al., 2017; Muerdter et al., 2012). Therefore, dual-strand clusters have evolved a separate mechanism to promote export and guide their transcripts to sites of piRNA production outside the nucleus. The recently discovered factor, Bootlegger (Boot), has been shown to be recruited by the RDC-complex (ElMaghraby et al., 2019; Kneuss et al., 2019). This leads to further recruitment of the THO complex and the DEAD-box helicase U2AF65-associated protein (UAP56) (Zhang et al., 2012), components of the TREX complex. Additionally, Nuclear export factor 3 (Nxf3) in

complex with Nxt1 binds to the transcript-export complex and leaves the nucleus in a Crm1 (chromosomal maintenance 1) pathway-dependent manner. The cluster transcript is exported together with the Nxf3, Nxt1 and Boot complex and subsequently guided to sites of piRNA production at the nuage (ElMaghraby et al., 2019; Kneuss et al., 2019) (Figure 1.3).

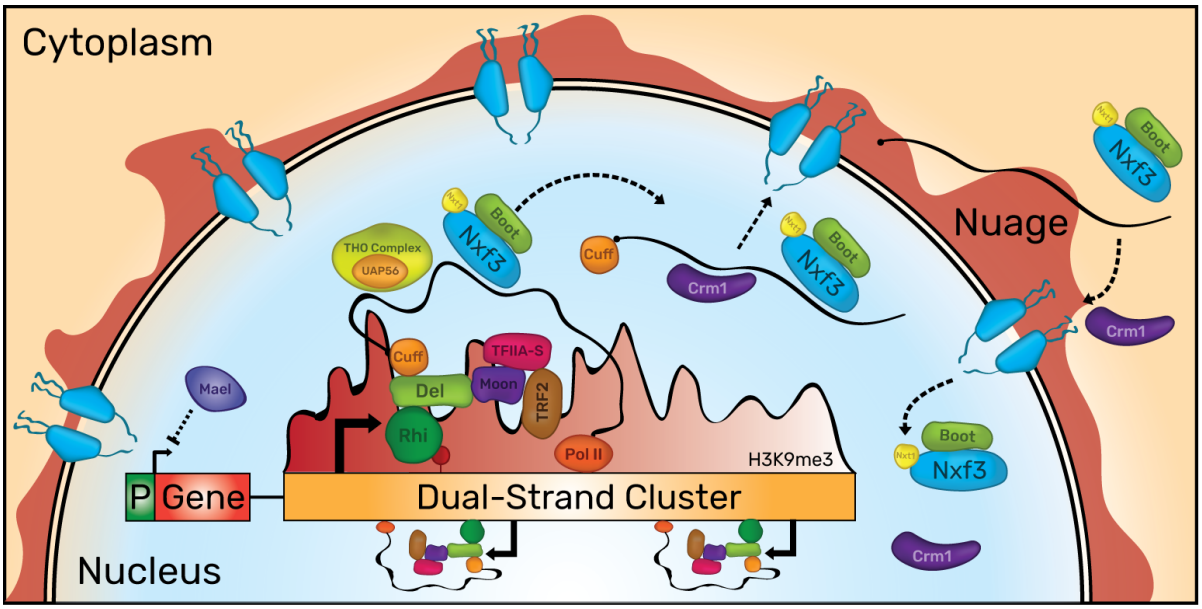

Figure 1.3: Dual-strand cluster transcription and export

Dual-strand clusters are covered in H3K9me3 marks that are bound by Rhi. Mael is believed to repress canonical transcription at nearby promoters (P) of genes. Rhi binds Del and serves as a binding platform for additional proteins. Moon is bound by Del and recruits further transcription initiation proteins such as TFIIA-S (transcription factor II A) and TRF2, which recruit Pol II resulting in transcription initiation. Del additionally binds Cuff, which is likely required to shield the 5' end from degradation. Rhi-dependent transcription recruits Boot as well as Nxt1 and Nxf3 together with TREX complex components including UAP56. The transcript leaves the nucleus in a Crm1-denpendent way and is transported to the nuage. The export complex than relocates to the nucleus.

1.5.5 piRNA biogenesis

Following export of cluster transcripts from the nucleus, the piRNA precursors have to find the sites of piRNA biogenesis in the cytoplasm. piRNA precursor processing occurs at specialised cellular compartments. In ovarian follicle cells of somatic origin, precursors are processed and loaded into Piwi protein at Yb bodies (Dennis et al., 2016; Dennis et al., 2013; Murota et al., 2014). Biogenesis in germ cells, such as ovarian nurse cells, occurs at perinuclear structures called nuage (Brennecke et al., 2007; Findley et al., 2003; Lim and Kai, 2007; Malone et al., 2009). Little is known about the precise mechanisms required for specifically guiding precursors to sites of biogenesis, while excluding other RNAs. However, it is possible that most exported RNAs are screened for base complementarity with Aub/Ago3-bound piRNAs. Once piRNAs engage a target, the transcript is funnelled into the biogenesis machinery, while canonical mRNA transcripts are unaffected (Senti et al., 2015). Maturation of piRNA precursors can be divided in Zucchini (Zuc)-dependent biogenesis, utilising the mitochondria associated endonuclease Zuc (Ipsaro et al., 2012; Nishimasu et al., 2012; Pane et al., 2007), and the ping-pong cycle that uses transposon mRNA in a feed-forward loop to generate piRNAs on demand (Brennecke et al., 2007; Gunawardane et al., 2007). However, both processes are interconnected (Han et al., 2015; Mohn et al., 2015; Senti et al., 2015; Wang et al., 2015).

1.5.6 Zuc-dependent piRNA biogenesis

The initial cleavage event of cluster transcripts is thought to be generated by either Aub/Ago3-dependent cleavage as part of the ping-pong cycle (see chapter 1.5.7) or by the endonuclease Zuc (Han et al., 2015; Ipsaro et al., 2012; Mohn et al., 2015; Senti et al., 2015; Wang et al., 2015). This generates a 5’ monophosphate (5’-P) that is bound by Piwi at Yb bodies or Piwi/Aub at nuage (Han et al., 2015; Mohn et al., 2015; Senti et al., 2015; Wang et al., 2015). Following binding, the RNA helicase Armitage (Armi) is recruited to the precursor-Piwi/Aub complex. This in term leads to translocation of the Armi-bound complexes to mitochondria, where several additional

biogenesis factors reside (Ge et al., 2019; Ishizu et al., 2019; Munafo et al., 2019). On the mitochondria outer membrane, two transmembrane proteins, Daedalus (Daed) and Gasz assemble into homopolymeric and heteropolymeric complexes. Together, both might tether precursor-bound Armi to the mitochondrial outer membrane via binding to precursor RNA (Czech et al., 2013; Handler et al., 2013; Munafo et al., 2019). Active Zuc forms homodimers on the mitochondria membrane and cleaves piRNA precursors in a Piwi-dependent manner (Ipsaro et al., 2012; Nishimasu et al., 2012). Piwi binding of the 5'-P as well as Armi association is believed to shield approximately 25-nt of the precursor RNA. The precursor RNA likely further translocates towards Zuc for cleavage. Zuc-dependent cleavage occurs downstream of the Piwi-bound RNA, thus generating a Piwi footprint of ~25nt fragments (Han et al., 2015; Mohn et al., 2015; Nishimasu et al., 2012). Zuc shows a strong cleavage bias towards downstream uridine residues, also known as 1U bias, which is reflected in the abundance of a 5' terminal U in most piRNA populations across species (Aravin et al., 2006; Brennecke et al., 2007; Kawaoka et al., 2009). However, it is currently unknown how Zuc specifically recognises uridines and if this bias has any implication on the downstream function of generated piRNAs. The recently reported structure of *Drosophila* Piwi in complex with piRNAs suggest a binding pocket for the 5' U of piRNAs between the MID and PIWI domain of Piwi proteins (Yamaguchi et al., 2020). However, it remains unclear if this pocket binds preferentially 5' uridine residues or if other nucleotides are able to bind with similar affinities. Additionally, studies have suggested discrimination against all nucleotides except U during piRNA biogenesis (Stein et al., 2019). Cleavage downstream of the Piwi bound piRNA precursor creates a new 5' terminus that can be bound by unloaded Piwi. Zuc, again cleaves downstream of Piwi thereby generating the 3' end of the previous cut piRNA as well as a new 5' terminus of the precursor. This process creates phased piRNAs loaded into Piwi from a common piRNA precursor and is known as trail piRNA biogenesis (Han et al., 2015; Mohn et al., 2015). Zuc-mediated cleavages can occur several times downstream of the initial cleavage site. Unloaded Piwi protein associates with the biogenesis factor Shutdown (Shu) in the cytoplasm. Shu is believed to facilitate loading of Piwi together with Heat shock protein 90 (Hsp90) (Olivieri et al., 2012; Preall et al.,

2012). The 3' ends of Piwi-bound piRNAs are defined by Zuc cleavage and trimming seems to be unnecessary in *Drosophila* (Han et al., 2015; Hayashi et al., 2016; Mohn et al., 2015). 3' ends are further methylated by Hen1 thus generating mature piRNAs (Horwich et al., 2007; Saito et al., 2007). The mature piRNA-Piwi complex is presumably undergoing a conformational change that exposes a bipartite nuclear localisation signal (NLS), ultimately licensing the complex for nuclear import (Yashiro et al., 2018) (Figure 1.4).

1.5.7 The ping-pong amplification loop

The ping-pong cycle represents an additional route of piRNA biogenesis. This amplification loop utilises two of the five Argonaute proteins present in *Drosophila*: Aubergine (Aub) and Argonaute 3 (Ago3). Aub-bound piRNAs are mostly antisense to transposon transcripts, while Ago3 predominantly binds sense piRNAs. Antisense piRNAs loaded into Aub recognise sense transposon transcripts in the perinuclear nuage of germ cells by recruitment of Krimper (Krimp), a Tudor-domain protein. Krimp also interacts with unloaded Ago3 and is responsible for the sense bias of Ago3-bound piRNAs (Sato et al., 2015; Webster et al., 2015). Aub exhibits slicer activity and cleaves transposon RNA. The 3' cleavage product is then loaded into Ago3 aided by the DEAD-box helicase Vasa (Vas). The Tudor-domain protein, Qin, prevents loading of Aub cleavage products into unloaded Aub, thus contributing to the antisense bias of Aub-bound piRNAs (Wang et al., 2015; Zhang et al., 2011). The newly generated 5' end bound by Ago3 overlaps 10nt with the Aub-bound piRNA. This results in an Adenine (A) at position 10 of the Ago3-bound piRNA also known as the ping-pong signature (Brennecke et al., 2007; Gunawardane et al., 2007). The 3' end can be further processed by Zuc and/or trimmed by Nibbler (Nbr) (Hayashi et al., 2016). The piRNA is methylated by Hen1, thus giving rise to a mature Ago3-piRNA complex (Horwich et al., 2007; Saito et al., 2007). The sense piRNA bound by Ago3 can recognise piRNA cluster transcripts and initiate precursor RNA slicing. The generated 3' product with a 5' terminal U is then loaded into Aub. Further processing can give rise to mature Aub-piRNA complexes that can then restart the cycle by cleaving a

complementary transposon transcript (Brennecke et al., 2007; Gunawardane et al., 2007). Alternatively, Aub-bound precursors can be funnelled into the Zuc-dependent piRNA biogenesis pathway, thereby further contributing to Piwi loading (see chapter 1.5.6) (Figure 1.4).

This germ cell-specific mechanism is able to reduce transposon mRNA levels and prevent translation. Additionally, transposon transcripts are utilised in a feed-forward loop that generates piRNAs and thereby increases the defence ability of germ cells in response of increased transposon transcription.

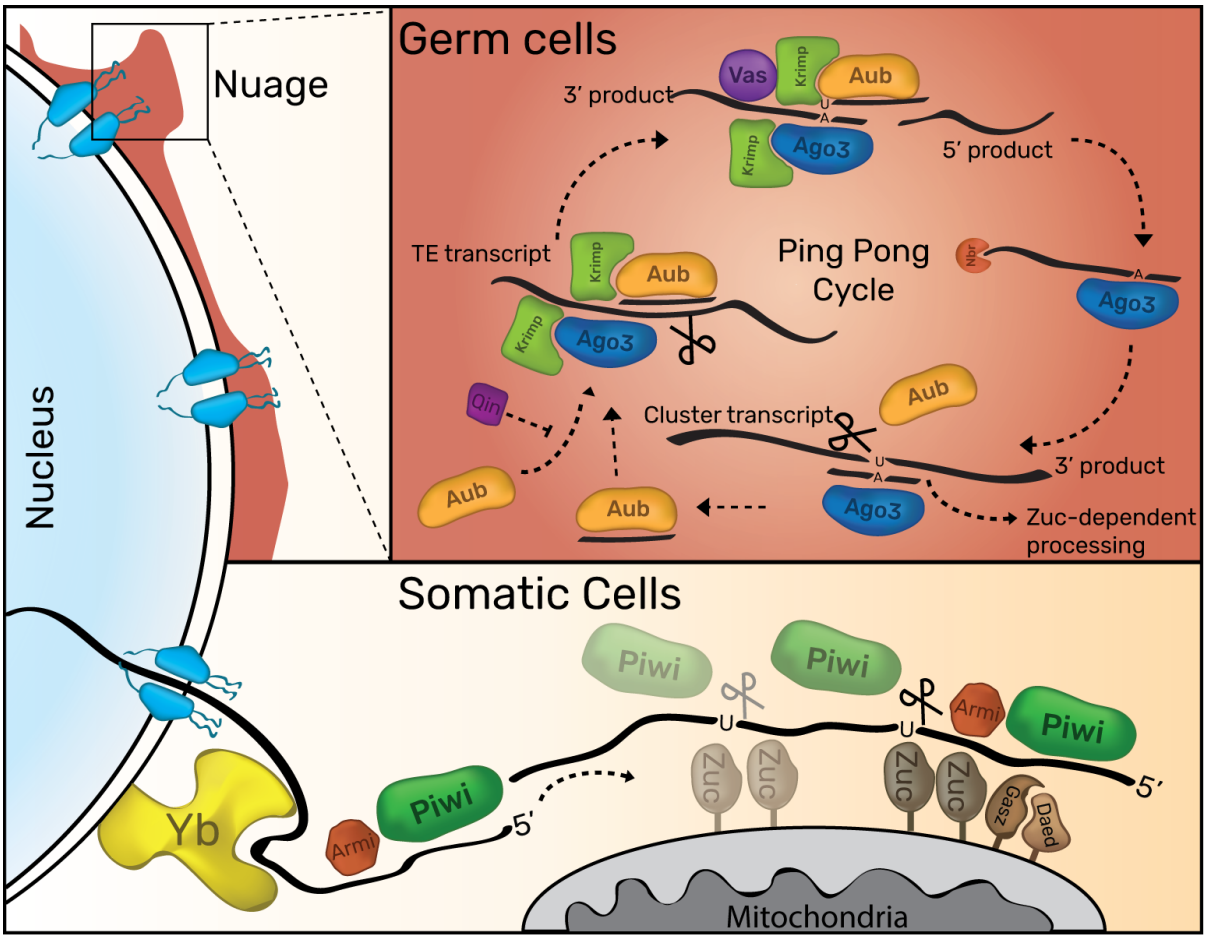

Figure 1.4: piRNA biogenesis in somatic and germ cells

In somatic cells, piRNA biogenesis occurs at Yb bodies (bottom). Piwi binds the 5' phosphate of transcripts derived from clusters such as *flam*. Together with Armi, the complex translocates to the mitochondria and is tethered there by Gasz and Daed. Zuc forms a homodimer on the mitochondria outer membrane and cuts the Piwi/Armi-bound precursor transcript upstream of a uridine residue. This generates a new 5' P that is bound by unloaded Piwi and Armi leading

to the generation of trail piRNAs loaded into Piwi. The ping-pong cycle is active in germ cells at the perinuclear nuage (top). piRNAs loaded in Aub recognise TE transcripts by base complementarity. Krimper (Krimp) recruits Ago3 and Aub slices the TE transcript. Ago3 is loaded with the 3' product generated by Aub slicing. Qin prevents loading of Aub cleavage products into unloaded Aub. The 3' end of Ago3-bound piRNAs is further trimmed by Nibbler (Nbr) and methylated by Hen1. Ago3-bound mature piRNAs can recognise and slice cluster transcripts. Aub is loaded with the 3' cleavage product. Simplified, the piRNA intermediate bound by Aub is further processed in Zuc-dependent trail piRNA biogenesis and gives rise to Aub-piRNA complexes that are able to restart the cycle.

1.5.8 piRNA pathway-dependent transcriptional gene silencing

Mature piRNA-Piwi complexes are essential for transcriptional gene silencing of transposons in the nucleus (Saito et al., 2010). The majority of piRNAs loaded into Piwi are antisense to transposon transcripts and can target these through sequence complementarity. Once Piwi is loaded with a piRNA, the complex undergoes a conformational change and exposes a nuclear localisation sequence (NLS) (Yashiro et al., 2018), resulting in the translocation of the complex to the nucleus, which is critical for its function in TGS (Saito et al., 2010). Almost all research aiming to understand piRNA-mediated TGS in *Drosophila* has been conducted in ovaries or *in vitro* systems such as ovarian somatic cells (OSCs) (Donertas et al., 2013; Sienski et al., 2015; Sienski et al., 2012; Yu et al., 2015). The process of piRNA-guided TGS can be divided into three stages: Target engagement, establishment of silencing and Piwi-dependent maintenance.

Target engagement relies on the guiding piRNA loaded into Piwi. Nascent transposon transcripts are the likely targets for piRNA-Piwi complexes and transposon splicing has been shown to be directly affected by Piwi-bound piRNAs (Teixeira et al., 2017). piRNA-mediated binding is believed to induce another conformational change of Piwi that allows downstream effectors to interact (Yu et al., 2015). This is further supported by experiments showing piRNA-independent binding to nascent RNA is not sufficient to induce TGS at reporter loci (Fabry et al., 2019; Le Thomas et al., 2013; Sienski et al., 2015; Yu et al., 2015). Interestingly, while Piwi was previously reported

to exhibit slicer activity (Saito et al., 2006), the recently published crystal structure of *Drosophila* Piwi and *in vitro* assays suggest otherwise (Yamaguchi et al., 2020). The PIWI domain of Piwi adopts an RNase H-like fold including a catalytical DVDK tetrad but purified Piwi was unable to slice target RNA efficiently *in vitro*, suggesting that Piwi lost its catalytic activity (Yamaguchi et al., 2020). Indeed, previous studies suggest that the putative slicing ability of Piwi is dispensable for transposon silencing (Saito et al., 2010; Sienski et al., 2012). Since Piwi's nuclear localisation is crucial for its function in transposon silencing but the slicer activity is dispensable (Saito et al., 2010; Yamaguchi et al., 2020), Piwi might have evolved a nucleus-specific function likely as a binding platform for additional components of the piRNA pathway, rather than a function in RNA slicing such as Aub and Ago3 (see chapter 1.5.7). Interestingly, slicing ability is dispensable for the function of nuclear Piwi proteins in other species as well. For instance, the nuclear Piwi protein PRG-1 in *C. elegans* has been shown to function independent of its slicing ability (Bagijn et al., 2012).

In recent years, several proteins identified in genetic screens as novel piRNA pathway factors (Czech et al., 2013; Handler et al., 2013; Muerdter et al., 2013) have been shown to interact with Piwi. The small zinc-finger proteins Asterix (Arx/Gtsf1) is necessary for TGS induction. Arx harbours a putative RNA binding domain and is expressed at low levels specifically in gonads. Its localisation to the nucleus is Piwi-dependent, however only a small fraction of Piwi is bound to Arx and Piwi's localisation is independent of Arx expression. The majority of Arx is bound by the nuclear Piwi pool, which is crucial for Piwi's function (Donertas et al., 2013; Muerdter et al., 2013; Ohtani et al., 2013). Little is known about the precise function of Arx in TGS but given Arx's RNA binding domain, it is possible that Arx engages Piwi-bound nascent RNA targets and facilitates the recruitment of downstream factors further required for silencing by stabilising the complex of nascent RNA and piRNA-bound Piwi. However, experimental evidence that Arx associates with RNA has not been generated to this date.

Silencing establishment has been shown to involve a piRNA pathway-specific complex. The main effector of this complex is Panoramix (Panx), which interacts with Piwi and is a strong inducer of epigenetic chromatin state changes (Sienski et al.,

2015; Yu et al., 2015). Loss of Panx results in deregulation of TEs in germ cells and somatic cells. Additionally, repressive H3K9me3 marks are lost from target loci. Both Panx and Piwi localise to the nucleus in an independent manner. Artificial tethering of Panx to DNA or RNA reporters leads to strong transcriptional gene silencing, which is accompanied by the deposition of H3K9me3. While Panx is able to induce TGS, it is currently not known how it connects to the general silencing machinery of the cell that is required for deposition of H3K9me3 and heterochromatin formation (Sienski et al., 2015; Yu et al., 2015).

Recent studies identified further binding partners of Panx (Batki et al., 2019; Fabry et al., 2019; Murano et al., 2019; Zhao et al., 2019). The nuclear export factor 2 (Nxf2) together with its co-factor Nuclear Transport Factor 2 Like Export Factor 1 (Nxt1/p15) bind Panx. Nxt1 is involved in canonical export of mature mRNAs (see chapter 1.5.4) and also interacts with Nxf1 that mediates RNA binding and translocation to the nuclear pore followed by export (Braun et al., 2001; Fribourg et al., 2001; Levesque et al., 2001). However, Nxf2 is not involved in bulk RNA export but has been co-opted by the piRNA pathway for transcriptional gene silencing. While both Nxf2 and Nxt1 bind Panx and, in complex, can induce TGS, the silencing ability resides in Panx (Batki et al., 2019; Fabry et al., 2019). Nxf2's localisation and stability are dependent on Panx, which harbours a nuclear localisation sequence (Batki et al., 2019; Fabry et al., 2019; Murano et al., 2019; Zhao et al., 2019). While Panx has been shown to engage with transposon transcripts (Sienski et al., 2015; Yu et al., 2015), whether Nxf2 binds nascent transposon RNA is still debated. *In vitro* studies have shown that the RNA-binding domain of Nxf2 is able to bind RNA in general. The same study, however, was unable to find evidence for transposon RNA binding *in vivo* (Batki et al., 2019). On the other hand, two independent studies reported some evidence for TE transcript binding to Nxf2 under artificial conditions (Murano et al., 2019; Zhao et al., 2019). The transcriptional silencing complex has been named PICTS (Panx-induced co-transcriptional silencing), SFiNX, PPNP or Pandas complex by the four labs involved in the discovery (Batki et al., 2019; Fabry et al., 2019; Murano et al., 2019; Zhao et al., 2019) and will be referred to as PICTS complex from hereafter (Figure 1.5).

Several piRNA-dependent TGS factors have been characterised throughout the years. However, the mechanism and hierarchy within the pathway is often poorly understood. For instance, Mael, in addition to its role in cluster biology (see chapter 1.5.4) (Chang et al., 2019) has also been shown to act as TGS factor. Mael harbours a high mobility group (HMG) domain as well as a MAEL domain (Sienski et al., 2012). In germ cells, Mael is detected in the cytoplasm, the nuage and the nucleus (Findley et al., 2003). Overexpression of Mael results in primarily nuclear localisation in an *in vitro* system such as OSCs, which express a functional piRNA-dependent TGS pathway but have no dual-strand cluster transcription or ping-pong cycle, (Sienski et al., 2012). However, depletion of Mael in OSCs has been shown to reactivate transposon expression similar to Piwi loss, with little to no effect on H3K9me3 occupancy at affected transposon insertions. This data suggests that Mael might have evolved multiple functions within the piRNA pathway beyond TGS.

While the PICTS complex explains the connection to Piwi and the piRNA pathway-specific mechanism of transposon silencing, the downstream effectors involved in translating the information carried by the guiding piRNA into precise silencing events have not been identified yet. However, several components involved in general heterochromatin formation have been implicated in transposon control and maintenance of silent epigenetic states. A hallmark of piRNA-dependent TGS is epigenetic conversion of histone states by removal of activating chromatin marks from promoter regions and the deposition of repressing marks at transposon bodies. Lsd1, in complex with CoRest, is involved in the removal of methylation marks from H3K4 and emerged as strong candidate in genetic screens for transposon regulators in *Drosophila* gonads (Czech et al., 2013; Handler et al., 2013; Muerdter et al., 2013) (see also chapter 1.3). Additionally, Egg has been proposed as the relevant histone methyltransferase responsible for H3K9me3 deposition (Sienski et al., 2015; Yu et al., 2015) (see chapter 1.3). How the general silencing machinery connects to the PICTS complex or other components of the piRNA pathway is still debated.

A recent study suggests a connection between Egg and the SUMO E3 ligase Su(var)2-10, which shows deregulation of transposons upon depletion. Su(var)2-10 has been proposed to physically interact with several TGS components including Piwi,

Arx and Panx (Ninova et al., 2020a; Ninova et al., 2020b). Following H3K9me3 deposition, the previously active euchromatin is converted into heterochromatin. In another study, Histone 1 (H1) is actively recruited by piRNA-Piwi complexes and decreases target loci accessibility (Iwasaki et al., 2016). Last, HP1a is well-known for its function in heterochromatin formation and maintenance (see chapter 1.3). Loss of HP1a leads to deregulation of transposons (Ohtani et al., 2013; Wang and Elgin, 2011) and is believed to be required for heterochromatin formation following H3K9me3 deposition at transposon insertions (Figure 1.5). Of note, a potential direct interaction of HP1a dimers with a PxVxL-type motif in the amino-terminal domain of Piwi was reported, which would provide a mechanism that recruits Piwi loosely to chromatin in a piRNA-independent way (Brower-Toland et al., 2007).

Maintenance of a silenced transposon state is Piwi- and piRNA pathway-dependent (Sienski et al., 2015; Sienski et al., 2012; Yu et al., 2015). Whether occasional transcription of silenced transposons is required to maintain silencing is currently under investigation. In addition, silencing is heavily dependent on the presence of multiple factors that are not piRNA pathway-specific as listed above. Disturbing either of the general silencing machinery factors not only affects transposon expression, but has a much broader effect on the transcriptome (Ninova et al., 2020b).

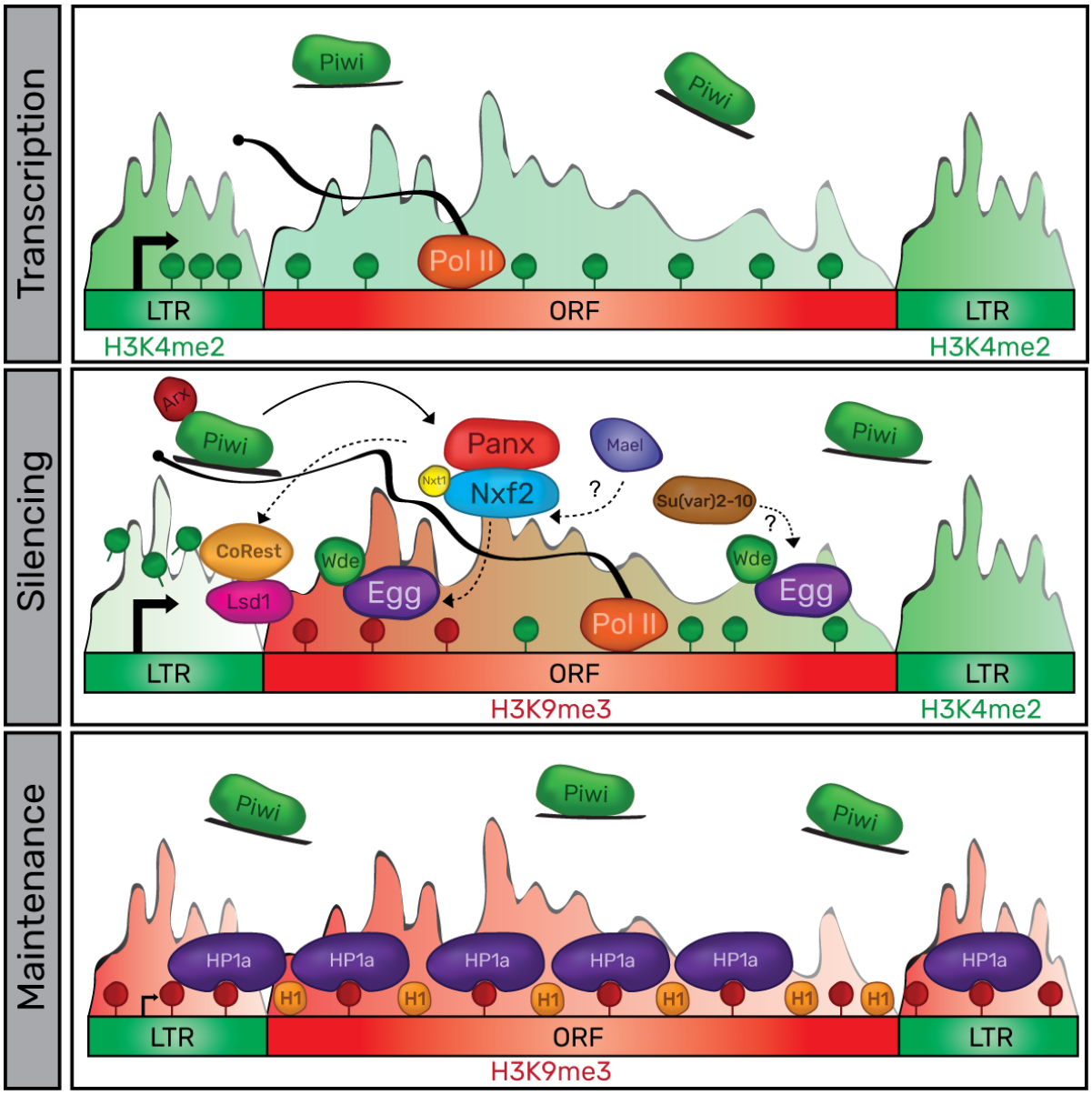

Figure 1.5: piRNA pathway-dependent transcriptional gene silencing

Transcription: Transposons are transcribed by Pol II and permissive chromatin marks are deposited at LTRs and gene body encoding the TE's open reading frame (ORF). **Silencing**: Nascent transcripts are recognised by Piwi through complementary piRNAs. Target engagement recruits the PICTS complex including Nxf2, Panx and Nxt1, leading to the recruitment of the general silencing machinery. Lsd1 removes permissive chromatin marks such as mono-, di- and tri-methylation of H3K4, while Egg and its co-factor Wde methylate H3K9. **Maintenance**: Repressive chromatin marks are bound by HP1a and the genomic locus is transformed into condensed heterochromatin thus shutting down transcription. The presence of Piwi is required to maintain the silenced state.

1.6 Transgenerational epigenetic inheritance

Embryos inherit one set of chromosomes from each of their parents that combine to form the zygotic genome. However, not only information in form of DNA is inherited by the next generation. Several molecules and even entire organelles have been indicated to be transmitted primarily through the oocyte and carry epigenetic information indispensable for the development of offspring (see also chapter 1.7.1) (Boskovic and Rando, 2018). Early embryogenesis represents a phase of cell totipotency and pluripotency (Hu, 2019; Wu et al., 2016). However, the freedom to form all cell lineages required by an organism comes with a great price. Regulatory adaptations such as genome organisation, definition of heterochromatin and euchromatin as well as DNA methylation are partially lost in newly formed zygotes and have to be restored during development (Borsos and Torres-Padilla, 2016; Hug et al., 2017). Transgenerational epigenetic inherited factors carry this memory and ensure that features not coded by DNA are remembered by offspring.

1.6.1 Transgenerational epigenetic inheritance of small RNAs

Small RNAs represent a distinct class of transgenerationally inherited factors carrying epigenetic information required during embryogenesis and beyond. For instance, inheritance of piRNAs from one generation to the next has profound implications for transposon silencing and the fertility of offspring in *Drosophila* (Brennecke et al., 2008) (see also chapter 1.6.2). The piRNA pathway in mammals is primarily active during male spermatogenesis (Chuma and Nakano, 2013). The impact of inherited piRNAs in mammals, however, is poorly understood. In fact, research suggests that in mature sperm, piRNAs are almost entirely absent. Instead, tRNA fragments and microRNAs have been identified to be transmitted (Peng et al., 2012). While the majority of inherited small RNAs are inherited from the female parent through the oocyte, paternal inheritance of small RNAs has been shown to be essential for developmental potential in mice (Yuan et al., 2016).

C. elegans is an ideal model organism to study multigenerational effects of small RNAs due to its short generation time (~3 days). Stable and completely penetrant multigenerational transgene silencing has been described in *C. elegans* (Ashe et al., 2012; Luteijn et al., 2012; Shirayama et al., 2012). piRNA pathway-dependent transgene silencing can be maintained in the next generations. The initial silencing event depends on the piRNA-associated Argonaute PRG-1, therefore indicating an involvement of the piRNA pathway. However, the maintenance of silencing in the next generation is independent of PRG-1 but requires other factors involved in transcriptional silencing. PRG-1-mediated silencing is slicing-independent and piRNAs in complex with PRG-1 have the ability to recruit RNA-dependent RNA polymerases (RdRPs) that uses targeted transcripts as a template for the generation of 22G RNAs (Bagijn et al., 2012; Lee et al., 2012). piRNA-triggered secondary 22G RNAs act downstream of piRNAs and are loaded into nuclear Argonaute protein HRDE-1/WAGO9 (Buckley et al., 2012). Indeed, evidence suggests that 22G RNAs are the inherited triggers maintaining multigeneration silencing. Considering that both pre-mRNA and mature mRNA levels of sensor constructs are reduced following inheritance of specific silencing triggers, it is likely that both transcriptional and post-transcriptional gene silencing mechanisms could be involved in the silencing response in *C. elegans*. Silencing is dependent on HRDE-1 and NRDE-1-mediated H3K9me3 deposition at target sites guided by 22G RNAs, emphasising the requirement for changes in the chromatin state for efficient target silencing (Ashe et al., 2012; Luteijn et al., 2012; Shirayama et al., 2012).

Transgenerational inheritance of small RNAs has also been extensively studied in plants especially in the context of DNA methylation. While CG and CHG methylation levels (see chapter 1.3) of sperm cells in early development of *Arabidopsis* remain relatively constant (Jullien et al., 2012), CHH methylation is lost at retrotransposons. Loss of methylation has been shown to be restored over generations by an RNAi-dependent mechanism (Teixeira et al., 2009). Re-establish of CHH methylation *de novo* during embryogenesis is likely directed by inherited siRNAs (Calarco et al., 2012). Similarly, epigenetic inheritance of small RNAs in the fission yeast *Schizosaccharomyces pombe* (*S. pombe*) is crucial for maintenance of gene silencing

and epigenetic chromatin states. Secondary siRNAs spreading to target genes are transmitted during mitosis and meiosis (Yu et al., 2018). siRNAs are involved in recruitment of the RNAi-induced transcriptional silencing (RITS) complex that recruits the Su(var)3-9 homolog Clr4, ultimately leading to H3K9 methylation and target silencing (Buhler et al., 2006; Jih et al., 2017; Volpe et al., 2002).

1.6.2 Effects of maternally deposited piRNAs in *Drosophila*

In *Drosophila*, transgenerational epigenetic inheritance of small RNAs is necessary for fitness and fertility of offspring. Inherited piRNAs are involved in silencing transposons in adult gonads of the next generation.

Interestingly, of the three PIWI-clade proteins expressed in *Drosophila* ovaries, only Piwi and Aub are maternally deposited. Very little, if any, Ago3 is detectable in early embryos (Brennecke et al., 2008). Little is known about the function of maternally deposited piRNAs and piRNA pathway factors such as Piwi during early embryogenesis. However, evidence suggests that Piwi is responsible for re-establishment of epigenetic states at dual-strand clusters in early embryos ultimately enabling precursor transcription (see chapter 1.5.3). Maternal inheritance of piRNAs derived from a transgene-derived source locus were able to silence complementary sequences in adult gonads depending on the deposition of sequence specific piRNAs (Le Thomas et al., 2014). This set of inherited piRNA species is often referred to as the cytotype. Further research revealed that cluster redefinition occurs during early development. Embryos, depleted of Piwi using genetic tools, showed a decrease of Rhi binding at dual-strand clusters in adult ovaries, while flies depleted of Piwi during later stages of development did not show a reduction (Akkouche et al., 2017). This suggests that the epigenetic chromatin state necessary for recruiting Rhi is likely defined during embryogenesis (Figure 1.6). Whether inherited Piwi and Aub or other components of the piRNA-pathway have functions beyond their involvement in re-establishing piRNA clusters is currently unknown.

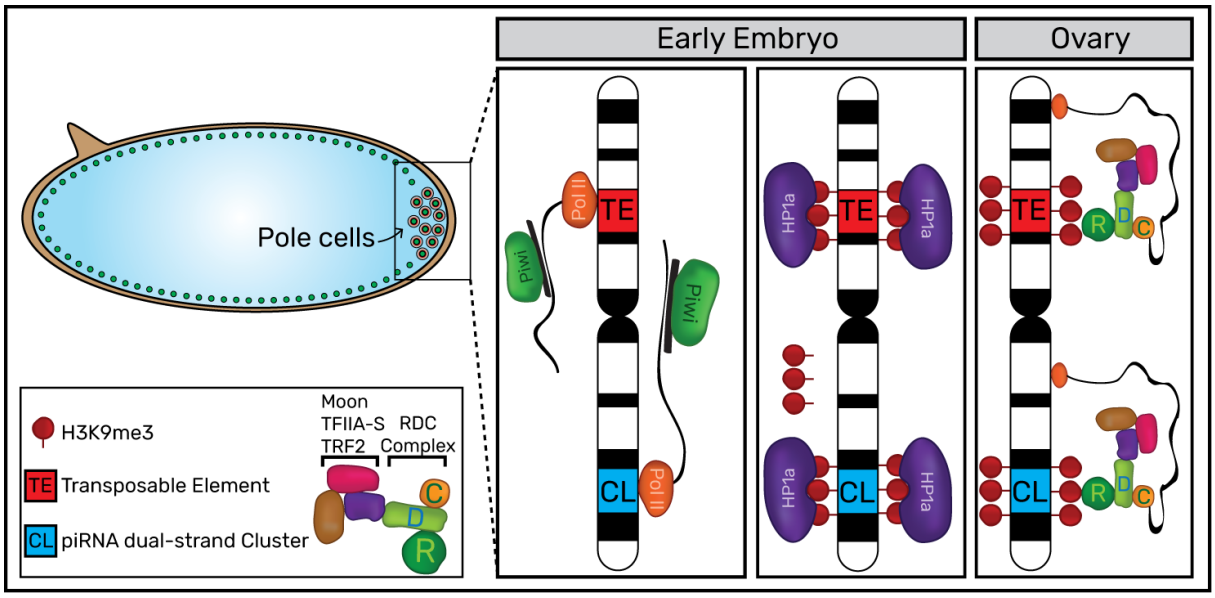

Figure 1.6: Maternally inherited piRNAs redefine piRNA source loci

piRNAs in complex with Piwi protein are maternally inherited through the oocyte. In the developing embryo, Piwi localises to somatic nuclei as well as the pole plasm surrounding future germ cells. Following entry of Piwi in the pole cell nuclei and cellularisation, Piwi is believed to engage cluster transcripts based on sequence complementarity licenced by the piRNA guide. This leads to the re-establishment of chromatin modification at future piRNA source loci. During oogenesis, mature germ cells transcribe dual-strand clusters marked by maternally inherited Piwi thus giving rise to piRNA precursor transcripts used in transposon silencing and for maternal inheritance of the next generation.

Several long-standing observations regarding the loss or decrease of fertility in flies have been attributed to maternal transmission of specific piRNA species. For instance, hybrid dysgenesis, a phenomenon characterised by transposition events and genome instability has been shown to depend on inherited piRNAs (Brennecke et al., 2008). Crosses of female wild-caught flies with certain laboratory strains produce fertile progeny, while the reciprocal cross yields sterile offspring (Bingham et al., 1982; Bucheton et al., 1984; Kidwell et al., 1977). This phenomenon has been described more than 50 years ago and was shown to be caused by the mobilisation of transposable elements, present in the genome of the wild-caught strain but not of the lab strain, in ovaries. The mechanism behind hybrid dysgenesis, however, was only discovered more recently. piRNAs, able to target a specific transposon, are only

produced in the wild-caught flies that carry it. The female parent maternally deposits those piRNA species into the developing egg, thereby protecting the future ovaries from TE expression. If, however, the transposon is inherited from the male parent, females lacking the transposon are not able to produce specific piRNAs targeting this TE. Thereby, the maternally deposited piRNA population is not able to prevent transposon expression in the offspring (Figure 1.7) (Brennecke et al., 2008). However, through currently unknown molecular mechanisms, fertility of dysgenic offspring can improve throughout adulthood (Khurana et al., 2011).

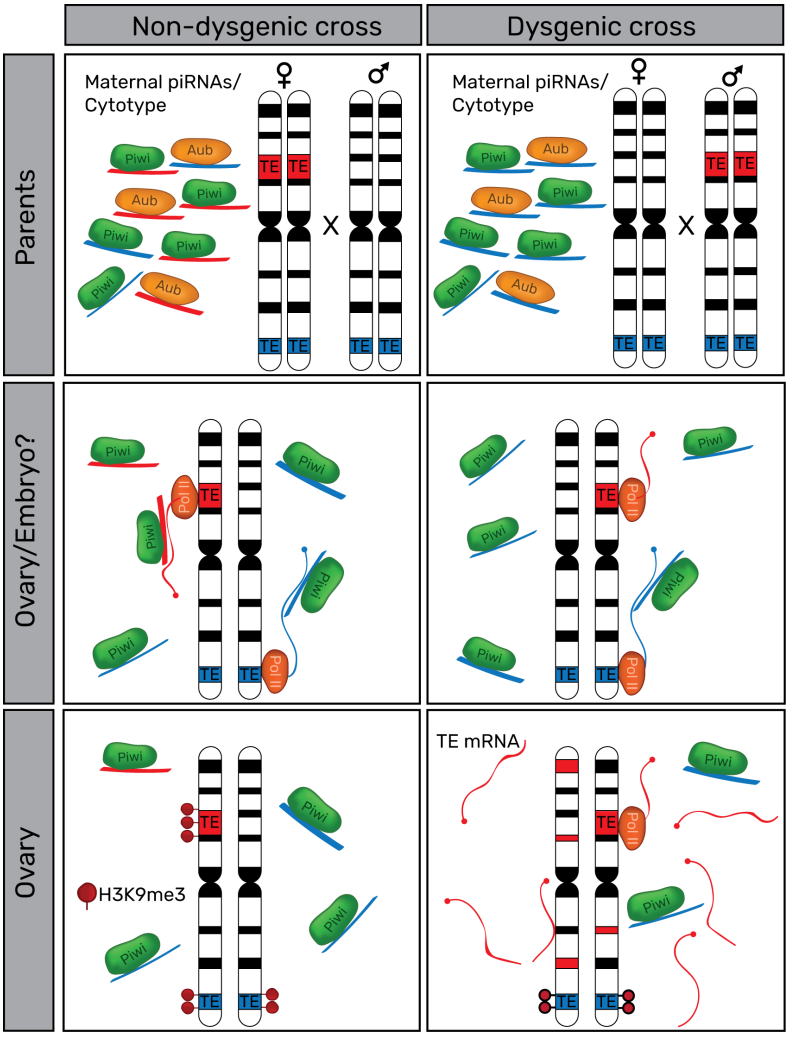

Figure 1.7: Hybrid dysgenesis in *Drosophila*

Source loci giving rise to piRNAs and target transposons in ovaries, are established during embryogenesis. Maternally deposited piRNAs target only transposons present in the female parent during oogenesis. Target engagement through piRNA guides bound by Piwi leads to transcriptional gene silencing of transposons. Transposons not present in the female parent are not silenced due to a lack of complementary piRNAs produced during oogenesis.

Similarly, paramutations in flies can be explained by inherited piRNAs (de Vanssay et al., 2012; Hermant et al., 2015). Paramutations are interactions between two alleles of a specific locus and were first described in maize (Brink, 1956; Brink and Weyers, 1957). One allele induces an epigenetic and inheritable change of the other allele without affecting the DNA sequence. The epigenetic change of the allele is dominant and able to induce changes at another susceptible locus. Alleles that share the same sequence but differ in their epigenetic state are called epialleles. Several epialleles have been discovered in plants. Most notably are epialleles inducing hereditary changes in expression of transcriptional activators of pigment biosynthesis in maize (Dooner et al., 1991; Hollick, 2017).

Maternally deposited piRNAs have been shown to be responsible for certain paramutagenic observations in *Drosophila* (de Vanssay et al., 2012; Hermant et al., 2015). Complementary piRNAs are able to target specific loci and induce transcriptional gene silencing by embedding the genomic locus in heterochromatin, while also amplifying the initial piRNA pool. This process is further dependent on factors involved in piRNA cluster transcription and biogenesis such as Aub, Rhi, Cuff and Zuc (Hermant et al., 2015) (see also chapter 1.5). Amplified piRNA species are then again maternally inherited by the next generation and able to convert a previously unsilenced allele. This process is highly dependent on the piRNA pathway and was reported to penetrate as much as 50 generations independently of the initial triggering allele (de Vanssay et al., 2012; Hermant et al., 2015).

A recent study further reported paramutation of an epiallele mediated by the Polycomb pathway (Ciabrelli et al., 2017). This observation might be explained by the germline-inherited H3K27me3 mark, a histone modification strongly associated with repression and a hallmark of Polycomb-mediated silencing (Zenk et al., 2017) (see also chapter 1.3 and 1.6.3).

1.6.3 Inheritance of histone marks

Chromatin marks encode crucial epigenetic information important for various functions in organisms. Modifications involved in transcriptional control such as H3K4 and H3K9 methylation are lost during *Drosophila* oogenesis and spermatogenesis and start to reappear at the zygotic genome activation (ZGA) (Rudolph et al., 2007; Yuan and O'Farrell, 2016), indicating the requirement for a memory of chromatin organisation leading to re-establishment of euchromatic and heterochromatic domains during early development. The mechanisms and factors in place to facilitate this are currently poorly understood.

Interestingly, not all histone modifications are lost between generations. Evidence in several organisms suggests that certain histone modifications can be inherited by future generations. For instance, in *S. pombe*, histone acetylation-dependent gene activation has been shown to be maintained for over 200 generations (Ekwall et al., 1997).

Recent evidence has emerged that suggests intergenerational inheritance of certain chromatin modifications. H3K27me3 covers certain parts of the genome and is generally associated with transcriptional repression (see chapter 1.3). Oocyte chromatin does not lose H3K27me3 during maturation and is maintained throughout early embryogenesis. In sperm, most nucleosomes are replaced by protamines and therefore cannot easily transfer chromatin modifications from one generation to the next. However, some research suggests that this process is not universal for paternal DNA and that certain regions of the parental genome retain histones specific for centromere identity that seem to be involved in re-establishment of chromatin states in the developing zygote (Raychaudhuri et al., 2012). The transmission of chromatin modifications is essential for development and, in the case of H3K27me3, is believed to prevent premature activation of lineage-specific genes by restricting accumulation of the activating chromatin mark H3K27 acetylation (Zenk et al., 2017). Inheritance of H3K27me3 could also explain the effect of Polycomb-mediated paramutation in flies (Ciabrelli et al., 2017) (see chapter 1.6.2).

1.7 *Drosophila* oogenesis and embryogenesis

Drosophila oogenesis and embryogenesis are convenient models for studying developmental processes *in vivo*. Compared to the size of flies, the female ovary is a large organ that is not vital for survival, therefore allowing tissue-specific genetic manipulation. Additionally, the ovary is comprised of somatic and germline tissue in different developmental stages, thereby providing an easy and convenient tool for exploration of mechanisms required for genome integrity such as the piRNA pathway. Embryogenesis takes ~ 24hrs to complete and the rapid progression allows studying fundamental processes in real-time. Additionally, the small size and transparency of *Drosophila* embryos are ideal for live imaging with only a low risk of disturbing the developing organism (Icha et al., 2017).

1.7.1 *Drosophila* oogenesis

Drosophila oogenesis describes the maturation process of germ line progenitors into mature egg cells primed for fertilisation. This process takes approximately one week to complete. Each ovary pair is comprised of about 16-18 ovarioles (Figure 1.8). The germarium is found at the anterior tip of the ovariole and contains germline and somatic stem cells. The oocyte progenitor leaves the germarium posteriorly and establishes an egg chamber. The egg chamber is comprised of different cell types. Most notably are the oocyte, the nurse cells and the follicle cells surrounding the egg chamber (Bate and Arias, 1993). Both, the oocyte and nurse cells are of germline origin. Nurse cells contain sites of piRNA production at the perinuclear nuage (see chapter 1.5.7). Follicle cells are derived from somatic tissue and piRNA production occurs at Yb bodies (see chapter 1.5.6). Factors required for dual-strand cluster transcription or ping-pong cycling are not expressed in follicle cells. However, follicle cells express a functional piRNA-dependent TGS pathway (Li et al., 2009a; Malone et al., 2009). The maturing oocyte is transcriptionally inactive after leaving the germarium, and chromosomes condense into a karyosome-like state by stage 3 of oogenesis (Bate and Arias, 1993; Von Stetina and Orr-Weaver, 2011).

Therefore, the oocyte is dependent on the support of surrounding cells in order to ensure proper growth. Polyploid nurse cells are highly transcriptionally active and produce many factors necessary for oocyte growth. The maturing oocyte is interconnected with nurse cells through ring canals that allow transport of molecules such as mRNAs and proteins (Robinson et al., 1994). Starting at oogenesis stage 11, the nurse cells transport RNAs, proteins and other growth factors to the oocyte in a process termed oocyte dumping or maternal deposition. This process is completed by stage 12, with most nurse cells disappearing and oogenesis completing at stage 14. (Bastock and St Johnston, 2008; Bate and Arias, 1993). Deposited RNAs and proteins are not randomly distributed within the oocyte or mature egg. Many germline defining components such as nanos (nos) or oskar (osk) mRNA are actively transported by microtubules to the posterior pole, the future site of pole cell formation, and anchored by F-Actin (Lasko, 2012) (Figure. 1.8). Studies suggests, that this mechanism has been hijacked by retrotransposons to ensure transport of virus-like particles into oocytes (Wang et al., 2018).

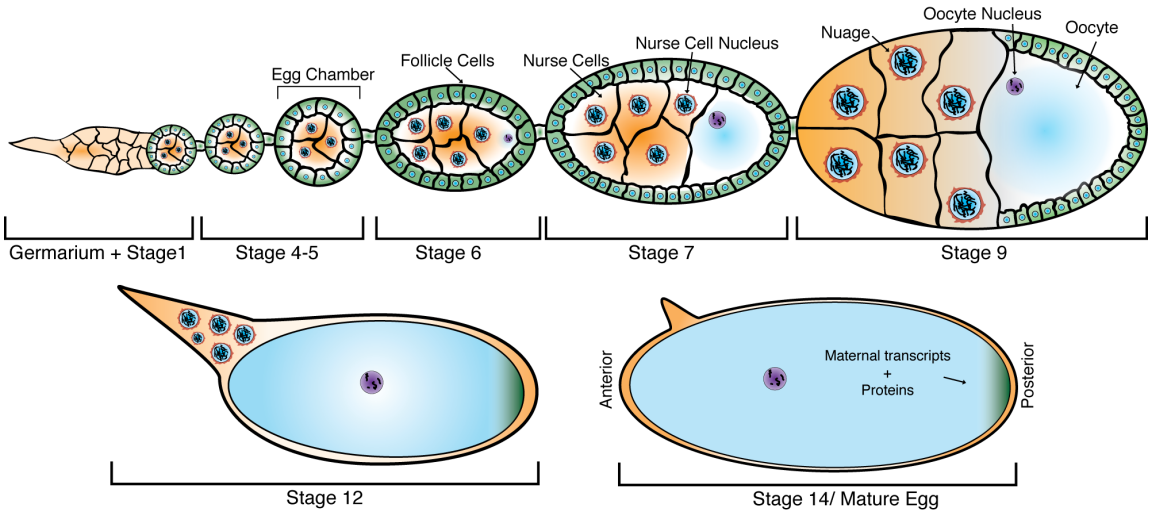

Figure 1.8: Schematic of *Drosophila* oogenesis

Germline progenitor cells reside in the germarium and migrate through the ovariole posteriorly, thus establishing an egg chamber. Egg chambers are comprised of 15 nurse cells and the oocyte, which are engulfed by a monolayer of somatic follicle cells. The perinuclear structure surrounding nurse cell nuclei is called nuage and is the production site of piRNAs. Maternal factors are deposited from stage 11-12. mRNAs and proteins are actively arranged within the developing oocyte. Oogenesis concludes at stage 14 and the mature egg is ready for fertilisation.

1.7.2 *Drosophila* early embryogenesis

Mature oocytes are arrested in meiotic metaphase I (Jang et al., 1995; King, 1970). The fertilised egg is initially transcriptionally quiescent and translation of maternal transcripts is inhibited probably by poly(A) tail-length shortening during oogenesis (Eichhorn et al., 2016). Following ovulation and fertilisation in the uterus, the oocyte completes meiosis and protein synthesis is activated (Horner and Wolfner, 2008). While fertilisation and continuation of meiosis occur at the same time, it is likely not sperm penetration that activates resumption of meiosis but hydration and Ca^{2+} influx into the egg during ovulation (Kaneuchi et al., 2015).

Early development can be subdivided into morphologically distinct timed stages as proposed by Bownes (Bownes, 1975). Pronuclear fusion occurs at the end of the first zygotic nuclear cycle (NC) during stage 1 (0-10min after egg laying; AEL). The embryo then enters the pre-blastoderm stage (mitotic cycle 1-9, stage 2). Nuclei further replicate in stage 2 and form a syncytial blastoderm. Mitotic cycles are on average 8.4-8.8min and occur synchronously. Gap phases (G2) are omitted to further accelerate mitosis (Bate and Arias, 1993). Despite the fast speed of the nuclear divisions, some zygotic transcripts have been reported to appear as early as NC 7 (Kwasnieski et al., 2019). Interphase 9 marks the beginning of stage 3. Nuclei start to migrate from the interior of the embryo to the posterior pole. These nuclei give rise to pole cells, the germ cell progenitors. Only a small number of nuclei form pole cells; on average 24. Pole cells are initially transcriptionally quiescent and polyadenylated transcripts are not detected until 3.5h AEL (stage 8) (Van Doren et al., 1998; Zalokar, 1976). Transcription is actively repressed by the germ plasm component germ cell-less (gcl) (Leatherman et al., 2002). The remaining nuclei reach the periphery during interphase 10 which marks the beginning of stage 4, the syncytial blastoderm. Some nuclei leave the periphery and fall back into the interior of the embryo. Those nuclei stop dividing and develop into polyploid yolk cells. While somatic nuclei continue to divide synchronously at the periphery, the posterior pole nuclei are packaged during telophase of cycle 10 and form pole cells. The newly formed syncytial blastoderm

completes three additional cycles (NC11-13). The cycle length decreases to 10, 13 and 16-17.5min respectively (Bate and Arias, 1993) (Figure 1.9).

Once the embryo reaches cycle 14, mitosis slows down significantly. NC 14 is the first nuclear cycle with a prolonged interphase. The slowdown of the cell cycle further allows the zygotic genome activation (ZGA) and marks the beginning of the maternal to zygotic transition (MZT). During this phase, maternal transcripts and some proteins are degraded, while zygotic transcription strongly increases. This process is dependent on several mechanisms. The RNA-binding protein Smaug (Smg) is involved in clearing hundreds of maternal transcripts. Smg binds specific sequence motifs of mRNAs and recruits the CCR4-NOT complex, leading to deadenylation of targeted mRNAs, thus destabilising them, which results in degradation (Semotok et al., 2005; Zaessinger et al., 2006). Of note, one report suggests the involvement of the piRNA pathway factor Aub and complementary piRNAs in recruitment of Smg to target mRNAs (Rouget et al., 2010). Interestingly, there is a high enrichment for DNA binding sites of early transcripts that are recognised by the protein Zelda (Zld). Zld is a pioneer transcription factor that facilitates transcription from nucleosome rich regions and has a critical role in early genome activation (Hug et al., 2017; Li et al., 2014; McDaniel et al., 2019). Early zygotic transcripts dependent on Zld are short in length and intron depleted (Kwasnieski et al., 2019). Early embryogenesis is also the phase of 3D genome restructuration. The genome is largely unstructured before zygotic transcription starts. However, topologically associated domain (TAD) boundaries are established at early expressed genes in a transcription-independent way (Hug et al., 2017).

Coordination of the first 14 nuclear cycles is the result of precise deposition, localisation, and activation of maternal factors. Small changes in any of these parameters can have severe implications on the development of the embryo. For instance, alteration of the pool size of maternally deposited histone mRNA and protein can lead to cell cycle elongation and additional nuclear cycles before gastrulation (Chari et al., 2019), disturbing the fine balance leading up to the MZT. Additionally, environmental conditions have a major impact on embryogenesis duration. Temperature is the most prominent factor controlling development time in *Drosophila*.

While changes in temperature can decrease or increase the duration of embryogenesis by almost 2-fold, transitions between major events are proportionally decreased or increased depending on temperature (Kuntz and Eisen, 2014).

Nuclear cycle 14 also marks the cellularisation process (stage 5). This process is the first morphological change of the embryo that is under control of zygotic transcripts (Bate and Arias, 1993). Membranes start to grow inwards between nuclei and a cellularised blastoderm forms a single epithelial sheet comprised of 6,000 somatic cells, pole cells at the posterior pole and yolk cells in the interior. This phase is very transient and gastrulation shortly follows cellularisation (stage 6, ~3h AEL).

Gastrulation establishes cell patterns that will differentiate into specific cell populations. The three primary germ layers mesoderm, endoderm and ectoderm are reflected by mitotic domains that formed during the early nuclear divisions. The boundaries of the germ layers are set by the expression of only few transcription factors. For instance, *twist* (*twi*) is expressed throughout the anterior and posterior tip of the developing embryo, while *snail* (*sna*) expression overlaps with *twist* but is not found in the posterior tip. Both set the cell fate for mesodermal differentiation (Bate and Arias, 1993).

Following *Drosophila* gastrulation, most somatic cells will only divide one or two additional times (NC 15 and 16) before the embryo hatches. Divisions conclude 7.5h AEL. However, some cells, such as neuronal cells and macrophages, continue to divide (Bate and Arias, 1993).

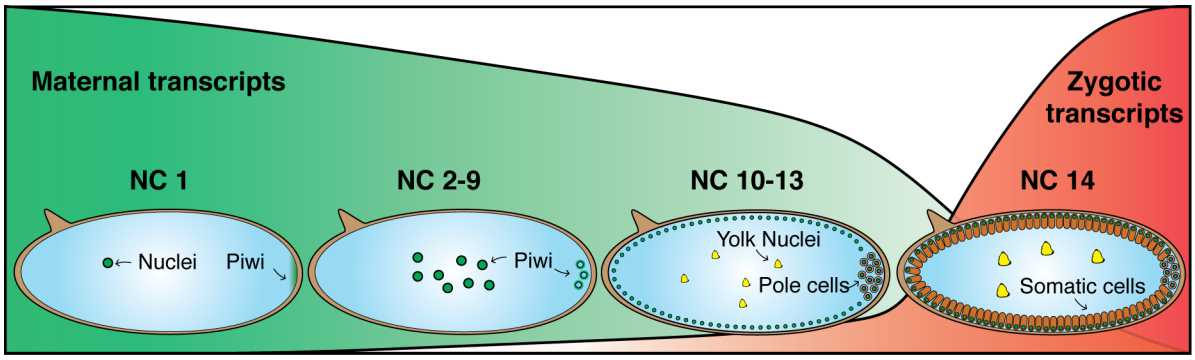

Figure 1.9: *Drosophila* early embryogenesis

Piwi protein is maternally deposited along mRNAs and transported to the posterior pole of the embryo. Following pronuclear fusion during mitotic nuclear cycle (NC) 1, pole cells form at the posterior pole at NC 9 and Piwi localises to the pole plasm but does not initially enter the pole cell nuclei. Somatic nuclei show high abundance of maternal Piwi. Following NC 9, somatic nuclei migrate to the cortex and form the syncytial blastoderm. Early transcripts are detected from NC 8 onwards. NC 14 marks the slowdown of mitotic divisions. The maternal to zygotic transition (MZT) is initiated by degradation of maternal transcripts and the zygotic genome activation (ZGA). Somatic nuclei cellularise and gastrulation follows shortly afterwards.

2 Methods and Materials

2.1 Fly Stocks and handling

Fly experiments were carried out at 25°C. Flies expressing OsTIR1, AtTIR1, JabbaTrap or deGradFP under the *Drosophila melanogaster Ubiquitin-63E* (*Ub*) promoter were generated by phiC31 integrase-mediated transgenesis by injection of plasmids containing expression cassettes for proteins (see also chapter 2.2) in embryos of genotype y w P{y[+t7.7]=nos-phiC31\int.NLS}X #12;; P{y[+t7.7]=CaryP}attP2 resulting in transgene integration on chromosome 3. Nxf2-mCherry-LEXY, GFP-AID-Piwi and CRY2-mCherry-Piwi knock-in flies were generated by CRISPR/Cas9 genome engineering (see also Table 2.2, chapter 2.2). Embryos of genotype y[1] M[vas-Cas9]ZH-2A w[1118] were injected with sgRNA-coding and donor plasmids. Founders were screened by genotyping for successful insertion of the donor sequence and verified by Sanger sequencing. All embryo injections were kindly performed by the Fly Facility at the Department of Genetics, University of Cambridge. A comprehensive list of fly stocks can be found in Table 2.1.

Table 2.1: Fly stocks used for experiments

Fly Name	Genotype	Origin	ID/ Publication
GFP-Piwi (BAC)	w; +; EGFP_Piwi [attP2];	Brennecke Lab	VDRC 313319
GFP-Panx (BAC)	+; P{EGFP-Panx.H}	Brennecke Lab	(Handler et al., 2013)
GFP-Nxf2	*GFP-Nxf2*	Hannon Lab	(Fabry et al., 2019)
panx^{M1}	panx^{M1}	Hannon Lab	(Yu et al., 2015)
panx^{M4}	panx^{M4}	Hannon Lab	(Yu et al., 2015)
nxf2^{F10*}	*Nxf2[F10*]*	Hannon Lab	(Fabry et al., 2019)
nxf2$^{\Delta1*}$	*Nxf2[Δ1*]*	Hannon Lab	(Fabry et al., 2019)
GFP-AID-Piwi	*GFP-AID-Piwi*	Hannon Lab	This study
NXF2-mCherry-LEXY	*NXF2-mCherry-LEXY*	Hannon Lab	This study
CRY2-mCherry-Piwi	*CRY2-mCherry-Piwi*	Hannon Lab	This study
w^{1118}	*w^{1118}*	Dep. of Genetics	Donation
D. simulans	*nos-Cas9*	Dep. of Genetics	Donation
His2AV-RFP	+; P{His2Av-mRFP}	St Johnston Lab	Donation
JabbaTrap	w[*];+;P{w[+]=ubi::JabbaTrap9}attp2	Hannon Lab	This study

OsTIR1	w[*]; +; P{w[+]=ubi::OsTIR19}attp2	Hannon Lab	This study
AtTIR1	w[*]; +; P{w[+]=ubi::AtTir19}attp2	Hannon Lab	This study
deGradFP	w[*]; +; P{w[+]=ubi:: deGradFP 9}attp2	Hannon Lab	This study

2.2 Generation of plasmids for microinjection and transfection

All plasmids were generated using standard molecular cloning techniques such as PCR and Gibson assembly (Gibson et al., 2009). Constructs were electroporated into Endura ElectroCompetent Cells (Lucigen) according to the manufacturer's recommendation and bacteria plated on ampicillin-containing LB agar plates (100 *µ*g/ml). Single colonies were picked and used for inoculation of ampicillin-containing LB liquid medium. Bacteria were grown for no longer than 16h at 37°C while shaking. Plasmids were isolated using QIAprep Miniprep Kit (Qiagen) or Plasmid Plus Maxi Kit (Qiagen).

sgRNA-coding plasmids for CRISPR/Cas9 injections were generated by inserting guide sequences in pCFD3 (Addgene plasmid #49410 (Port et al., 2014)) using Gibson assembly (see Table 2.2 for guide sequences). Donor plasmids were constructed by inserting a cassette coding for the desired tag flanked by 1kb up- and downstream of the targeted genomic region into pUC19. Constructs used for phiC31-mediated transgenesis and cell culture transfections were generated using pUBI_attB (Hannon lab) for *ubiquitin* promoter-driven constructs or Dsim_pUbi_attB (Hannon lab) for *D. simulans ubiquitin* promoter-driven constructs. Insert cassettes were generated by amplifying protein coding sequences from cDNA generated using w^{1118} ovaries (as described in chapter 2.17). AID, deGradFP, OsTIR1 and AtTIR1 sequences were ordered as gBlocks (IDT). CRY2 and mCherry were amplified from pCRY2PHP-mCherryN1 (Addgene #26866). LEXY was amplified from pDN122 NLS-mCherry-LEXY (Addgene #72655). JabbaTrap was constructed by fusing vhh4GFP nanobody sequences amplified from deGradFP gBlock to the 5' and 3' end of *jabba* isoform B amplified from cDNA as previously described (Seller et al., 2019).

Table 2.2: List of DNA oligos used for FISH, qRT-PCR and cloning

Oligo Name	DNA Sequence 5' -> 3'
RF_EK23_Gypsy_probe 2	GCCCTTACTCCCAATTCCaaaaa AAGCCGACTCAGCATTTCTTGCAGCGTGAAGCAACACTCCCGGTAGGAAGTG
RF_EK24_Gypsy_probe 3	GCCCTTACTCCCAATTCCaaaaa ctgaggttcgtcttagacactgtttatggagattaggttggagggcttgact
RF_EK25_Gypsy_probe 4	GCCCTTACTCCCAATTCCaaaaa TGACAAAGTGTGTTAAATTAGATTGGTGGGTTCAGATTGTTGGTTGGGCGCC
RF_EK26_Gypsy_probe 5	GCCCTTACTCCCAATTCCaaaaa CCAACTCATTGGTTGTTGGTTGGCACACCACAAATATACTGTTGCCGAGCAC
RF_EK27_Gypsy_probe 6	GCCCTTACTCCCAATTCCaaaaa tgctcctcctcccagctatcctcgctttcgtattcgaccttaacctttctgt
RF_EK28_Gypsy_probe 7	GCCCTTACTCCCAATTCCaaaaa ATCTGATTGGGGTCCATGGTAATATCTACCGTGGCACTATCTAACGGCCGAC
RF_EK29_Gyspsy_probe8	GCCCTTACTCCCAATTCCaaaaa ttccaatccttgcctcaacaggcgtaaggatgttttgtccgagtacgtgcag
RF_EK30_Gypsy_probe 9	GCCCTTACTCCCAATTCCaaaaa ctctgacctcagcgttaagcaggtcagcaccctcttgttcatgcgtcattac
RF_EK31_Gypsy_probe 10	GCCCTTACTCCCAATTCCaaaaa cctctgcttctctagctaaagccagtgcagatggcaggtcttttggttggg
RF_EK42_mdg1_probe 1	GCCCTTACTCCCAATTCCaaaa aaaagatctactagggtgaccctaaggaattagggtggtcctaagtttactta
RF_EK43_mdg1_probe 2	GCCCTTACTCCCAATTCCaaaaa cgagtccgacccctaaaggcgtacatcctgaattcgcatatttagtattagg
RF_EK44_mdg1_probe 3	GCCCTTACTCCCAATTCCaaaaa ctacagggagatctcttgtgacagccaagtgctgacactagcaaattctgca
RF_EK45_mdg1_probe 4	GCCCTTACTCCCAATTCCaaaaa gcgagtgtacattccagatctattcctgatttataactgatcttagtgtgg
RF_EK46_mdg1_probe 5	GCCCTTACTCCCAATTCCaaaaa gcggtttcctttagcttctgctacaacttataacgatcctttgtcacggtcg
RF_EK47_mdg1_probe 6	GCCCTTACTCCCAATTCCaaaaa tttcttcataattgccgccgtcttggtggatgagcatcgaagttgaaaatta
RF_EK48_mdg1_probe 7	GCCCTTACTCCCAATTCCaaaaa gtcgaactccgcccctgtttttttcccatcacactgacactctactcactcaga
RF_EK49_mdg1_probe 8	GCCCTTACTCCCAATTCCaaaaa ttggggtggtggagtgttcgtatataatagccactttatgtgcgctattctc
RF_EK50_mdg1_probe 9	GCCCTTACTCCCAATTCCaaaaa cgtatgttcacaatgcggtgtgaacagtggtccctcgcagtcgttcgggcat
RF_EK51_mdg1_probe 10	GCCCTTACTCCCAATTCCaaaaa cgggctgctcaccacgttgatgatgatttcctcatttaggggttatgtggtg

oMF0413_roo_hcr_probe1	GCCCTTACTCCCAATTCCaaaaa GGGAACGATCTCAAGTGACTGACTCATGTAGTGTGCACTTAAATTACATGTT
oMF0414_roo_hcr_probe2	GCCCTTACTCCCAATTCCaaaaa CGGGCACATCTGCCTATCTTGAGCGGCGAGGACCTTATCTGTGGTCTCCCAC
oMF0415_roo_hcr_probe3	GCCCTTACTCCCAATTCCaaaaa TTAAAGTAAATGGCCTACGCAGAGGCCTACGTAAATAGTCCCCGCCTTATCG
oMF0416_roo_hcr_probe4	GCCCTTACTCCCAATTCCaaaaa GGGAAACTGCAGAGTCGATTAAAGGCTCGATTGACCAAATGTAAAATCCCAA
oMF0417_roo_hcr_probe5	GCCCTTACTCCCAATTCCaaaaa TTTTTGCTACCTTTAGCTGTAAGATGCTTAAAGGAGCTGGCCTTTCTCTGAG
oMF0418_roo_hcr_probe6	GCCCTTACTCCCAATTCCaaaaa TATGGCCTCAAGCACGCCTTACCACAATTTATAATGGTACACAAAGCAACCT
oMF0419_roo_hcr_probe7	GCCCTTACTCCCAATTCCaaaaa TGCTTCTGCTGCTGGTAGAGGCTCCTTTGAATTTGACTTCCTTCTCTTCTTT
oMF0420_roo_hcr_probe8	GCCCTTACTCCCAATTCCaaaaa ATTCTTGTTTTGACTTAGCTGATGTCGTTGTTGTTGCTGCTGCTGTTGCTGC
oMF0421_roo_hcr_probe9	GCCCTTACTCCCAATTCCaaaaa GGAGGGTTTGATTTAGGGACAGTGTTTGATTTAGGGAAAGTGTTTTCTACCG
oMF0422_roo_hcr_probe10	GCCCTTACTCCCAATTCCaaaaa TCACCAAAGAAGGTGGGAATCTGTATTTTAGGCAGGGTTGGTAACTCCTCCG
act5c_qPCR_fwd	GCATCCACGAGACCACCTACAAC
act5c_qPCR_rev	CGGTGATCTCCTTCTGCATACGG
rp49_qPCR_fwd	GCATACAGGCCCAAGATCGTGAA
rp49_qPCR_rev	CGCACTCTGTTGTCGATACCCTT
gypsy_qPCR_fwd	AGAAAGTCGCCGTCTACCCTGTA
gypsy_qPCR_rev	GTGTGACATTGAGCAGCGTTTCC
mdg1_qPCR_fwd	TATACGAACACTCCACCACCCCA
mdg1_qPCR_rev	GGCTTTTCGGATTGGGAGTTGGA
HetA_qPCR_fwd	CGCGCGGAACCCATCTTCAGA
HetA_qPCR_rev	CGCCGCAGTCGTTTGGTGAGT
sensor_ZsGreen1_qPCR_fwd	CTACTTCAAGAACTCCTGCCCCG
sensor_ZsGreen1_qPCR_rev	GGTACATGCAGTTCTCCTCCACG
sensor_ZsGreen2_qPCR_fwd	CCCCGTGATGAAGAAGATGACCG
sensor_ZsGreen2_qPCR_rev	CGTCCTTCAGCAGCAGGTACATG
sensor_nascent_qPCR_fwd	GCAGCAGCAAGTACAAGCAAAAAG

sensor_nascent_qPCR_rev	TGGCCGAACAAAGACCTTGAAATG
oLig3new	/5rApp/NNNNAGATCGGAAGAGCACACGTCTGAACTCCAGTCA/3ddC/
oRT	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
oPCR5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
oPCR3_11	CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
oPCR3_12	CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
oPCR3_13	CAAGCAGAAGACGGCATACGAGATTTGACTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
oPCR3_14	CAAGCAGAAGACGGCATACGAGATGGAACTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
oPCR3_21	CAAGCAGAAGACGGCATACGAGATCGAAACGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
oPCR3_22	CAAGCAGAAGACGGCATACGAGATCGTACGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
oPCR3_23	CAAGCAGAAGACGGCATACGAGATCCACTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
oPCR3_24	CAAGCAGAAGACGGCATACGAGATGCTACCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
sg_Nxf2	GATATAGAGATCACATATTCT
sg_Piwi	GTAACAATGGCTGATGATCA

2.3 Cell culture

Drosophila Schneider 2 (S2) cells (Thermo Fisher, R69007) and Ovarian Somatic Cells (OSCs, gift from Mikkiko Siomi) were cultured at 26°C. S2 cells were grown in Schneider's *Drosophila* Media (Gibco) supplemented with 10% heat inactivated FBS (Gibco). OSCs were grown as described before (Niki et al., 2006; Saito, 2014; Saito et al., 2009) in 10% heat inactivated FBS, 60 mg/ml glutathione and 10 mU/ml insulin (Sigma Aldrich) in M3 Insect Medium (Sigma Aldrich). Both, OSCs and S2 cells, tested negative for mycoplasma contamination (in-house testing).

siRNA-mediated knockdowns and co-transfections with rescue plasmids in OSCs were performed according to published protocols (Saito, 2014). 10 x 10^6 OSCs were nucleofected with 200pmol annealed siRNAs (see Table 2.3) using the Amaxa

Cell Line Nucleofector Kit V (Lonza) with program setting T-29. Following 48h incubation, cells were nucleofected as described above again and, for rescue experiments, 5*µ*g plasmid DNA was co-transfected. Cells were allowed to grow for an additional 48h for knockdowns, rescues and RNA-tethering experiments. Cells used for DNA tethering were incubated for 72h following the second transfection pulse. OSCs used for *in vitro* degron experiments were transiently transfected with 10*µ*g plasmid DNA using Xfect (Clontech) according to manufacturer's instructions. S2 cells used for IF and co-IP experiments for domain mutants were transfected with 2*µ*g of plasmid DNA using Effectene (Qiagen) according to manufacturer's instructions. Transfected cells were allowed to express constructs for 48h prior to downstream experiments.

Table 2.3: List of RNA oligos used for siRNA-mediated knockdowns and small RNA-Seq Library preparation

Oligo Name	RNA Sequence (5'-3')
siGFP_passenger	rArGrCrUrGrGrArGrUrArCrArArCrUrArCrArArCrA
siGFP_guide	rUrUrGrUrArGrUrUrGrUrArCrUrCrCrArGrCrUrUrG
siPiwi_passenger	rCrGrGrUrCrArUrGrCrUrGrCrArGrArCrGrArArCrU
siPiwi_guide	rUrUrCrGrUrCrUrGrCrArGrCrArUrGrArCrCrGrGrG
siPanx_passenger	rCrGrArUrGrArArGrCrUrArGrArGrGrUrCrArArArG
siPanx_guide	rUrUrGrArCrCrUrCrUrArGrCrUrUrCrArUrCrGrGrA
siNxf2_passenger	rGrGrUrArCrUrUrCrArCrGrGrArArArUrArArArCrU
siNxf2_guide	rUrUrUrArUrUrUrCrCrGrUrGrArArGrUrArCrCrArG
oLig5new	rArCrArCrUrCrUrUrUrCrCrCrUrArCrArCrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrN

2.4 Co-immunoprecipitation from S2 cell lysates

Transfected S2 cells were allowed to express proteins for 48h before lysing. Cells were washed twice in 1ml ice-cold PBS and pelleted at 800g for 3min at 4°C. Lysis was carried out in 250*µ*l Co-IP Lysis Buffer (Pierce) supplemented with 1x Protease inhibitor cocktail (PI) (Roche) at 4°C for 30min while rotating. Lysate was spun down at full speed for 15min at 4°C and supernatant transferred to new tube. Input sample was saved for western blot analysis. Protein concentration was

measured using the Direct Detect Infrared Spectrometer (Millipore). 200µg protein was diluted in 1ml total volume in Co-IP Lysis Buffer + PI. 20µl anti-Flag M2 Magnetic Beads (Sigma) were washed twice in Co-IP Lysis Buffer + PI and added to diluted lysate. Following incubation for 2h at 4°C, supernatant was removed and saved for western blot analysis. Beads were washed three times with 500µl ice-cold TBS. Elution from beads was performed using 20µl Elution Buffer (10µl 4x NuPage LDS Sample Buffer (Thermo Fisher Scientific), 8µl Co-IP Lysis Buffer, 2µl NuPage 10x Sample Reducing Agent (Thermo Fisher Scientific)) and boiled for 10min at 95°C. Supernatant was recovered and used for western blot analysis.

2.5 Piwi-IP from *Drosophila* ovaries and embryos for smallRNA-Seq

This protocol was adapted from a previously published method (Hayashi et al., 2016; Mohn et al., 2015). 100µl of w^{1118} ovaries were dissected in PBS on ice. 100µl of 0-8h AEL w^{1118} embryos were collected on grape juice agar plates and transferred to a mesh strainer. Following dechorionation in 50% bleach, embryos were washed under running tap water for at least 1min or until bleach smell disappeared. Ovary and embryo samples were washed twice with ice-cold PBS and homogenised in 1ml lysis buffer (10mM HEPES ph 7.3, 150mM Nacl, 5mM $MgCl_2$, 10% glycerol, 1% Triton x-100, 1mM DTT, 1mM EDTA, 0.1mM PMSF, 1x PI, 1:1000 RNasin (Promega)) using a 2ml Dounce homogeniser. Material was lysed with 5 strokes with a loose pestle and 5 strokes with a tight pestle on ice. Lysate was incubated for 1h at 4°C while rotating and centrifuged at full speed for 10min to pellet debris. Supernatant was transferred to a new tube and protein concentration determined by Direct Detect (Millipore). 1mg of lysate per IP was used for the following steps. 50µl Protein A Dynabeads (Thermo Fisher Scientific) were washed with lysis buffer 3 times for 3min each. Washed beads were resuspended in 400µl lysis buffer and 5µl anti-Piwi (Hannon Lab) or rabbit IgG antibodies (Abcam, ab37415) added. Following overnight incubation at 4°C while rotating, beads were washed 3 times for 5min in 500µl lysis buffer. Antibody-coupled beads were added to lysates and volume brought up to 1ml with lysis buffer. The

solution was incubated at 4°C overnight while rotating. Supernatant was removed and saved for quality control western blotting analysis. Beads were washed 6 times for 10min with 1ml wash buffer (10mM HEPES ph 7.3, 150mM NaCl, $MgCl_2$, 10% glycerol, 1% Empigen BB Detergent (Merck), 1x PI). For the first wash, 1μl RNasin was added to the wash buffer and tubes were changed between each wash. 10% of beads were set aside for quality control and 90% resuspended in 1ml Trizol (Thermo Fisher Scientific) and stored at -80°C until further processing (see chapter 2.13).

In order to evaluate the purity of isolated Piwi proteins, I performed western blot analysis (see chapter 2.6) on 15μl supernatant following the Piwi-IP (unbound fraction) and on 10% beads. Western blot was probed with anti-αTubulin to indicate cellular protein contaminants and anti-Piwi to indicate purified protein. A band corresponding to Piwi's expected protein weight was detected in the IP while tubulin was absent. Therefore, samples were considered pure and used for small RNA-Seq library preparation (see chapter 2.14.3)

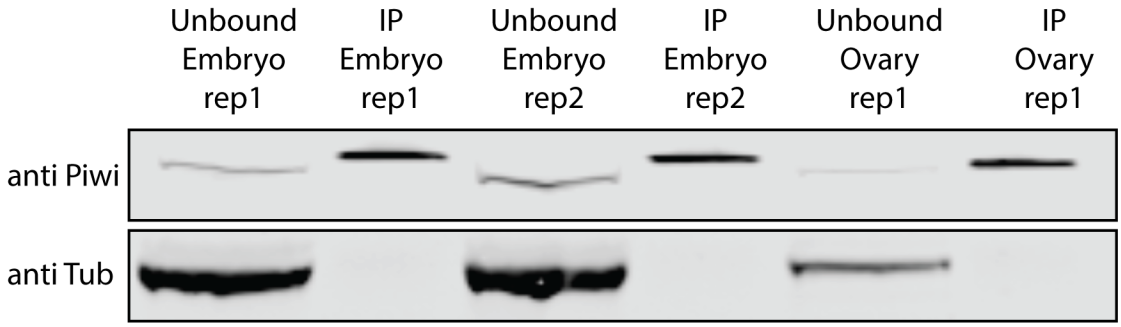

Figure 2.1: Quality control of purified piRNA-Piwi complexes by western blot

Western blot analysis of purified piRNA-Piwi complexes from *w*^{1118} embryos and ovaries. Unbound fraction represents proteins in 15μl supernatant following Piwi IP. IP indicates 10% of total isolated bead-bound proteins. Western blot was probed with anti-Piwi and anti-αTubulin antibodies (see Table 2.4)

2.6 Western Blot

Cells were lysed in 50µl Pierce RIPA buffer (Thermo Fisher Scientific) by pipetting. Embryos and ovaries were lysed in 100µl RIPA buffer using a rotating pestle. Lysates were incubated on ice for 10min and centrifuged at full speed at 4°C. Supernatants were transferred to new tubes. Protein concentrations were determined by Direct Detect. 20µg of protein was mixed with NuPage LDS Sample Buffer and NuPage Sample Reducing Agent (Thermo Fisher Scientific) and boiled at 95°C for 10min. Proteins were separated using a NuPage 4-12% Bis-Tris gel (Thermo Fisher Scientific) and transferred on iBlot 2 Transfer Stack nitrocellulose membranes (Invitrogen) using the iBlot 2 device (Invitrogen). Proteins for GFP-AID-Piwi ovary or embryo lysates were transferred using wet transfer in order to increase transfer efficiency of high molecular weight proteins. An Immun-Blot Low Fluorescent PVDF Membrane (BioRad) was soaked in 100% methanol for 5min. Transfer was carried out in NuPage Transfer Buffer (Thermo Fisher Scientific) according to manufacturer's recommendation. Transfer occurred at 100V, 400mA, 100W for 2h at RT with ice packs placed in the transfer chamber. Nitrocellulose and PVDF membranes were blocked in 1x Intercept (TBS) Blocking Buffer (Licor) for 1h while shaking at room temperature. Primary antibodies (see Table 2.4) were diluted in 1x Blocking buffer and incubated at 4°C while shaking overnight. Western blots were washed 3 times for 5min with TBST (0.1% Tween20 in TBS). Incubation of secondary antibodies (see Table 2.4) was performed in 1x Blocking buffer supplemented with 0.1% Tween20. Additionally, 0.01% SDS was added to the secondary antibody solution for PVDF membranes only. Incubation was carried out for 1h while shaking and followed by washing as described above. Western blots were stored at 4°C in TBS until imaging using an Odyssey Clx (Licor).

Table 2.4: Antibodies used for western blot experiments

Antibody	Dilution	Origin	Provided by	Publication
anti-Piwi	1:1,000	Rabbit	Hannon Lab	(Brennecke et al., 2007)
anti-HA	1:1,000	Rabbit	Abcam ab9111	-
anti-3xFLAG	1:2,000	Mouse	Cell Signalling 14793S	-
anti-GFP	1:5,000	Chicken	Abcam ab13970	-
anti-Nxt1	1:500	Rabbit	Gift from E. Izaurralde	(Herold et al., 2001)
anti-His3	1:500	Rabbit	Santa Cruz sc-10809	-
anti-αTubulin	1:1,000	Mouse	Abcam 18251	-
iRDye 800CW Goat anti-Mouse	1:20,000	Goat	Licor	-
iRDye 680RD Goat anti-Mouse	1:20,000	Goat	Licor	-
iRDye 680RD Goat anti-Chicken	1:20,000	Goat	Licor	-

2.7 Immunofluorescence (IF) and RNA fluorescence *in situ* hybridisation (RNA-FISH)

2.7.1 OSC/ S2 cell IF

Glass cover slips were sterilised with 70% ethanol and washed with PBS. For OSC IFs, 0.1% fibronectin solution (Sigma Aldrich) was diluted 1:50 in PBS and added to sterilised cover slips for 2h at 37°C. OSCs were grown on glass coverslips coated with fibronectin (Sigma Aldrich) overnight. For S2 cell IFs, sterilised cover slips were coated with 0.5 mg/ml concanavalin A solution (Sigma Aldrich) overnight and cells grown for 1h until firmly attached. Growth medium was removed, and cells washed once with PBS. Cells were fixed in 4% paraformaldehyde (PFA) diluted in PBS for 10min followed by three washing steps with PBS for 3min each. Following permeabilisation with 0.1% Triton X-100 in PBS for 10min, cells were washed three times with PBS for 3min. Cells were blocked in PBST (0.1% Tween20 in PBS) with 1% bovine serum albumin (BSA) added for 30min at RT. Primary antibodies (see Table 2.5) were diluted in blocking buffer and incubated at 4°C overnight followed by three washes in PBST for 5min. All following steps were performed in the dark.

Secondary antibodies (see Table 2.5) were diluted in blocking buffer and incubated 1.5h at RT. Cells were washed twice with PBST for 5min and stained with 0.5 µg/ml 4′,6-diamidino-2-phenylindole (DAPI) (Thermo Fisher Scientific) for 10min followed by two washes in PBS for 5min. Cells were mounted in ProLong Diamond Antifade Mountant (Thermo Fisher Scientific) and imaged on a Leica SP8 confocal microscope using a 100x Oil objective.

2.7.2 *Drosophila* ovary IF

Fly ovaries were dissected in ice-cold PBS and fixed in 4% PFA diluted in PBS for 15min at room temperature while rotating. Following 3 rinses and three 10min washing steps in PBS-Tr (0.3% Triton X-100 in PBS), ovaries were blocked for 2h at RT while rotating in PBS-Tr +1% BSA. Primary antibody incubation (see Table 2.5) was carried out in blocking buffer overnight at 4°C while rotating, followed by three washing steps for 10 min each in PBS-Tr. All following steps were performed in the dark. Secondary antibodies (see Table 2.5) were diluted in blocking buffer and incubated overnight at 4°C while rotating. Ovaries were washed 4 times for 10min in PBS-Tr and stained with 0.5µg/ml DAPI (Thermo Fisher Scientific) for 10min. Following two additional washing steps for 5min in PBS, ovaries were mounted in ProLong Diamond Antifade Mountant (Thermo Fisher Scientific) and imaged on a Leica SP8 confocal microscope using a 40x Oil objective.

2.7.3 *Drosophila* embryo IF

Embryos were collected and dechorionated as described above (chapter 2.5). Embryos were transferred into 1ml fixing solution (600µl 4% PFA in PBS, 400µl n-heptane) and fixed for 20min at RT while rotating. The lower aqueous phase was removed and 600µl methanol added. The tube was vortexed vigorously for 1min to remove vitelline membranes. Embryos were allowed to sink to the bottom of the tube and all liquid was removed including embryos floating in interphase, followed by two washes with methanol for 1min each. Embryos were stored at -20°C at least overnight

or until further processing. In order to rehydrate embryos, three washes each 5min with PBST (0.1% Tween20 in PBS) were performed and embryos blocked for 1h at RT in PBST + 5% BSA. Primary antibodies (see Table 2.5) were incubated overnight at 4°C while rotating in blocking buffer followed by 3 washes for 15min each with PBST. All following steps were performed in the dark. Secondary antibodies (see Table 2.5) were diluted in blocking buffer and incubated at RT for 2h. Embryos were rinsed 3 times and washed 2 times for 15min. Nuclei were stained with 0.5 μg/ml DAPI (Thermo Fisher Scientific) for 10min. Following two additional washing steps for 5min in PBS, embryos were mounted in ProLong Diamond Antifade Mountant (Thermo Fisher Scientific) and imaged on a Leica SP8 confocal microscope using a 40x Oil objective.

Table 2.5: List of antibodies used for immunofluorescence experiments

Antibody	Dilution	Origin	Provided by	Publication
antiPiwi	1:500	Rabbit	Hannon Lab	(Brennecke et al., 2007)
anti-HA	1:500	Rabbit	Abcam ab9111	-
anti-3xFLAG	1:1,000	Mouse	Cell Signalling 14793S	-
anti-GFP	1:1,000	Chicken	Abcam ab13970	-
anti-Lamin	1:200	Mouse	DSHB ADL67.10	-
anti-mCherry	1:500	Rabbit	Abcam ab167453	-
anti-Chicken-488	1:1,000	Goat	Abcam 150169	-
anti-Rabbit-555	1:1,000	Goat	Thermo Fisher	-
anti-Mouse-647	1:1,000	Goat	Thermo Fisher	-

2.7.4 OSC Hybridisation Chain Reaction (HCR) RNA-FISH

HCR RNA-FISH was performed using an adapted protocol reported previously (Ang and Yung, 2016; Choi et al., 2014). OSCs were grown and fixed as described in chapter 2.7.1 followed by additional fixation in 70% ethanol overnight at -20°C. Cells were rinsed twice with 2x Ambion saline-sodium citrate (SSC) buffer (Invitrogen) followed by two washing steps for 5min each. Cells were incubated for 15min at RT with 15% formamide (Invitrogen) in 2x SSC and probes hybridised overnight in hybridisation buffer (15% formamide, 10% dextran sulphate in 2x SSC) with 1nM of each probe added (see Table 2.2) at 37°C in a humidified chamber. Following two

rinses with 2x SSC, cells were washed with 30% formamide in 2x SSC for 10min at 37°C. All following steps were performed in the dark. HCR hairpins conjugated to Alexa Fluor 647 (IDT) were heat-denatured and diluted to 2μM in 5x SSC + 0.1% Tween20. Oligo-dT probes conjugated to ATTO-488 (IDT) were heat-denatured and, together with prepared hairpins, incubated in 5x SSC and 10% dextran sulphate at a final concentration of 120nM each. HCR amplification was performed for 2h at RT. Cells were rinsed twice with 5x SSC + 0.1% Tween20 and washed 3 times for 10min. Nuclei were stained with 0.5 μg/ml DAPI in 2x SSC for 10min and washed twice with 2x SSC for 10min. Following mounting in ProLong Diamond Antifade Mountant (Thermo Fisher Scientific), cells were imaged on a Leica SP8 confocal microscope using a 100x Oil objective.

2.7.5 Embryo RNA-FISH-Co-IF

Embryos were collected and dechorionated as described above (chapter 2.5) and processed as described in chapter 2.7.3 until secondary antibody incubation. For all steps containing BSA addition, RNAsin Plus RNase inhibitors were added (1:1,000, Promega). Following secondary antibody incubation, cells were washed 3 times for 15min in PBST at RT while rotating. Embryos were fixed in 4% PFA in PBST solution for 25min and rinsed 3 times with PBST for 5min each. Embryos were pre-hybridised in 100μl hybridisation buffer (50% formamide, 5x SSC, 9mM citric acid pH 6.0, 0.1% Tween20, 50μg/ml heparin, 1x Denhardt's solution (Sigma Aldrich), 10% dextran sulphate) for 2h at 65°C. Probes were hybridised in hybridisation buffer supplemented with 2nM of each FISH probe (see Table 2.2) at 45°C overnight. Following washing twice with probe wash buffer (50% formamide, 5x SSC, 9mM citric acid pH 6.0, 0.1% Tween20, 50 μg/ml heparin) for 5min and twice for 30min at 45°C, embryos were incubated in amplification buffer (5x SSC, 0.1% Tween20, 10% dextran sulphate) for 10min at RT. Hairpins were prepared as described above (chapter 2.7.4) and embryos incubated in fresh amplification buffer with 120nM of each probe at RT overnight in the dark. Embryos were washed twice with 5x SSC for 5min. Nuclei were stained with 0.5 μg/ml DAPI diluted in 2x SSC for 15min. Following washing twice with 2x SSC for

10min, embryos were mounted in ProLong Diamond Antifade Mountant (Thermo Fisher Scientific) and imaged on a Leica SP8 confocal microscope using a 100x Oil objective.

2.8 Spinning-Disk Microscopy of *Drosophila* embryos

Embryos were collected on agar plates for 30min, dechorionated as described above (chapter 2.5), and mounted on a μ-Dish 35mm, high Glass Bottom (Ibidi) using UltraPure Low Melting Point agarose (LMA) (Thermo Fisher Scientific). 1% LMA was prepared with deionised water and centrifuged to remove undissolved particles. 1ml of 1% LMA was warmed up to 37°C and transferred to the dish. Embryos were carefully placed in liquid agarose and gently moved to the bottom of the dish. Agarose was allowed to solidify for 3min and 2ml RT PBS was added to the dish. Mounted embryos were imaged immediately on a Dragonfly High Speed Confocal Platform using a 100x Oil objective (Andor). Generated data was further processed using the Imaris Cell Imaging Software (Bitplane).

2.9 Light Sheet Fluorescent Microscopy (LSFM) of *Drosophila* embryos

Embryos were collected and dechorionated as described above (chapter 2.5). 1ml 1% LMA (chapter 2.8) was prepared and embryos transferred into capillaries (catalogue number 100003476381, Brand) using a fitting plunger. Embryos were attempted to be positioned vertically in the capillary by twisting until agarose solidified. Capillaries were stored in PBS at RT until imaging. LSFM was performed on a Zeiss Lightsheet Z.1 (Zeiss) at 25°C with a 20x objective. Embryos were lowered carefully out of the capillary into the imaging chamber filled with PBS and positioned directly between the light sheet objectives. Z-stack images for GFP and RFP were acquired every 2min for >10h with the lowest possible laser intensity (2.5% for GFP and 10% for RFP) for GFP-Piwi; His2AV-RFP, GFP-Panx; His2AV-RFP, GFP-Nxf2; His2AV-

RFP embryos. Z-stack images for GFP-AID-Piwi; JabbaTrap and GFP-AID-Piwi; AtTIR1 embryos were acquired as fast as possible for 10h and 3h respectively. Only GFP signal was imaged with laser power settings as described above. For imaging of Nxf2-mCherry-LEXY embryos see chapter 2.10. Generated data was analysed in ZEISS ZEN Imaging Software (Zeiss) and Fiji (ImageJ).

2.10 Treatment of OSCs and *Drosophila* embryos expressing LEXY constructs

OSCs were grown on a *μ*-Dish 35mm, high Glass Bottom (Ibidi) coated with fibronectin (see chapter 2.7.1). Light-induced delocalisation was performed on a SP8 Confocal Microscope (Leica) with 20% 488nm laser power for 30sec. Recovery was imaged with a 555nm laser and images acquired every 30sec. Light induced delocalisation of Nxf2-mCherry-LEXY in living embryos was achieved by LSFM imaging RFP with 10% laser power initially followed by 100% GFP laser power and 10% RFP laser power for one Z-stack acquisition (30sec exposure in total). Recovery was measured with 10% RFP laser power every minute. Generated data was analysed in ZEISS ZEN Imaging Software (Zeiss) for LSFM and Fiji for confocal microscopy (ImageJ).

2.11 Chromatin immunoprecipitation (ChIP)

2.11.1 ChIP for OSCs

The OSC ChIP protocol was established with the help of Aled Perry (Narita group, CRUK CI)

ChIP for OSCs was performed by using a modified protocol as previously described (Schmidt et al., 2009). Cells were trypsinised and washed once with 10ml PBS and centrifuged at 500g for 5min. The pellet was resuspended in 10ml PBS and cells counted. 10x10^6 cells were aliquoted into 15ml tubes and pelleted as described

above. Pellets were resuspended in 2.5ml room temperature PBS and mixed with 2.5ml 2x crosslinking solution (2% formaldehyde (Pierce) in PBS). The tubes were rotated slowly (10rpm) for exactly 10min to prevent cell clumping. 500μl of 2.5M glycine solution was added and tubes inverted twice to quench remaining formaldehyde and stop the crosslinking process. Cells were immediately pelleted at 1,000g for 5min at 4°C and the supernatant discarded. Cells were washed twice with 5ml ice-cold PBS and incubated on ice for 5min before centrifugation as described above. Following the second wash, the cell pellet was resuspended in 1ml ice-cold PBS and transferred into 1.5ml DNA lo-bind tubes (Eppendorf). Following centrifugation at 6,200g for 5min at 4°C, the supernatant was entirely removed and cell pellets either flash-frozen in liquid nitrogen or used directly for chromatin extraction.

Frozen or freshly cross-linked cell pellets were resuspended in 1ml ice-cold LB1 (50mM Hepes-KOH, pH 7.5, 140mM NaCl, 1mM EDTA, 10% glycerol, 0.5% Igepal CA-630, 0.25% Triton-X 100), 1x Protease Inhibitor cocktail (PI, Roche)) by vigorous pipetting and incubated on ice for 10min with occasional inversion of the tube. Cells were spun down at 2,000g for 5min at 4°C. The supernatant was discarded and the pellet resuspended in 1ml LB2 (10mM Tris-HCl, pH8.0, 200mM NaCl, 1mM EDTA, 0.5M EGTA, 1x PI) and incubated with occasional inversion of the tube for 5min on ice. Nuclei were pelleted at 2,000g for 5min at 4°C and resuspended in sonication buffer LB3 (10mM Tris-HCl, pH 8, 100mMNaCl, 1mM EDTA, 0.5mM EGTA, 0.1% Na-Deoxycholate, 0.5% N-lauroylsarcosine, 1x PI) and transferred into Diagenode sonication tubes (Diagenode). Sonication was carried out using the Bioruptor pico (Diagenode) for 16 cycles (30sec on, 30sec off settings). Debris was removed from the chromatin containing supernatant by spinning down at full speed for 10min at 4°C. Prepared chromatin was either frozen down in liquid nitrogen and stored at -80°C or used immediately. 5% of the chromatin fraction was flash-frozen as an input sample. 100μl magnetic Protein A-coupled Dynabeads (Thermo Fisher Scientific) were washed three times in 1ml blocking solution (0.2% BSA in PBS). The blocking solution was removed using a magnetic rack. 5μl of anti-H3K9me3 or anti-H3K4me2 polyclonal antibody (see Table 2.6) was diluted in 250μl blocking solution and incubated with

100µl washed beads by rotating at 4°C for at least 4h up to overnight. The supernatant was removed and beads washed three times in blocking solution as described above. The chromatin solution was added to the beads and incubated at 4°C while rotating overnight. Following 4 washing steps for 2min each using ice-cold Lysis Buffer (15mM HEPES, 140mM NaCl, 1mM EDTA, 0.5mM EGTA, 1% Triton, 0.5mM DTT, 10mM Sodium Butyrate, 0.1% Sodium Deoxycholate, 1x PI, 0.05% SDS), beads were washed two additional times with ice-cold 1x TE buffer. All liquid was removed and beads resuspended in 200µl Elution buffer (50 mM Tris-HCl, pH 8; 10 mM EDTA; 1% SDS). Input samples were thawed and brought up to 200µl with Elution buffer. Samples were transferred into 200µl Maxymum Recovery PCR tubes (Axygen) and incubated at 65°C for 16-18h for reverse crosslinking. RNA contamination was removed by adding 200µl 1x TE buffer and 8µl of 1 mg/ml RNase A (Ambion) to ChIP and input samples followed by incubation at 37°C for 30min. Proteins were digested using 4µl Proteinase K (800U/ml, NEB) and incubation at 55°C for 2h. Reverse crosslinked DNA was recovered using the MinElute PCR purification Kit (Qiagen) according to the manufacturer's recommendation and eluted in 15µl nuclease-free water. DNA recovery was verified and quantified using 1µl for Bioanalyzer (Agilent) electrophoresis.

2.11.2 ChIP for *Drosophila* embryos

50µl of embryos were collected and dechorionated as described above (chapter 2.5) and transferred in 1ml Crosslinking solution (1% formaldehyde in PBS, 50% n-heptane) and vortexed on high speed for precisely 15min. 90µl 2.5M glycine solution was added to quench excess formaldehyde and incubated for 5min at RT while rotating. Embryos were allowed to sink to the bottom of the tube and all liquid was removed. Embryos were washed three times for 4min with ice-cold buffer A (60mM KCl, 15mM NaCl, 4mM $MgCl_2$,15mM HEPES pH 7.6, 0.5% DTT, 1x PI) supplemented with 0.1% Triton X-100 (A-Tx buffer). All liquid was removed and embryos flash-frozen and stored at -80°C until further processing. Crosslinked embryos were transferred to a 2ml Dounce homogeniser in 1ml A-TBP (Buffer A + 0.5% Triton X-100). Following

an additional washing step with A-TBP, embryos were lysed in 1ml A-TBP using 10 strokes with a tight-fitting pestle. Lysate was centrifuged at 3,200g for 10min at 4°C and supernatant removed. The pellet was resuspended in 1ml Lysis buffer (15mM HEPES, 140mM NaCl, 1mM EDTA, 0.5mM EGTA, 1% Triton, 0.5mM DTT, 10mM Sodium Butyrate, 0.1% Sodium Deoxycholate, 1x PI) and incubated at 4°C for 15min while rotating. Following centrifugation at 3,200g for 10min at 4°C, the pellet was washed twice with Lysis buffer and centrifuged again. All liquid was removed, and the pellet resuspended in 300µl LB3 (see chapter 2.11.1). Following steps were performed as described above for OSC ChIP (chapter 2.11.1). However, sonication was performed with 6 cycles (30sec on, 30sec off) instead of 16.

2.11.3 ChIP for *Drosophila* ovaries and heads

The ChIP ovary protocol was developed by Dr. Filippo Ciabrelli (CRUK CI) with inputs from myself.

120-150 *Drosophila* ovaries were dissected in ice-cold PBS. Heads were dislodged by pouring liquid nitrogen over whole flies in a dish followed by shaking and collecting 50µl broken-off heads in 1.5ml tube. Samples were homogenised in 100µl Buffer A1 (60mM KCl, 15mM NaCl, 4mM $MgCl_2$,15mM HEPES pH 7.6, 0.5% DTT, 0.5% Triton X-100, 1x PI) using a rotating pestle. The volume was brought up to 1ml with buffer A1 and formaldehyde added to a final concentration of 1.8% for crosslinking. Samples were rotated for exactly 15min at RT and glycine solution added to a final concentration of 225mM. Samples were allowed to rotate for an additional 5min and were centrifuged at 4,000g for 5min at 4°C. The supernatant was removed and the pellet washed twice with buffer A1 and once with buffer A2 (140 mM NaCl, 15 mM HEPES pH 7.6, 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.5 mM DTT, 0.1% sodium deoxycholate, 10 mM sodium butyrate, 1x PI) at 4°C. The pellet was then resuspended in 100µl A2 buffer supplemented with 1% SDS and 0.5% N-laurosylsarcosine and incubated at 4°C for 2h while shaking vigorously. Lysate was sonicated using the Bioruptor pico for 16 cycles (30sec on, 30sec off). The sonicated

lysate was spun at full speed for 10min at 4°C and the supernatant transferred to a new tube. The volume was brought up to 1ml with A2 buffer supplemented with 0.1% SDS. Chromatin used for ChIP was precleared with 15µl washed Protein A Dynabeads and incubated with antibody coated beads as described above (chapter 2.11.1). Further steps were performed as described above (chapter 2.11.1).

Table 2.6: List of antibodies used in ChIP experiments

Antibody	Lot #	Origin	Manufacturer and ID
anti-H3K9me3	15617003	Rabbit	Active Motif #39161
anti-H3K4me2	2971019	Rabbit	Millipore # 07-030

2.12 Formaldehyde RNA immunoprecipitation (fRIP) from embryos

fRIP from embryos was performed using an adapted protocol previously published (Kneuss et al., 2019). 100µl of embryos were collected and dechorionated for 1min in 50% bleach followed by washing under tap water for 1min. Embryos were transferred into 2ml tube and washed twice with PBS. Crosslinking was carried out in 750µl 0.1% PFA in PBS and 750µl n-heptane for 10min while vortexing vigorously. Supernatant was removed and crosslinking quenched with addition of glycine solution at a final concentration of 125mM in PBS for 10min on ice. Embryos were rinsed twice with ice-cold PBS and washed for 5min at 4°C while rotating. Washing buffer was completely removed and embryos flash-frozen in liquid nitrogen. Samples were stored at -80°C until further processing. Dry embryos were resuspended in 200µl Pierce RIPA buffer (Thermo Fisher Scientific) supplemented with 1x PI and RNasin (1:1,000, Promega) and homogenised using a rotating pestle for 1min on ice. The volume was brought up to 1ml with RIPA buffer and samples incubated for 20min at 4°C while rotating. The solution was split into three Bioruptor sonication tubes (Diagenode) and samples sonicated for 3 cycles (30sec on, 30sec off) on a Bioruptor pico. Tubes were centrifuged at full speed for 10min at 4°C and supernatants pooled in a new tube. 1ml of fRIP binding/washing buffer (150mM KCl, 25mM Tris (pH 7.5), 5mM EDTA, 0.5%

NP-40, 0.5mM DTT, 1× PI, 1:1,000 RNasin) was added. 40*µ*l Protein A Dynabeads were washed twice with fRIP buffer and added to supernatant. The solution was incubated for 1h while rotating at 4°C. 100*µ*l of supernatant was transferred and stored at -80°C and further processed as input sample. Beads were removed and 10*µ*l anti-Piwi antibody (see Table 2.5) or 10*µ*l rabbit IgG antibodies (Abcam, ab37415) added to the remaining supernatant. Antibody binding was carried out overnight at 4°C while rotating. The supernatant was removed, and beads washed 3 times for 5min with 1ml fRIP buffer at 4°C while rotating. Tubes were changed during each washing step to minimise carry-over contaminations. Beads were resuspended in 56*µ*l RNase-free water and 33*µ*l reverse-crosslinking buffer (3x PBS, 6 % N-lauroysarcosine, 30mM EDTA, 15mM DTT, 10µl of Proteinase K, 1:1,000 RNasin). Protein digestion and reverse-crosslinking was carried out for 1h at 42°C followed by 1h at 55°C. Solution was allowed to cool completely and supernatant was removed. Beads were resuspended in 1ml Trizol (Thermo Fisher Scientific) and stored at -80°C until RNA extraction (see chapter 2.13).

2.13 RNA extraction

RNA for mRNA-Seq and qRT-PCR experiments was isolated using the RNeasy Mini kit (Qiagen) with on-column DNA digestion (RNase-free DNase Set, Qiagen) according to the manufacturer's recommendations. In short, cells were resuspended in 1ml Trizol by pipetting. Embryos or ovaries were lysed in 1ml Trizol using a rotating pestle. 200*µ*l chloroform was added to all samples and vortexed vigorously for 10sec, followed by incubation at RT for 5min. Samples were centrifuged at 12,000g for 15min at 4°C and the aqueous phase transferred to a new tube. 500*µ*l ethanol was added and mixed well. The solution was transferred to a RNeasy mini column and processed as recommended by the manufacturer. RNA was eluted in 50*µ*l RNase-free water and samples stored at -80°C.

RNA for small RNA-Seq and fRIP-Seq experiments were isolated using RNA precipitation. Samples were homogenised as described above in 1ml Trizol. 200*µ*l chloroform was added, vortexed for 10sec and incubated at RT for 5min. Following

centrifugation at 12,000g for 10min at 4°C, the aqueous phase was transferred to a new tube (500µl) and 500µl isopropanol together with 1µl GlycoBlue (Thermo Fisher Scientific) added and incubated at -20°C for 30min. Samples were centrifuged at full speed for 30min at 4°C and pellets washed with 80% ice-cold ethanol. Ethanol was completely removed, and pellet dried for 2min at RT followed by resuspension in 20µl RNase-free water. RNA concentrations were measured using a NanoDrop Spectrophotometer (Thermo Fisher Scientific). RNA samples were stored at -80°C.

2.14 Library preparation for Next-Generation Sequencing

2.14.1 mRNA-Seq Library Preparation

mRNA was isolated from 1µg total RNA using the NEBNext Poly(A) mRNA magnetic Isolation Module (NEB). Libraries were generated with the NEBNext Ultra Directional RNA Library Prep kit for Illumina (NEB) according to manufacturer's recommendations with 12 PCR amplification cycles. Multiplexing was achieved by using unique indexes from the NEBNext Multiplex Oligos for Illumina (Index Primer Set 1-3; NEB). Libraries were quantified by Bioanalyzer (Agilent) electrophoresis and pooled at equal molarity. Pooled libraries were quantified further using the KAPA Library Quantification Kit for Illumina (Kapa Biosystems) before sequencing.

2.14.2 fRIP-Seq Library Preparation

Ribosomal RNAs (rRNAs) in Piwi-IP samples and inputs (see chapter 2.12) were depleted prior to library preparation using the RiboGone Mammalian kit (Clontech) using all isolated RNA (20µl) as input. fRIP-Seq libraries were generated with the SMARTer stranded RNA-Seq Kit (Clontech) according to the manufacturer's recommendations with 18 PCR amplification cycles for all samples. Samples were pooled and quantified as described in chapter 2.14.1

2.14.3 small RNA-Seq Library Preparation

Small RNA libraries from ovary and embryo Piwi-IPs were generated essentially as previously described (McGinn and Czech, 2014). All RNA isolated from Piwi-IP (20µl) experiments was used for the library preparation and purified by polyacrylamide gel electrophoresis (PAGE). 18-29 nt long RNAs were extracted from the gel by incubating excised gel fragments in 425µl 400mM NaCl overnight. RNA was recovered by precipitation and eluted in 12µl RNase-free water. 2µl of DMSO, 2µl of 10x ATP-free T4 RNA ligase buffer (NEB), 1µl RNaseOUT (Thermo Fisher Scientific), 1µl of 100µM 3' adapter (oLig3new) containing four random nucleotides at the 5' end (Jayaprakash et al., 2011) (see Table 2.2) and 2µl T4 RNA ligase (NEB) were mixed and incubated for 2h at RT. Ligated RNA was purified by PAGE and precipitated as described above. 12µl of 3' ligated RNA was mixed with 2µl DMSO, 2µl 10x T4 RNA ligase buffer, 1µl RNAseOUT, 2µl T4 RNA ligase and 1µl of 100µM 5' adapter (oLig5new) (see Table 2.3) and incubated at 37°C for 2h. Ligated RNA was isolated as described above and reverse transcribed using SuperScript III Reverse Transcriptase (Thermo Fisher Scientific). cDNA was amplified by PCR using 18 amplification cycles and purified by gel electrophoresis followed by gel extraction using the Wizard SV Gel and PCR Clean-up System (Promega). Libraries were pooled and quantified as described in chapter 2.14.1.

2.14.4 ChIP-Seq Library Preparation

ChIP-Seq input material for library preparation varied between 1.5ng and 5ng depending on the recovered DNA amount. ChIP-Seq libraries were generated using the NEBNext Ultra II DNA Library Prep kit (NEB) according to manufacturer's recommendation. For input material of less than 3ng, 12 PCR amplification cycles were performed and for input more than 3ng 9 cycles. Libraries were pooled and quantified as described in chapter 2.14.1.

2.14.5 Whole Genome Sequencing (WGS) of *w*^{1118} flies

100 flies were collected in a 1.5ml tube and frozen at -80°C for at least 1h. High molecular weight genomic DNA (HMW gDNA) was isolated using the Blood and Cell Culture DNA Mini kit (Qiagen). 2ml G2 buffer with 4µl 100mg/ml RNase A (Qiagen) was prepared fresh. Frozen flies were split in two 1.5ml tubes and 250µl G2 added to each tube. Flies were homogenised using a rotating pestle on ice for 1min. 700µl G2 and 50µl Proteinase K (800U/ml, NEB) was added to each tube and incubated at 50°C for 2h with occasional tube inversions. Tubes were spun at 5,000g for 10min at 4°C and supernatant transferred to new tube avoiding debris. A Qiagen Genomic-tip 20/G was equilibrated with 1ml QBT buffer and allowed to empty by gravity flow. The supernatant containing digested proteins and genomic DNA was added to the column and allowed to flow through. The column was washed 3 times with 1ml QC buffer. Elution was carried out with 1ml QF elution buffer and repeated once. Flow through was transferred to two new tubes (1ml each) and 700µl isopropanol added. Tubes were inverted 10 times and centrifuged at full speed for 15min at 4°C. The pellet was washed with 70% ethanol twice and air-dried for 5min. 25µl RNase-free water was added and DNA resuspended by flicking tube gently several times while incubating at 37°C for 2h. DNA was stored at 4°C.

DNA was sheered using a Covaris S220 (Covaris). 3µg of genomic DNA was diluted in RNase-Free water and transferred to a AFA Fiber Crimp-Cap (PN520052, Covaris) microtube. Sonication was carried out with the following settings: Peak incident Power (W) 105, Duty Factor 5%, Cycles per Burst 200, Treatment time 80sec. This resulted in sheared DNA fragments peaking at 500bp. DNA was recovered using the QIAquick PCR Purification Kit (Qiagen). Libraries were generated using the NEBNext Ultra II DNA Library Prep kit (NEB) according to manufacturer's recommendation with 1µg input material. 3 PCR amplification cycles were performed. Libraries were quantified as described in chapter 2.14.1.

2.15 Next-generation Sequencing (NGS)

Sequencing was performed by the Genomics Core facility at CRUK CI.

mRNA-Seq, ChIP-Seq, fRIP-Seq and small RNA-Seq libraries were sequenced on an Illumina HiSeq4000 according to manufacturer's recommendations using single-end 50bp runs. WGS libraries were sequenced with paired-end 150bp runs.

2.16 Protein isolation from whole embryos and quantitative mass spectrometry

Protein isolation from embryos was performed with Evangelia Papachristou and TMT-labelling as well as mass spectrometry was performed by the Proteomics Core facility at CRUK CI.

100μl of w^{1118} embryos for time points 0-2h, 5-7h and 10-12h AEL were collected in three biological replicates on agar plates and dechorionated as previously described (chapter 2.5). Embryos were lysed in lysis buffer (0.1% SDS, 0.1M triethylammonium bicarbonate (TEAB), 1x Halt Protease and Phosphatase Inhibitor (Thermo Fisher Scientific)) using a rotating pestle on ice for 2min or until entirely homogenised. Lysate was heated for 5min at 90°C and probe sonicated for 20sec (20% power with pulse of 1sec). Debris was pelleted by centrifugation at full speed for 10min at 4°C and supernatant transferred to a new tube. Protein concentration was measured using Bradford Assay (Bio Rad). 100μg protein was digested with trypsin overnight at 37°C. TMT chemical isobaric labelling was performed as described previously (Papachristou et al., 2018). Peptide fractions were analysed on a Dionex Ultimate 3000 UHPLC system coupled with the nano-ESI Fusion Lumos mass spectrometer (Thermo Scientific).

2.17 Quantitative reverse transcription polymerase chain reaction (qRT-PCR)

Reverse transcription (RT) was performed using the SuperScript IV Reverse transcriptase Kit (Thermo Fisher Scientific). 1μg of total RNA was used as input material for qRT-PCR analysis (chapter 2.13) and mixed with 1μl 50μM Oligo $d(T)_{20}$ primer (IDT), 1μl 10mM dNTP mix (10mM each) (Thermo Fisher Scientific) and volume brought up to 14μl with RNase-free water. RNA was denatured for 5min at 65°C and incubated immediately on ice for 2min. 4μl 5x SSIV Buffer, 100mM DTT, 1μl RNaseOUT and 1μl SuperScript IV Reverse Transcriptase (200U/μl) was added and mixed well by pipetting. The reaction was incubated at 55°C for 15min, followed by inactivation at 80°C for 10min. cDNA was diluted in RNase-free water to a final volume of 100μl.

qRT-PCR was performed on a QuantStudio Real-Time PCR Light Cycler (Thermo Fisher Scientific) in technical triplicates. Each triplicate consisted of 5μl Fast SYBR Green Master Mix, 2.8μl RNase-free water, 0.2μl primer pairs (10μM each) (see Table 2.2) and 2μl of diluted cDNA. Expression of targets was quantified using the $\Delta\Delta$CT method (Livak and Schmittgen, 2001). Fold change was calculated as indicated in figure legends and normalised to *rp49* or *act5c*.

2.18 Treatment of embryos for auxin-induced degradation

Embryos were collected for 1h and dechorionated as described above (chapter 2.5). Control embryos were transferred into a fine mesh strainer placed in a plastic dish and submerged in PBS. 1M auxin solution was generated by diluting Indole-3-acetic acid (IAA) in water and stored at -20°C protected from light. Auxin-treated embryos were submerged in PBS with indicated auxin concentrations. Embryos were placed at 25°C for appropriate times and harvested for RNA experiments by transferring into 1ml Trizol followed by RNA extraction (chapter 2.13). Embryos used for ChIP-Seq were processed as described in chapter 2.11.2.

2.19 Determination of hatching rate

Embryos were collected for 1h and dechorionated as described above (chapter 2.5). Treatment was performed as described above (chapter 2.18) for 2h. Embryos for each condition were then transferred individually with a fine paint brush on grape juice agar plates and counted. Agar dishes were wrapped in parafilm to prevent dehydration and incubated at 25°C overnight or until larvae emerged. Hatching rate was determined by counting hatched larvae and additionally embryos that failed to develop.

2.20 Computational Data Analysis

2.20.1 RNA-Seq analysis

RNA-Seq library fastq files contained 50bp reads. The first and the last two bases of all reads were trimmed using fastx_trimmer (http://hannonlab.cshl.edu/fastx_toolkit/). Reads were first aligned to the consensus sequence for all *Drosophila melanogaster* transposons using STAR (Dobin et al., 2013) allowing random allocation of multimappers. Unmapped reads were further uniquely aligned to *Drosophila melanogaster* genome release 6 (dm6). Generated bam files were split in reads originating from sense and antisense genomic strands using samtools view options -f 0x10 and -F 0x10 for sense and antisense reads respectively (Li et al., 2009b). Indexes were generated using samtools index function. Coverage files were generated using bamCoverage with normalisation mode --normalizeUsing CPM (Ramirez et al., 2014) and applying a scaling factor (--ScaleFactor). Scaling factors for individual files were calculated by dividing the sum of mapped reads contained in the file by the sum of all transposon and dm6 mapping reads of the corresponding library. Reads mapping to protein-coding genes were counted with htseq (Anders et al., 2015) using a feature file downloaded from Ensembl (https://www.ensembl.org/Drosophila_melanogaster/Info/Index). Reads mapping to

individual transposons were counted with a custom script using samtools idxstats function to extract reads mapping to individual sequences of the reference genome/transposon consensus sequence. Code used to analyse RNA-Seq libraries can be found in the appendix (Code 8.1).

Biological replicates for RNA-Seq libraries in chapter 3 were collapsed and differential expression calculated using custom scripts (see also Code 8.2). Count files generated as described above were normalised to reads per million (RPM) and plotted using ggplot2 (Wickham, 2016). Bar graphs displaying transposon expression profiles and heatmap during embryogenesis were calculated from collapsed replicates. Bar graphs were plotted using ggplot2 in R and heat map was plotted using the seaborn package in Python. Differential expression analysis for RNA-Seq data from Degron experiments (chapter 5.3.8) was performed using DeSeq2 (Love et al., 2014) (see also Code 8.3).

2.20.2 fRIP-Seq analysis

fRIP-Seq libraries were analysed according to chapter 2.20.1 with the following exceptions. Trimmed reads were first aligned to a genome file containing rRNA contamination sequences using STAR. Mapped reads were discarded, and unmapped reads further analysed as described above (see also Code 8.4). Differential expression analysis for fRIP-Seq was performed using DeSeq2 (see also Code 8.3)

2.20.3 Small RNA-Seq

Reads from small RNA-Seq libraries were adapter clipped using fastx_clipper with settings -Q33 -l 15 -a AGATCGGAAGAGCACACGTCT. The first and last 4 bases of clipped reads were trimmed using seqtk trimfq (https://github.com/lh3/seqtk) and trimmed reads aligned as described above (see chapter 2.20.1) to the transposon consensus sequence. Reads mapping in sense direction were extracted using samtools view options -F 0x10 and total mapping reads to individual transposons counted as described above (chapter 2.20.1) (see also Code 8.5). Fraction of

antisense mapping transposons reads to total antisense transposons reads was determined for all identified transposons and plotted as a bar graph using ggplot2.

2.20.4 ChIP-Seq

ChIP-Seq was analysed according to chapter 2.20.1. However, reads were not separated in sense or antisense files (see also Code 8.6).

Metaplots flanking euchromatic transposon insertion sites were calculated using deepTools2's computeMatrix scale-region function with bin size 10 (see also Code 8.7) and plotted in R (see also Code 8.8).

2.20.5 Generation of annotation files

Mappability track for dm6 with 50bp resolution was calculated according to previously published instructions (Derrien et al., 2012). Bed files for random genomic windows for box plots of H3K9me3 ChIP-Seq data were calculated using Bedtools' random function with seed number 800, bin number 1,000 and bin size 5,000bp (Quinlan and Hall, 2010). 100 random windows (200-300) were chosen. Random euchromatic insertion sites were determined as described above with bin size 1bp. Bedtools output bed files were converted into GTF files using custom scripts.

Euchromatic transposon insertions reported previously (Sienski et al., 2012) for OSCs were updated from genome release dm3 to dm6 using the UCSC liftOver tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver). *De novo* transposon insertion calling for w^{1118} strain was performed using the TEMP algorithm (Zhuang et al., 2014). In short paired-end reads generated by WGS (chapter 2.14.5) were aligned to dm6 using BWA (Li and Durbin, 2009). Reads with only one mate aligned to dm6 were extracted and the unmapped mate aligned to transposon consensus sequences. Calculated insertion sites were extracted from generated GTF if they were supported by reads at both sides (1p1).

Lists of transcripts maternally deposited or zygotically transcribed were generated using counting outputs from htseq for collapsed RNA-Seq data from 0-2h

and 3-4h w^{1118} embryos (see chapter 2.20.1). Reads were normalised to reads per million (RPM). Transcripts detected with less <2 RPM in the sum of both conditions were removed. Maternal transcripts were defined as >10-fold more abundant in maternal transcriptome (0-2h AEL) in comparison to zygotic transcriptome (3-4h AEL). Zygotic transcripts were defined as >10-fold enriched in zygotic versus maternal transcriptome.

Protein database used to identify peptides from *Drosophila* genes and transposons was generated by merging an existing database downloaded from flybase (dmel-r6.24.fa) with translated ORFs of transposons. ORFs were predicted and translated using prodigal (https://github.com/hyattpd/Prodigal). ORFs with less than 300 amino acids were removed using seqtk -L 300 and file converted in fasta format.

2.20.6 Protein domain prediction

Functional analysis of protein sequences was performed using the InterPro web application (https://www.ebi.ac.uk/interpro/). Protein domains and families for open reading frame encoded by *roo* transcripts were predicted using standard settings.

2.20.7 Gene ontology (GO) enrichment analysis

Gene ontology (GO) enrichment analysis was performed using the PANTHER Classification System searching for biological processes in *Drosophila melanogaster*. Fold enrichments reported correspond to results obtained by the PANTHER Overrepresentation Test (Released 20200407).

2.20.8 Mass spectrometry raw data processing

Raw data processing was performed by the Proteomics Core facility at CRUK CI.

Raw data files were processed according to previous reports (Papachristou et al., 2018). Spectral .raw files were analysed with the SequestHT search engine on Thermo Scientific Proteome Discoverer 2.1 for peptide and protein identification. Data was searched against a modified FlyBase protein database ("dmel-r6.24-te_fused3.fasta", also see chapter 2.20.5) with following parameters: Precursor Mass Tolerance 20 ppm, Fragment Mass Tolerance 0.5 Da. Dynamic Modifications were oxidation of methionine residues (+15.995 Da), deamidation of asparagine and glutamine (+0.984 Da) and Static Modifications were TMT6plex at any amino-terminus, lysine (+229.163 Da) and methylthio at cysteine (+45.988). The Reporter Ion Quantifier node included a TMT 6plex (Thermo Scientific Instruments) Quantification Method, for MS3 scan events, HCD activation type, integration window tolerance 20 ppm and integration method Most Confident Centroid. Peptides with an FDR >1% were removed. The downstream workflow included signal to noise (S/N) calculation of TMT intensities. 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.

2.20.9 Bioinformatics Analysis of proteomics data

qPLEXanalyzer R package was applied to files generated in chapter 2.20.7 by Kamal Kishore (Bioinformatics Core facility at CRUK CI).

Processed data files were analysed as described in a previous publication (Papachristou et al., 2018) using qPLEXanalyzer in R with multimapping peptides included in the analysis. Bar graphs showing protein intensities for Piwi and volcano plots with indicated comparisons were plotted using ggplot2 in R.

2.20.10 Image analysis

Script estimating nuclear and cytoplasmic signal for OSC HCR FISH experiments was written by Alessandro Passera (CRUK CI).

Confocal microscopy images were analysed in Fiji (ImageJ). Custom scripts were used to count cells (see Code 8.9). DAPI channel was used to outline nuclei and oligo dT channel was used to outline cell boarders. Signal intensities of nuclei and cytoplasm as well as dimensions of cells and nuclei were measured and results saved as csv files. Further processing was carried out in R (see Code 8.10). Cells with RNA FISH total cell signal <300 (arbitrary units) were removed as well as cells with an area <50 pixels and nuclei with an area <10 pixels. Nuclear and cytoplasmic signal intensity was calculated from means of 4 biological replicates. Nuclear fractions were calculated for each detected cell individually by dividing nuclear signal by total signal (see also 2.9.10 for statistical analysis).

2.20.11 Statistical analysis

Statistical analyses were performed in R and results as well as applied method are indicated in figure legends and accompanying text. Statistics for boxplots depicting H3K9me3 occupancy at genomic regions or over consensus sequence was performed using Welch two sample t-test. qRT-PCR and western blot results were statistically evaluated using unpaired t-test. Distribution of transposon RNA signal was tested for normality using a Shapiro-Wilk test. Identification of statistically significant differences in signal distribution was further tested by Analysis of Variance (ANOVA). Individual conditions were compared using a post hoc Tukey test.

2.20.12 Data visualisation

Sequencing data was plotted in R or Python as indicated in previous chapters (chapter 2.20). Microscopy images were processed in Fiji (ImageJ) and programs indicated in chapter 2.7. Figures were assembled in Adobe Illustrator.

3 piRNA guided co-transcriptional gene silencing co-opts nuclear export factors

Parts of this chapter have been published in a similar form in eLife (Fabry, Ciabrelli, Munafò et al., 2019). The project was conceived by Ben Czech and Greg Hannon with inputs from myself, Filippo Ciabrelli, Marzia Munafò and Evelyn Eastwood. Evelyn Eastwood generated *nxf2* knock-out flies. Filippo Ciabrelli performed *Drosophila* crosses for knockout experiments followed by RNA extraction and H3K9me3 ChIP, as well as cloning of RNA-Seq and ChIP-Seq libraries for *in vivo* experiments and qRT-PCRs for RNA and DNA tethering experiments. Cloning of rescue constructs was assisted by Federica Falconio. Marzia Munafò established the RNA-tethering system in OSCs, performed RNA-tethering for qRT-PCR experiments as well as S2 cell immunofluorescence imaging and assembled figures for RNA-FISH confocal images. ImageJ script counting nuclear and cytoplasmic signal was written by Alessandro Passera. Ben Czech and Greg Hannon supervised the project.

3. 1 Introduction

Co-transcriptional gene silencing (TGS) is one of the effector mechanisms of the piRNA pathway for genome defence against transposable elements in animal gonads (Czech et al., 2018). piRNAs, in complex with Piwi are believed to bind nascent transcripts complementary to the piRNA guide sequence (Brennecke et al., 2007; Klenov et al., 2011; Saito et al., 2009; Sienski et al., 2012). Binding of the piRNA-Piwi complex to nascent transposon RNA targets has been proposed to result in conformational changes of the complex thereby recruiting downstream effectors such as gonad-specific protein Panoramix (Panx) (Sienski et al., 2015; Yu et al., 2015). Panx is believed to connect to the general transcriptional silencing machinery of the cell. Recruitment of downstream factors results in the deposition of repressive chromatin marks at targeted genomic regions as well as removal of marks indicative of active transcription (Sienski et al., 2015; Yu et al., 2015). Targeted loci are then transformed into transcriptionally-silenced heterochromatin.

Immunoprecipitation followed by mass spectrometry of Panx protein from *Drosophila* ovary lysate performed in our lab revealed two highly enriched binding partners, Nxf2 and Nxt1 (Fabry et al., 2019). We named this protein complex the Panx-induced co-transcriptional silencing (PICTS) complex. Nxf2 is a gonad-enriched protein with no previously assigned function. Interestingly, both proteins, Nxf2 and Nxt1 were also highly scoring candidates in screens for piRNA pathway components (Czech et al., 2013; Handler et al., 2013; Muerdter et al., 2013).

3.2 Aim

The putative binding partners of Panx, Nxf2 and Nxt1, are predicted to be associated with mRNA export in cells. However, both have not been linked to transposon regulation before. The fact that Nxf2 and Nxt1 appeared as positive hits in previous screens, as well as interacting with Panx, suggested their direct involvement in piRNA-mediated transposon control. Therefore, I aimed to further characterise the interaction of Panx with Nxf2 and Nxt1 and analyse their impact on transposon regulation both *in vivo* and in ovarian somatic cells (OSCs), a model cell line harbouring a functional piRNA-dependent TGS pathway.

3.3 Results

3.3.1 Nxf2 is necessary for transposon control *in vivo* and *in vitro*

Previous studies revealed Panx's function in TGS. Loss of Panx in *Drosophila* ovaries or OSCs was accompanied by deregulation of transposons and female sterility. Additionally, the repressive chromatin mark H3K9me3 was lost from transposon bodies, indicating disruption of epigenetic transcriptional control at affected genomic transposon insertions (Sienski et al., 2015; Yu et al., 2015).

Binding of the previously uncharacterised Nuclear Export Factor 2 (Nxf2) to Panx suggests its involvement in transcriptional gene silencing. Nxf2 is one of the four nuclear export factors found in *Drosophila* (Figure 3.1 A). We assessed the impact of Nxf2 loss on transposon control by generating two *nxf2* knockout strains ($nxf2^{F10*}$, $nxf2^{\Delta 1*}$) (Figure 3.1 B) and analysed steady-state RNA levels in ovaries of trans-heterozygous mutants ($nxf2^{F10*/\Delta 1*}$, from now on called homozygous mutants) or heterozygous control flies. RNA-Seq analysis of ovaries revealed deregulation of transposon expression in *nxf2* homozygous mutants in comparison to heterozygous flies (Figure 3.1 C). Transposons commonly affected by piRNA pathway disruption in ovaries including *HeT-A* and *TAHRE* showed strong upregulation (268-fold and 191-fold upregulation respectively). In total 28 out of 60 expressed transposon families showed upregulation of at least 4-fold. Similar results were obtained for *panx* homozygous mutants (*HeT-A* 109-fold upregulation, *TAHRE* 136-fold upregulation) with 16 out of 60 transposons > 4-fold upregulated as previously reported (Sienski et al., 2015; Yu et al., 2015) (Figure 3.1 D). piRNA pathway-specific factors are well known to cause deregulation of transposons. Gene expression, however, remains mostly unchanged. While transposon mRNA expression was highly affected by both *nxf2* and *panx* knockouts in comparison to heterozygous controls, gene expression was only mildly changed as measured by correlation between conditions ($nxf2^{F10/\Delta 1*}$vs. $nxf2^{het}$ $R^2 = 0.963$, $panx^{M1/M4}$ versus $panx^{het}$ $R^2 = 0.991$) (Figure 3.1 C-D). This suggests a transposon-specific effect for both proteins on transcriptional control.

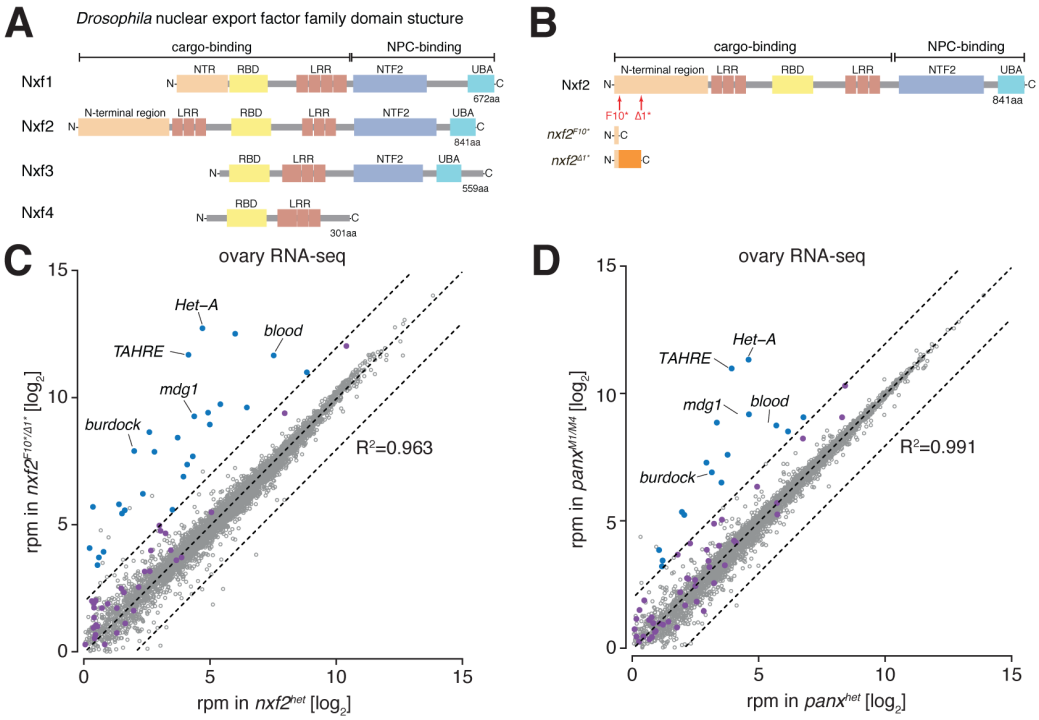

Figure 3.1: Nxf2 controls transposon expression *in vivo*

A) Cartoon displaying domain structure of *Drosophila* nuclear export factor family. NTR, amino-terminal region; LRR, leucine rich repeats; RBD, RNA-binding domain; NTF2, NTF2-like domain; UBA, Ubiquitin associated domain. **B)** Cartoon displaying *nxf2* knockout mutations (*nxf2*^{F10*} and *nxf2*$^{\Delta 1*}$) **C**) Scatter plot showing expression levels in RPM (reads per million sequenced reads) of genes (in grey) and transposons (in purple) from total RNA of ovaries for indicated *nxf2* genotypes (n= 3; R^2 values represent expression of genes only). Transposons whose abundance changed more than four-fold compared to heterozygotes are highlighted in blue. **D)** Same as (C) but for *panx* genotypes.

Loss of piRNA-dependent transposon control is commonly accompanied by reduction of repressive H3K9me3 histone modifications. Therefore, we measured epigenetic changes by H3K9me3 chromatin immunoprecipitation followed by sequencing (ChIP-Seq) from homozygous or heterozygous knockout ovaries for either *nxf2* or *panx*. In order to systematically assess changes on the chromatin level of *nxf2* mutants at transposon bodies, I categorised TE families into two separate groups. H3K9me3 ChIP-Seq reads from TEs that were upregulated > 4-fold in our RNA-Seq experiment were summed up as well as reads derived from TE families with a < 4-fold increase in *nxf2* mutant ovaries. As a control 100 5-kb random genomic intervals were

assigned and ChIP-Seq reads counted for each group. TEs > 4-fold upregulated in *nxf2* mutant ovaries showed significant reduction of H3K9me3 signal in comparison to random intervals (Welch Two Sample t-test, t= -5.8339, df= 39.325, p-value= 8.566e-07) and to TE families < 4-fold upregulated (Welch Two Sample t-test, t= -3995, df= 34.138, p-value= 0.0003269) (Figure 3.2 A). H3K9me3 levels strongly decreased at the germline-specific transposon *HeT-A* in both *nxf2* and *panx* mutant fly ovaries compared to heterozygous knockout controls (Figure 3.2 B), thus confirming the involvement of both Nxf2 and Panx in epigenetic transposon control.

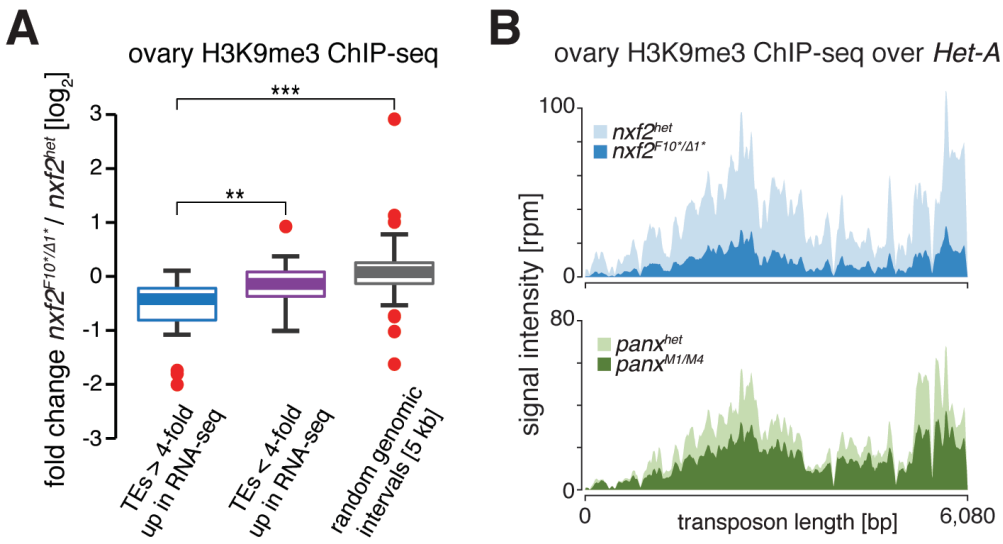

Figure 3.2: Nxf2 is involved in transcriptional gene silencing

A) Box plot showing changes in H3K9me3 levels between *nxf2* heterozygotes and mutants in ovarian ChIP-Seq for the indicated categories (colouring of transposons is based on their fold change in RNA-Seq in Figure 3.1 C-D; 100 random genomic intervals of 5-kb bin sizes are shown in grey). ** denotes p-value < 0.001; *** denotes p-value < 0.0001 (Welch two sample t-test). **B**) Density plots representing ChIP-Seq signal from ovaries mutant for *nxf2* and *panx* over the *Het-A* transposon consensus sequence are shown along with heterozygote flies in RPM (n= 2 biological replicates).

Drosophila ovaries are complex organs comprised of several cell types including germ cells, somatic sheet cells as well as muscle and connective tissue. For this reason, experiments performed on whole ovaries contain contaminations of cells not expressing transposons or the piRNA pathway. OSCs are derived from precursors of ovarian somatic follicle cells that surround the germline and express a functional piRNA-dependent TGS pathway (Saito et al., 2009). Thus, they represent a convenient and clean model for mechanistic studies without contamination from undesirable cells.

In order to assess whether loss of Nxf2 induced a similar deregulation of transposon expression *in vitro*, I performed siRNA-mediated knockdowns for either *panx* or *nxf2* mRNA in OSCs. As a positive control for piRNA-dependent loss of transposon control, depletion of Piwi was included (*siPiwi*) along with a *siGFP* knockdown as a negative control. Cells were harvested and prepared for RNA-Seq following two pulses of siRNA transfection each 48h apart and a total 4 days of incubation. Transposons were deregulated in Piwi, Panx and Nxf2-depleted OSCs in comparison to *siGFP* controls (Figure 3.3). As previously published (Sienski et al., 2012), depletion of Piwi in OSCs led to strong derepression of certain transposon families including *mdg1*, *gypsy* and *297,* while transposons that are usually not affected such as *invader2* showed no change in mRNA levels. Depletion of Panx showed a similar, yet less severe impact on transposon control. Interestingly, loss of Nxf2 resulted in deregulation of transposon families similar to Panx depletion and was considerably different from cells depleted of Piwi. Comparing RNA-Seq data from Nxf2 or Panx depleted OSCs (*nxf2* versus *panx*), revealed that both showed near identical deregulation of TE families with similar fold changes.

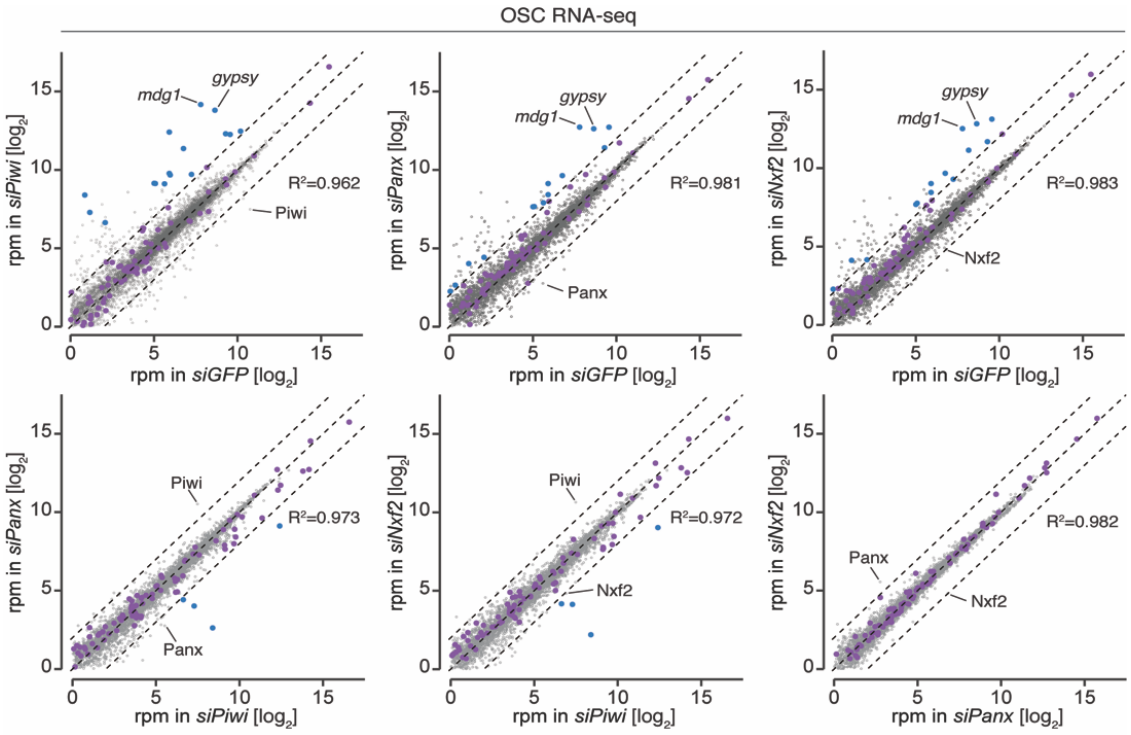

Figure 3.3: Nxf2 and Panx control similar transposon families

Scatter plots showing expression levels (reads per million sequenced reads, RPM) of genes (in grey) and transposons (in purple) of OSCs for indicated knockdowns (n= 3; R^2 values represent expression of genes only). Transposons whose abundance changed more than 4-fold are highlighted in blue.

Upregulation of transposon expression in cells depleted of piRNA pathway factors is believed to be a consequence of a loss of H3K9me3 marks at targeted transposon insertions. I performed ChIP-Seq using H3K9me3-specific antibodies for knockdowns described above in order to assess the abundance of repressive chromatin marks at targeted transposon insertions. As previously reported, loss of Piwi was associated with a strong reduction of H3K9me3 marks at transposon sequences in comparison to *siGFP* controls (Figure 3.4). Loss of H3K9me3 was inversely correlated with the upregulation of transposon mRNA following depletion of Piwi, while transposons not affected at the transcript level, such as *invader2*, did not show any loss of H3K9me3. Depletion of either Panx or Nxf2 had a similarly strong impact on H3K9me3 deposition compared to Piwi depletion.

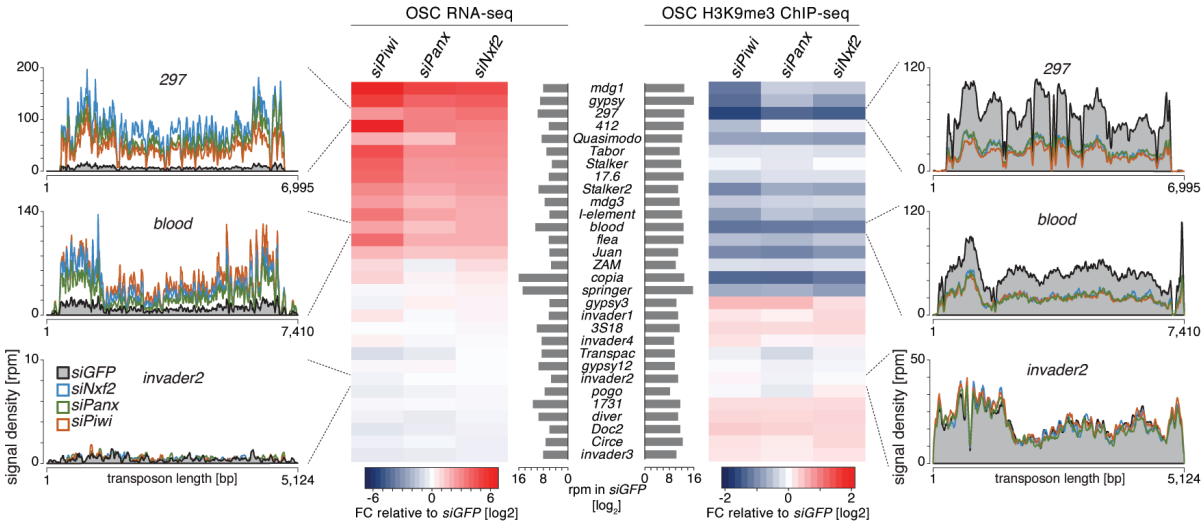

Figure 3.4: Loss of Nxf2 affects the epigenetic silencing state at transposons

Heat maps showing RNA-Seq (left) and H3K9me3 ChIP-Seq (right) of the 30 most expressed transposons in OSCs (compared with *siGFP*) upon indicated knockdowns (n= 3 biological replicates for RNA-Seq and 2 biological replicates for ChIP-Seq). Density profiles show normalised reads from RNA-Seq (left) and H3K9me3 ChIP-Seq (right) experiments mapping to the indicated transposons.

TGS-induced H3K9me3 deposition at TE insertions has been shown to have an impact on nearby genomic regions (Sienski et al., 2012). Although I was able to identify a loss of H3K9me3 analysing the consensus sequence of specific transposon families (Figure 3.4), an accurate characterisation of histone modifications at individual insertions is still challenging due to the repetitiveness of transposon insertions within the genome. The abundance of mostly identical transposon sequences at multiple genomic locations prevents mapping of unique short sequencing reads (50bp), therefore impeding exact evaluation of the epigenetic state of individual transposon insertions. However, the deposition of H3K9me3 is not limited to the transposon sequence, instead spreading of this mark in both directions of genomic TE insertions can be observed regularly.

In order to identify changes of chromatin marks for individual insertions and to characterise changes of affected nearby genomic regions, I revisited a list of published insertion coordinates of TEs in OSCs (Sienski et al., 2015; Sienski et al., 2012). Following updating genomic locations from *Drosophila melanogaster* genome release

dm3 to dm6, I characterised insertions affected by loss of Piwi (> 2-fold decrease in H3K9me3 signal 5-kb up and downstream of insertions of *siPiwi* treated cells in comparison to *siGFP*) as piRNA pathway dependent and used this list for subsequent analyses. I mapped unique H3K9me3 reads to dm6 and probed the epigenetic state 15kb up- and downstream of piRNA pathway-dependent TE insertion sites. Insertions were sorted by the intensity of H3K9me3 marks deposited in undisturbed, *siGFP* treated OSCs. Following the loss of Piwi, H3K9me3 signal strongly decreased at affected TE insertions. Loss of Panx or Nxf2 lead to a similar reduction of H3K9me3 signal, however, less severe in comparison to Piwi depletion (Figure 3.5 A). Collapsing H3K9me3 signal at piRNA pathway-dependent TEs clearly indicates a global loss of H3K9me3 at affected insertions (Figure 3.5 B).

Loss of H3K9me3 and the subsequent reactivation of transcription is accompanied by an increase of the histone mark H3K4me2, associated with active transcription. H3K4me2 ChIP-Seq revealed strong deposition of this mark in cells treated with siRNAs against Piwi, Panx or Nxf2 in comparison to *siGFP* controls, strongly correlating with the increased transcriptional output of the respective transposons (Figure 3.5 C). Collapsing all H3K4me2 signal at piRNA-dependent TE insertions, again, showed a global increase of H3K4me2 deposition (Figure 3.5 D).

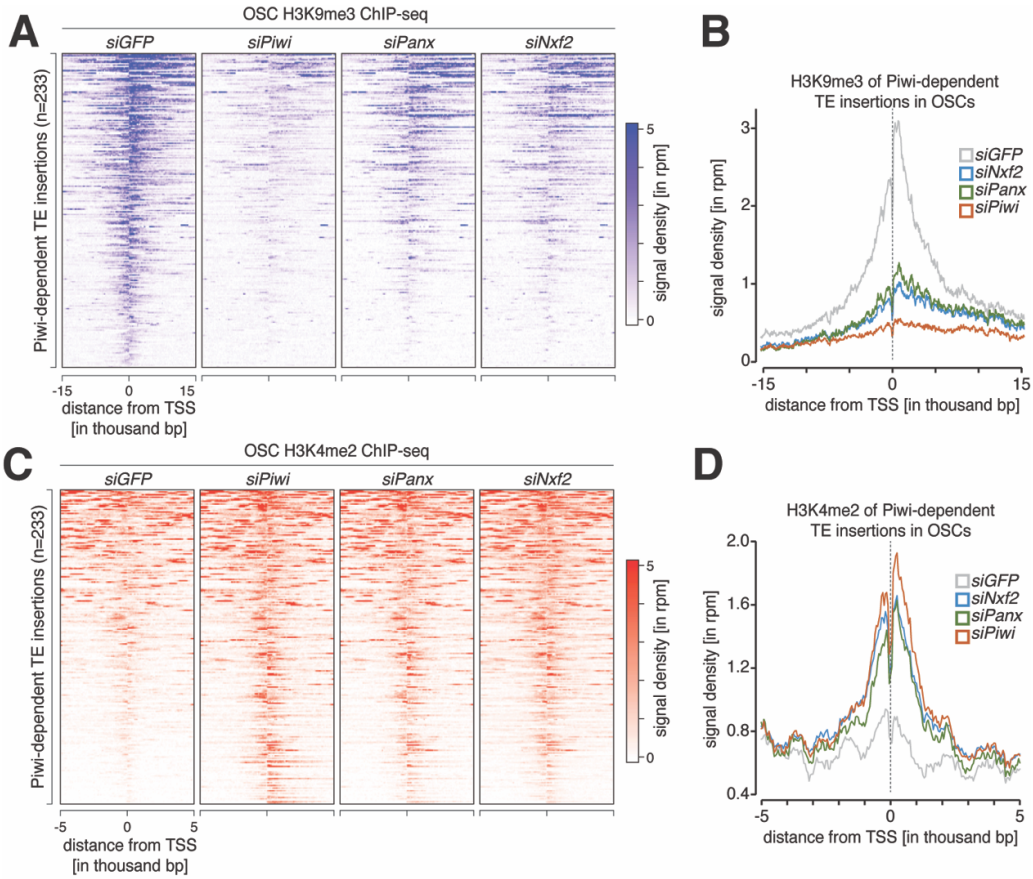

Figure 3.5: Global change of epigenetic state upon loss of Nxf2

A) Heat maps showing H3K9me3 levels calculated from 233 euchromatic and Piwi-dependent transposon insertions in the indicated knockdowns in OSCs (sorted for decreasing intensity in *siGFP*). Genomic regions flanking the transposon insertions (15-kb upstream and downstream) are shown. **B**) Shown are metaprofiles of the average H3K9me3 signals 15-kb up- and downstream of 233 euchromatic transposon insertions in *gfp*, *piwi*, *panx* and *nxf2* knockdowns shown in A). **C**) Same as A) but displaying H3K4me2 signal for 5-kb up and downstream of TE insertions. **D**) Same as A) but displaying metaplots for H3K4me2 signal 5-kb up and downstream of TE insertions.

The depletion of piRNA pathway factors does not only affect the regulation of targeted transposon insertions in the genome of OSCs but can also have an effect on transcription of nearby genes. This is believed to be a consequence of loss of H3K9me3 marks that spread into regulatory regions such as promoters of nearby genes. I was able to find transcriptional deregulation of several genes that were in close proximity to piRNA pathway-dependent TE insertions. The gene *expanded* (*ex*) harbours a sense *gypsy* insertion in its first intron in OSCs. Due to the piRNA pathway-

dependent deposition of H3K9me3 and the subsequent shutdown of transcription, the mRNA expression of the gene is impaired. Loss of Piwi resulted in upregulation of *ex* and removal of H3K9me3 marks surrounding the *gypsy* insertion, as previously reported (Sienski et al., 2015). Interestingly, H3K9me3 signal predominantly trails in the transcriptional direction of transposons (5' to 3'), thereby producing an orientation-dependent epigenetic signature, characteristic of piRNA pathway-controlled TEs. Additionally, H3K4me2 signal increased at the transcription start site (TSS). Similar results were obtained for cells depleted of Panx or Nxf2 (Figure 3.6).

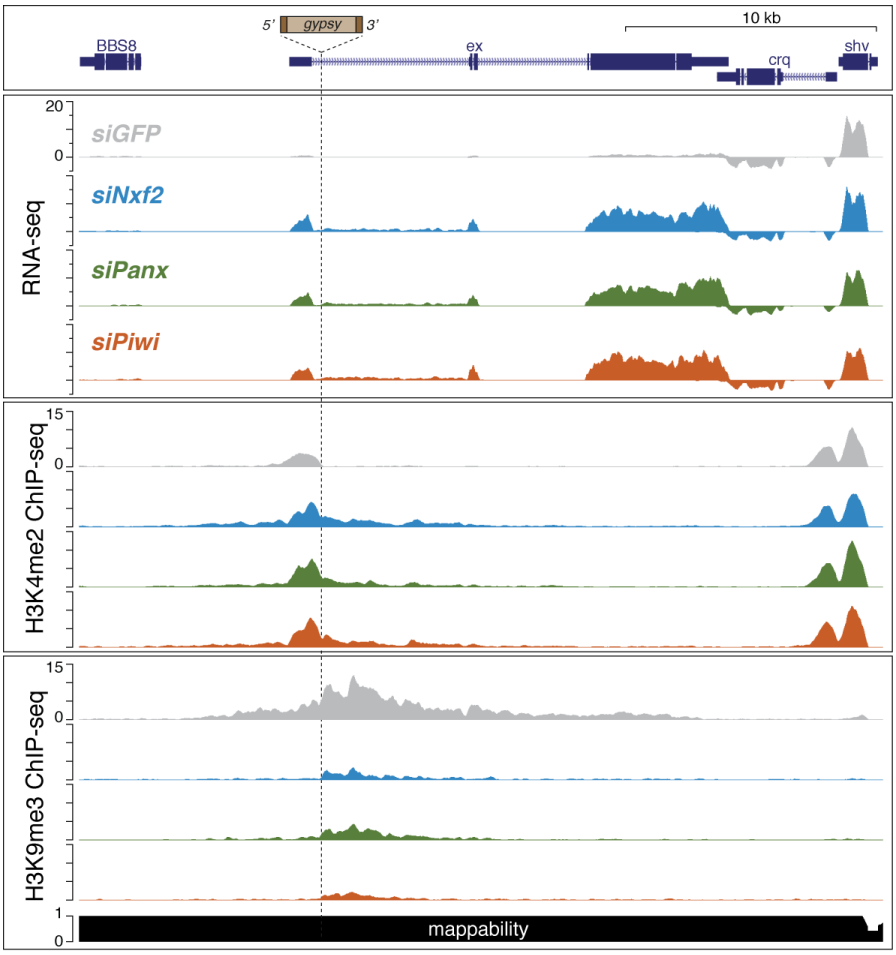

Figure 3.6: Loss of Nxf2 leads to deregulation of genes in close proximity of transposon insertions

UCSC genome browser snapshot displaying profiles of RNA-Seq levels and the density of H3K4me2 and H3K9me3 chromatin marks in OSCs upon the indicated knockdowns. Shown is an euchromatic *gypsy* insertion located within an intron of the 5' UTR of the gene *ex,* located on chromosome 2L. Signal intensity is depicted in RPM.

3.3.2 PICTS complex assembly is required for transcriptional gene silencing

The results presented above strongly support the hypothesis that both, Panx and Nxf2, are necessary to induce transcriptional gene silencing *in vivo* and *in vitro*. However, whether binding of Panx to Nxf2 and Nxt1 is necessary to silence transposon insertions remained unclear. Therefore, I further characterised the interactions between the three proteins.

First, I wanted to identify the domains of Panx and Nxf2 that are responsible for their interaction. In order to study the direct interaction between Nxf2 and Panx, I overexpressed full-length Nxf2 or full-length Panx with truncated or mutated versions of the respective binding protein in S2 cells that lack a functional piRNA pathway. Co-immunoprecipitation (coIP) followed by western blot analysis was used to probe potential interactions between the full-length protein and the truncated binding partner. To quantify binding efficiency, we calculated a coIP index (Figure 3.7 C). Additionally, the mutated versions of Panx or Nxf2 were tested for their ability to induce TGS in OSCs in a rescue approach. To do so, OSCs were depleted of endogenous Nxf2 or Panx by siRNA-mediated knockdowns followed by reintroduction of siRNA-resistant mutant versions of the respective protein by transfection. The transposon repression ability of the mutated proteins was assessed by qRT-PCR measuring the expression of the transposon *mdg1*, which is sensitive to piRNA pathway disturbances in OSCs (Figure 3.3, 3.4, 3.6). We also probed for the ability of combinations of mutated and full-length Nxf2 and Panx pairs to localise to the nucleus by immunofluorescence in S2 cells.

Panx harbours a largely unstructured amino-terminal region with two coiled-coil domains (CC1 and CC2) and a carboxy-terminal domain (CTD) (Figure 3.7 A). In order to understand binding of Panx to Nxf2, we generated hemagglutinin (HA)-tagged Panx mutants that either contained the unstructured region (HA-Panx-ΔC) or the carboxy-terminal domain of full-length Panx (HA-Panx-ΔN) (Figure 3.7 A). I expressed full-length Nxf2 with an amino-terminal 3xFLAG-tag along with HA-ZsGreen as a negative binding control in S2 cells and performed co-immunoprecipitation with antibodies

raised against FLAG epitopes. As expected, ZsGreen did not co-precipitate with Nxf2. Expression of full-length Panx, however, showed strong co-precipitation with Nxf2 (Figure 3.7 B, D) confirming previous findings by mass spectrometry (Fabry et al., 2019). However, Nxf2 failed to co-precipitate with the amino-terminal region of Panx (Panx-ΔC), while the carboxy-terminal region including the second coil-coil domain (CC2) (Panx-ΔN) showed strong co-precipitation with full-length Nxf2. Removing only Panx's CTR (Panx-ΔCTR) abolished binding to Nxf2, while removing CC1 (PanxΔCC1) only very mildly affected binding. Interestingly, removing CC2 (PanxΔCC2) as well as the unstructured region between CC1 and CTR reduced binding strongly (Figure 3.7 B, D). This either suggests that the interaction between Nxf2 and Panx is dependent on the carboxy-terminal domain of Panx with the 41 amino acids between CC2 and CTR being crucial for binding or multiple regions contributing to the interaction. The amino-terminal domain of the protein is dispensable for binding. Neither Panx domain mutant was able to rescue transposon silencing in Panx depleted OSC except for PanxΔCC1 (Figure 3.7 E). Panx-ΔN was not able to rescue transposon silencing, suggesting that binding of the carboxy-terminal region of Panx to Nxf2 alone is insufficient for the function of the PICTS complex.

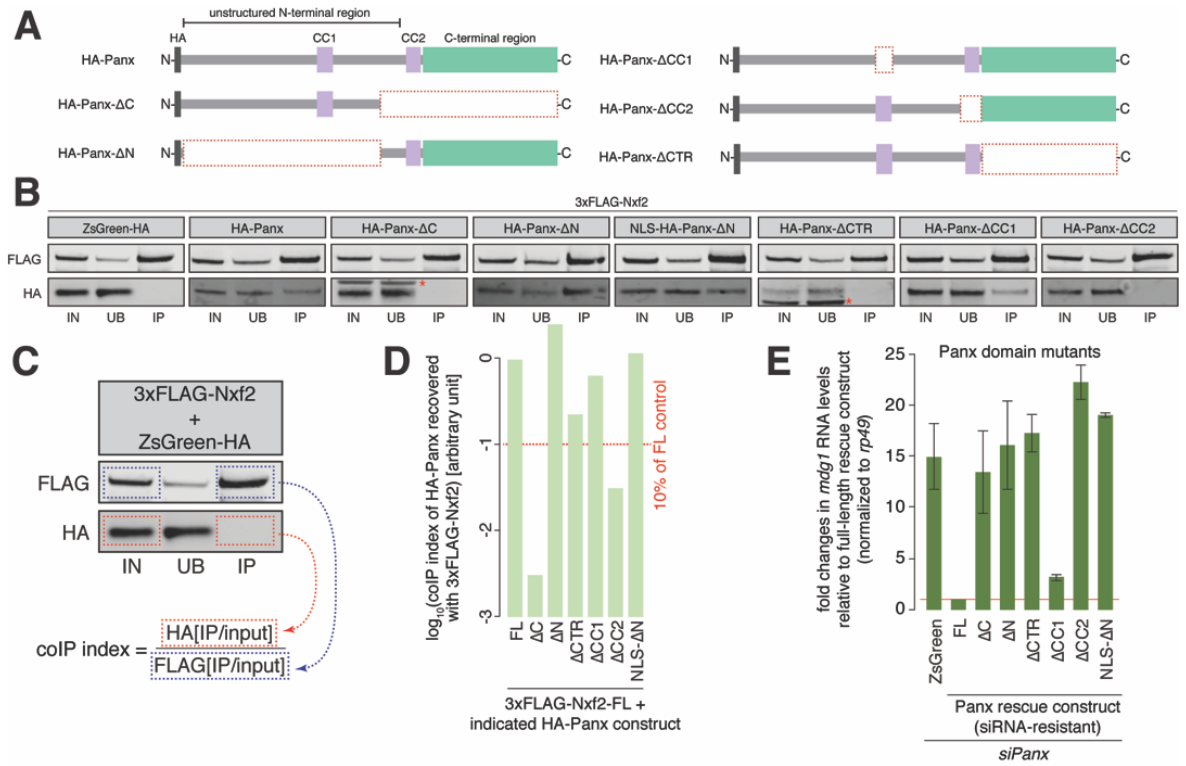

Figure 3.7: Panx interacts with Nxf2 via its carboxy-terminal region

A) Cartoons displaying the Panx protein structure and domain mutants used. CC, coiled coil domain; HA, Hemagglutinin tag. **B)** Western blot analyses of FLAG-tag co-immunoprecipitation from lysates of S2 cells transfected with the indicated expression constructs (IN, input; UB, unbound; IP, immunoprecipitate). **C**) Representative western blot displaying calculation of Co-immunoprecipitation (coIP) index from HA IP and input signals as well as FLAG IP and input signals. Co-immunoprecipitation (coIP) index was calculated from HA[IP/input] over FLAG[IP/input]. **D**) Bar graphs showing quantification of B). **E**) Bar graphs showing fold changes in steady-state RNA levels of *mdg1* in total RNA from OSCs transfected with siRNAs against Panx and the indicated siRNA-resistant expression constructs relative to full-length (FL) rescue construct and normalised to *rp49*. Error bars indicate standard deviation (n= 2).

We next probed the ability of Panx mutants to localise to the nucleus. Overexpression of full-length Nxf2 and Panx resulted in localisation of both proteins to the nucleus. Removing the carboxy-terminal region of Panx (Panx-ΔC) had no effect on Panx localisation to the nucleus. However, full-length Nxf2 was delocalised from the nucleus and became mainly cytoplasmic (Figure 3.8). Given that Panx-ΔC is unable to bind to Nxf2, this observation suggests that import of Nxf2 is dependent on

interaction with Panx. Deletion of the amino-terminal region of Panx (Panx-ΔN) resulted in cytoplasmic localisation of Nxf2 and Panx indicating a nuclear localisation sequence (NLS) present in the amino-terminal region of Panx. Binding to Nxf2, however, was not affected (Figure 3.8). As expected, Panx-ΔN was unable to rescue *mdg1* expression in OSCs depleted of Panx, probably due to the delocalisation from the nucleus, where the PICTS complex mediates its function. Forcing nuclear localisation of Panx-ΔN using a SV40-derived NLS (NLS-Panx-ΔN) lead to translocation of both NLS-Panx-ΔN and Nxf2 to the nucleus (Figure 3.8). However, silencing ability was not restored by artificial nuclear transport of Panx-ΔN (Figure 3.7 E).

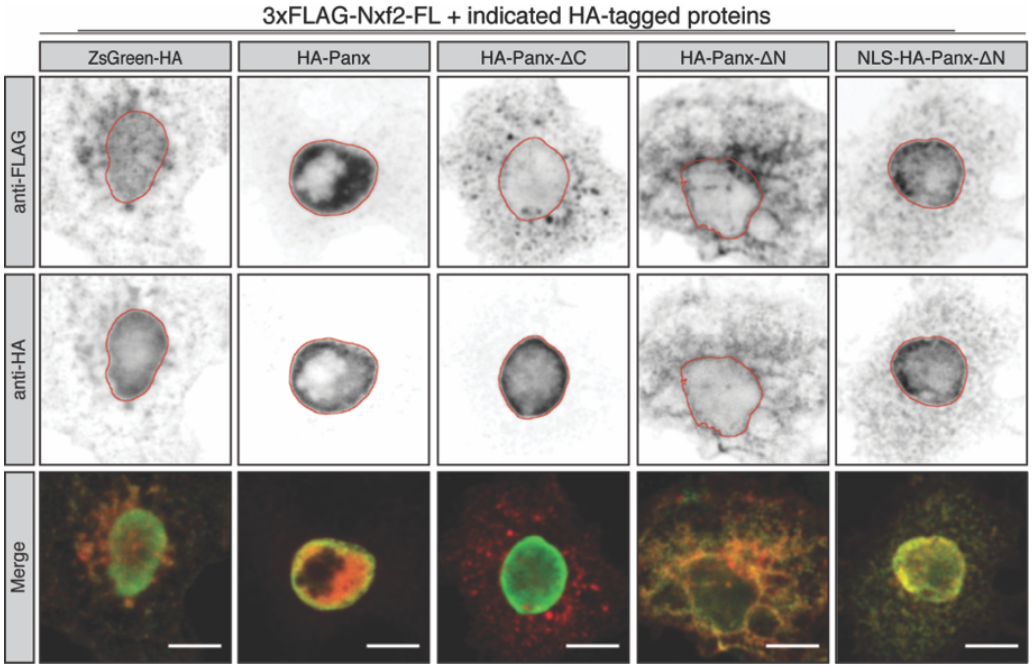

Figure 3.8: Nxf2 depends on Panx for its nuclear localisation

Expression and localisation of the indicated Nxf2 (3xFLAG-tagged, top, shown in red in the merge) and Panx (HA-tagged, bottom, shown in green in the merge) expression constructs in transfected S2 cells are shown by immunofluorescence. Lamin staining (red lines) was used to draw the outline of the nuclear envelope. Scale bar, 5 μm.

The amino-terminus of Nxf2 is comprised of an amino-terminal region (NTR) followed by leucin-rich repeats (LLR1-3), an RNA-binding domain (RBD) and additional leucin-rich repeats (LLR4-6). The amino-terminus is associated with cargo

binding in Nxf2's homologue Nxf1. At its carboxy-terminal region, an NTF2-like domain (NTF2) and a ubiquitin associated (UBA) domain can be found, which is believed to mediate nuclear pore binding of Nxf1 (Figure 3.9 A).

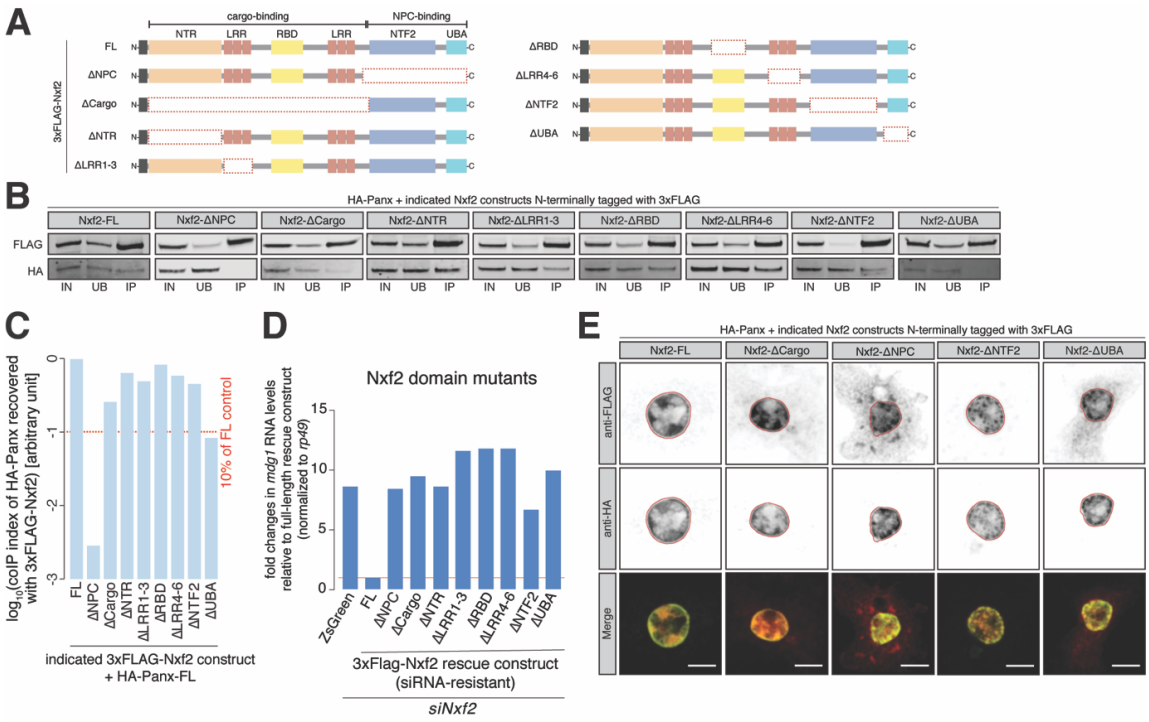

Figure 3.9: Requirements for the formation and function of the PICTS complex

A) Cartoons displaying the Nxf2 protein structure and domain mutants used. NTR, amino-terminal region; LRR, leucine rich repeats; RBD, RNA-binding domain; NTF2, NTF2-like domain; UBA, Ubiquitin associated domain. **B**) Western blot analyses of FLAG-tag co-immunoprecipitation from lysates of S2 cells transfected with the indicated expression constructs (IN, input; UB, unbound; IP, immunoprecipitate). **C**) Bar graphs showing quantification of B). **D**) Bar graphs showing fold changes in steady-state RNA levels of *mdg1* in total RNA from OSCs transfected with siRNAs against Nxf2 and the indicated siRNA-resistant expression constructs relative to full-length (FL) rescue construct and normalised to *rp49*; n= 1. **E**) Expression and localisation of the indicated Nxf2 domain mutant (3xFLAG-tagged, top, shown in red in the merge) and full-length Panx (HA-tagged, bottom, shown in green in the merge) expression constructs in transfected S2 cells are shown by immunofluorescence. Lamin staining (red lines) was used to draw the outline of the nuclear envelope. Scale bar, 5 μm.

Next, I investigated the dependency for protein domains of Nxf2 on binding to full-length Panx. Nxf2 was split in two parts, the cargo binding domain (Nxf2-ΔNPC) and the nuclear pore complex binding domain (Nxf2-ΔCargo) (Figure 3.9 A). The cargo binding domain of Nxf2 (Nxf2-ΔNPC) failed to bind full-length Panx, while Nxf2-ΔCargo was able to bind, indicating that the NPC-binding region is facilitating this interaction (Figure 3.9 B, C). Again, neither Nxf2-ΔNPC or Nxf2-ΔCargo was able to rescue *mdg1* silencing in OSCs depleted of endogenous Nxf2 (Figure 3.9 D). Together with the results described above, the data suggest a direct interaction between the carboxy-terminal region of Panx and the NPC domain of Nxf2, while the amino-terminal part of Panx and the cargo binding domain of Nxf2 are dispensable for binding.

We next systematically deleted individual domains of Nxf2 to narrow down the Panx binding site in Nxf2 (Figure 3.9 A). Deleting LRR1-3 (Nxf2-ΔLRR1-3) or LRR4-6 (Nxf2-ΔLRR4-6) had no effect on the ability of Nxf2 to bind Panx. Similarly, deleting the RBD (Nxf2-ΔRBD) or NTF2 (Nxf2-ΔNTF2) domain had no effect on binding to full-length Panx. Removing the UBA domain (Nxf2-ΔUBA), however, abolished binding to Panx, suggesting a direct interaction of Panx with the UBA domain of Nxf2.

We further narrowed down the interaction of Nxf2 with Panx by introducing point mutations in the UBA domain of Nxf2. Changing four amino acids (aa811-814$^{IVEE->KRGG}$, Nxf2 UBA mut1) (Figure 3.10 A), reduced binding to Panx strongly (<10% compared to full-length Nxf2). An independent mutation further downstream (Nxf2 UBA mut2) had only a weak effect on binding (Figure 3.10 B, C). Nxf2 UBA mut1 was not able to repress *mdg1* expression, while Nxf2 UBA mut2 rescued silencing (Figure 3.10 D).

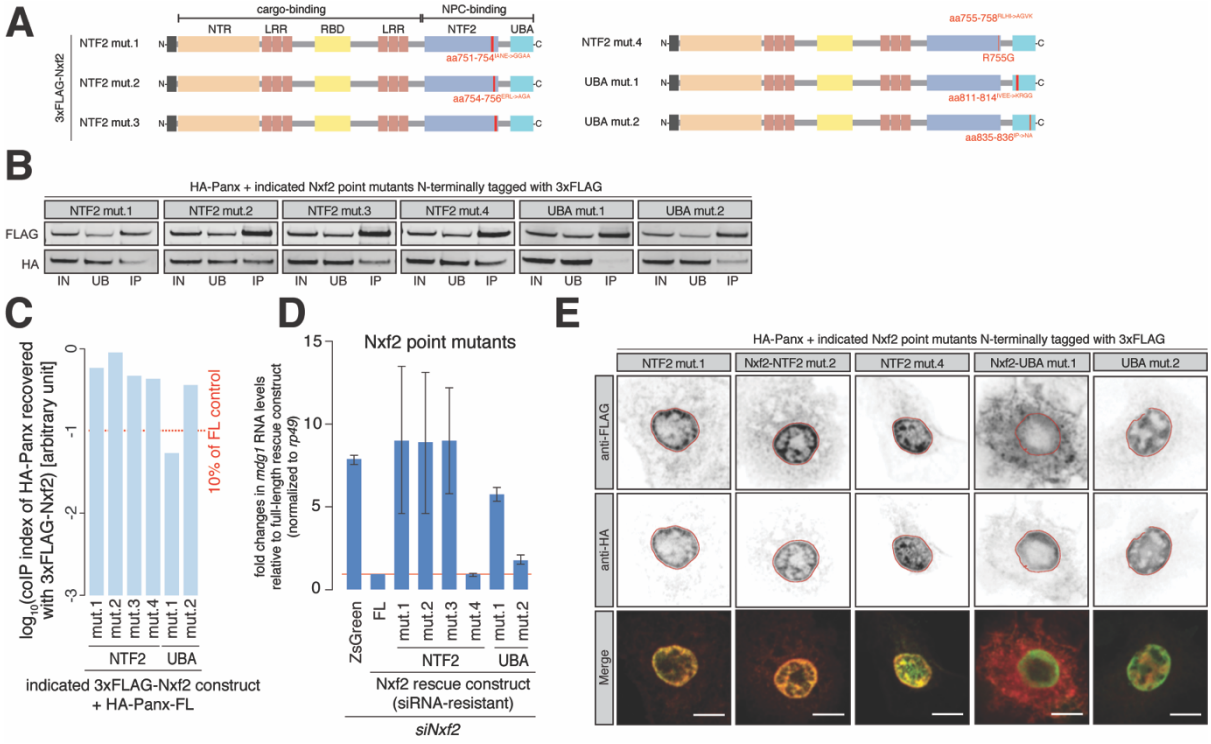

Figure 3.10: Nxf2 UBA domain is crucial for Panx binding and PICTS function

A) Cartoon displaying altered residues in Nxf2 domains used for experiments in B-E **B**) Western blot analyses of FLAG-tag co-immunoprecipitation from lysates of S2 cells transfected with the indicated expression constructs (IN, input; UB, unbound; IP, immunoprecipitate). **C**) Bar graphs showing quantification of B). **D**) Bar graphs showing fold changes in steady-state RNA levels of the *mdg1* transposon in total RNA from OSCs transfected with siRNAs against Nxf2 and the indicated siRNA-resistant expression constructs relative to full-length (FL) rescue construct and normalised to *rp49*; n= 2. **E**) Expression and localisation of the indicated Nxf2 domain mutant (3xFLAG-tagged, top, shown in red in the merge) and full-length Panx (HA-tagged, bottom, shown in green in the merge) expression constructs in transfected S2 cells are shown by immunofluorescence. Lamin staining (red lines) was used to draw the outline of the nuclear envelope. Scale bar = 5 μm.

Nxf2 and Panx binding is dependent on specific domains. However, whether their binding is dependent on Nxt1, and if this binding is necessary for transcriptional gene silencing, remains unknown. Therefore, I studied Nxt1 binding in context of the Nxf2-Panx complex. Previous studies examining the binding of Nxf1 and Nxt1 in various animals suggest that the NTF2 domain is important for the assembly of the Nxf1-Nxt1 complex (Herold et al., 2000; Kerkow et al., 2012; Suyama et al., 2000). By

analysing the indicated binding site of Nxf1 and Nxt1, I generated various small mutations (2-4 amino acids) in the NTF2 domain of Nxf2 altering highly conserved amino acids in order to impair Nxt1 binding to the Nxf2-Panx complex. NTF2 mutant 1-3 were not able to rescue *mdg1* silencing, while mutant 4 showed levels of *mdg1* repression similar to the full-length control (Figure 3.10 D). CoIP for Nxf2 NTF2 mutants confirmed that binding to full-length Panx was only mildly affected by either mutant. However, binding to endogenously expressed Nxt1 was abolished in NTF2 mutant 2 (Figure 3.11 A). This result was confirmed by testing the coIP efficiency of FL-Nxf2 with HA-tagged Panx and HA-tagged Nxt1. Again, mutation of the NTF2 domain (mutant 1) abolished binding to Nxt1, while Panx-Nxf2 binding was not impaired (Figure 3.11 B). NTF2 domain mutant 4, however, showed no reduction in Nxt1 binding, in accordance with its silencing ability by rescue experiments (Figure 3.10 D).

These data together with results from rescue experiments suggests that binding of Panx and Nxf2 is independent of Nxt1. However, Nxf2 binds to Nxt1 via its NTF2 domain and Nxt1 binding is indispensable for induction of transcriptional gene silencing at transposon insertions.

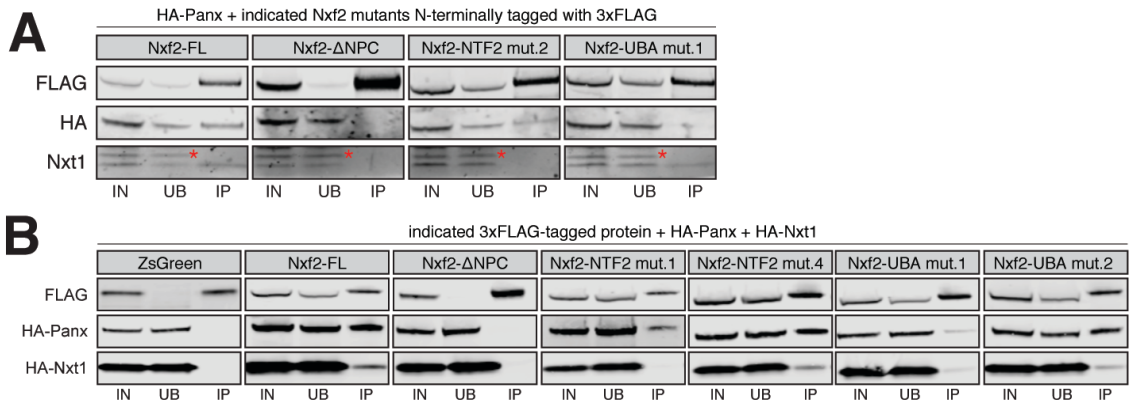

Figure 3.11: The NTF2 fold of Nxf2 stabilises Nxt1 binding

A) Western blot analyses of FLAG-tag co-immunoprecipitation from lysates of S2 cells transfected with the indicated expression constructs (IN, input; UB, unbound; IP, immunoprecipitate). Asterisks indicate an unspecific band from anti-Nxt1 antibody. **B**) As in A) but showing co-immunoprecipitation of HA-Panx and HA-Nxt1 recovered with the indicated Nxf2 expression constructs.

3.3.3 PICTS induces TGS through epigenetic changes of chromatin states

Previous studies showed that artificial tethering of Panx to nascent RNA and DNA induced transcriptional gene silencing of a reporter gene (Sienski et al., 2015; Yu et al., 2015). In order to investigate if components of the PICTS complex are able to induce TGS, we established an artificial DNA and RNA tethering reporter system in OSCs.

Tethering of candidate proteins directly to DNA was achieved by transiently transfecting a sensor construct comprised of eight LacO sites followed by the ubiquitin promoter of *Drosophila simulans* controlling downstream expression of a ZsGreen cassette (Figure 3.12 A). The protein of interest (POI) tagged with LacI, which is able to specifically bind LacO sites (Robinett et al., 1996; Straight et al., 1996), was introduced by transfection together with the sensor construct. Consistent expression of ZsGreen as well as the tethering construct was verified by western blot (Figure 3.12 B).

Tethering of Piwi to DNA has previously been shown to be insufficient to induce TGS at sensor sequences integrated in the *Drosophila* genome, probably due to the missing conformational change of Piwi caused by piRNA-target engagement (Figure 3.13 C, D) (Sienski et al., 2015; Yu et al., 2015). Overexpression of LacI-Piwi had a weak impact on mRNA and protein output of ZsGreen as measured by qRT-PCR and western blot analysis in comparison to cells transfected with LacI-Renilla as a negative control. Overexpression of LacI-Panx however, reduced expression of ZsGreen mRNA and protein levels strongly, thus confirming previous studies. Tethering of Nxf2 to DNA resulted in strong reduction of ZsGreen transcription and protein abundance comparable to Panx tethering. However, tethering of Nxt1 did not change expression or protein abundance of ZsGreen indicating that Nxt1 on its own is not sufficient to silence transcription when tethered directly to DNA.

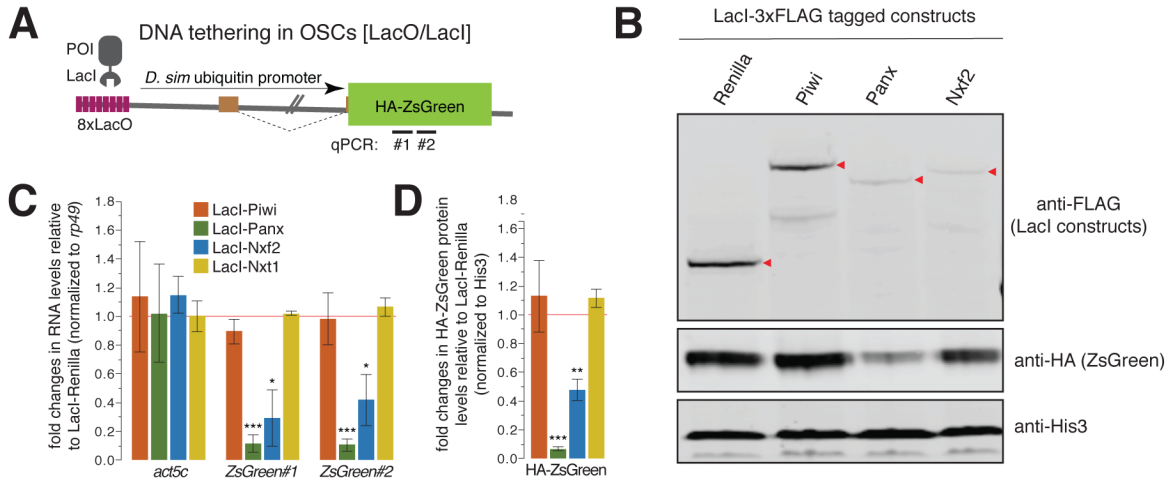

Figure 3.12: DNA tethering of PICTS complex components silences sensor expression

A) Cartoon displaying the DNA tethering sensor construct used in OSCs. The construct was transiently transfected and features eight consecutive LacO sites followed by the *Drosophila simulans ubiquitin* promoter (which contains an intron), the ZsGreen coding sequence (fused to the HA-tag and a nuclear localisation sequence). Amplicons for qRT-PCR are indicated. **B)** Representative western blot analyses of LacI-3xFLAG-fusion constructs used for DNA tethering in the DNA HA-ZsGreen sensor and a loading control (His3) from lysates of OSCs transfected with the indicated constructs, used for quantification of DNA tethering in D). Asterisks indicate calculated size of LacI fusion protein **C)** Bar graphs showing fold changes in steady-state RNA levels of the DNA sensor and *act5c* in total RNA from OSCs transfected with the indicated LacI-fusion expression constructs (relative to a LacI-Renilla construct and normalised to *rp49*), * denotes p-value <0.01, ** denotes p-value <0.001; *** denotes p-value <0.0001 (unpaired t-test). Error bars indicate standard deviation (n= 3). **D)** Bar graphs showing fold changes in protein levels of HA-ZsGreen in lysates from OSCs transfected with the indicated LacI expression constructs (relative to a LacI-Renilla construct and normalised to His3), ** denotes p-value <0.001; *** denotes p-value <0.0001 (unpaired t-test). Error bars indicate standard deviation (n= 3).

A hallmark of transcriptional gene silencing is the deposition of H3K9me3 marks at targeted sites. Therefore, I evaluated the epigenetic state of cells transfected with the DNA sensor and a protein of interest fused to LacI using ChIP-Seq for H3K9me3 epitopes. I then mapped ChIP-Seq reads back to the sequence of the sensor plasmid. Using the *Drosophila simulans* promoter enabled me to characterise

the epigenetic state of most of the sensor construct by uniquely mapping ChIP-Seq reads (50-bp) back to the reporter (see mappability track, Figure 3.13).

Tethering of Renilla or Piwi showed no deposition of H3K9me3 marks at the sensor. HP1a is known to bind and spread H3K9me3 marks. Therefore, I included LacI-HP1a as a positive control in my ChIP-Seq experiments. Tethering of HP1a lead to a strong increase of H3K9me3 signal at the sensor. Interestingly, H3K9me3 signal was highest at the gene sequence coding for ZsGreen. Similarly, LacI-Nxf2 tethering showed equally strong deposition of H3K9me3 marks, while tethering Panx led to a strong increase of signal to higher levels than either Nxf2 or HP1a (Figure 3.13 A).

While TGS is accompanied by the deposition of H3K9me3 chromatin marks, it also impacts histone marks associated with active transcription such as H3K4me2. Therefore, I performed ChIP-Seq with chromatin lysate from the same cells used above, probing for H3K4me2 epitopes. Cells transfected with either LacI-Renilla or LacI-Piwi showed strong H3K4me2 signal at transcription start sites (TSS). Overexpression of LacI-HP1a and LacI-Nxf2 resulted in a strong reduction of H3K4me2 signal, recapitulating the results presented above. Tethering Panx to DNA had the strongest effect on H3K4me2 signal and signal intensity was severely decreased in comparison to LacI-Renilla cells (Figure 3.13 B).

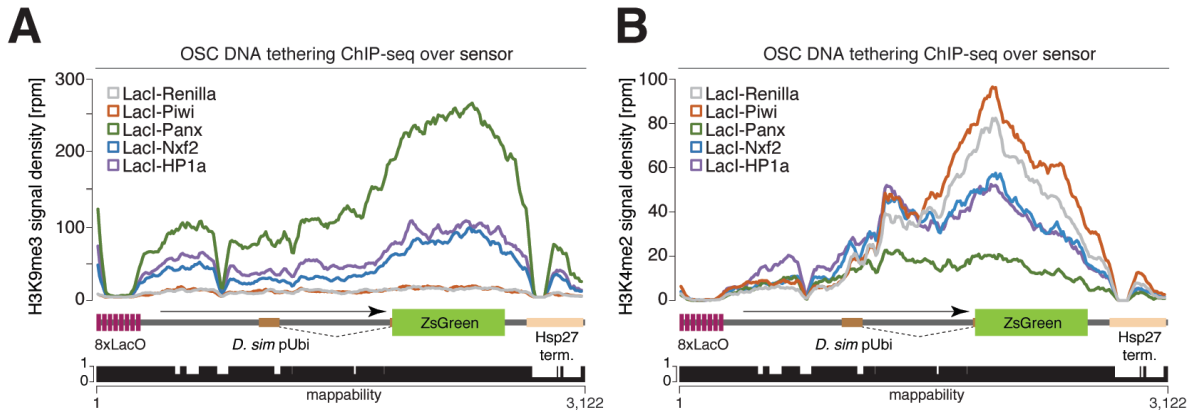

Figure 3.13: Recruitment of the PICTS complex to DNA results in epigenetic changes on chromatin level

A) Density profiles of normalised reads from H3K9me3 ChIP-Seq experiments mapping to the DNA tethering reporter as indicated in the cartoon below. The mappability of 50-bp reads is shown below. **B)** same as in A) but results of H3K4me2 signal mapped to DNA sensor sequence.

We next inquired whether tethering of the PICTS complex directly to nascent RNA had a similar impact on the deposition of repressive chromatin marks by establishing an RNA-tethering sensor system comprised of the identical *D. simulans ubiquitin* promoter followed by the *ZsGreen* expression cassette used in our DNA-tethering approach. Nine consecutive BoxB sites were introduced in the 3' UTR of *ZsGreen*. BoxB sites fold into characteristic hairpins following transcription and are specifically bound by the λN protein derived from the bacteriophage *lambda* (Baron-Benhamou et al., 2004; Keryer-Bibens et al., 2008). The *Drosophila simulans ubiquitin* promoter contains an intron, which is spliced out during mRNA maturation (Figure 3.14 A). This enabled us to probe for unspliced/nascent transcripts. We stably integrated this sensor into OSCs by co-transfection with a puromycin resistance plasmid. Cells were selected with puromycin and our proteins of interest fused to λN were introduced in the stable sensor cell line by transient transfection. qRT-PCR probing for ZsGreen mRNA production revealed similar findings in comparison to DNA tethering. Again, tethering Piwi had no effect on mRNA output. However, tethering Panx or Nxf2 directly to RNA resulted in a significant decrease of nascent RNA levels, while Nxf2 tethering showed stronger silencing abilities than Panx, in contrast with the results of the DNA tethering experiment for both factors (compare Figure 3.13 A). Interestingly, tethering Nxt1 to RNA was able to induce strong silencing comparable to Nxf2, while DNA tethering showed no indication of silencing. Western blot analysis reflected the qRT-PCR results (Figure 3.14 B).

Next, I performed H3K9me3 ChIP-Seq from cells transfected with several POI fusions for the RNA tethering sensor. Tethering Piwi directly to RNA had little effect on H3K9me3 signal in comparison to λN-Renilla control. However, tethering of Nxf2 or Panx to RNA resulted in strong deposition of H3K9me3 at the sensor construct. Interestingly, again, tethering Nxt1 to RNA showed strongest deposition of H3K9me3 marks, while it had no effect on RNA or protein output in DNA tethering experiments (Figure 3.14 C).

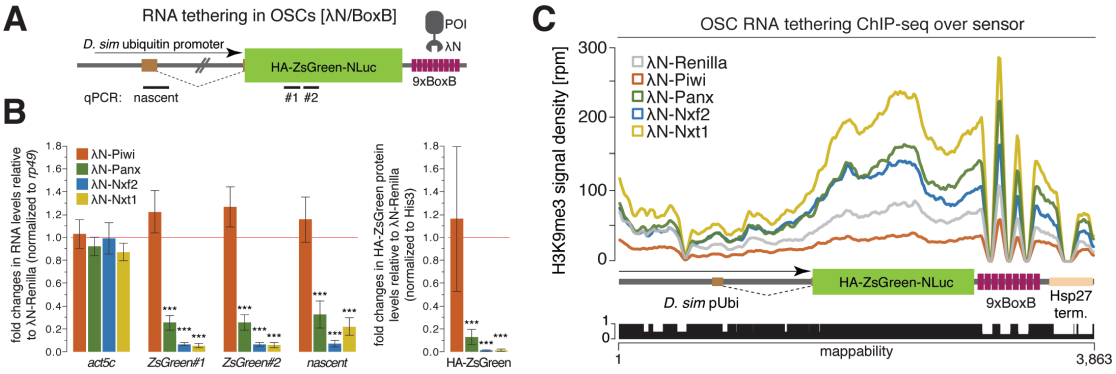

Figure 3.14: Recruitment of PICTS components to nascent RNA results in chromatin silencing

A) Cartoon displaying the RNA tethering sensor construct used in OSCs. The construct was stably integrated and features the *Drosophila simulans* ubiquitin promoter (which contains an intron), the ZsGreen coding sequence (fused to the HA-tag, a nuclear localisation sequence and the NLuc luciferase), and a 3' UTR containing nine BoxB sites. Amplicons for qRT-PCR are indicated. **B)** Left: Bar graphs showing fold changes in steady-state RNA levels of the sensor and *act5c* in total RNA from OSCs transfected with the indicated λN-fusion expression constructs (relative to a λN-Renilla construct and normalized to *rp49*). Right: Bar graphs showing fold changes in protein levels of HA-ZsGreen in lysates from OSCs transfected with the indicated λN expression constructs (relative to a λN-Renilla construct and normalized to His3). *** denotes p-value < 0.0001 (unpaired t-test). Error bars indicate standard deviation (n= 4). **C)** Density profiles of normalised reads from H3K9me3 ChIP-Seq experiments mapping to the tethering reporter as indicated in the cartoon below. The mappability of reads is shown below.

The results presented above suggest that all PICTS components are able to induce transcriptional gene silencing. However, whether those individually tethered factors work alone or are in complex with full-length proteins is unclear. Therefore, I used the DNA tethering system described above to probe domain mutants of Panx and Nxf2. Panx-ΔN is still able to bind to Nxf2. However, tethering failed to affect RNA output or protein levels of ZsGreen. Nxf2-ΔCargo is not able to bind full-length Panx and tethering also failed to induce TGS. Tethering the amino-terminal region of Panx (Panx-ΔC), however, which is unable to bind Nxf2, was able and sufficient to induce TGS at levels comparable to a LacI-HP1a control suggesting that the amino-terminal region alone is responsible for the induction of silencing (Figure 3.15 A, B).

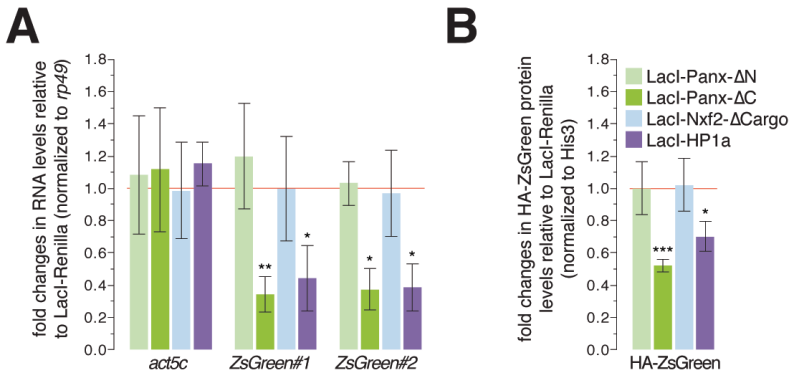

Figure 3.15: Tethering of Panx amino-terminus is sufficient to induce transcriptional gene silencing

A) Quantification of qRT-PCR results for indicated DNA tethering constructs transfected together with the DNA sensor in OSCs normalised to *rp49* (* denotes p-value < 0.01; ** denotes p-value <0.001 unpaired t-test). Error bars indicate standard deviation (n= 3). **B)** Quantification of protein abundance of same experiment as A) relative to His3 expression.

3.3.4 Nxf2 or Panx depletion affects transposon mRNA localisation

Nxf1 is known to affect mRNA localisation through mediating export to the cytoplasm. In order to understand whether Nxf2 is involved in the localisation of transposon mRNA, I probed TE RNA distribution by fluorescence *in situ* hybridisation (FISH) in OSCs.

The transposon *mdg1* was one of the most de-repressed TEs upon disruption of the piRNA pathway (Figure 3.3, 1.4). I was unable to detect *mdg1* signal in OSC treated with *siGFP,* confirming low expression measurements obtained in our RNA-Seq data. Knockdown of Piwi, however, resulted in strong signal accumulation. 52.2% of signal was detected in the nucleus, while 47.8% was cytoplasmic. *mdg1* RNA localisation was significantly different in OSCs depleted of either Panx or Nxf2 in comparison to Piwi. Strong signal was detected in the nucleus but little signal was recovered from the cytoplasm. Nuclear signal of *siPanx* or *siNxf2* treated cells was 79.0% and 84.9% of the total signal respectively (Figure 3.16 A, B). Additionally, *siPanx* or *siNxf2* treated cells showed distinct nuclear foci indicating accumulation of nascent transcripts (Figure 3.16 A). One-way ANOVA testing showed that the mean of the nuclear fraction was significantly different between the three tested siRNAs

targeting Piwi, Panx or Nxf2 (F=185.75, df=2,285, p=<2.2e-16). A post hoc Tukey test further revealed significant changes of the nuclear signal fraction between *siPiwi* and *siPanx* (p=4.84e-13) or *siPiwi* and *siNxf2* (p=4.84e-13) while there was also a significant difference between *siPanx* and *siNxf2* (p=0.00198) (Figure 3.16 C).

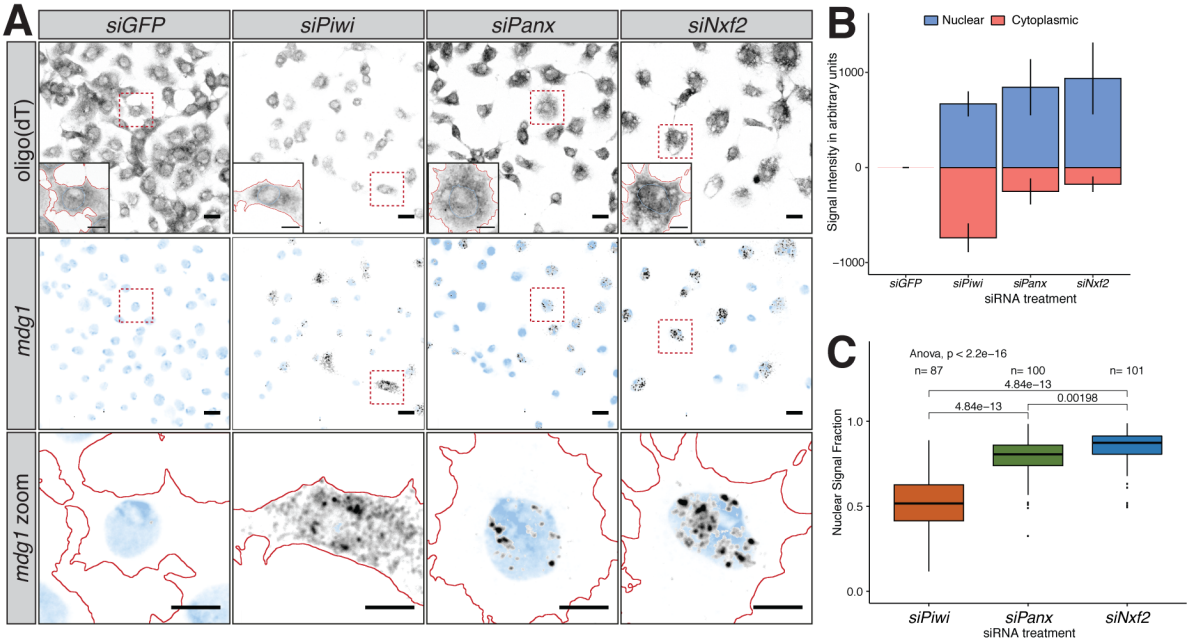

Figure 3.16: Depletion of the PICTS complex retains *mdg1* transcripts in the nucleus

A) Representative image showing FISH signal for total mRNAs (oligo(dT)) used to draw cell outlines and *mdg1* transcripts in OSCs treated with indicated siRNAs. Blue, DNA; red, cell outline; grey/black, *mdg1* signal (scale bar 10*µ*m, in zoomed images 5*µ*m). **B)** Quantification of A) showing *mdg1* signal intensity in cytoplasm and nucleus (n= 4 biological replicates, error bars showing standard deviation). **C)** Bar graph showing the mean nuclear/cytoplasmic fraction of *mdg1* signal for the indicated number of measured cells from 4 biological replicates (error bar showing standard deviation; one-way ANOVA (all conditions) and Post hoc Tukey (individual conditions) testing with indicated p-values).

Next, I tested if this localising behaviour of transposon mRNA was consistent for TE families affected by piRNA pathway disruptions. *gypsy* is a commonly deregulated transposon in OSCs depleted of Piwi and showed strong upregulation in *siPanx* or *siNxf2* treated cells (Figure 3.3, 1.4). RNA-FISH signal was detectable in OCSs treated with *siGFP* using FISH probes complementary for *gypsy* transcripts. Most transcripts localised to the nucleus with only little detected in the cytoplasm.

Knockdown of Piwi increased signal strongly. However, in contrast to *mdg1* transcripts, signal was confined mostly to the nucleus (89.6%). Similarly, *gypsy* transcripts in *siPanx* or *siNxf2* treated cells were mainly nuclear with 94.2% and 93.2% respectively (Figure 3.17 A, B). Testing again using a one-way ANOVA indicated a significant difference between the means of the nuclear fraction of *siPiwi*, *siPanx* and *siNxf2* treated cells (F=9.5275, df=2,343, p=9.4e-05). Post hoc Tukey testing further revealed significant differences between *siPiwi* or *siPanx* treated cells (p=0.000338) as well as *siPiwi* or *siNxf2* treated cells (p=0.00364). There were no significant differences in the nuclear fraction mean between *siPanx* and *siNxf2* samples (p=0.706) (Figure 3.17 C).

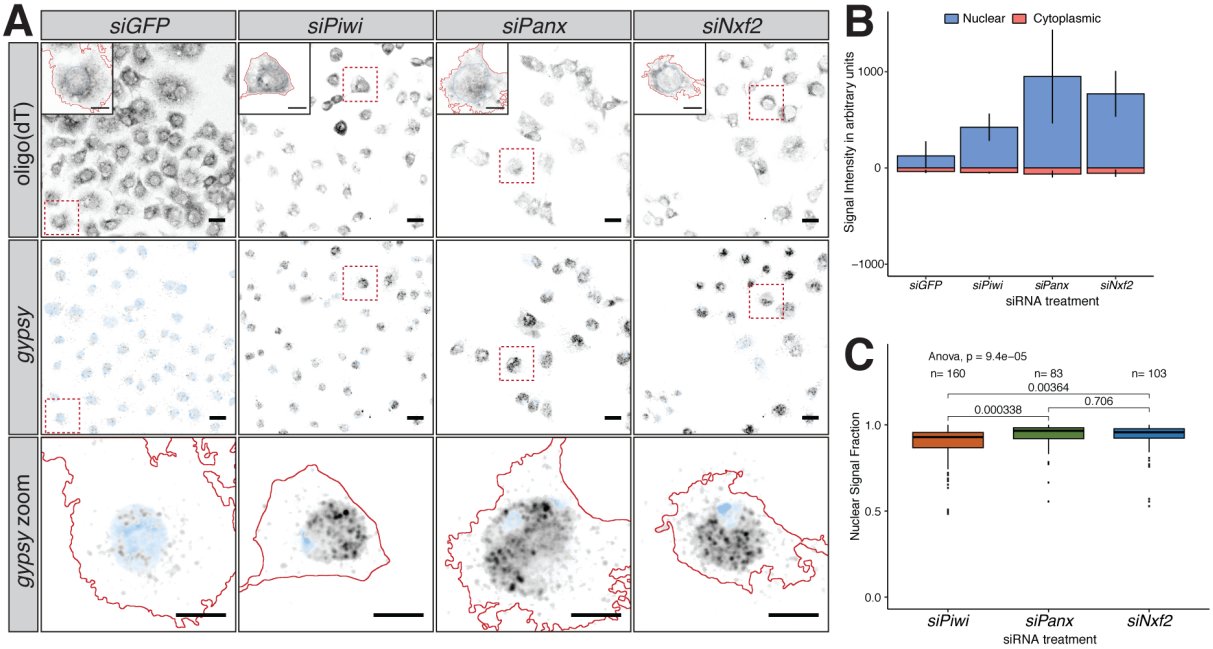

Figure 3.17: Depletion of the PICTS complex only mildly affects *gypsy* transcript localisation

A) Representative image showing FISH signal for total mRNAs (oligo(dT)) used to draw cell outlines and *gypsy* transcripts in OSCs treated with the indicated siRNAs. Blue, DNA; red, cell outline; grey/black, *gypsy* signal (scale bar 10*μ*m, in zoomed images 5*μ*m). **B)** Quantification of A) showing *gypsy* signal intensity in cytoplasm and nucleus (n= 4 biological replicates, error bars showing standard deviation). **C)** Bar graph showing the mean nuclear/cytoplasmic fraction of *gypsy* signal for indicated number of cells from 4 biological replicates (error bar showing standard deviation; one-way ANOVA (all conditions) and Post hoc Tukey (individual conditions) testing with indicated p-values).

3.4 Conclusion

Our work identified two novel factors involved in piRNA-guided transcriptional gene silencing of transposons in *Drosophila* gonads. The nuclear export family protein Nxf2 and its co-factor Nxt1 bind the previously identified piRNA pathway-dependent TGS factor Panx and form a complex we named PICTS.

We showed that loss of PICTS results in deregulation of piRNA pathway-dependent transposons *in vitro* and *in vivo* and is accompanied by loss of repressive chromatin marks. We further mapped the molecular interaction of the individual PICTS components. The UBA domain of Nxf2 interacts directly with the carboxy-terminal region of Panx. Nxt1 binds the NTF2 domain of Nxf2. Rescue experiments further revealed that proper assembly of all three components of the PICTS complex is required for transposon regulation. Panx likely harbours a nuclear localisation sequence in its amino-terminal domain, which is necessary for the translocation of the complex into the nucleus.

RNA tethering experiments suggest that all components of the PICTS complex are able to induce transcriptional gene silencing of a reporter construct. However, tethering individual components likely induces assembly of functional PICTS complexes, thereby resulting in tethering of all components. Tethering of PICTS to chromatin and nascent RNA further demonstrated the ability of the complex to induce transcriptional silencing by epigenetic conversion of targeted genomic loci. We further found that the unstructured amino-terminal domain of Panx is required and sufficient to induce TGS independent of Nxf2 or Nxt1. Therefore, the silencing ability of the PICTS complex resides in Panx.

While loss of PICTS components led to the accumulation of transposon transcripts in the nucleus, Piwi depletion resulted in the release and export of transposon RNA, suggesting a possible Piwi-dependent fail-safe mechanism for TE control in ovaries lacking the PICTS complex.

4 Conditional protein inactivation and degradation systems in *Drosophila melanogaster*

OSC rescue experiments using LEXY constructs were performed by Emily Lythgoe. The western blot experiment for degradation of GFP-Piwi in ovaries using deGradFP was performed by Federica Falconio. OsTIR1 expressing flies were generated by Dr. Ben Czech.

4.1 Introduction

piRNA-guided transcriptional gene silencing is crucial for protecting the germ line from transposon activity. This process is dependent on Piwi serving as a nuclear binding platform for the PICTS complex (see chapter 3). While this process has been studied extensively in follicle and germ cells of *Drosophila* ovaries, it remains an open question, whether this pathway is also active in other tissues or developmental time points.

Studies previously reported maternal deposition of the two PIWI-clade Argonaute proteins Piwi and Aub into embryos (Brennecke et al., 2008; Megosh et al., 2006). Inheritance of piRNAs by the future generation has been implicated in the re-establishment of dual-strand clusters and has a major impact on the ability to silence transposons in adult gonads (Akkouche et al., 2017; Brennecke et al., 2008; Le Thomas et al., 2014). Interestingly, inherited Piwi protein is not only present in germ cell progenitors but was also shown to be highly enriched in somatic nuclei of early embryos (Brennecke et al., 2008). This suggests that the piRNA pathway might have evolved a function beyond genome maintenance in gonads and acts in more tissues than previously reported. In fact, studies on other arthropods confirmed the widespread existence of somatic piRNAs even in related species such as *Drosophila virilis* (Lewis et al., 2018).

Studies of the function(s) of the piRNA pathway in early embryogenesis are, however, difficult due to limited genetic tools available. Disrupting the genomic locus of *piwi* or other piRNA pathway-specific genes in *Drosophila* female parents leads to

oogenesis defects and often results in sterility (Cox et al., 1998; Klenov et al., 2011). Similarly, removing Piwi transcripts during oogenesis using RNAi approaches such as shRNAs leads to transposon reactivation and infertility (Czech et al., 2013). piRNA pathway factor loss has been shown to impair stem cell maintenance as well as embryonic axis specification (Cook et al., 2004; Cox et al., 2000; Klattenhoff et al., 2007). This is likely due to deregulation of transposons and DNA damage in piRNA-pathway expressing cells (Chen et al., 2007), thus complicating experiments or their interpretation in embryos derived from mutant ovaries.

Targeting maternally deposited Piwi transcripts directly in embryos is difficult due to several technical restrictions. Most maternal Piwi transcripts are likely actively translated following fertilisation. RNAi-mediated knockdowns in embryos, however, require transcription of small RNA guides, therefore relying on the transcriptionally active zygotic genome. Since the genome is mostly inactive for the first 2h of embryogenesis (Kwasnieski et al., 2019), it is difficult to remove maternal transcripts using RNAi before translation occurs. Microinjection of small RNAs can overcome some zygotic transcription restrictions. However, this is time consuming and large-scale experiments that require substantial amounts of material, such as ChIP-Seq, are not feasible with this approach. Many maternally deposited transcripts are degraded at the maternal to zygotic transition and Piwi transcription following ZGA is limited to germ line progenitor cells thereafter (Brennecke et al., 2008). This limits the efficiency of RNAi-mediated approaches to a very narrow time window during embryogenesis. Most importantly, however, genetic approaches like RNAi in embryos would not affect maternally deposited Piwi protein already in complex with piRNAs.

The problems outlined above have prevented the exploration of the piRNA pathway during early embryogenesis in the past. However, recent technological advances provided me with several options to address those issues.

4.2 Aim

Understanding the role of the piRNA pathway during early development requires a new set of approaches. The aim of this project was to established various tools that are able to disrupt the function of the piRNA pathway in a controlled manner during early embryogenesis. To do so, I established different systems capable of delocalising Piwi or its downstream effectors from the nucleus, therefore preventing piRNA-dependent TGS. Alternatively, I devised strategies to directly degrade maternally deposited Piwi in a conditional manner during early embryogenesis.

4.3 Results

4.3.1 Protein delocalisation

Transcriptional gene silencing is a nuclear process. Multiple studies have shown that the nuclear localisation of Piwi is crucial for its function (Klenov et al., 2011; Saito et al., 2009). Additionally, we showed that the downstream effector of the TGS pathway, the PICTS complex, is required for transcriptional gene silencing in the nucleus (chapter 3). In order to investigate the function of the piRNA pathway during embryogenesis, I employed several tools to delocalise Piwi or a component of the PICTS complex. A number of different approaches are available to change the intracellular localisation of proteins.

Genetic methods involve tethering the protein of interest to a specific cellular compartment, thus preventing endogenous localisation. To do so, a protein trap is expressed that is able to recognise the protein of interest. Binding of the trap leads to delocalisation to the desired subcellular compartment. However, this method requires the zygotic expression of the protein trap during early embryogenesis or microinjection of mRNA coding for the protein trap. This method has several advantages in comparison to RNAi-mediated approaches. It directly targets proteins, rather than RNA, thus affecting both maternally deposited and zygotically expressed proteins. However, maternal deposition of the trap along with the target protein could impact localisation during oogenesis.

A different delocalisation approach is the use of optogenetic tools. A light stimulus can force the delocalisation of a protein tagged with a light-sensitive domain (Niopek et al., 2016; Park et al., 2017). This method has the distinctive advantage that the delocalisation is conditional and can be easily reversed by omitting the light stimulus, thus granting nearly instantaneous temporal and spatial control of protein localisation. While phototoxicity using modern illumination methods, such as light sheet and spinning disk microscopy, is relatively low, an impact on the development

of embryos cannot be ruled out, therefore those experiments require stringent controls.

3.3.1.1 Trapping GFP-tagged Piwi at lipid droplets using JabbaTrap

Piwi protein localises to the nucleus and is believed to function at chromatin. In order to remove Piwi from its canonical location, I established a tool that recognises GFP-tagged proteins and traps them at a different cellular compartment. The protein trapping system, called JabbaTrap (Seller et al., 2019), is comprised of two GFP-targeting nanobodies (vhhGFP4) (Saerens et al., 2005) fused to the amino- and carboxy-termini of the lipid droplet-associated protein Jabba. The endogenous *Drosophila* Jabba localises to lipid droplets through a transmembrane domain (Li et al., 2012). Expression of the JabbaTrap allows tethering of GFP-tagged proteins to cytoplasmic lipid droplets and effectively removes nuclear proteins (Figure 4.1).

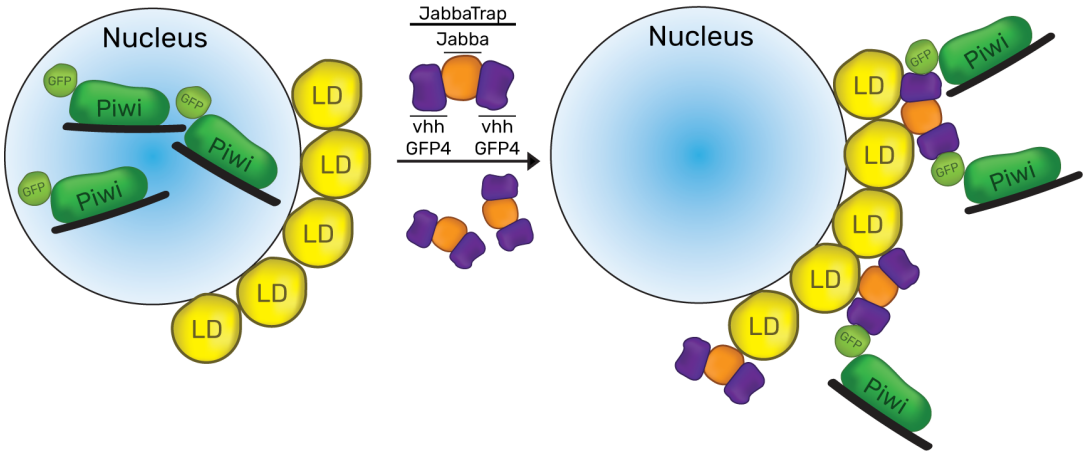

Figure 4.1: JabbaTrap-mediated Piwi delocalisation

Piwi protein localises to the nucleus and is involved in TGS at transposon loci. The protein trap, JabbaTrap, is comprised of two vhhGFP4 nanobodies fused to the lipid droplet-associated protein Jabba. Overexpression of JabbaTrap captures endogenous Piwi (that was modified with a GFP tag) though GFP-specific nanobodies and tethers the protein complex at cytoplasmic lipid droplets (LD).

JabbaTrap-mediated tethering requires two components: a GFP-tagged target protein and expression of JabbaTrap. Therefore, I genetically engineered the *piwi* locus using the CRISPR/Cas9 system by inserting a GFP-AID tag at Piwi's amino-terminus. GFP can be targeted by nanobodies, while the AID-tag (Auxin-induced degradation) can be utilised in a different degradation approach (see chapter 4.3.2.2). Additionally, I generated flies expressing HA-tagged JabbaTrap under the control of the *Drosophila ubiquitin* promoter (for details see chapter 2.2).

JabbaTrap was developed for *Drosophila* and has been demonstrated to be functional in early embryogenesis (Seller et al., 2019). However, I first evaluated whether GFP-tagged proteins are efficiently trapped in an *in vitro* system using OSCs. I transiently expressed different nuclear GFP-tagged proteins in cells with and without JabbaTrap co-expression. As a negative control, I co-transfected a puromycin selection plasmid instead of the JabbaTrap. Overexpression of GFP-Nxf2 in OSCs resulted in strong nuclear accumulation as expected (Figure 4.2) (chapter 3). However, co-expression of JabbaTrap was sufficient to delocalise Nxf2 from the nucleus and trap it at prominent foci in the cytoplasm, which likely represent lipid droplets. Similarly, the nuclear protein HP1a was trapped at lipid droplets in OSCs expressing both GFP-HP1a and JabbaTrap. The dynamics of this process, however, were not entirely clear from this experiment. For instance, whether trapping requires target proteins to first translocate to the cytoplasm during mitosis when the nuclear envelope breaks down is not known.

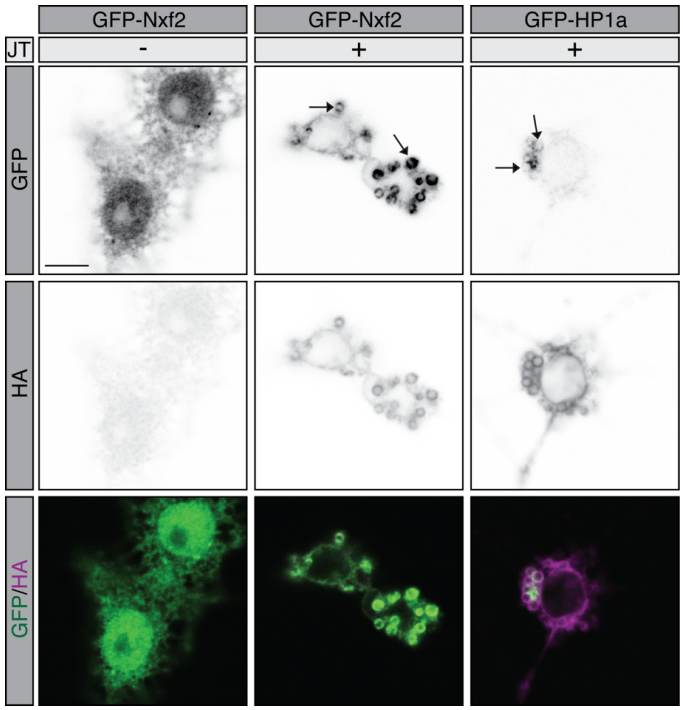

Figure 4.2: JabbaTrap efficiently tethers nuclear proteins to lipid droplets

OSCs were transiently transfected with the indicated, normally nuclear GFP-tagged proteins (GFP channel). Co-expression of JabbaTrap (JT) (HA channel) trapped most GFP-tagged protein at lipid droplets (arrows). Scale bar is 5μm.

Encouraged by the results from *in vitro* experiments, I generated flies expressing JabbaTrap and GFP-AID-Piwi by crossing homozygous GFP-AID-Piwi flies with homozygous JabbaTrap flies. The progeny of this cross was heterozygous for GFP-AID-Piwi and JabbaTrap. Thereby wild type, untagged Piwi was expressed in progeny gonads along with GFP-Piwi and JabbaTrap thus preventing potential deleterious effects during oogenesis caused by delocalisation of GFP-tagged Piwi. As a negative control, GFP-AID-Piwi flies were crossed to w^{1118} flies. In order to evaluate the effect of JabbaTrap on Piwi localisation, I examined the localisation of Piwi and JabbaTrap in ovaries using immunofluorescence. While control ovaries showed nuclear GFP-AID-Piwi localisation, GFP-AID-Piwi was delocalised from the nucleus of late stage egg chambers in flies expressing both, GFP-AID-Piwi and JabbaTrap. GFP-tagged Piwi signal colocalised with HA-tagged JabbaTrap signal at lipid droplets and was not detectable in somatic or nurse cell nuclei of late stage egg chambers (Figure 4.3).

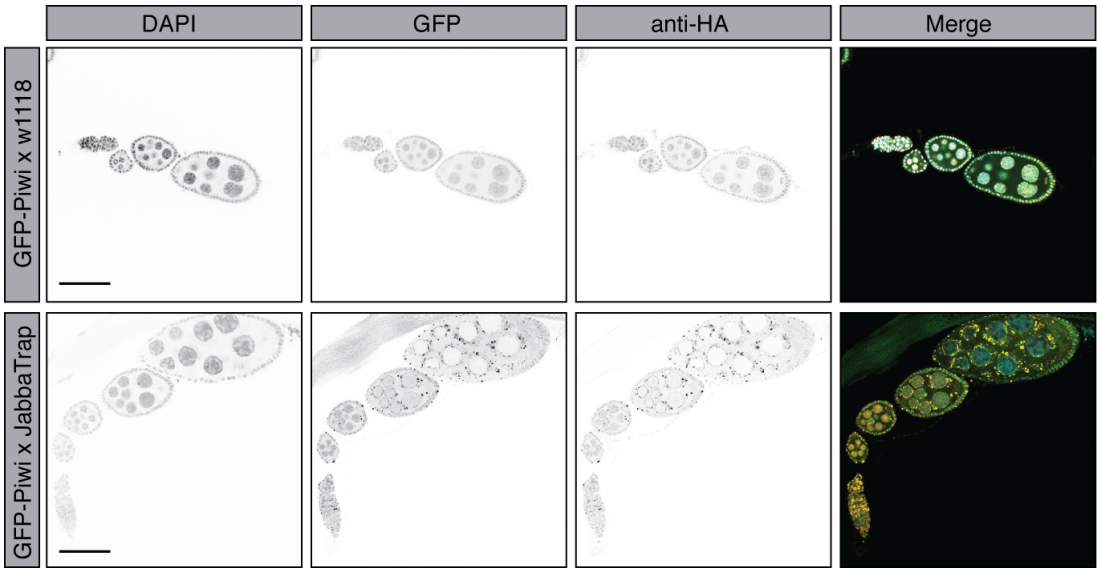

Figure 4.3: JabbaTrap delocalises Piwi from nurse cells in ovaries

GFP-AID-Piwi flies crossed to w^{1118} controls showed uniform expression of Piwi in both somatic and nurse cell nuclei (note that GFP-AID-Piwi is tagged with HA as well as JabbaTrap). Overexpression of JabbaTrap leads to delocalisation of Piwi in late stage egg chambers, while early egg chambers are only mildly affected. Scale bar is 50μm.

The *in vitro* and *in vivo* experiments suggested that nuclear proteins can be efficiently delocalised from the nucleus and trapped in the cytoplasm at lipid droplets. Importantly, I was able to delocalise proteins in complexes, such as Piwi and Nxf2, that are generally closely associated with chromatin, thereby indicating that even strong interactions between piRNA pathway components and chromatin can be overcome by using JabbaTrap-mediated delocalisation.

The implementation of JabbaTrap in embryos, however, was challenging. As discussed above, the JabbaTrap cannot be maternally deposited along homozygous GFP-AID-Piwi because delocalisation of Piwi in ovaries could have an impact on transposon control and genome integrity during oogenesis. Additionally, effects arising already in the developing oocyte would be difficult to separate from effects during embryogenesis. Therefore, I crossed homozygous GFP-AID-Piwi females with homozygous JabbaTrap embryos. In this crossing scheme, GFP-AID-Piwi is maternally deposited and JabbaTrap has to be zygotically expressed by the embryo from the paternally inherited chromosome. The *ubiquitin* promoter is active

approximately 5h after egg laying (Graveley et al., 2011). Therefore, JabbaTrap-mediated delocalisation should only affect late stage embryos. This approach, however, allows GFP-AID-Piwi to function normally during early embryogenesis.

Embryos generated by the outlined cross above were dechorionated and used for light sheet microscopy (see chapter 2.9). As expected from the expression timeline of the *ubi* promoter, GFP-AID-Piwi was localising to somatic and pole cell nuclei in both control and JabbaTrap embryos during early development. However, after around 5.5h AEL, Piwi signal started to migrate out of the nucleus and formed prominent foci of dense signal intensity, reminiscent of the results obtained in ovaries (Figure 4.3). Following an additional hour, most Piwi protein was delocalised from the nucleus and concentrated at lipid droplets, while the negative control showed continuous nuclear localisation of Piwi (Figure 4.4). This data confirms that JabbaTrap is an efficient trapping tool, usable in multiple situations. By selecting various promoters, spatio-temporal resolution can be adjusted to study the function of GFP-tagged proteins at desired developmental stages or different tissues. However, due to the late expression of zygotically transcribed JabbaTrap under the control of the *ubiquitin* promoter, examining the function of maternally deposited Piwi during very early embryogenesis is not possible.

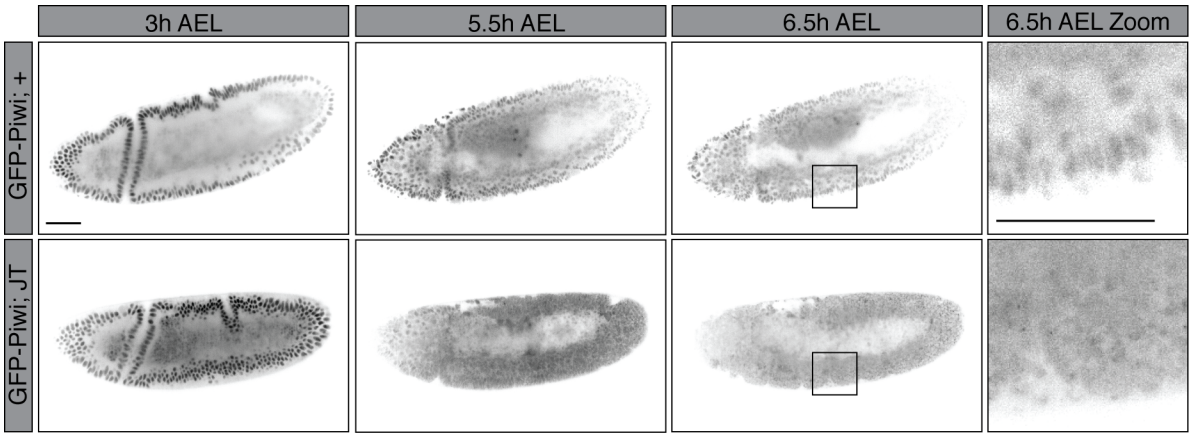

Figure 4.4: JabbaTrap delocalises maternally deposited Piwi in embryos

Light sheet microscopy images of live *Drosophila* GFP-AID-Piwi embryos during early development. GFP-AID-Piwi female flies were crossed to either w^{1118} males (GFP-AID-Piwi; +) or to JabbaTrap (JT) expressing males (GFP-AID-Piwi; JabbaTrap). Embryos were collected for imaging. Signal displayed correspond to GFP fluorescence. Time point images were selected from a 10h time course (AEL, after egg laying). Scale bar is 50μm.

3.3.1.2 Optogenetic inactivation of CRY2-tagged Piwi

Optogenetic approaches use photosensitive domains to alter the physical and biochemical properties of proteins. One commonly used domain is the *Arabidopsis thaliana*-derived protein cryptochrome 2 (CRY2). CRY2 is a blue light-absorbing photosensor and can be fused to either the amino- or carboxy-terminus of a protein of interest (Park et al., 2017). In recent years CRY2 has been used together with a truncated version of the *Arabidopsis* basic helix-loop-helix (bHLH) protein CIB1, termed CIBN for photo-switchable delocalisation (Kennedy et al., 2010). CRY2 harbours a light sensitive domain and blue light illumination results in a conformational change that allows the selective binding of CIBN. This process shows fast kinetics with binding occurring seconds after illumination. The interaction between photoactivated CRY2 and CIBN is disrupted once the light stimulus is stopped, therefore allowing for fast spatio-temporal application. CIBN can be fused to multiple proteins to change their localisation or dimerisation properties (Kennedy et al., 2010). For instance, fusing CIBN to transmembrane domains localising to the cytoplasmic cell membrane enables CRY2-tagged proteins to localise to those structures upon blue light stimulus, similar to trapping approaches such as JabbaTrap (see chapter 3.3.1.1). However, recent studies further showed that CRY2-tagged chromatin associated factors are inactivated by light stimuli without the presence of CIBN such as transcription factors Bicoid and Zelda in *Drosophila* (Huang et al., 2017; McDaniel et al., 2019). The mechanism leading to inactivation is not entirely clear. However, it might be possible that the photoactivated conformation of the CRY2 domain impacts the overall fold of the protein of interest or sterically hinders proper chromatin engagement.

Piwi has been shown to act at chromatin level. While it does not bind DNA directly it is closely associated with chromatin and might be similarly affected by a photoactivated CRY2 tag. Therefore, I genetically engineered the genomic *piwi* locus using CRISPR/Cas9 in order to introduce a CRY2-HA-mCherry tag at the amino-terminus of Piwi. Following genotyping of flies potentially positive for the knocked-in tag, I performed immunofluorescence to ensure proper nuclear CRY2-mCherry-Piwi localisation in the absence of blue light. While CRY2-mCherry-Piwi was expressed in

germ cells of the ovaries of genetically altered flies, Piwi was absent from follicle cell nuclei entirely (Figure 4.5). Nurse cells showed strong signal for tagged Piwi in the cytoplasm especially at perinuclear structures but not in nuclei. Interestingly, localisation of CRY2-mCherry-tagged Piwi in nurse cells strongly resembled endogenous localisation of Armi in germ cells (Ge et al., 2019; Munafo et al., 2019). Piwi is associated with Armi during piRNA biogenesis. It is possible that the CRY2-mCherry tag interferes with loading of piRNA precursors into Piwi and therefore prevents translocation into the nucleus of mature piRNA-Piwi complexes. However, it is possible that the light used to grow flies already had an impact on localisation of CRY2-tagged proteins. Therefore, growing fly stocks in the dark could restore endogenous localisation of tagged Piwi.

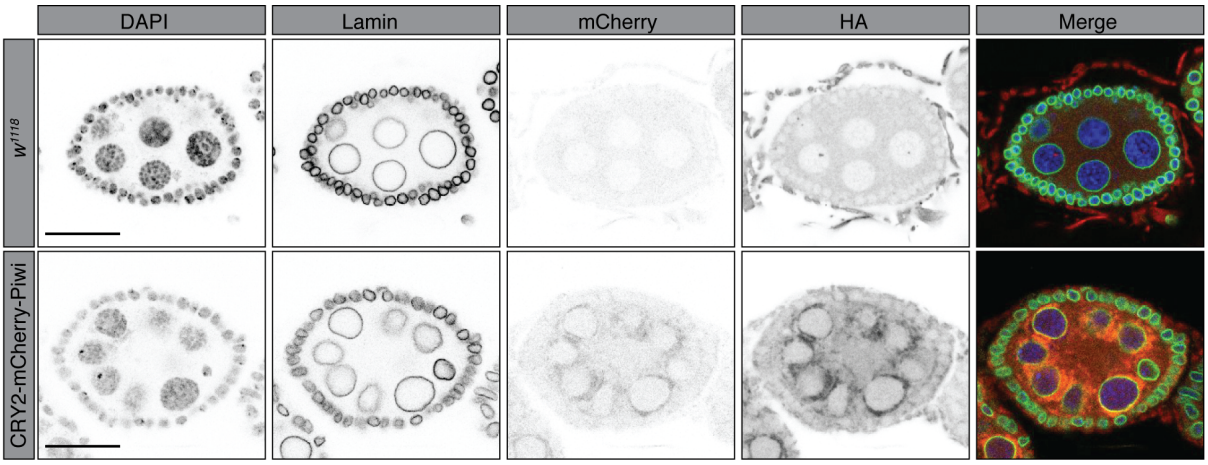

Figure 4.5: Cry2-mCherry-Piwi forms cytoplasmic aggregates in ovaries

Ovary immunofluorescence staining from CRY2-HA-mCherry-Piwi expressing flies. Shown are representative egg chambers. *w*^{1118} was used as a negative control. DAPI staining marks nuclei, Lamin staining marks nuclear envelopes and mCherry fluorescence as well as HA staining marks CRY2-HA-mCherry-Piwi. Scale bar is 25μm.

3.3.1.3 Light-induced Nuclear Export System (LEXY)

Delocalisation of Piwi using JabbaTrap was very effective in ovaries and embryos. However, the late expression of JabbaTrap during embryogenesis has several disadvantages. There is no precise temporal control of Piwi disturbance,

making it difficult to study the possible dynamic functions of the piRNA pathway during very early embryogenesis. Therefore, I established a different optogenetic delocalisation approach. The light-induced Nuclear Export System (LEXY) tag contains a nuclear export signal (NES) fused to a modified Jα helix (Niopek et al., 2016). The NES is caged by an engineered light oxygen voltage (LOV2) domain derived from *Avena sativa*. Blue light exposure results in a conformational change of the Jα helix and exposes the NES, enabling the export of the tagged protein. Omitting the light stimulus leads to re-caging of the NES and subsequent re-import of the protein depending on an amino-terminal nuclear localisation signal (NLS) (Figure 4.6).

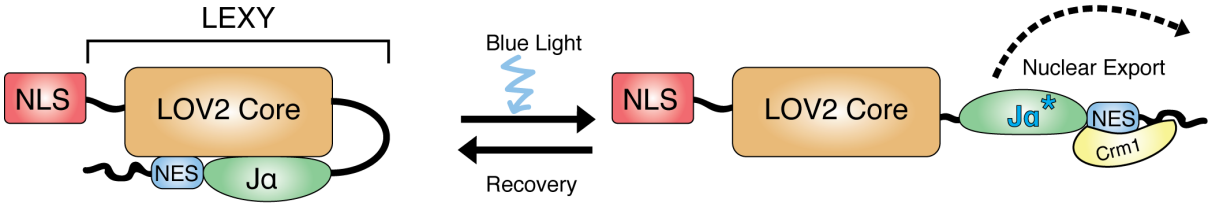

Figure 4.6: Light-induced Nuclear Export System (LEXY)

The LEXY tag is comprised of a LOV2 core domain, Jα helix and nuclear export signal (NES) fused to the carboxy terminus of a nuclear protein of interest or a nuclear localisation signal (NLS). Upon blue light exposure, the LOV2 core domain changes its conformation and the Jα helix (Jα*) unfolds, thereby making the NES accessible for proteins involved in export such as Crm1. The protein is exported and remains cytoplasmic until blue light exposure is stopped. During recovery, the NES is re-caged, allowing nuclear import depending on the NLS of the protein of interest.

In order to study the dynamics of LEXY-induced nuclear export, I engineered a LEXY-tagged fluorescent protein and overexpressed it in OSCs. Cells expressing mCherry with a nuclear localisation signal derived from the *Drosophila* Serendipity δ protein (sry) fused to LEXY (sry-mCherry-LEXY) showed rapid export of the fusion protein upon blue light exposure for 30sec. In the absence of light, mCherry-LEXY was quickly re-imported into the nucleus within ~4min post exposure (Figure 4.7) (see chapter 2.10). This result demonstrates that the LEXY system can be efficiently employed in a *Drosophila* cell culture system.

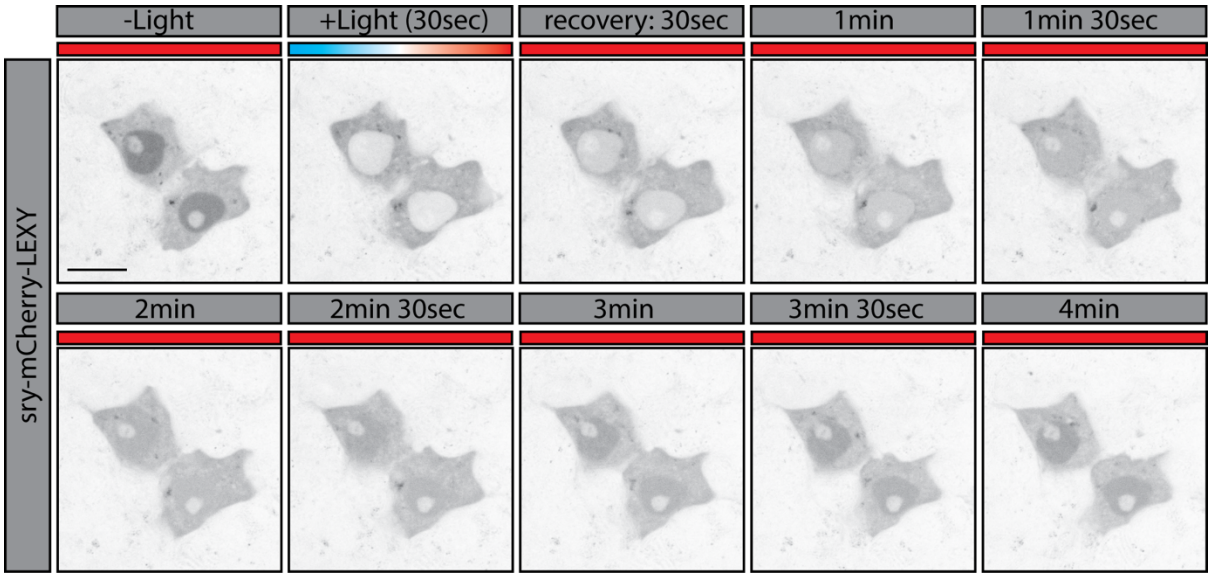

Figure 4.7: Blue light-induced nuclear export of LEXY-tagged proteins in OSCs

Live confocal microscopy of OSCs expressing sry-mCherry-LEXY. Signal displayed represents endogenous fluorescence of mCherry. mCherry was imaged with 587nm laser (red bars). Cells were illuminated with a 488nm laser (blue light) for 30sec (blue-red bar) and allowed to recover for indicated time intervals. Scale bar is 10μm.

While this method offers a precise spatio-temporal regulation of protein localisation within cells, LEXY is only functional when fused to the carboxy-terminus of a protein of interest (Niopek et al., 2016). Previous studies, however, suggested that Piwi is non-functional when tagged at its carboxy-terminus. Therefore, instead of tagging Piwi, we decided to modify one of its downstream effectors. We recently showed that the PICTS complex acts downstream of Piwi and induces transcriptional gene silencing (TGS) at transposable elements. The PICTS complex is comprised of Panx, Nxf2 and Nxt1. piRNA-dependent TGS fails upon removal of either component of the PICTS complex (see chapter 3). Disturbing Nxf2, therefore, should result in the disruption of TGS.

While we previously showed that amino-terminal tagging of Nxf2 did not alter its functionality within the piRNA pathway (Fabry et al., 2019), whether carboxy-terminal tagging would be similarly tolerated had to be tested. In order to evaluate if a Nxf2-LEXY fusion protein behaves similar compared to wild type Nxf2, we performed a rescue assay. OSCs were first depleted of endogenous Nxf2 using siRNAs. Cells

depleted of components of the PICTS complex show strong deregulation of transposons such as *mdg1* or *gypsy* (see chapter 3). Therefore, measuring the transposon expression in OSCs was used as a readout for compromised piRNA pathway-dependent TGS. As expected, reintroduction of siRNA-resistant Nxf2 (rNxf2) rescued transposon regulation compared to a negative control (ZsGreen). I first tested if amino-terminal modification of Nxf2 with a LEXY tag (rLEXY-Nxf2) was able to rescue *mdg1* and *gypsy* expression and observed similar rescue ability compared to rNxf2. Similarly, siRNA-resistant carboxy-terminally tagged Nxf2-LEXY (rNxf2-LEXY) was able to reduce transposon expression comparable to rNxf2. Therefore, we concluded that Nxf2 can be modified at either the amino- or carboxy-terminus without altering its function within the piRNA pathway, in agreement with recent studies (Batki et al., 2019; Fabry et al., 2019; Murano et al., 2019; Zhao et al., 2019).

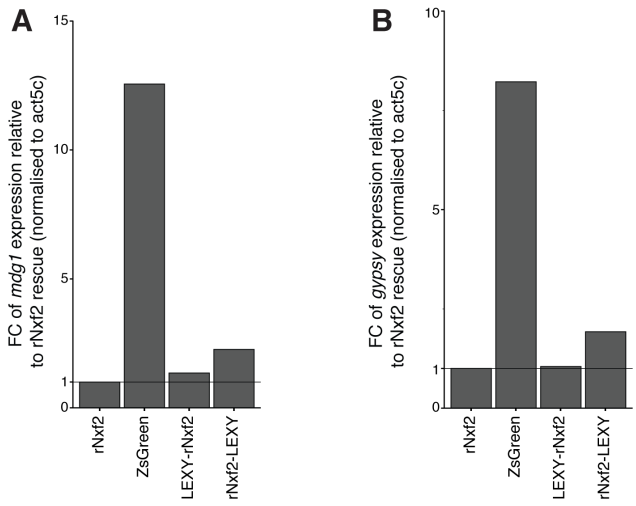

Figure 4.8: Nxf2 can be tagged at its carboxy terminus without loss of functionality

qRT-PCR results for transposon expression in OSCs depleted of endogenous Nxf2 with indicated rescue construct. **A)** Fold change (FC) of *mdg1* expression relative to rNxf2 (set to 1, see line) and normalised to *act5c*. n=1. **B)** Same as in A) but measuring *gypsy* expression.

These *in vitro* results suggested that the LEXY system is functional in *Drosophila*. LEXY-mediated nuclear export upon blue light illumination was nearly instantaneous in OSCs (30sec) and reimport occurred within 5min, thereby allowing great temporal control (Figure 4.7). Additionally, using lasers would allow the precise

delocalisation of LEXY-tagged proteins in selected embryogenetic compartments such as pole cells or somatic nuclei during early development. Carboxy-terminally tagged Nxf2-LEXY showed nearly wild type behaviour in OSCs, thereby indicating that the tag is tolerated. Therefore, I genetically engineered the endogenous locus of *nxf2* by inserting a carboxy-terminal HA-mCherry-LEXY tag using CRISPR/Cas9 (see chapters 2.1, 2.2). Flies expressing Nxf2-mCherry-LEXY were homozygous viable and fertile, indicating that the Nxf2 function was not affected by the tag. Embryos showed strong maternal deposition of Nxf2-mCherry-LEXY. The localisation of modified Nxf2 was identical with GFP-Nxf2 embryos (see chapter 5). Strong signal was detected in somatic nuclei as well as pole cells (Figure 4.9).

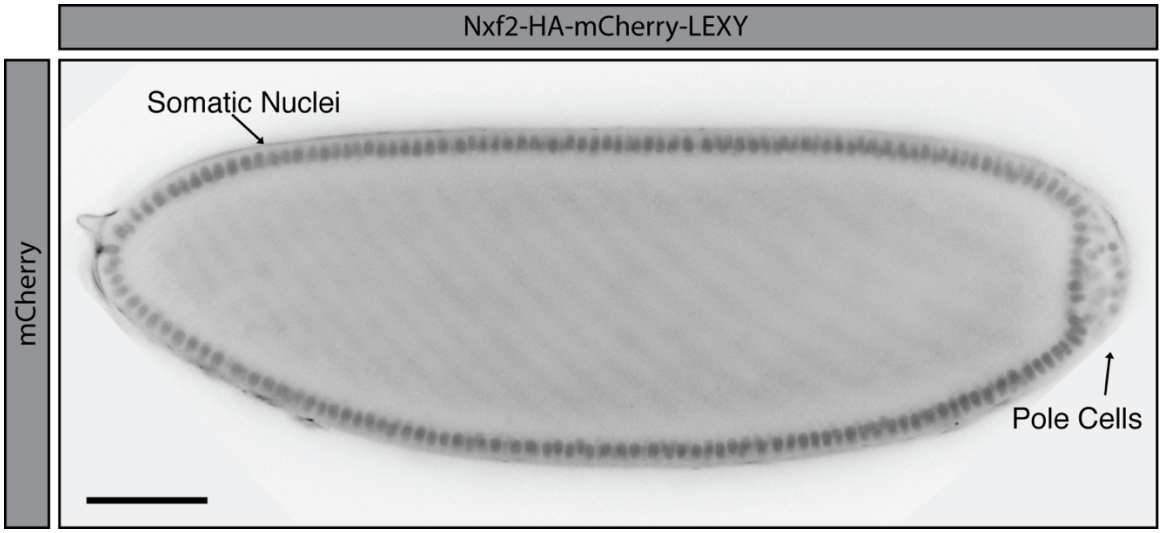

Figure 4.9: Nxf2-mCherry-LEXY localisation during early embryogenesis

Light sheet imaging of a living *Drosophila* embryo expressing Nxf2-HA-mCherry-LEXY. Detected signal represents mCherry endogenous fluorescence. Modified Nxf2 localises to somatic nuclei as well as pole cells (see arrows). Scale bar is 50μm.

I further tested the LEXY-mediated nuclear export ability by using light sheet microscopy. Embryos were immobilised in agarose and imaged in the dark. The embryo was first illuminated with a light sheet generated by a 561 nm laser to measure mCherry signal distribution. A Z-stack with small focal plane thickness was acquired in order to evenly illuminate the embryo. This was followed by one pulse of 488nm at full intensity for 30sec and 561nm light sheet imaging (see chapter 2.10). As previously

described, modified Nxf2 predominantly localised to nuclei. However, already during blue light illumination, Nxf2 was exported from the nuclei and was mostly undetectable (Figure 4.10). Following blue light illumination, recovery was measured by 561nm illumination every minute. As expected, modified Nxf2 was reimported into the nucleus within a short time and already visible after 2min similar to *in vitro* observations. However, while delocalisation kinetics resembled experiments performed *in vitro* (Figure 4.7), fluorescence intensity of tagged Nxf2 decreased strongly and recovery was prolonged. This could be due to slower re-import kinetics of the PICTS complex compared to recovery observed using strong NLS such as sry *in vitro* or due to delocalisation of Nxf2 only, which is unstable outside of the PICTS complex (Fabry et al., 2019). Additionally, photobleaching during the imaging procedure could partially contribute to a decrease of total signal intensity in somatic nuclei.

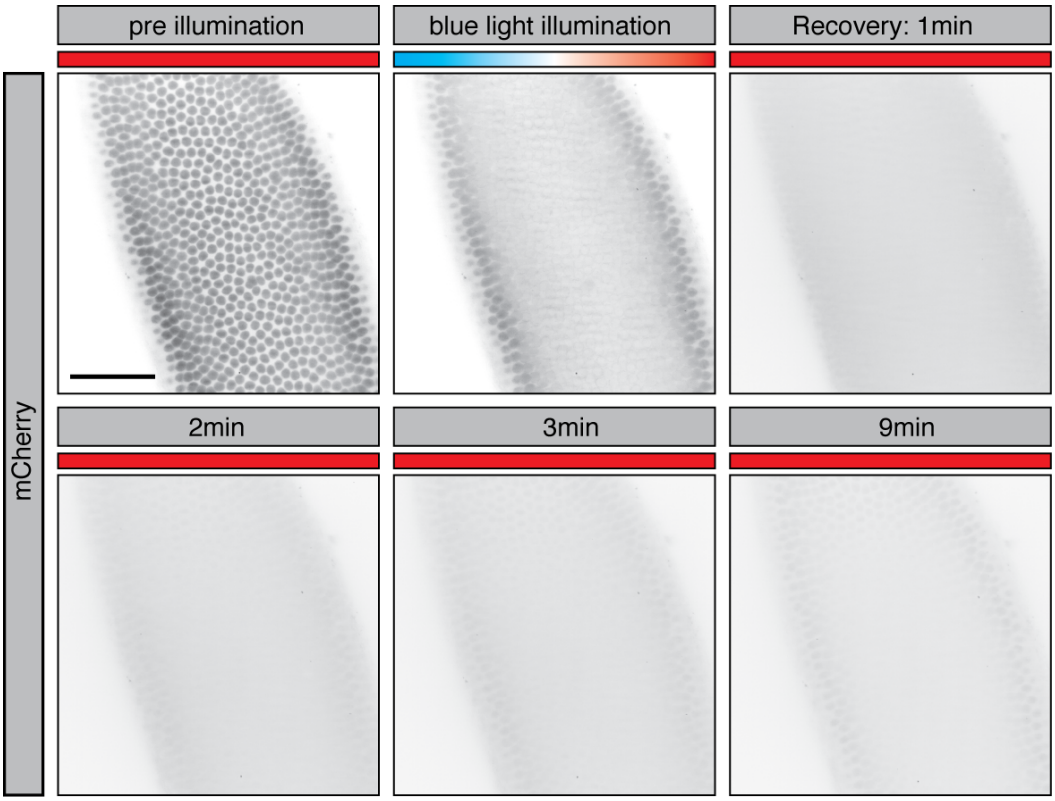

Figure 4.10: Blue light illumination delocalises embryonic Nxf2-mCherry-LEXY

Intensity projection of embryos expressing Nxf2-mCherry-LEXY using light sheet live imaging. Embryos were first imaged with 561nm light sheet (red bar, mCherry channel) followed by a 30sec pulse using a 488nm light sheet (blue-red bar). Recovery was measured every minute for 9min using a 561nm light sheet.

4.3.2 Conditional protein degradation

The approaches described above are aiming at delocalising piRNA pathway components from chromatin. However, in order to study the effect of maternally inherited piRNAs on development, Piwi or its downstream effectors have to be tethered away from the nucleus constantly for a prolonged period of time.

Novel tools are available to address this problem by degrading the protein of interest directly, therefore preventing the function of a specific protein indefinitely. Several different methods have been described in recent years. Most rely on the fusion of a degradation (degron) sequence to the protein of interest and utilise the general ubiquitin-mediated protein degradation machinery of the cell.

4.3.2.1 deGradFP, a GFP nanobody-driven degradation system

Targeting a wide range of proteins for degradation is time consuming and costly due to the requirement of engineering each corresponding genomic locus with a specialised tag for recognition. However, the deGradFP system makes use of the broad availability of GFP-tagged proteins. For many organisms endogenously tagged proteins have been generated and are widely accessible. Therefore, targeting GFP as a universal structure for degradation allows studying the function of already available GFP-tagged proteins. deGradFP is a fusion protein comprised of a F-Box protein and the vhh4GFP nanobody. The GFP nanobody targets GFP-tagged proteins and links these to the ubiquitination machinery of the cell via the F-Box protein domain derived from the *Drosophila* protein supernumerary limbs (Slmb). Binding of deGradFP to target proteins leads to ubiquitination of the protein of interest and the subsequent degradation facilitated by the proteasome (Caussinus et al., 2011). Previously generated GFP-AID-Piwi flies in combination with deGradFP expressing flies were used to target maternally deposited Piwi during embryogenesis (Figure 4.11).

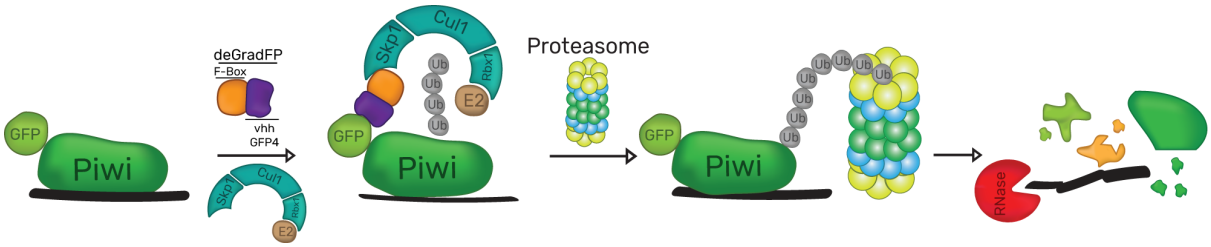

Figure 4.11: deGradFP licences GFP-tagged proteins for degradation

GFP-tagged Piwi is recognised by the vhhGFP4 nanobody domain of deGradFP. The F-Box protein domain recruits the ubiquitination machinery leading to accumulation of ubiquitin chains at lysine residues of GFP-Piwi. The proteasome degrades ubiquitinated proteins thereby exposing piRNAs for degradation by RNases.

deGradFP was established for *Drosophila* and was shown to be effective in degrading GFP-tagged proteins (Caussinus et al., 2011). In order to test if deGradFP is suitable to degrade maternally deposited Piwi, I first tested the general degradation efficiency *in vitro*. OSCs were transfected with a highly expressed cytoplasmic target protein (miniAID-GFP). Additionally, either a negative control or a HA-tagged deGradFP under the control of the *Drosophila ubiquitin* promoter were co-expressed. Cells were allowed to express for 48h followed by protein extraction and western blotting for protein quantification (see chapters 2.3, 2.8). miniAID-GFP expressing cells without deGradFP co-expression showed strong protein abundance. However, co-expression of deGradFP had only little effect on protein levels (Figure 4.12 A). deGradFP co-expression resulted in ~30% reduction of protein levels in comparison to a negative control (Figure 4.12 B). The expression and protein abundance of miniAID-GFP was very high and might have oversaturated the capabilities of the deGradFP system. Therefore, I performed the same experiment described above with a GFP-HP1a target instead of miniAID-GFP. GFP-HP1a localises to the nucleus and is a chromatin associated protein, thereby mirroring some properties of nuclear Piwi. Co-expression of deGradFP with HP1a-GFP resulted in ~94% degradation (Figure 4.12 A-B).

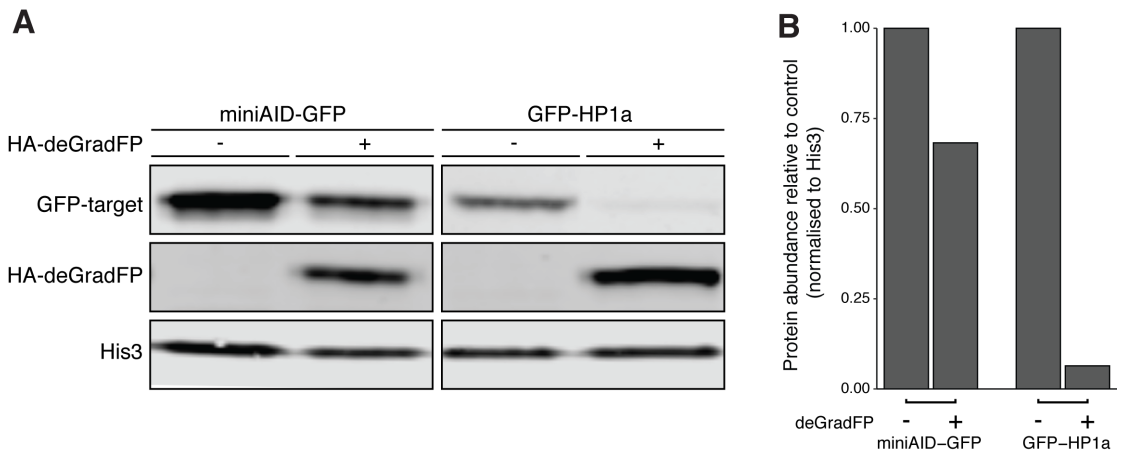

Figure 4.12: deGradFP efficiently degrades chromatin associated GFP-tagged proteins *in vitro*

A) Western Blot assay of protein lysate form OSCs expressing GFP-tagged target proteins (miniAID-GFP or GFP-HP1a) together with either a negative control or deGradFP. **B**) Quantification of A) using His3 as loading control for normalisation (n=1). Negative control was set to 1.

Highly abundant target proteins such as miniAID-GFP were not efficiently degraded by deGradFP. However, the chromatin associated protein HP1a-GFP was greatly reduced in OSCs overexpressing deGradFP with little target protein detected. Therefore, we next tested degradation efficiency *in vivo*. GFP-Piwi expressing flies were crossed with flies overexpressing deGradFP or a negative control. Ovaries were dissected from offspring and Piwi abundance was evaluated by western blot as described above (see chapter 2.6). GFP-Piwi was highly abundant in ovaries of control crosses. Flies co-expressing deGradFP showed strong degradation of GFP-Piwi. However, the reduction was only ~50% in comparison to the control (Figure 4.13 A-B).

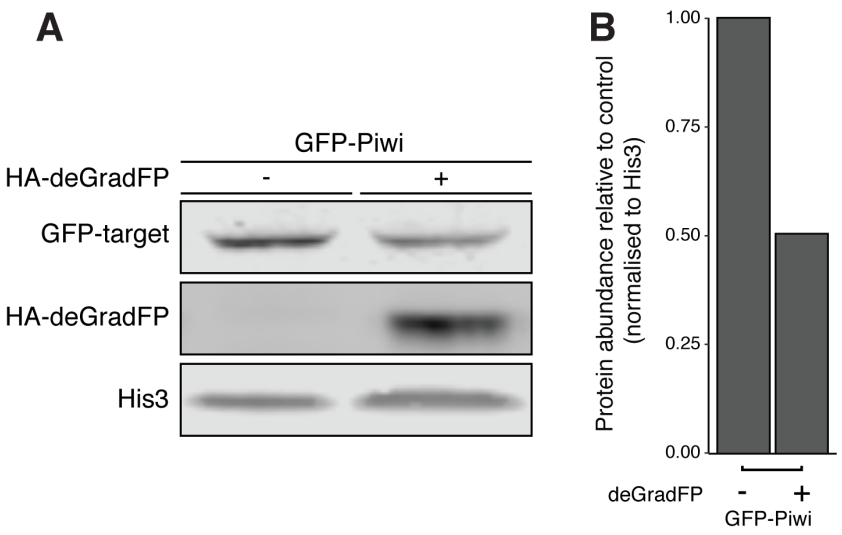

Figure 4.13: deGradFP degradation efficiency of GFP-Piwi in ovaries

A) Western blot of GFP-Piwi and HA-tagged deGradFP from *Drosophila* ovary lysate. His3 was used as loading control. **B**) Quantification of A). His3 was used for normalisation and negative control was set to 1 (n=1).

4.3.2.2 Degradation of maternally deposited Piwi in embryos using the Auxin inducible degron (AID)

The Auxin Inducible Degron (AID) system allows the degradation of proteins of interest by treatment with the small molecule, auxin. The system is comprised of two components. An AID-tag fused to the protein of interest recognises auxin, a plant hormone. The plant-derived F-box protein transport inhibitor response 1 (TIR1) binds to the AID-auxin complex. TIR1 is part of a SCF E3 ubiquitin ligase complex that binds AID specifically in the presence of auxin. This recruits the cellular ubiquitin machinery similar to the Slmb F-box domain used in deGradFP (see chapter 4.3.2.1). Following ubiquitination, the protein of interest is degraded by the proteasome. While the AID system is highly dependent on auxin, small concentrations found in plant-derived animal food might induce mild degradation in the absence of added auxin. However, potent TIR1 inhibitors have been developed recently to prevent such undesired degradation (Yesbolatova et al., 2019). Protein degradation is irreversible. However, halting auxin treatment stabilises protein levels of newly translated AID-tagged protein. This conditional degradation system has been successfully applied in several model

organisms including *Drosophila* (Nishimura et al., 2009, Bence et al., 2017) and has been recently shown to efficiently degrade the germ line associated protein Vasa (Bence et al., 2017), thereby making it an ideal system to test for degradation of maternally deposited Piwi (Figure 4.14).

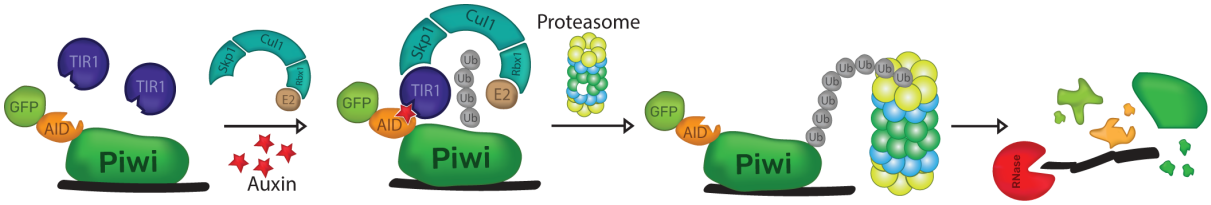

Figure 4.14: Auxin induced degradation of GFP-AID-Piwi

Endogenously tagged GFP-AID-Piwi is stable in untreated flies. Treatment with auxin leads to binding of the AID-tag and TIR1, thereby recruiting the SCF E3 ubiquitin ligase complex. GFP-AID-Piwi is ubiquitinated and degraded by the proteasome. The exposed piRNA is unprotected and likely degraded by endogenous RNases.

I used CRISPR/Cas9 to edit the *Drosophila piwi* locus and knocked in an amino-terminal GFP-AID tag (see chapters 2.1, 2.2). Homozygous founders were screened for expression of GFP-AID-Piwi by western blot with an expected size of 150kDa (Figure 4.15). Additionally, proper localisation and maternal deposition was verified by immunofluorescence performed on ovaries and embryos respectively (Figure 4.16 and Figure 4.18 respectively).

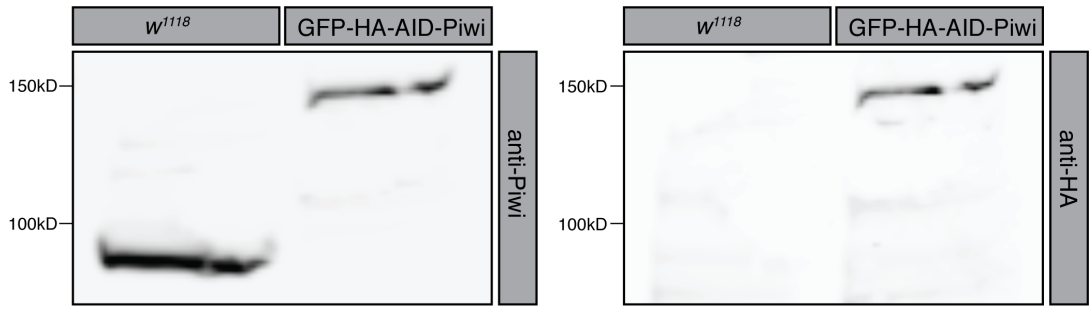

Figure 4.15: GFP-AID-Piwi is expressed and detected at the predicted molecular weight in ovaries

Western blot assay probing for Piwi or HA from ovaries of w^{1118} or homozygous GFP-HA-AID-Piwi flies. GFP-HA-AID-Piwi has a calculated molecular weight of 150kD, while wild type Piwi is expected at 97kD.

GFP-AID-Piwi flies were homozygous viable and showed no reduction in fertility. The second component of the AID system is TIR1 that is necessary to connect the AID-auxin complex to the cellular ubiquitination machinery. Therefore, we generated fly strains expressing either of two versions of TIR1: OsTIR1 derived from *Oryza sativa* and AtTIR1 derived from *Arabidopsis thaliana*. Both OsTIR1 and AtTIR1 were expressed under the control of *Drosophila ubiquitin* promoter. AtTIR1 was modified with two amino acid substitution (D170E and M473L) that have been shown to increase auxin affinity, while not increasing auxin-independent binding (Yu et al., 2013; Zhang et al., 2015). For the following experiments, I generated fly stocks homozygous for GFP-AID-Piwi; OsTIR1 as well as GFP-AID-Piwi; AtTIR1.

In order to evaluate the efficiency of the AID system *in vivo*, I treated flies homozygous for GFP-AID-Piwi and OsTIR1 with 5mM auxin containing yeast paste or standard paste. Flies were allowed to feed on yeast for 16h before dissection of ovaries for immunofluorescence. Ovaries of auxin fed flies were mostly depleted of GFP-AID-Piwi and showed only traces of GFP signal in early stage egg chambers. However, the ovary morphology was not affected by auxin treatment (Figure 4.16).

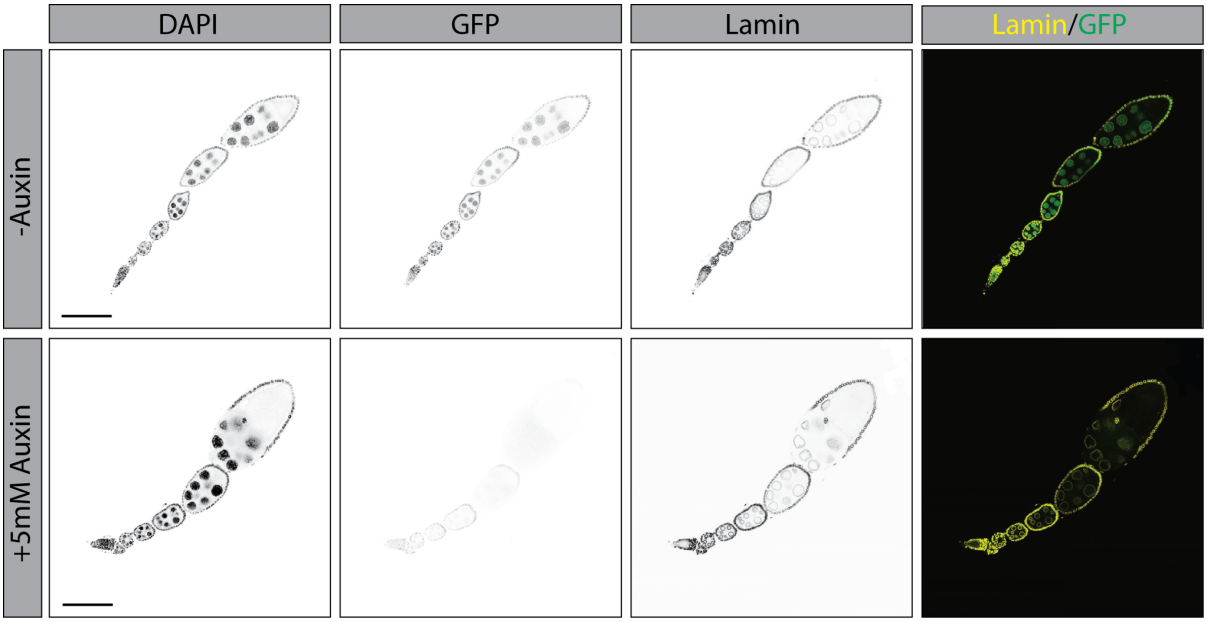

Figure 4.16: Auxin treatment degrades GFP-AID-Piwi in ovaries within 16h

Immunofluorescence of ovaries from flies treated with or without 5mM auxin for 16h. DAPI staining represents DNA, GFP signal is derived from GFP-AID-Piwi and Lamin staining shows the nuclear envelope. Scale bar 100μm.

Treatment of female flies lead to significant reduction of GFP-AID-Piwi in ovaries. Next, I examined a potential impact on transposon control. Transposon expression during oogenesis might lead to maternal deposition of transposon mRNA into developing embryos. This could affect the development of embryos and therefore would complicate the evaluation of effects directly caused by maternally deposited Piwi during embryogenesis. Therefore, I measured the expression of commonly deregulated transposon upon piRNA pathway disruption in ovaries. Most transposons showed little to no change in treated samples in comparison to controls. However, *mdg1* expression was more than 20-fold upregulated in auxin-treated flies, indicating that auxin treatment for 16h had an impact on Piwi-dependent transposon silencing (Figure 4.17 A). *mdg1* is predominantly expressed in somatic follicle cells of the ovary. This indicates that transposon expression detected in treated ovaries was derived from somatic cells rather than nurse cells. Ovary morphology was not impaired in auxin-treated females (data not shown). This is likely due to varying degradation efficiencies during oogenesis stages. The germarium and early egg chamber showed only a mild degradation of Piwi, while late-stage egg chambers were almost entirely lacking Piwi signal (Figure 4.16). 0-4h embryos collected from flies treated for 16h showed some deregulation of transposons. This could be due to maternal deposition of transposon mRNA or zygotic expression. Interestingly, *mdg1* transcripts did not change compared to control embryos, while the germ line-specific transposon *Het-A* was > 2-fold upregulated (Figure 4.17 B). This is likely due to expression of *mdg1* in somatic follicle cells rather than nurse cells that are not contributing to maternal deposition during oogenesis.

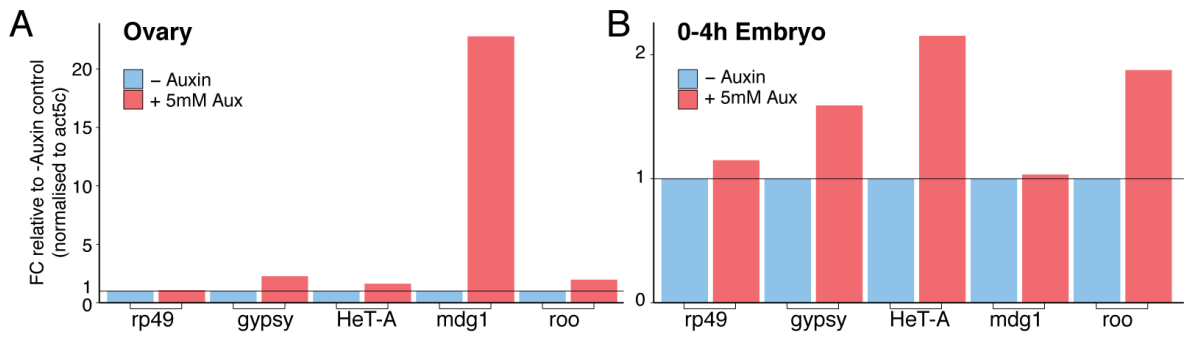

Figure 4.17: Auxin treatment leads to deregulation of TEs in ovaries and embryos

qRT-PCR measurements of transposon transcripts from ovaries and embryos of flies homozygous for GFP-AID-Piwi and AtTIR1. Fold change (FC) of expression is relative to no Auxin control and normalised to *act5c*. *rp49* was included as additional housekeeping gene. **A)** Results for ovaries of flies treated for 16h (n=1). **B)** Results for 0-4h embryos collected from flies treated for 16h (n=1).

Due to the deregulation of transposons in ovaries of treated flies, I decided to treat early embryos directly instead. This has the advantage that all effects observed in treated embryos are due to degradation of maternally deposited Piwi, rather than side effects of degradation occurring already in ovaries. First, I examined the efficiency of embryo treatment. Embryos from parents homozygous for both GFP-AID-Piwi and OsTIR1 were collected for 1h and dechorionated in bleach followed by treatment with various concentrations of auxin in PBS or control treatment without auxin (see chapter 2.18). Immunofluorescence of fixed embryos treated with 25mM auxin indicated an efficient degradation of maternally deposited Piwi (Figure 4.18 A-B).

The immunofluorescence results obtained from treatment of whole flies or embryos suggested strong degradation. However, to quantify the efficiency, I performed treatments of embryos with different concentrations of auxin as described above and measured protein abundance. Western blot assays revealed that treated embryos showed little Piwi protein remaining after 2h treatment (~8-2%) compared to PBS control embryos (Figure 4.18 C-D). The optimal auxin concentration for embryo treatment was 5mM with ~98% of degradation of maternally deposited GFP-AID-Piwi after 2h. Therefore, 5mM auxin concentration was used for subsequent experiments described below.

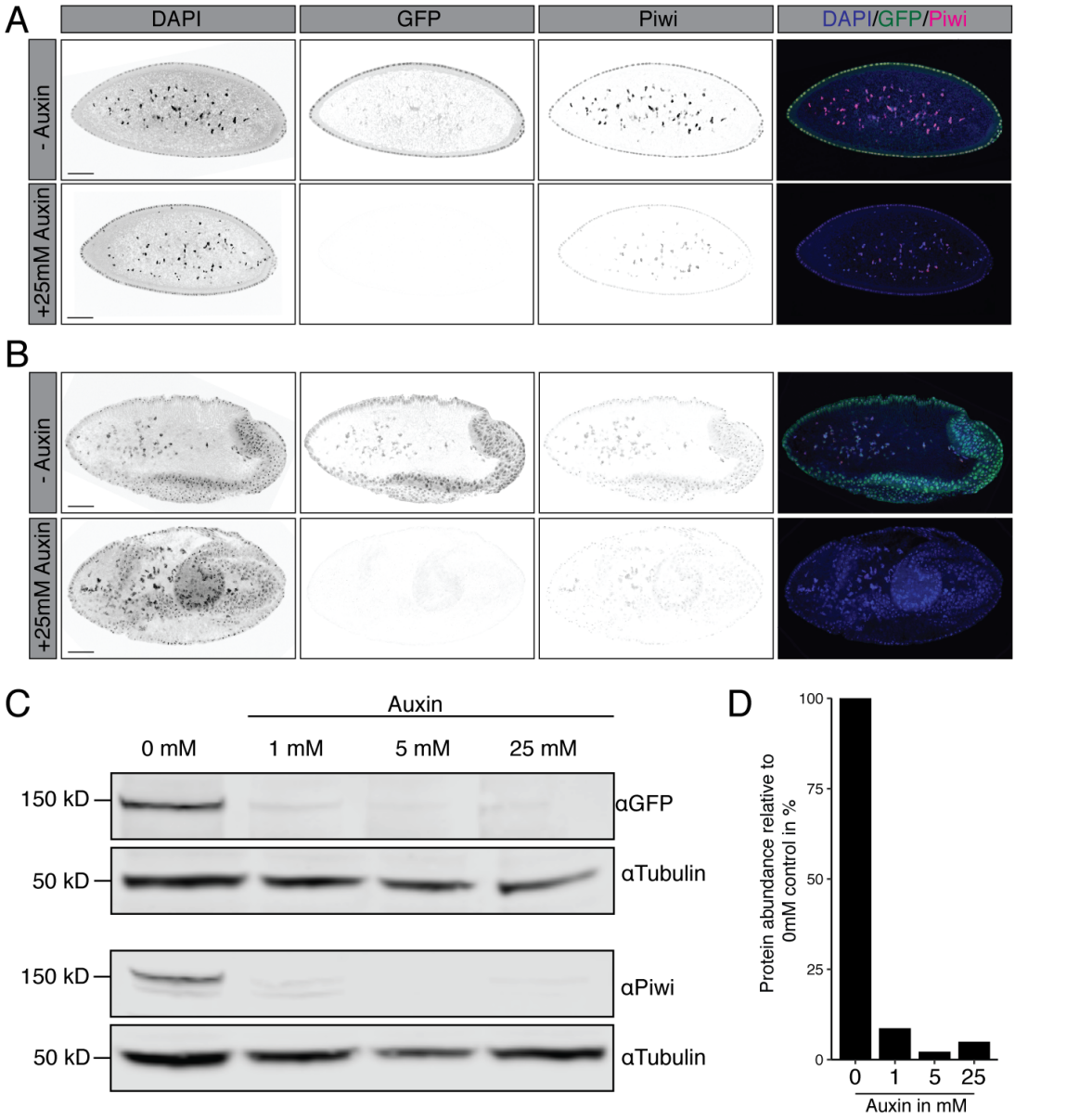

Figure 4.18: Treatment of embryos with auxin degrades maternally deposited Piwi efficiently

A) Immunofluorescence for early embryos (1h collection) treated with 25mM auxin for 2h. DAPI marks DNA, GFP and Piwi staining shows GFP-AID-Piwi. **B**) Same as A) but for embryos shortly after gastrulation (~3h AEL). Scale bar represents 50μm. **C**) Western blot from embryos treated as described above with different auxin concentrations. **D**) Quantification of C) using α-tubulin as a loading control for normalisation. Values in % are relative to 0mM treatment control (100%) (n=1).

Auxin treatment was efficient in degrading maternally inherited GFP-AID-Piwi in embryos. However, the precise dynamics of degradation were not measurable by the experiments described above. Therefore, I used light sheet microscopy to image living embryos treated with auxin. Embryos treated with 5mM auxin showed a steady decrease in GFP-AID-Piwi signal and GFP was not detectable after < 25min (Figure 4.19).

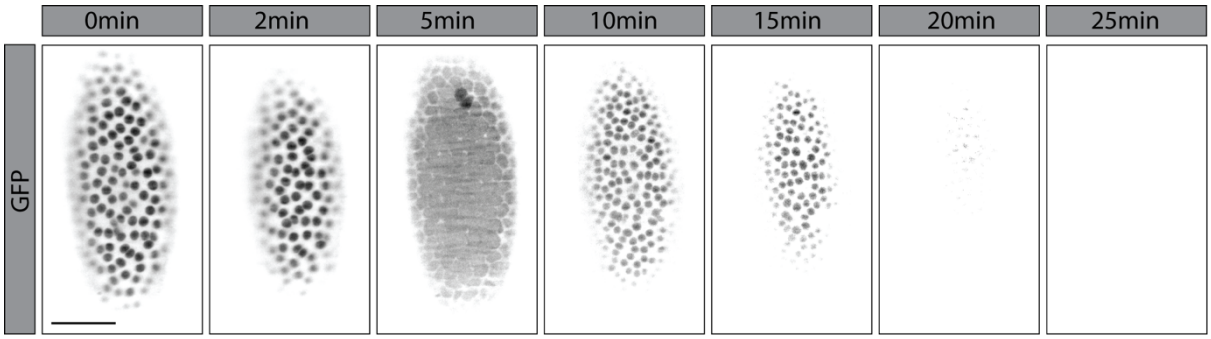

Figure 4.19: Maternally deposited GFP-AID-Piwi is rapidly degraded upon auxin treatment

Selected time points for light sheet imaging of somatic nuclei of living early embryo expressing GFP-AID-Piwi and AtTIR1. Treatment of embryo was started at 0min with 5mM auxin and GFP signal measured for 25min post treatment. 5min time point shows nuclear division of blastoderm embryo. Scale bar is 50μm.

4.4 Conclusion

This work established different approaches aiming to disrupt maternally deposited Piwi or downstream TGS factors during *Drosophila* embryogenesis. Cytoplasmic tethering of GFP-tagged nuclear proteins using JabbaTrap was efficient *in vitro* and *in vivo*. While cell culture experiments and ovary experiments suggested a complete removal of piRNA pathway factors from the nucleus, embryo experiments revealed a delayed onset of transcription of JabbaTrap without affecting early embryogenesis. Therefore, more work is necessary to improve this technique for studies of maternally deposited proteins.

An optogenetic approach using LEXY was able to delocalise the PICTS complex component Nxf2 during early embryogenesis using light exposure in a highly dynamic and spatially-regulated manner. However, further experiments are needed to optimise light treatment conditions to prevent phototoxic effects on embryogenesis and increase the throughput of this approach.

Degradation of maternally deposited Piwi was achieved by employing the AID system during early embryogenesis. Auxin treatment led to rapid and highly conditional degradation of GFP-AID-tagged Piwi during early development therefore representing a powerful tool to study Piwi's function during embryogenesis.

5 Maternally inherited piRNAs control transposons during *Drosophila* embryogenesis

Quantitative mass spectrometry was performed by the Proteomics Core facility at CRUK-CI and raw data was analysed by Kamal Kishore (Bioinformatics Core facility). Federica Falconio assisted with the generation of fly stocks for light sheet microscopy. Emily Lythgoe contributed to embryo collections for w^{1118} ChIP-Seq experiments.

5.1 Introduction

Transposon silencing during oogenesis is highly dependent on the piRNA pathway. Nascent transposon transcripts in the nucleus are recognised by piRNA-Piwi complexes. Following recruitment of the PICTS complex by Piwi, transcriptional gene silencing is induced by epigenetic changes at transposon loci. While the function and dynamics of this process are relatively well documented in *Drosophila* ovaries, little is known about the function of the piRNA pathway beyond germ cells.

In adults, expression of piRNA pathway components such as Piwi or the PICTS complex is believed to be limited to gonadal tissues, more specifically somatic follicle cells surrounding the oocyte and germline-derived nurse cells, both required for oogenesis. However, previous studies suggested the maternal deposition of PIWI-clade Argonaute proteins into embryos (Brennecke et al., 2008; Gunawardane et al., 2007; Megosh et al., 2006). While Aub exclusively localises to the pole plasm of the developing *Drosophila* embryo, Piwi is present in somatic and pole cell nuclei (Brennecke et al., 2008; Megosh et al., 2006). Aub was reported to contribute to clearing of maternal transcripts (Rouget et al., 2010). Additionally, maternally inherited Piwi was implicated in redefining piRNA source loci in adult ovaries (Akkouche et al., 2017; Le Thomas et al., 2014) and silencing of position-effect variegation (PEV) reporters in somatic cells of adult flies (Gu and Elgin, 2013).

While some data suggests an impact of Piwi on somatic cells (Gu and Elgin, 2013), these studies failed to remove maternally deposited Piwi protein during

embryogenesis. A precise function of the piRNA pathway during embryogenesis in somatic cells is therefore not known to this date. However, Piwi's abundance in somatic cells of the embryo suggests that the piRNA pathway might have evolved a function beyond maintenance of genome integrity of germ cells.

5.2 Aim

Studying the function of Piwi during embryogenesis has been challenging. Piwi protein is maternally deposited and can therefore not be targeted efficiently by conventional tools, such as RNAi and genetics, that are used to study the function of the piRNA pathway in *Drosophila* gonads. However, the recent improvement of protein degradation systems enables specific removal of maternal Piwi in a highly conditional manner.

The aim of this project was to dissect the function of maternally deposited Piwi and the piRNA pathway during embryogenesis in non-gonadal somatic cells by examining molecular changes throughout development and studying Piwi's function by degradation during embryogenesis.

5.3 Results

5.3.1 Piwi is maternally deposited and localises to somatic and pole cell nuclei during early embryogenesis

Several years ago, the Hannon lab reported maternal deposition of Piwi and its localisation to somatic nuclei during embryogenesis (Brennecke et al., 2008). While this study provided insight into the localisation of Piwi at selected stages, only little information about spatial dynamics could be inferred.

In order to understand the precise dynamics of maternally deposited Piwi protein during embryogenesis, I performed Light Sheet Fluorescence Microscopy (LSFM) on living embryos. LSFM induces low phototoxicity and has been reliably utilised to study the development of *Drosophila* embryos (Khairy et al., 2015). We first generated fly stocks expressing GFP-Piwi from its endogenous promoter as well as a transgene carrying His2Av-RFP to mark nuclei. Preblastoderm stage embryos derived from parents expressing GFP-Piwi and His2Av-RFP were collected and fluorescent signal tracked for several hours of embryogenesis (see chapter 2.9).

Piwi protein showed strong maternal deposition as previously reported (Brennecke et al., 2008). During the preblastoderm stage (NC1-9, ~30min after egg laying; AEL), Piwi localised to the posterior pole and formed a crescent-like structure (Figure 5.1 A). While Piwi signal was strongest at the posterior pole, a homogeneous signal was detected throughout the entire embryo. As embryogenesis progressed and somatic nuclei migrated to the surface (NC 9-13, 1.5-2h AEL), Piwi localised to somatic nuclei. Additionally, Piwi localised to the pole plasm surrounding the nucleus of germline progenitor cells (~1h AEL). However, Piwi was excluded from pole cell nuclei initially. This localisation pattern is similar to other germ cell-associated factors required for directing cell fate towards germline development such as Oskar or Nanos (Lasko, 2012). Piwi entered pole cell nuclei at embryogenesis stage 4 (NC 12 of somatic nuclei) and remained strongly enriched for the duration of embryogenesis. Piwi's enrichment in somatic nuclei was detectable throughout the first 6h of embryogenesis. However, signal intensity decreased over time, probably due to

protein degradation and/or bleaching of GFP fluorophores while imaging. Interestingly, nuclear Piwi signal strongly decreased during mitotic cycles and only a faint signal overlapping with His2Av-RFP remained until completion of mitosis (Figure 5.1 B).

In order to quantify the stability of maternally deposited Piwi, I performed western blot analyses of *w*^{1118} embryos probing for Piwi protein and α-tubulin as a loading control (Figure 5.1 C). Piwi levels were highest in pre-maternal zygotic transition (pre-MZT) embryos (0-2h AEL) and decreased throughout development. However, at 12h AEL Piwi was still detectable, indicating a high stability of maternally deposited piRNA-Piwi complexes.

To further quantify the abundance and stability of Piwi in a more quantitative way, we performed mass spectrometry from *w*^{1118} embryos aged 0-2h, 5-7h and 10-12h AEL (see chapter 2.16). Signal intensity for Piwi peptides between different time points suggested a mild reduction of Piwi protein levels between the initial pool of maternally deposited protein (0-2h) and 5-7h AEL embryos. However, Piwi abundance further decreased throughout development, thereby confirming previous results (Figure 5.1 D). While Piwi abundance remained constant throughout early development, it cannot be excluded that zygotically transcribed Piwi in germ line progenitors contributed partially to detected signal in both western blot and mass spectrometry experiments.

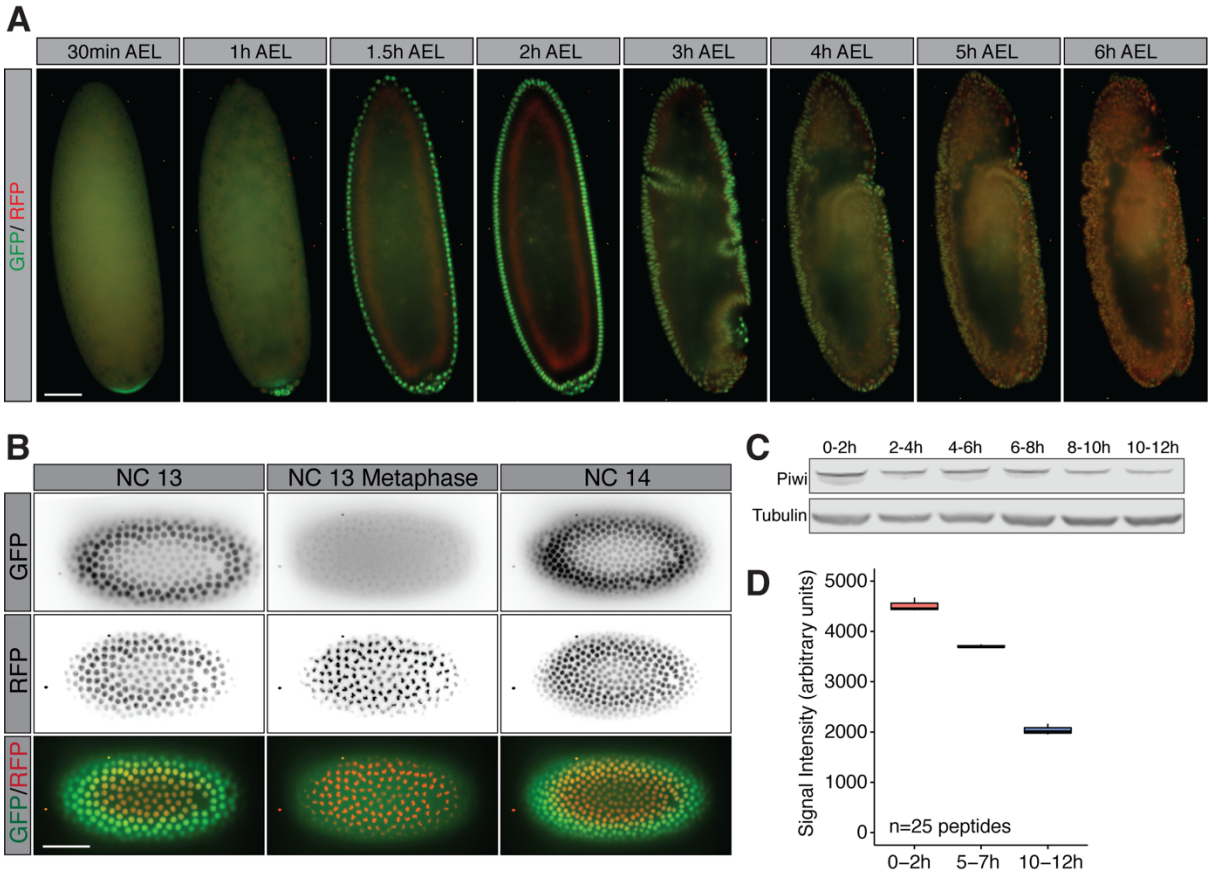

Figure 5.1: Piwi is maternally deposited and abundant throughout early development

A) LSFM live imaging of embryos derived from parents expressing GFP-Piwi and His2Av-RFP for the indicated time points. Green signal corresponds to Piwi and red signal marks nuclei. Scale bar is 50μm **B**) Same as A) but zoom-in at embryonic somatic cell sheet. Scale bar is 50μm. **C**) Western blot probing for Piwi and α-tubulin in *w*^{1118} embryos for various time points. **D**) Boxplot for Piwi abundance of *w*^{1118} embryos from quantitative mass spectrometry data for indicated time points after egg laying (arbitrary units). Number of detected Piwi peptides are indicated. Error bars represent standard deviation. n=3 biological replicates.

In order to assess the contribution of zygotically transcribed Piwi during embryogenesis, I performed immunofluorescent staining of early and late stage embryos. Embryos derived from female flies expressing GFP-AID-Piwi crossed to *w*^{1118} males showed strong maternal deposition of GFP-AID-Piwi during the syncytial blastoderm stage in accordance with LSFM data presented above (Figure 5.2 A). GFP fluorescence in late stage embryos (>12h AEL) was restricted to the germline. GFP-

AID-Piwi was detected in both nuclei and cytoplasm of developing germ cells. However, embryos derived from reciprocal crosses showed no GFP signal in early embryos due to a lack of maternal inheritance of GFP-tagged Piwi (Figure 5.2 B). Interestingly, while zygotically transcribed GFP-AID-Piwi was detected in late stage embryos, Piwi localised exclusively to the cytoplasm of germ cell progenitors and was not detected in nuclei. This indicates that zygotically transcribed Piwi is likely not relevant at this stage due to its cytoplasmic location.

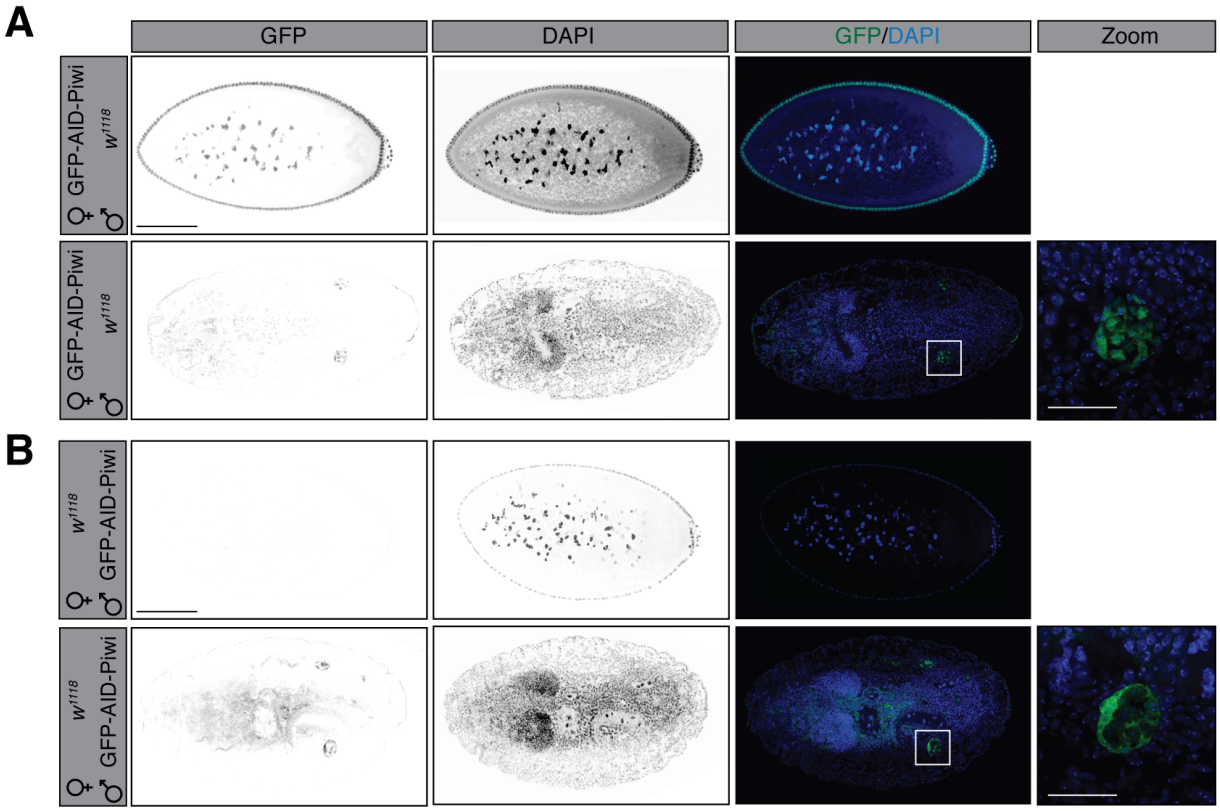

Figure 5.2: Zygotic Piwi transcription is limited to germ cells

A) Confocal fluorescent microscopy images of embryos derived from females expressing GFP-AID-Piwi crossed to *w*^{1118} males probing for GFP and DAPI. Upper panel shows early embryos (blastoderm stage) and bottom panel late stage embryos (>12h AEL). Zoom-in shows developing germline. Scale bar is 100μm, in zoom 10μm. **B**) Same as A) but showing embryos derived from *w*^{1118} females crossed to male GFP-AID-Piwi flies.

Maternally deposited Piwi primarily localised to somatic nuclei during early development. However, the resolution of LSFM was not sufficient to resolve the sub-nuclear localisation of Piwi. While previous studies showed Piwi localisation to nuclei and its known function on nascent RNA (Yashiro et al., 2018), little evidence has been generated that confirms those assumptions.

Therefore, I imaged living embryos derived from GFP-Piwi; His2Av-RFP expressing parents using spinning disk microscopy. Similar to LSFM, spinning disk microscopy allows the imaging of living samples while inducing low phototoxicity (Icha et al., 2017). Embryos during the blastoderm stage exhibited strong Piwi localisation to somatic nuclei (Figure 5.3 A). Interestingly, Piwi was absent in nuclei as early as mitotic prophase, indicating an active transport mechanism for Piwi out of the nucleus. Starting from nuclear cycle 12, Piwi appeared to accumulate at large defined foci which could indicate recruitment of Piwi by nascent RNA. In order to study Piwi's potential engagement with nascent RNA further, I used confocal microscopy to image fixed w^{1118} embryos during the syncytial blastoderm stage (NC 9-13) and probed for Piwi and DNA. Interestingly, while the light sheet and spinning disk data indicated that Piwi is transported out of the nucleus at the beginning of mitosis, a clear nuclear Piwi signal remained detectable during both prophase and anaphase. At the onset of mitotic prophase, chromosomes condensed and Piwi signal did not colocalise with DNA. During anaphase, chromosomes are separated at centromeres and daughter chromatids are moved to opposite poles of the nucleus. Piwi signal strongly resembled chromosome shape. However, again Piwi and DNA did not colocalise (Figure 5.3). Studies showed that some proteins and RNA remain associated with DNA during all phases of mitosis (Black et al., 2016). This suggests that Piwi might be attached to chromosomes probably by binding to chromosome-associated RNA during mitotic cycles.

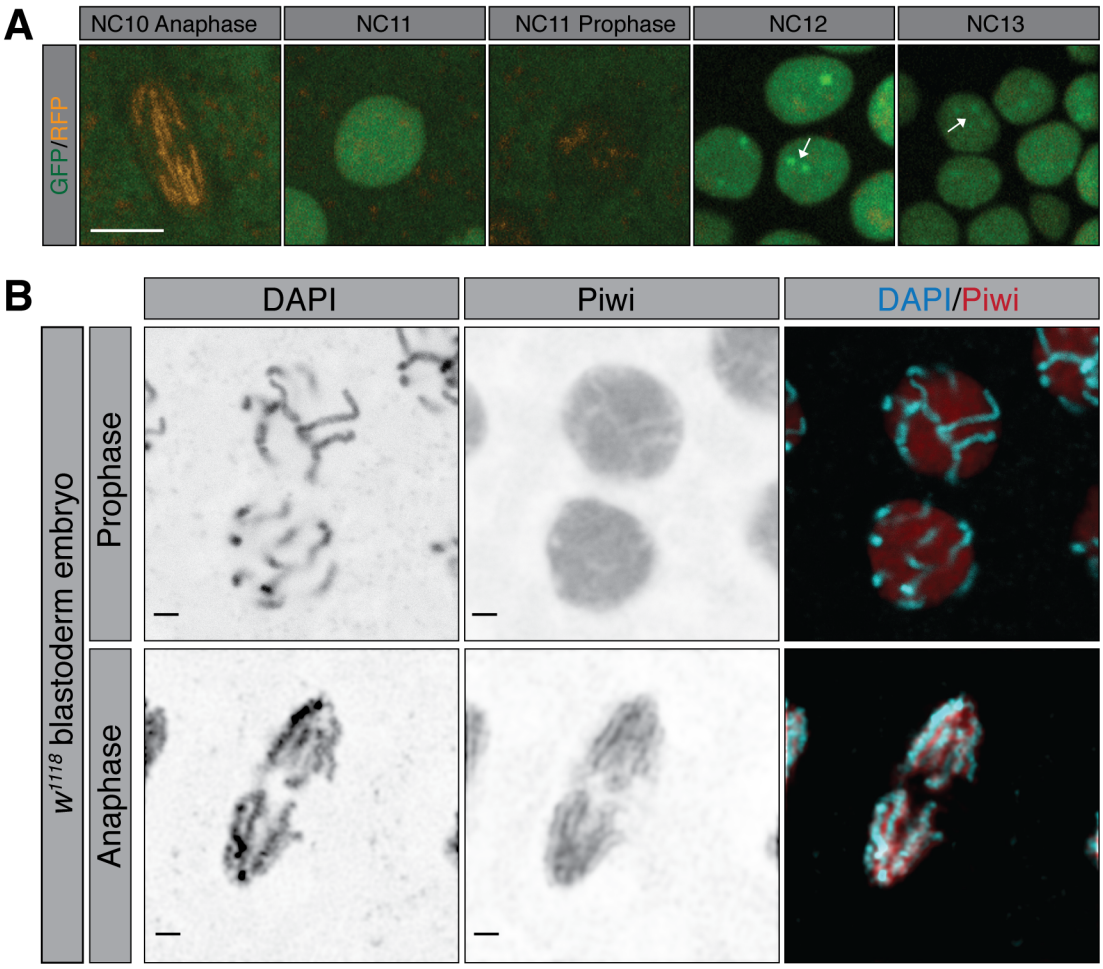

Figure 5.3: Piwi localises to somatic nuclei and forms defined foci but does not colocalise with DNA

A) Spinning disk microscopy images of blastoderm stage embryos derived from GFP-Piwi; His2Av-RFP parents for indicated nuclear cycles (NC). Arrows point towards Piwi foci. Green signal is derived from GFP-Piwi, orange signal from His2Av-RFP. Scale bar is 5μm. **B**) Confocal fluorescent microscopy images staining for DNA and Piwi of blastoderm stage *w*^{1118} embryos. Top panel shows Piwi and DNA localisation during prophase and bottom panel anaphase of mitosis. Scale bar is 1μm.

5.3.2 The PICTS complex and downstream effectors of the piRNA pathway are highly abundant during embryogenesis

Maternal deposition of Piwi suggests a potential function of the piRNA pathway during embryogenesis. However, studies on *Drosophila* ovaries have shown that nuclear Piwi likely serves as a binding platform for additional piRNA pathway-specific proteins. We previously reported that the PICTS complex is required for the induction of piRNA pathway-dependent TGS at transposon insertions in ovaries (Fabry et al., 2019) (see chapter 3).

In order to investigate whether the PICTS complex is maternally deposited together with Piwi or co-expressed during embryogenesis, I performed LSFM on early embryos derived from parents expressing either GFP-tagged Nxf2 or Panx as well as His2Av-RFP. While Piwi localised initially to the posterior pole, both Nxf2 and Panx were evenly distributed within the embryo during the preblastoderm stage (30min AEL) (Figure 5.4 A). Starting at the syncytial blastoderm stage, Nxf2 and Panx were detectable in somatic and pole cell nuclei, thereby co-localising with maternally deposited Piwi. Interestingly, both Panx and Nxf2 showed strong co-localisation with His2Av-RFP during mitotic cycles (Figure 5.4 B-C). This suggests that, like Piwi, the PICTS complex could associate with nascent RNA already during early stages of development. PICTS complex components were present for several hours during embryogenesis and reflected both temporal and spatial dynamics of Piwi.

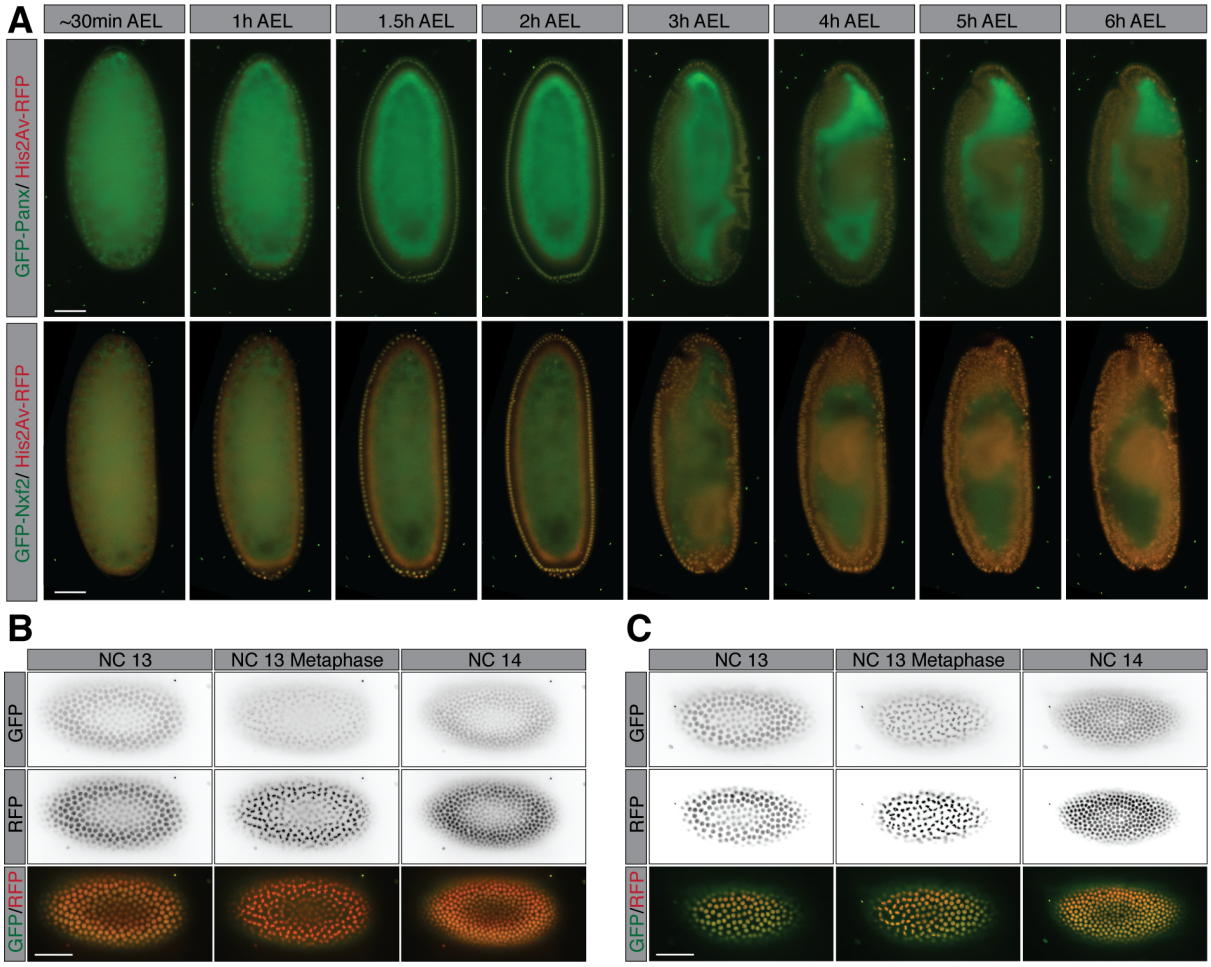

Figure 5.4: Nxf2 and Panx are maternally deposited and abundant throughout early development

A) Live imaging using LSFM of embryos derived from parents expressing GFP-Panx (top) or GFP-Nxf2 (bottom) and His2Av-RFP for several selected time points. Green signal corresponds to GFP, red signal marks nuclei. Scale bar is 50μm. **B**) Same as A) but zoom-in at somatic sheet for GFP-Panx. **C**) Same as B) but for GFP-Nxf2.

piRNA pathway-induced transcriptional gene silencing is not only dependent on Piwi and the PICTS complex but requires several key components of the general silencing machinery. Maternally deposited Piwi was detectable for more than 12h AEL. Therefore, a functional piRNA pathway would require co-expression of downstream effectors during the same time period.

I performed RNA-Seq for various embryogenesis time points in order to identify maternal deposition and zygotic expression of piRNA pathway factors as well as downstream effectors. Maternally deposited mRNAs were reflected by RNA extracted

from 0-2h AEL *w*^{1118} embryos. The maternal to zygotic transition (MZT) occurs during nuclear cycle 14 (2-3h AEL), which includes the degradation of most maternal transcripts as well as activation of the zygotic genome (Tadros and Lipshitz, 2009). MZT is complete by the start of gastrulation (3h AEL). Therefore 3-4h AEL embryos were selected to represent zygotic transcripts. Additionally, I chose 12-13h AEL embryos to represent the time point when 50% of maternal Piwi is degraded (see Figure 5.1 D) and a 17-18h AEL time point to represent late stage embryos, which lack maternal Piwi in somatic cells.

Piwi transcripts seemed to be exclusively maternally inherited as indicated by strong Piwi signal in 0-2h embryos (Figure 5.5 A). Following degradation of maternally deposited mRNAs, Piwi was detected at low levels only (3-4h, 12-13h AEL) and was almost undetectable in late stage embryos (17-18h AEL). However, as described above, zygotic Piwi transcription occurs in late stage embryos (Figure 5.2) but might be below the detection threshold of bulk RNA-Seq due to the low number of pole cells (n= ~24). Additionally, further transcripts of piRNA pathway factors involved in TGS such as Arx and Mael were detected in early embryos but were absent in later stages (Figure 5.5 A).

The PICTS complex acts downstream of Piwi (chapter 3), therefore, I next examined the expression profile for PICTS components. Panx and Nxf2 transcripts showed similar expression patterns throughout embryogenesis (Figure 5.5 B). Both mRNAs were maternally deposited as well as zygotically expressed in early embryos (3-4h AEL). Signal intensity fell sharply in 12-13h embryos and remained at low level in 17-18h embryos. The third component of the PICTS complex, Nxt1, showed similar expression patterns in comparison to Panx and Nxf2 (Figure 5.5. D). However, levels of zygotically expressed mRNAs were higher in comparison to maternally inherited mRNAs, probably reflecting its broader function beyond the piRNA pathway.

The PICTS complex is believed to recruit the general silencing machinery (Batki et al., 2019; Fabry et al., 2019; Murano et al., 2019; Zhao et al., 2019). Key proteins involved in heterochromatin formation that have been shown to be involved in piRNA pathway-dependent TGS are the histone methyltransferase Egg and its co-factor Wde as well as HP1a, Lsd1 and Su(var)2-10 (see chapter 1.3). All showed maternal

deposition of transcripts and similar expression patterns comparable to Nxt1, with signal intensity peaking in early embryos (3-4h AEL) and strongly declining in later stages (12-13h and 17-18h AEL) (Figure 5.5 C).

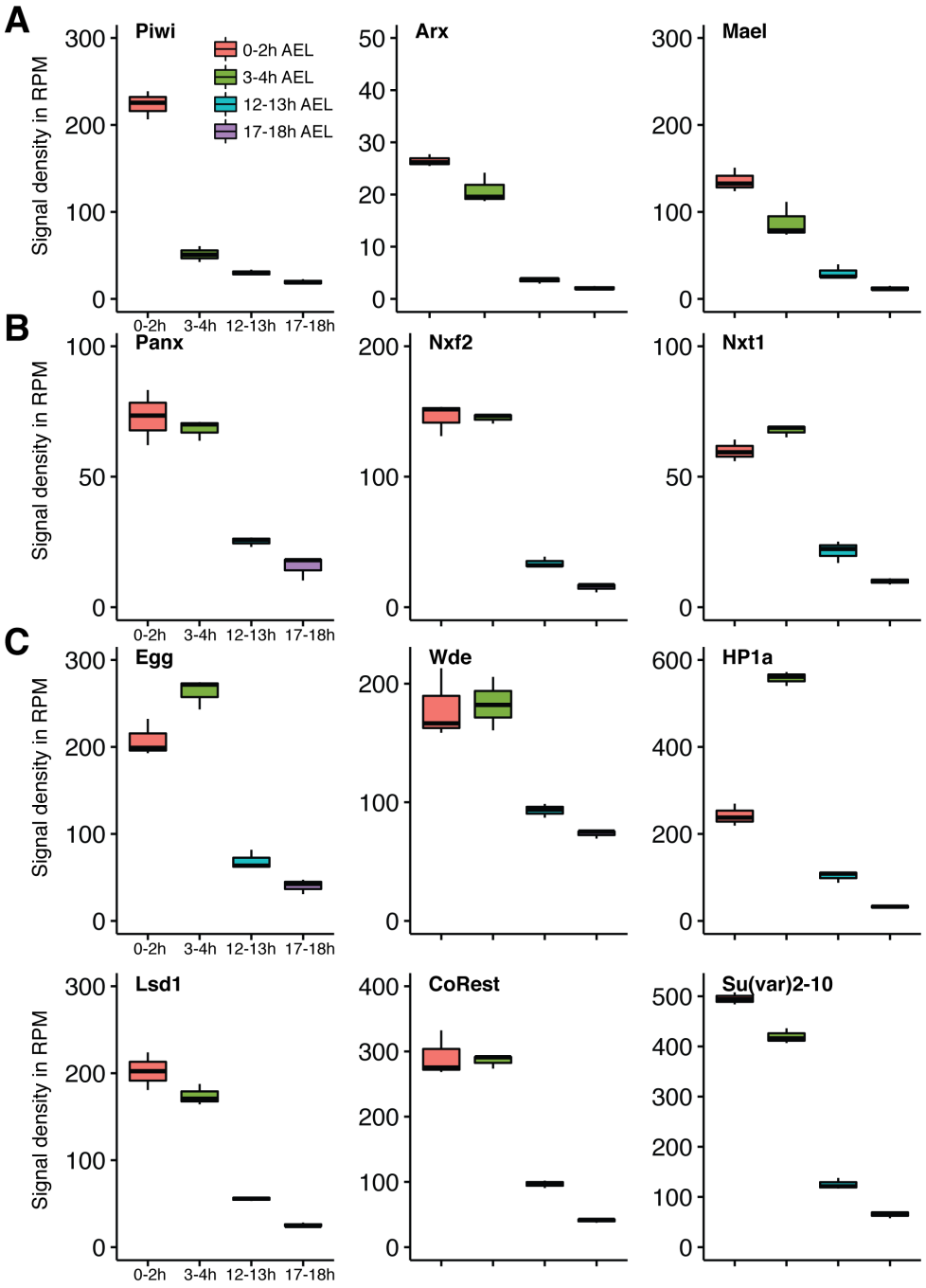

Figure 5.5: piRNA pathway factor transcripts are maternally inherited and expressed during embryogenesis

A) Expression profile for piRNA pathway factors involved in TGS in *w*^{1118} embryos of indicated time points. Signal density in reads per million (RPM). n=3 biological replicates. **B**) Same as A) but for PICTS components. **C**) Same as A) but for factors involved in general TGS.

This data strongly suggests that key effectors of the piRNA pathway-dependent TGS machinery are maternally inherited as mRNA and additionally, with the exception of Piwi, zygotically expressed in the early stages of embryogenesis (3-4h).

5.3.3 Transposons are highly expressed in somatic cells during embryogenesis

The presence of Piwi and the PICTS complex during embryogenesis suggests a function of the piRNA pathway beyond maintenance of genome integrity in adult gonads. However, the targets of an embryogenesis-associated piRNA pathway remained unknown. Previous studies showed that the piRNA pathway controls transposon expression in *Drosophila* ovaries. Therefore, I investigated whether transposon expression correlated with Piwi's presence during embryogenesis.

Twelve different time points of w^{1118} embryos were collected and polyA RNA extracted for sequencing. Time point selection was influenced by major developmental events such as maternal to zygotic transition including degradation of maternal transcripts and zygotic genome activation (stages 2-2.5h and 2.5-3h AEL) as well as gastrulation (3h AEL). RNA-Seq reads were mapped to transposon consensus sequences followed by mapping of unmapped reads to the dm6 reference genome (see chapter 2.20.1). Transposon and genome mapping reads (dm6 reads) were normalised and the ratio of TE/(TE+dm6) determined, thereby stating the contribution of transposon-derived reads to the entire transcriptome.

Pre-MZT embryos (0-2h AEL) showed low levels of transposon mRNA, as expected, probably due to low TE expression during oogenesis. (Figure 5.6 A). However, TE expression increased during development from 0.1% of the entire transcriptome in 0-2h embryos to over 1.4% in 5-6h embryos. Transposon levels remained stable in embryos between 4-8h AEL but sharply decreased in later time points to 0.7% and remained at this level until the end of embryogenesis (8-18h AEL), in accordance with previously published studies (Batut et al., 2013).

In order to evaluate the expression profile of individual transposon families, I further analysed the RNA-Seq data set described above. Interestingly, one transposon

family, *roo*, was highly expressed during embryogenesis. *roo*-derived reads contributed more that 1% to the entire embryonic transcriptome during its peak at 5-6h AEL and made up over 70% of total transposon-derived reads (Figure 5.6 B). In addition, several other transposons including *412*, *copia* and *297* showed highly dynamic expression patterns throughout embryogenesis.

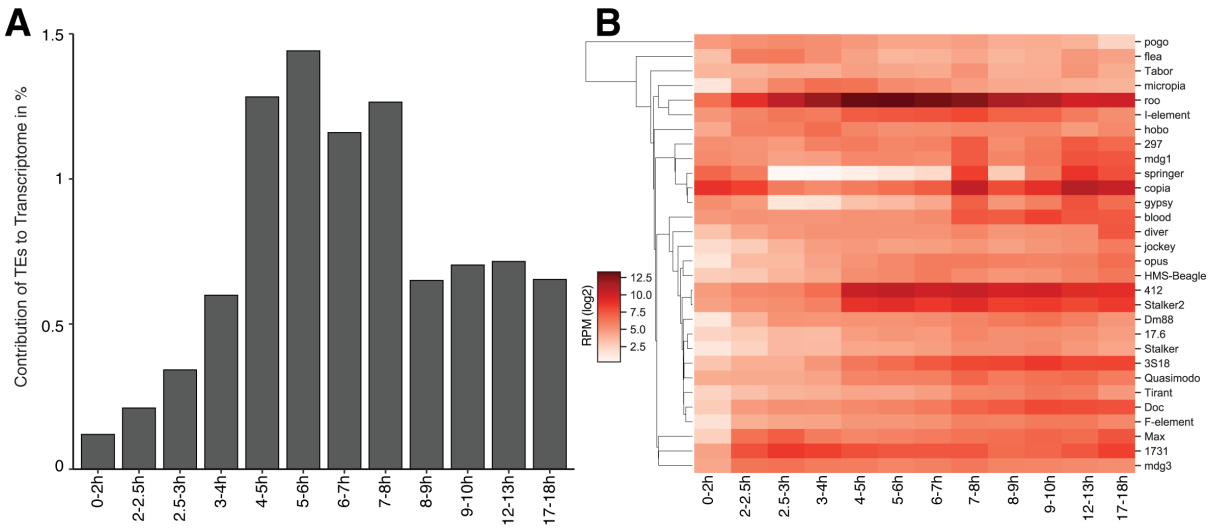

Figure 5.6: Transposon expression during *Drosophila* embryogenesis in w^{1118}

A) Contribution of transposon derived reads to the complete transcriptome of w^{1118} embryos of indicated time points in percent [%]. **B**) Expression data for top 30 expressed transposons during embryogenesis in log2 reads per million (RPM). n=3 biological replicates.

Transposon expression is often associated with germ cells. Only transposition events in cells giving rise to the next generation can be inherited by future generations. However, the presence of the piRNA pathway in somatic nuclei suggests expression of potential targets in somatic cells. While RNA-Seq analysis revealed the timing of transposon expression, it did not provide information on spatial distribution of mRNAs. In order to better understand the localisation of transposon transcripts during embryogenesis, I performed RNA fluorescence *in situ* hybridisation (RNA-FISH) probing for the highest expressed transposon, *roo.* I combined RNA-FISH with immunofluorescence imaging (FISH co-IF) for endogenous Piwi to correlate the spatio-temporal dynamics of the piRNA pathway with putative silencing targets (see chapter 2.7.5).

w^{1118} embryos of selected stages were examined for Piwi and *roo* localisation using confocal microscopy. As expected from LSFM microscopy data (Figure 5.1), Piwi localised to somatic nuclei of early embryos and was detectable throughout early embryogenesis (> 5h AEL) but gradually disappeared as development progressed (Figure 5.7 A). *roo* transcripts were first detectable in gastrulating embryos (stage 6, approx. 3h AEL) and localised to yolk cell nuclei. Stage 11 embryos (approx. 5h AEL) showed strong *roo* signal in somatic cells associated with the somatic mesoderm. This was in accordance with earlier reports (Bronner et al., 1995; Ding and Lipshitz, 1994). For both, stage 6 and 11 embryos, *roo* transcripts colocalised with Piwi positive cells. Late stage embryos (>10h AEL) showed reduced Piwi signal and *roo* transcripts were no longer detectable. The abundance of *roo* transcripts detected by RNA-FISH reflected the RNA-Seq expression profile of *roo*, with transcript abundance peaking around 5-6h AEL and disappearing in late stage embryos (Figure 5.7 B).

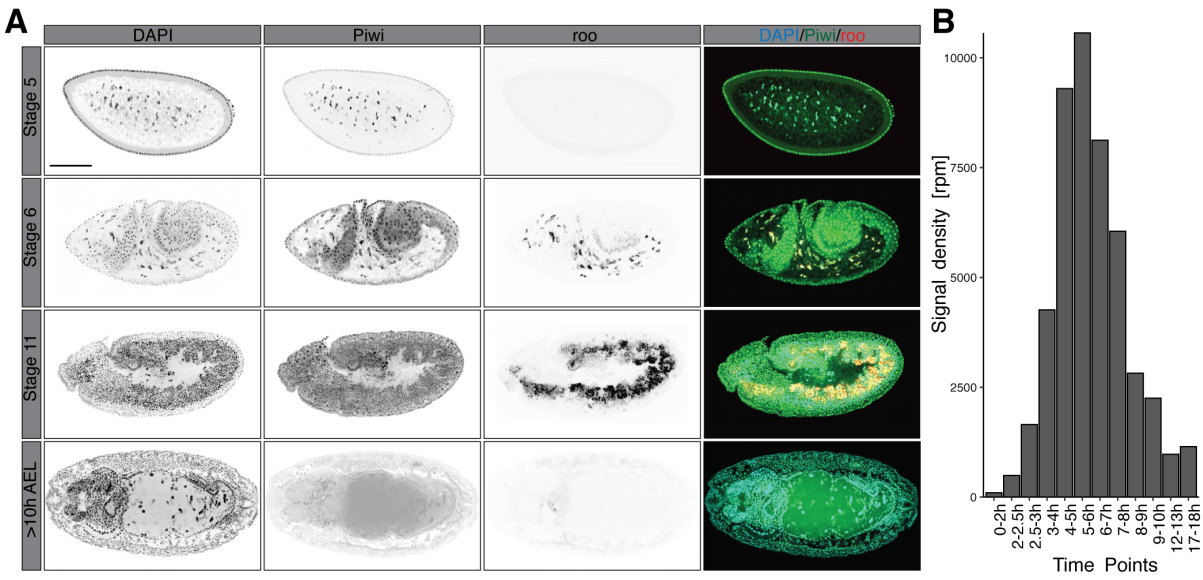

Figure 5.7: Piwi protein localisation overlaps with *roo* expression

A) RNA-FISH co-IF of *w*^{1118} embryos probing for Piwi protein and *roo* transcripts at the indicated stages of embryogenesis. Scale bar is 100μm. **B**) Expression profile for *roo* RNA during the indicated time points in reads per million (RPM) (n=3 biological replicates).

I further investigated whether both expression of *roo* and maternal deposition of Piwi was species specific or whether it is a common feature of other Drosophilidae.

I performed FISH co-IF experiments on *Drosophila simulans* embryos probing again for Piwi and *roo* transcripts. *D. simulans* Piwi shows high protein conversation compared to *D. melanogaster,* indicating cross-species reactivity of our polyclonal Piwi antibody. Similarly, *roo* has been shown to be present in *D. simulans* (de la Chaux and Wagner, 2009).

Staging was performed based on common morphological features of *D. melanogaster* embryos. Piwi protein was detectable in *D. simulans* early embryos. Piwi showed strong maternal deposition and prolonged abundance during embryogenesis highly reminiscent of *D. melanogaster*. As for *D. melanogaster*, Piwi was absent from embryos >10h AEL. *roo* transcripts appeared first in yolk cell nuclei during gastrulation (stage 6) and showed highest signal during stage 11. Late stage embryos were depleted of *roo* transcripts. Comparison of *D. melanogaster* and *D. simulans* revealed similar expression patterns for *roo*. Furthermore, maternal deposition of Piwi protein is conserved between both species, thus strongly indicating a similar relationship between transposons and the piRNA pathway during early embryogenesis in both species.

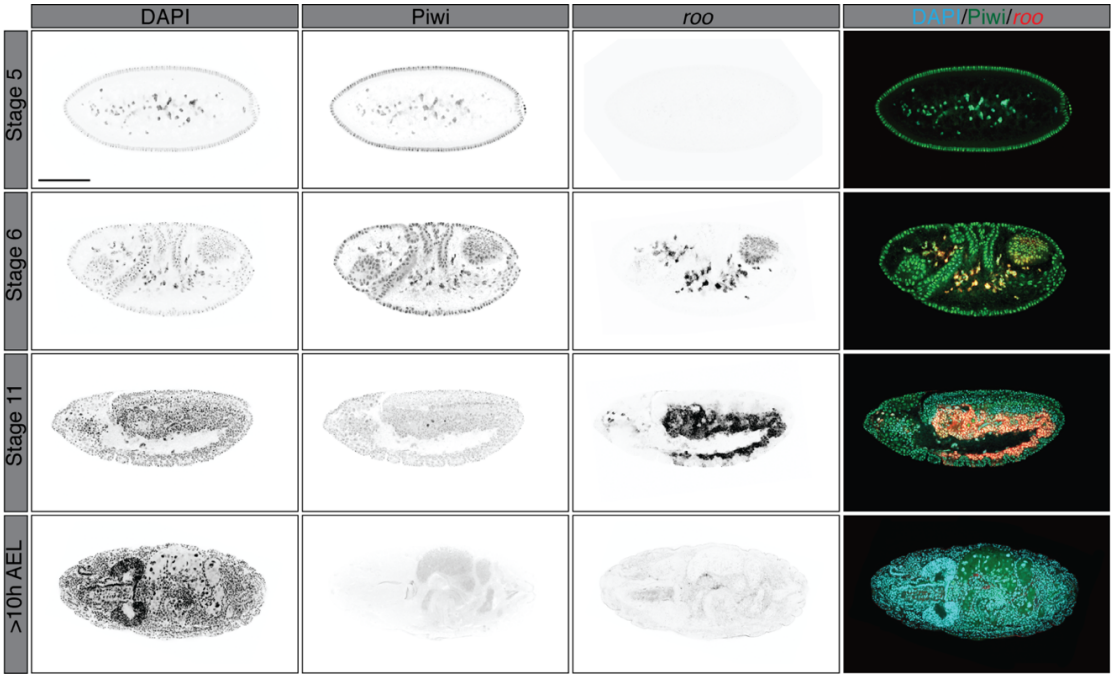

Figure 5.8: Piwi is maternally deposited and *roo* expressed in similar patterns in *Drosophila simulans* compared to *D. melanogaster*

RNA-FISH co-IF of *D. simulans* embryos probing for Piwi protein and *roo* transcripts for the indicated stages of embryogenesis. Scale bar is 100μm.

5.3.4 Transposons are translated during embryogenesis

Transposon transcription was detected by RNA-Seq and confirmed by RNA FISH in somatic cells during embryogenesis. The presence of transposon RNA is an indicator for the potential of effective transposition. However, it does not provide information about further factors required for the transposon lifecycle such as proteins assisting in the integration of transposon DNA into the host genome.

For transposition, most retrotransposons rely on proteins encoded in their open reading frames (ORFs) along with general cellular factors provided by the host cell. Many retrotransposons strongly resemble viruses in their protein-coding capacity. For example, *roo* contains LTRs at its 5' and 3' termini and codes for a single ORF with a predicted protein weight of 272kD. *roo* transcripts code for a Group-specific antigen-like protein (gag), reverse transcriptase (RT/pol) and an envelope protein (env) resembling the Baculovirus F-Protein (Figure 5.9). In addition, two peptidases-like domains (Pep), which might be involved in post-translational cleavage of the primary translation product into separate proteins, can be found. Interestingly, one domain resembles a zinc finger generally associated with binding of upstream activating sequence of histones (zf-H2C2).

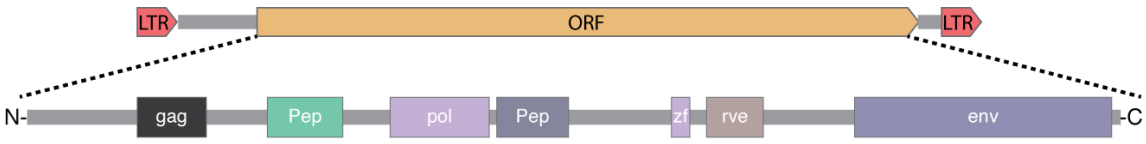

Figure 5.9: *roo* transcripts encode virus-like particles

5' and 3' termini of *roo* transcripts are comprised of LTRs. One open reading frame (ORF) coding for a 272kD protein is present. The encoded protein contains domains associate with Group-specific antigen-like protein (gag), reverse transcriptase (pol) and envelope protein (env) as well as two peptidases (Pep) and a zinc finger associated domain (zf).

We next tested if embryos produce transposon-derived proteins and whether transposon expression during embryogenesis poses a substantial threat to genome integrity. We performed whole proteome quantitative mass spectrometry for three w^{1118} embryo stages (see chapter 2.16). Time points were chosen based on

expression patterns of total transposon reads during embryogenesis (see Figure 5.6). The first time point (0-2h AEL) corresponds to the maternally inherited proteome of pre-MZT embryos, while the second (5-7h AEL) correlates with peak RNA expression of total transposons, including *roo*. The last time point represents the proteome of later stage embryos following the burst of transposon RNA expression (10-12h AEL).

We were able to detect peptides derived from multiple transposon families that showed RNA expression during embryogenesis. 5-7h embryos showed significant accumulation of *roo* particles (padj <0.01) as well as proteins encoded by *412, flea, copia* and *hobo* (Figure 5.10 A, D). Transposons expressed at later stages of embryogenesis such as *blood* (peak expression 9-10h AEL, Figure 5.6) were predominantly enriched in later stage embryos (10-12h) compared to 0-2h AEL embryos (Figure 5.10 B). Interestingly, while translation of transposon-derived proteins correlated with expression patterns, most proteins were still detectable several hours after peak expression, indicating high protein stability. However, while *roo* proteins were strongly detected, other proteins derived from transposon such as *297*, *412* and *mdg1* were less abundant (Figure 5.10 D). Reverse transcription of transposon RNAs performed by transposon-derived proteins produces DNA intermediates that can be integrated into the host genome (Finnegan, 2012). The abundance of *roo* protein suggests a potential continuous threat for genome integrity. However, whether required DNA intermediates required for TE propagation are present in later stages of embryogenesis remains to be examined (Figure 5.10 C).

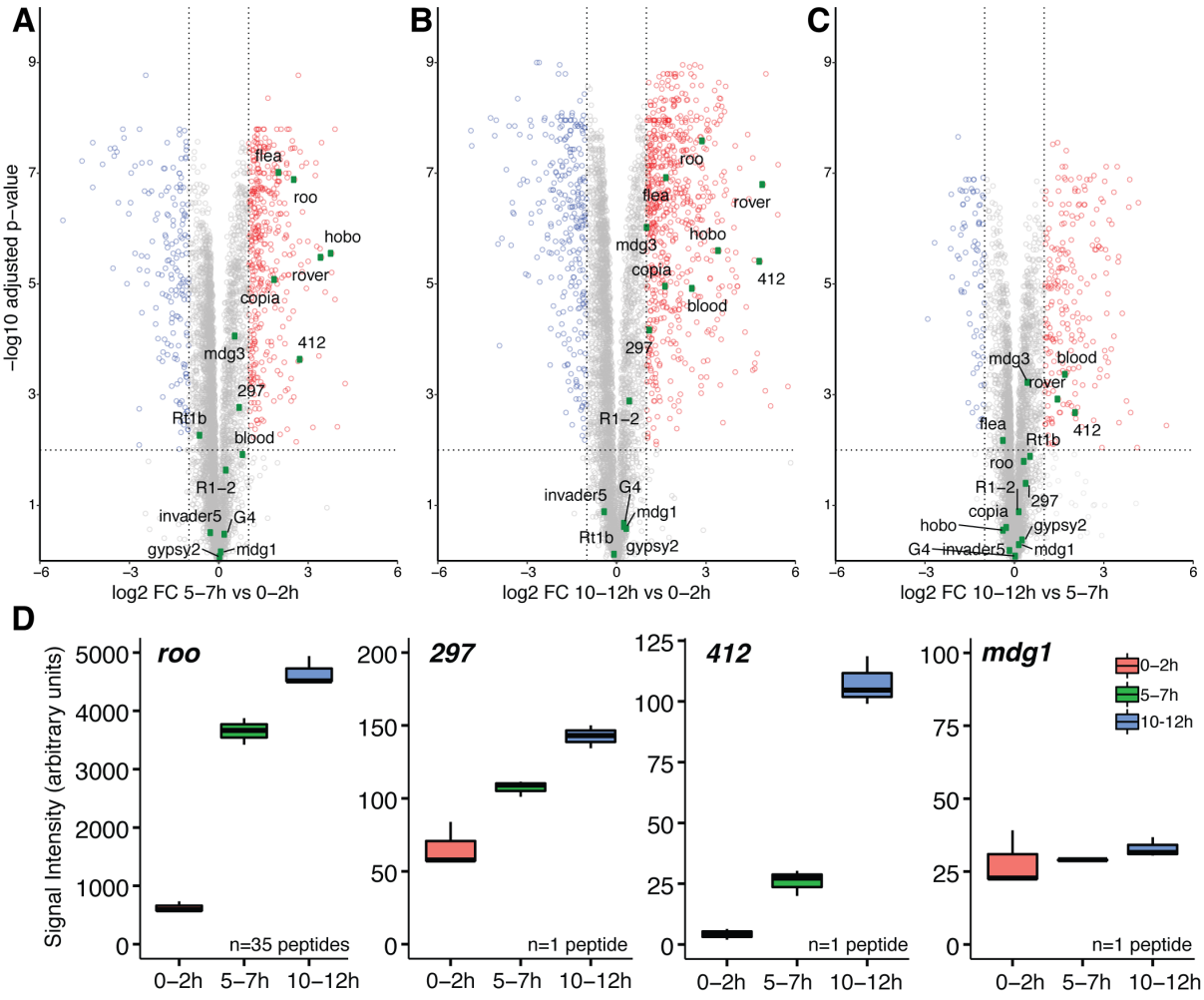

Figure 5.10: Transposon-derived proteins are translated in early and late stage embryos

A) Quantitative mass spectrometry whole proteome data of 5-7h *w*^{1118} embryos compared to 0-2h embryos. Blue and red dots show significance (adjusted p-value <0.01, y-axis) of de- or increased gene-coding proteins respectively. Grey dots represent proteins not changing significantly and/or with foldchange < 2. Green dots represent transposon-derived particles of the indicated family. n= 3 biological replicates. **B**) same as A) but comparing 10-12h versus 0-2h embryos. **C**) same as A) but comparing 10-12h versus 5-7h embryos. **D**) Protein abundance profiles for indicated transposons and time points in arbitrary units from 3 biological replicates. Indicated number of peptides corresponds to peptides for transposon detected by mass spectrometry.

5.3.5 Maternally deposited piRNAs are able to target transposons expressed during embryogenesis

The experiments presented above strongly suggest that a functional, maternally inherited piRNA pathway is active during the first 12h of embryogenesis in non-gonadal somatic cells. Transposons were highly expressed within this period and transcripts localised to Piwi-positive cells further implying a relationship between the piRNA pathway and transposon regulation. However, the piRNA pathway relies on base complementarity of bound piRNAs to target transcripts for its transcriptional gene silencing ability. Therefore, I examined whether piRNAs bound by maternal Piwi protein were in principle able to target transposons expressed during embryogenesis. Additionally, I compared the embryonic piRNA population, which is derived from maternal deposition, with piRNAs expressed during oogenesis to better understand the specific requirements for silencing during each developmental step.

piRNAs are tightly bound by Piwi protein and immunoprecipitation has been widely used to specifically isolate Piwi-bound piRNAs, thus excluding other small RNAs contaminations such as miRNAs or siRNAs. I isolated piRNA-Piwi complexes from 0-8h w^{1118} embryos as well as ovaries as a control using a Piwi-specific antibody followed by small RNA-Seq (see chapter 2.5, 2.14.3). Reads were then aligned to consensus sequences of all present transposons in *Drosophila melanogaster.*

Both maternally deposited and ovarian piRNAs were able to target transposon transcripts based on sequence complementarity as expected (Figure 5.11). Interestingly, antisense piRNAs targeting *roo* were the most abundant species loaded into maternally inherited Piwi with over 16% of total TE-targeting piRNAs. While small RNAs targeting *roo* have already been identified in ovaries and in embryos before (Brennecke et al., 2008), control of *roo* expression by the piRNA pathway has not been reported. The next most abundant piRNA species were able to target *297*, which has been shown to be regulated in ovaries, and the *F-element* with 9% and 6% of total TE-mapping piRNAs, respectively. As expected, piRNAs that target transposons highly regulated in ovaries, such as *mdg1* or *412,* were abundant in ovary-derived Piwi. Interestingly, both *mdg1* and *412* were relatively highly expressed during

embryogenesis. However, both showed only low levels of maternally deposited piRNAs.

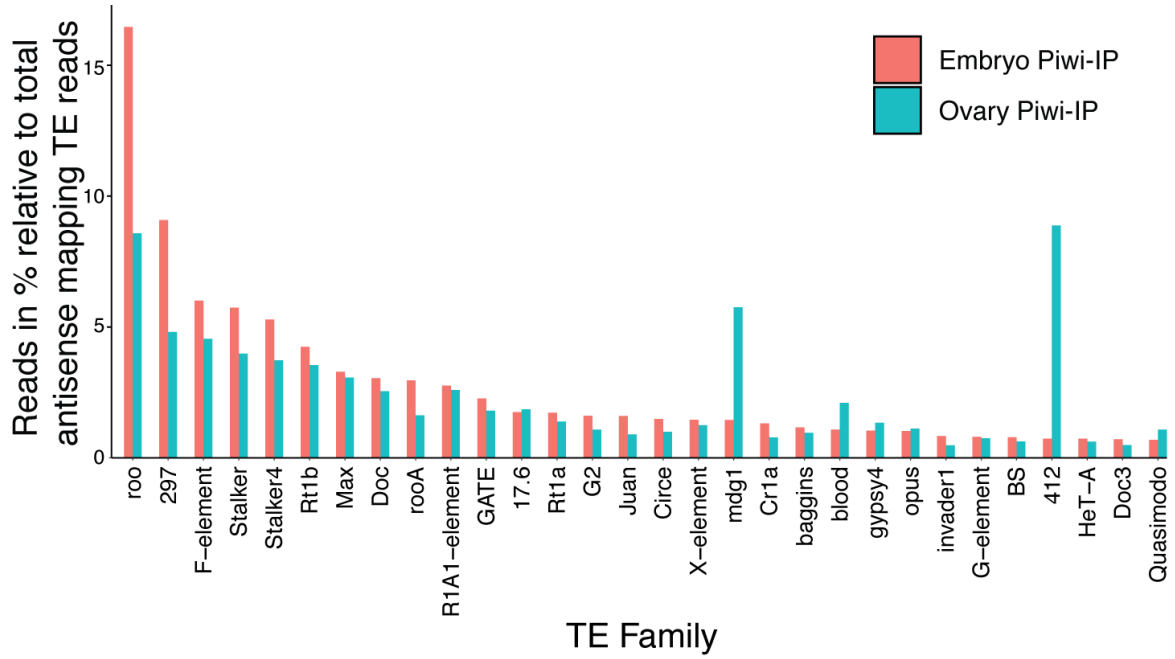

Figure 5.11: Maternally deposited piRNAs target transposons expressed during embryogenesis

Piwi-IP small RNA-Seq for 0-8h w^{1118} embryos (red) or ovaries (blue). Plotted are top 30 antisense piRNA populations targeting indicated TE families. Scale shows [%] of reads mapping to transposons relative to total antisense mapping TE-targeting piRNAs.

5.3.6 Maternally deposited Piwi binds transposon transcripts in pre-MZT embryos

Piwi is believed to interact with nascent transposon transcripts in ovaries. However, the targets of Piwi during embryogenesis remained unknown. Somatic nuclei of pre-MZT embryos show distinct Piwi foci during NC 12-13 (see Figure 5.3 A), potentially suggesting active binding of Piwi to newly transcribed RNA. The majority of zygotic transcripts start to emerge from NC 14. However, previous studies reported that certain transposons, including *297*, *roo*, *F-element* and *Doc*, are expressed during embryogenesis as early as NC 7 and throughout the syncytial blastoderm stage (Kwasnieski et al., 2019). Therefore, Piwi might target early zygotic transcripts including transposon RNA using antisense piRNAs as guides.

In order to identify the targets of Piwi during early development, I performed formaldehyde RNA immunoprecipitation followed by sequencing (fRIP-Seq) of maternally deposited Piwi. 0-2h AEL w^{1118} embryos were crosslinked and Piwi-associated RNAs were isolated using a Piwi polyclonal antibody targeting the amino-terminal domain of Piwi. Interestingly, Piwi-bound RNAs showed significant enrichment of transposon transcripts. ($p<0.05$). *297* was 3.7-fold enriched in comparison to the input control (Figure 5.12 A). Further transcripts including *roo*, *F-element* and *Doc* were significantly enriched but below the 2-fold threshold. Only one transposon, *R1A1*, showed a significant depletion, however, below the 2-fold threshold. As an additional control I performed fRIP-Seq on embryo lysates as described above using IgG antibodies instead of anti-Piwi to exclude technical artifacts arising from RNA retention on beads independently of the used antibody. fRIP-Seq of embryo lysates using IgG antibodies, however, showed no enrichment in transposon RNA when compared to the input, as expected (Figure 5.12 B).

This data suggests an interaction between Piwi and a defined set of transposon transcripts prior to MZT. However, whether Piwi binds early zygotically transcribed transposon RNAs or maternally deposited transcripts remains unknown.

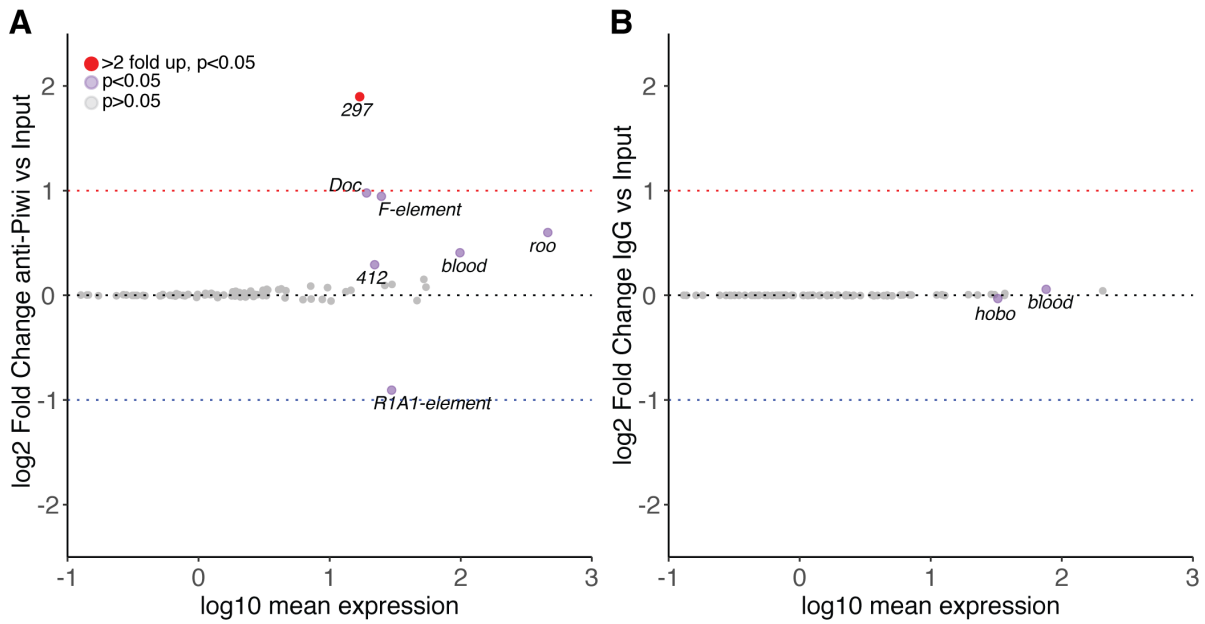

Figure 5.12: Piwi binds transposon RNA in early embryos before MZT

A) fRIP-Seq MA plot showing log2 fold change enrichment of transposon RNAs for Piwi pulldown versus input in 0-2h *w*^{1118} embryos. Y-axis shows log10 mean expression across replicates. Non-significantly changed transposons (p>0.05) are represented in grey, significantly changed transposons (p<0.05) with fold changes <2-fold in purple, and significantly changed transposons with fold changes >2-fold in red. n=3 biological replicates. **B**) Same as A) but fRIP-Seq data generated using IgG antibody instead of anti-Piwi.

The fRIP-Seq data suggested that Piwi binds to transposon transcripts in pre-MZT embryos. Thus, I next investigated if Piwi interacts with other RNAs by analysing enrichment of gene-derived transcripts. 46 transcripts were significantly (p<0.05) enriched at least 2-fold in Piwi pulldowns (Figure 5.13 A). Two enriched genes were associated with early zygotic transcription (CG6770, Hsp70) and none with maternally deposited mRNAs (see 2.20.5). Gene ontology (GO) analysis showed significant association of enriched genes with protein binding (fold enrichment 3.16, FDR 0.0323), however, 18 genes were uncharacterised, thus lacking GO annotations. 19 transcripts were depleted in Piwi pulldowns in comparison to the input control. GO analysis of depleted genes revealed a strong association with ribosomal constituents (fold enrichment 67.80, FDR 3.6E-16), probably representing rRNA contaminants in the input sample that were not present in the Piwi pulldown due to stringent washing conditions. fRIP-Seq performed on 0-2h embryo lysate using IgG antibodies instead

of Piwi-targeting antibodies showed no > 2-fold significantly enriched genes (Figure 5.13. B). Similarly, depleted genes were either of ribosomal origin or snoRNAs (not shown).

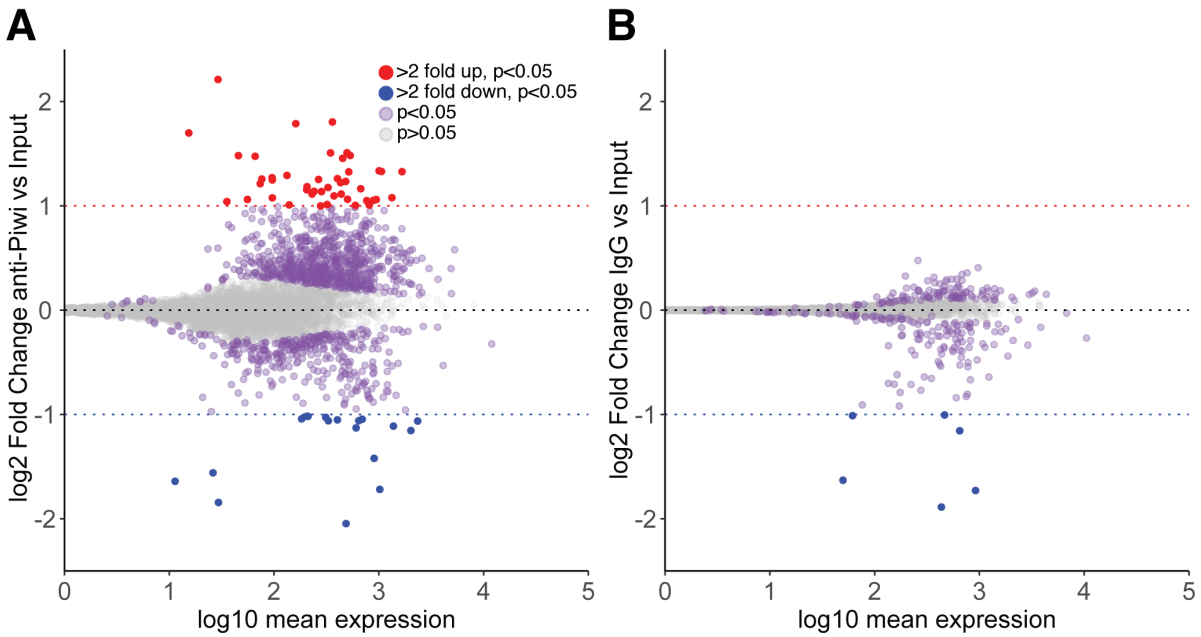

Figure 5.13: Piwi binds gene-derived transcripts in early embryos before MZT

A) fRIP-Seq MA plot showing log2 fold change enrichment of gene-derived RNAs for Piwi pulldown versus input in 0-2h w^{1118} embryos. Y-axis shows log10 mean expression across replicates. Non-significantly changed genes (p>0.05) are represented in grey, significantly changed genes (p<0.05) with fold changes <2-fold in purple, and significantly changed genes with fold changes >2-fold in red. n=3 biological replicates. **B**) Same as A) but fRIP-Seq data generated using IgG antibodies instead of anti-Piwi.

5.3.7 piRNA-targeted transposon insertions show epigenetic changes characteristic of transcriptional gene silencing

The data presented above strongly suggest the presence of a functional TGS-dependent piRNA pathway during embryogenesis. Maternal deposition of piRNA-Piwi complexes, which were able to target transposons expressed during early development, as well as the abundance of the PICTS complex in somatic cells further indicated that transposons might be regulated by the inherited piRNA pathway during embryogenesis. I further detected binding of maternal Piwi to transposon transcripts, including *roo* mRNA. However, evidence for transcriptional gene silencing of transposons by the piRNA pathway in embryos remained elusive.

Transcriptional gene silencing in gonads is accompanied by the conversion of euchromatic regions in close proximity to active transposons targeted by piRNAs into densely packed heterochromatin (Ohtani et al., 2013; Wang and Elgin, 2011). We previously showed that this process is dependent on the PICTS complex and correlates with the deposition of H3K9me3 marks at regulated transposon bodies and surrounding genomic loci (Fabry et al., 2019) (see chapter 3).

H3K9me3 occupancy at transposon insertions can be measured using ChIP-Seq. Since many transposon insertions reside in constitutive heterochromatin and are therefore decorated by H3K9me3 marks partially independently of the piRNA-pathway, only euchromatic insertions are usually examined for accumulation of repressive chromatin marks. However, the genomic location of transposon insertions is not conserved between different *Drosophila* strains. Therefore, I used whole genome sequencing (WGS) to *de novo* identify transposon insertions in w^{1118} flies used for most of my experiments (see 2.14.5). I was able to identify the location of over 600 transposons in euchromatin that were previously not annotated in the dm6 reference genome by using the TEMP algorithm (see chapter 2.20.5).

In order to understand whether the inherited piRNA pathway factors induce TGS at targeted transposon insertions, I performed H3K9me3 ChIP-Seq on w^{1118} embryos for 2h intervals covering the presence of maternal Piwi during early stages of development. Additionally, I included a late embryonic time point (16-18h AEL) as

well as ovaries and adult somatic tissue (heads) to compare the dynamics of TGS throughout the life of *Drosophila*.

I first evaluated the H3K9me3 occupancy of the highest expressed transposon during embryogenesis. piRNAs loaded into maternally inherited Piwi were highly enriched in antisense *roo* sequences and evidence suggested binding of Piwi to *roo* transcripts during early stages of embryogenesis (see Figure 5.12 A). Euchromatic *roo* insertions were highly abundant in w^{1118} flies. I was able to identified 117 individual insertion loci for *roo*, the highest number of all TE families present in *Drosophila melanogaster* (see 2.20.5). H3K9me3 is not only deposited at transposon sequences but spreads several kb from the insertion site. Therefore, I plotted the signal intensity at 10kb up- and downstream of each individual transposon insertion (see 2.20.4). Early embryos (2-4h AEL) showed little accumulation of H3K9me3 signal at *roo* insertions (Figure 514 A). However, as development progressed, H3K9me3 was highly and specifically enriched close to roo insertions. H3K9me3 occupancy at *roo* insertions peaked between 6-10h AEL. Deposition of repressive marks correlated well with the expression profile of *roo*. However, peak H3K9me3 signal lagged behind peak RNA expression by approximately 2h, revealing the dynamics of deposition and the likely requirement of transcription prior to conversion into heterochromatin. Interestingly, deposition of repressive marks trailed the direction of transcription and showed higher signal enrichments downstream of transposon insertions. I previously showed that this signature is characteristic for piRNA pathway-induced transposon silencing in *in vitro* models (see chapter 3).

Maternally deposited Piwi is highly stable but is not detectable in somatic nuclei of late stage embryos (chapter 5.3.1) (Brennecke et al., 2008). H3K9me3 signal at *roo* insertions of 16-18h AEL embryos showed greatly reduced signal intensities. Similarly, adult tissues, such as heads, that lack a functional piRNA pathway showed no enrichment of H3K9me3 at *roo* insertions. The same could be observed in adult ovary tissue which express all piRNA pathway components, thus suggesting an embryo-specific, transcription-dependent silencing process at *roo* insertions.

While ovaries showed no transcriptional silencing of *roo*, other transposons associated with piRNA-dependent regulation during oogenesis such as *Doc, mdg1*

and *412* showed a clear accumulation of H3K9me3 marks that were absent in somatic embryo tissues during all time points (Figure 5.14 B, 5.15 B). This suggests a highly tissue-specific and piRNA-dependent regulation of different transposon families throughout *Drosophila* development.

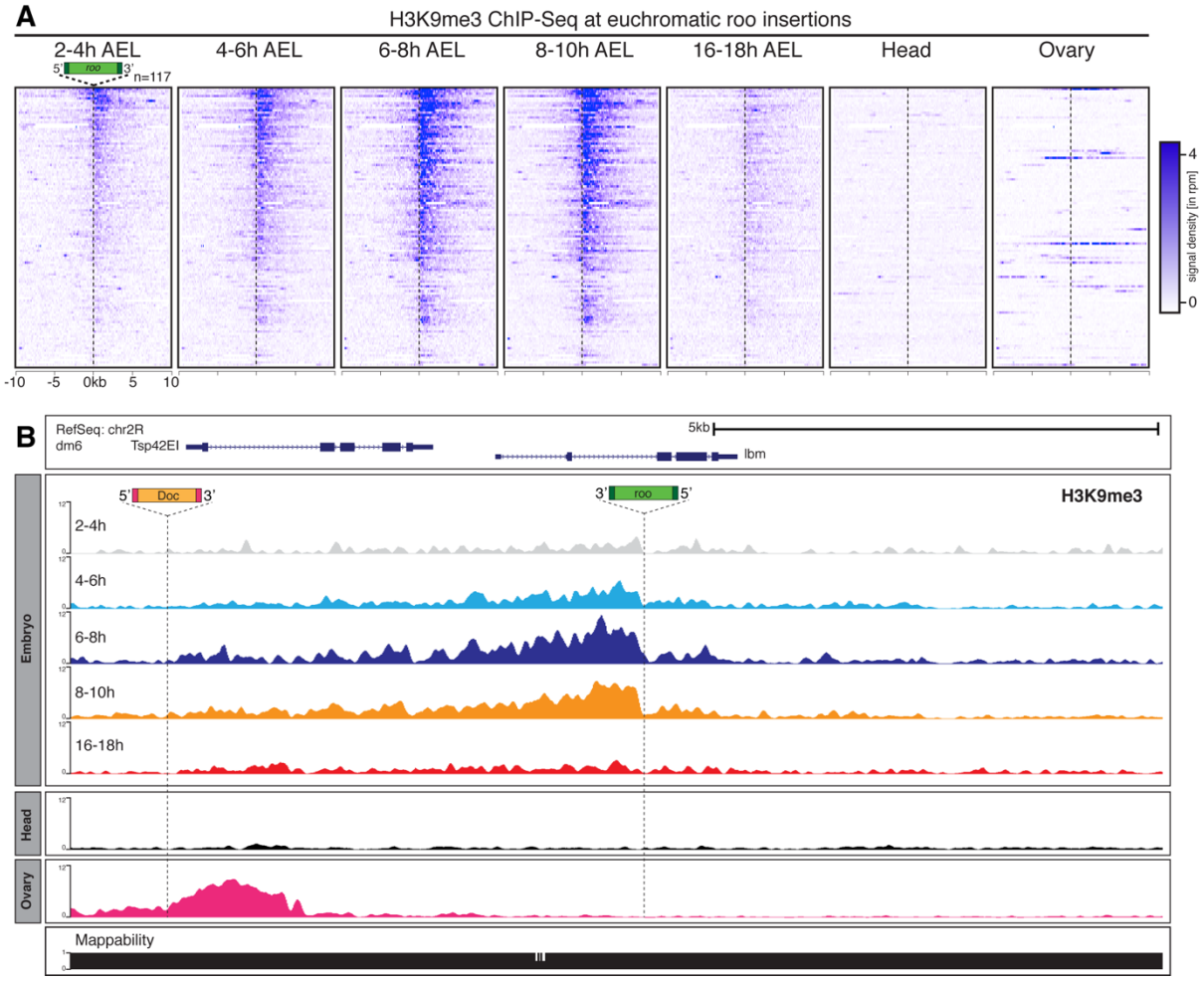

Figure 5.14: H3K9me3 marks are deposited at *roo* insertions in early embryos

A) H3K9me3 ChIP-Seq for indicated w^{1118} embryo stages and tissues at euchromatic *roo* insertions (n=117). Signal is plotted 10kb up- and downstream of transposon insertions. Direction of loci was arranged to reflect transcription (downstream) of transposon insertion. Signal density in reads per million (RPM), n=3 biological replicates. **B**) UCSC genome browser snapshot of H3K9me3 ChIP-Seq data for chromosome 2R. Indicated are w^{1118}-specific *roo* and *Doc* insertions.

The data presented above further suggests transcriptional gene silencing of *roo* during embryogenesis by the piRNA pathway. However, whether this is the case for other active transposons that are targeted by embryonic piRNAs remained unclear. Therefore, I systematically assessed H3K9me3 occupancy at transposon insertions that are highly expressed and targeted by piRNAs during embryogenesis. For this analysis, I again used the list of identified euchromatic transposon insertions described above. Additionally, I evaluated repressive marks at transposons that were highly expressed in embryos but showed little deposition of complementary piRNA species.

The transposon *297* was expressed during embryogenesis and showed high targeting potential by maternally inherited piRNAs (Figure 5.15 A). Genomic loci in close proximity to *297* insertions (n=20) showed similar deposition dynamics of H3K9m3 in comparison to *roo* (Figure 5.15 B). However, while H3K9me3 signal at *roo* insertions peaked between 6-10h AEL, H3K9me3 deposition was highest between 2-6h AEL suggesting earlier regulation in comparison to *roo*.

In contrast, *mdg1* and *412*, were highly expressed during embryogenesis but lacked maternal deposition of piRNAs (Figure 5.15 A). H3K9me3 occupancy remained on a low, constant level throughout embryogenesis and an enrichment signature comparable to *roo* or *297* insertions was absent. In fact, *mdg1* and *412* H3K9me3 resembled occupancy of randomly selected euchromatic loci (n=200). However, as expected, insertions in ovary genomes acquired a strong TGS signature (Figure 5.15 B).

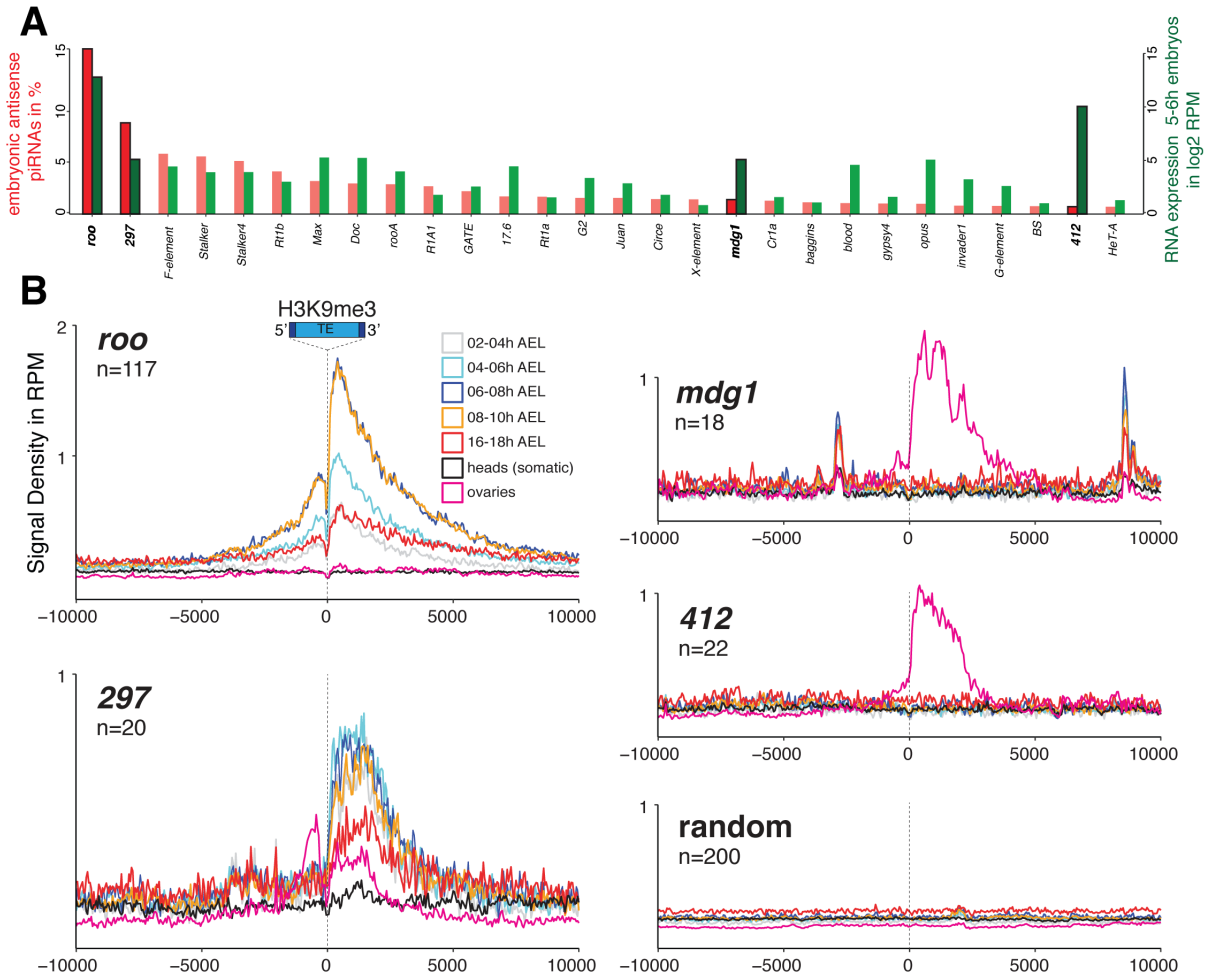

Figure 5.15: H3K9me3 accumulates at targeted transposon insertion during early embryogenesis

A) Bar graph summarising the abundance of maternally inherited Piwi-bound piRNAs (embryo Piwi-IP small RNA-Seq) targeting indicating transposon family in red (n= 1) as well as RNA expression for 5-6h *w*^{1118} embryos in green (n= 3 biological replicates). Both scales show reads per million (RPM). **B**) Line graphs represent collapsed H3K9me3 signal from all euchromatic transposon insertions for indicated time points, tissues and random euchromatic locations. n represents number of euchromatic insertions. Signal density in reads per million (RPM) averaged from 3 biological replicates.

5.3.8 Degradation of maternally deposited Piwi leads to deregulation of transposons and affects MZT

The data presented above further implied that the piRNA pathway is responsible for transposon silencing during embryogenesis. However, we still lacked direct evidence that maternally deposited piRNA-Piwi complexes control transposon expression by induction of transcriptional gene silencing.

Therefore, using the AID degron approach, I next investigated the impact of loss of maternally deposited Piwi on transposon silencing and gene expression in embryos during the MZT (2-3h AEL). I previously showed that treatment of embryos with the plant hormone auxin results in rapid degradation of AID-tagged Piwi in the presence of E3 ubiquitin ligase TIR1 (Figure 5.16 A) (see chapter 4.3.2.2). This system enables a comprehensive study of the function of maternally deposited Piwi by examining the impact of Piwi degradation on transposon expression and deposition of repressive marks.

Embryos derived from flies expressing GFP-AID-Piwi and AtTIR1 under control of the ubiquitin promoter were collected for 1h and treated for an additional 2h in the presence or absence of 5mM auxin (see chapter 2.18). This generated embryos 2-3h AEL, thereby corresponding to the maternal to zygotic transition (MZT) phase of embryogenesis, which includes the degradation of maternal transcripts as well as the major activation of zygotic transcription. Following auxin treatment, RNA was extracted and mRNAs sequenced. Reads were aligned first to consensus sequences of all transposon families, followed by unique alignment to the dm6 genome of unmapped reads (see chapter 2.20.1).

Transposon expression was mildly affected by depletion of maternal Piwi in 2-3h embryos (Figure 5.16 B). While most transposons showed no significant change between conditions, only *roo* was significantly (padj<0.001) upregulated more than 2-fold (2.8-fold). Additionally, the transposons *Stalker2*, *micropia*, *flea*, *mdg3* and *Doc* were significantly enriched in auxin treated embryos, however, fold changes were below a 2-fold threshold.

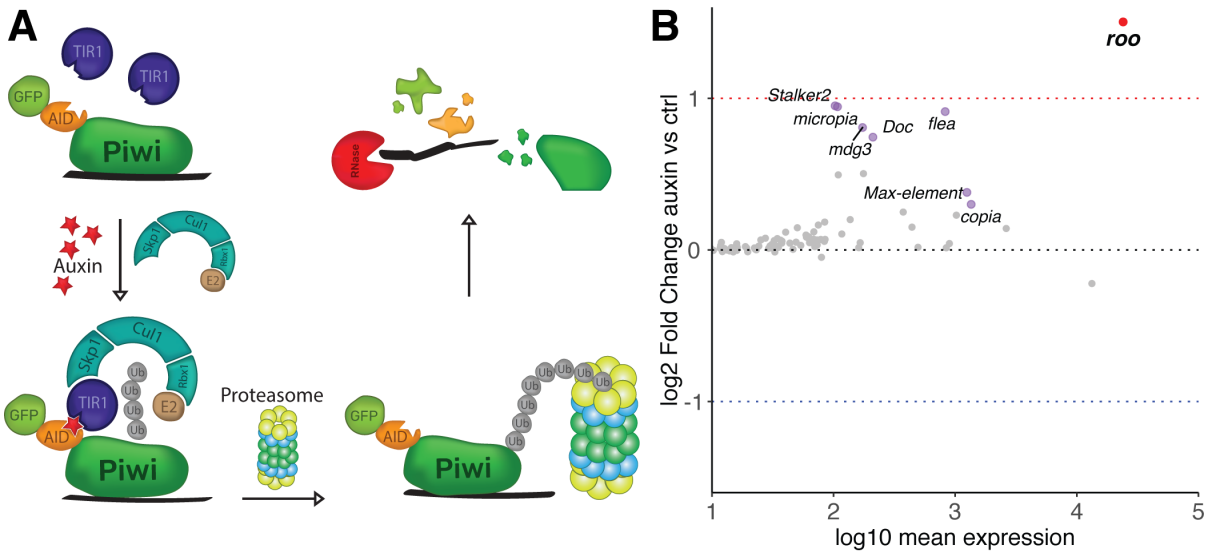

Figure 5.16: Depletion of maternally deposited Piwi leads to deregulation of transposons during MZT

A) Cartoon depicting the degron strategy. **B**) Differential expression of transposons for GFP-AID-Piwi; AtTIR1 embryos treated with 5mM auxin (auxin) compared to untreated control (ctrl). Grey dots show transposons not significantly changed (padj >0.05). Purple dots correspond to significantly deregulated transposons (padj <0.05). Red dots show significantly deregulated transposons with fold changes of >2. n=3 biological replicates.

Additionally, several genes were deregulated upon Piwi depletion (Figure 5.17 A). Most significantly changed genes (>2-fold) were downregulated (n=55), while upregulation of genes by more than 2-fold was observed less frequently (n=30). I further characterised deregulated genes by categorising their origin into two categories: (1) genes that were exclusively maternally deposited but lacked zygotic early transcription and (2) genes expressed by the early zygotic genome but lacked maternal deposition (see chapter 2.20.5). Interestingly, downregulated transcripts were highly enriched in exclusively maternally deposited RNAs (n=36, 66% of all downregulated genes) (Figure 5.17 B). In contrast, upregulated transcripts were primarily derived from genes only zygotically transcribed at MZT (n=27, 90% of all upregulated genes).

This data indicates a shift in MZT timing upon Piwi depletion and suggests accelerated turnover of maternally deposited mRNAs, while zygotic transcripts emerge earlier in comparison to untreated embryos.

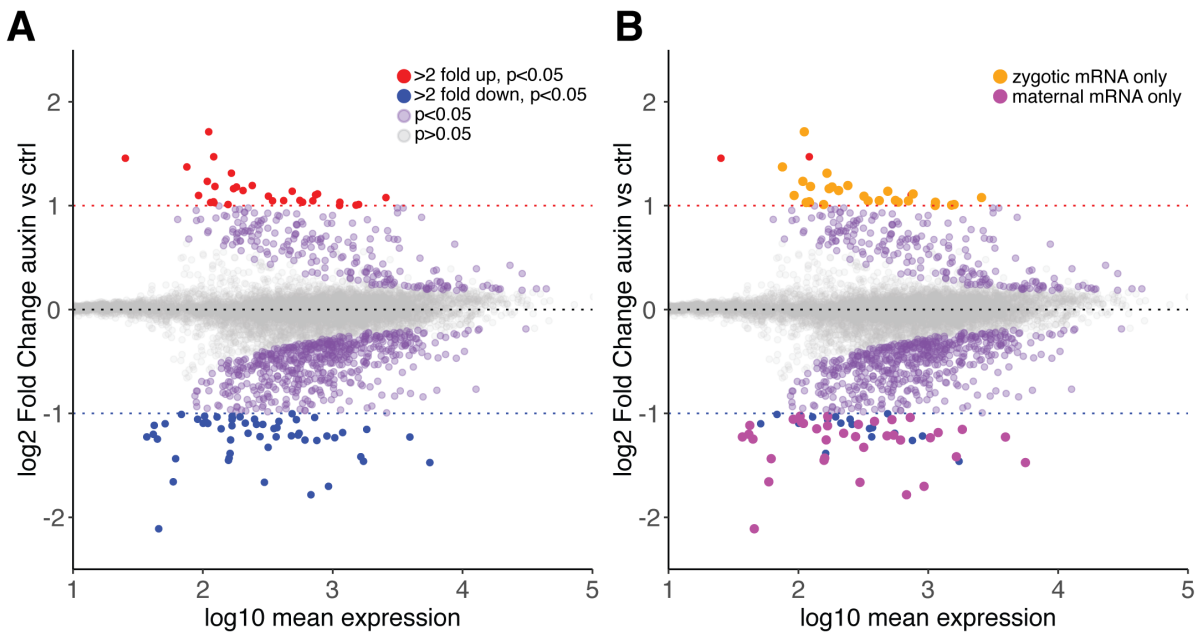

Figure 5.17: Maternally deposited Piwi impacts transcripts at MZT

A) Differential expression of transposons for GFP-AID-Piwi; AtTIR1 embryos treated with 5mM auxin (auxin) compared to untreated control (ctrl). Grey dots show genic transcripts not significantly changed (padj >0.05). Purple dots correspond to significantly deregulated transcripts (padj <0.05). Red dots show significantly deregulated transcripts with fold changes of larger than 2. n=3 biological replicates. Blue dots show transcripts significantly downregulated with fold changes of less than -2. **B**) Same as A). Yellow dots show genes associated with early zygotic expression and no maternal deposition. Magenta dots represent transcripts exclusively maternally deposited but not expressed during MZT.

Auxin is a plant-derived hormone. Although previous studies suggest that auxin in small concentrations has a neglectable impact on *Drosophila* development (Bence et al., 2017; Trost et al., 2016), I tested whether auxin treatment of w^{1118} embryos was responsible for the deregulation of transposons and genes rather than degradation of maternal Piwi. I evaluated gene and transposon expression in embryos of the same developmental stage and treatment as described above by RNA-Seq.

Treatment of w^{1118} 2-3h embryos with 5mM auxin showed a minor impact on gene expression compared to PBS treated controls (Figure 5.18). Overall, genes were expressed at similar levels across conditions and correlated greatly (R^2=0.99). Similarly, transposons were mostly unaffected. Eight transposons were slightly above the >2-fold threshold in treated embryos in comparison to control. However,

upregulated transposons were expressed at very low overall levels. Expression levels of moderately and highly expressed transposons including *roo,* did not change between conditions, indicating that auxin treatment in embryos lacking both GFP-AID-Piwi and AtTIR1 had minimal, if any, impact on gene and TE expression in accordance with the literature and is unlikely to be responsible for effects observed in the experiments presented above.

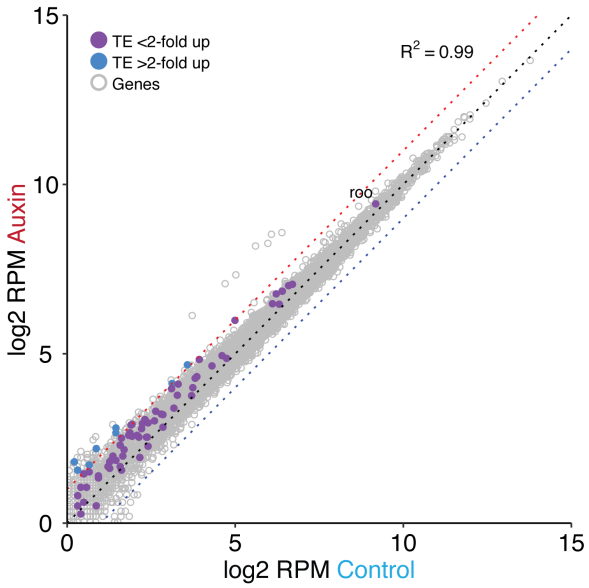

Figure 5.18: Auxin treatment has minor impact on gene and transposon expression of *w*^{1118} embryos during MZT

Scatter plot showing expression of transposons and genes of 2-3h *w*^{1118} embryos treated with 5mM auxin (Auxin) or PBS (Control). Grey dots correspond to expressed genes. R^2 value indicates correlation between auxin treated and control embryos for genes only. Blue dots represent transposons greater than 2-fold upregulated in auxin-treated embryos and purple dots transposons with differential expression less than 2-fold (n=1 biological replicate).

5.3.9 Maternally deposited Piwi induces epigenetic changes at targeted transposons and affects viability of embryos

Degradation of maternally deposited Piwi had a significant impact on gene and transposon expression. Especially *roo*, which showed highly dynamic accumulation of repressive chromatin marks during development in *w*^{1118} embryos, was deregulated

upon Piwi degradation (see Figure 5.16 A). However, whether deregulation was caused by transcriptional gene silencing and epigenetic changes required more careful examination.

In order to understand the impact of Piwi depletion on the epigenetic landscape at transposon insertions, I performed H3K9me3 ChIP-Seq on embryos derived from GFP-AID-Piwi and OsTIR1 expressing parents. OsTIR1 was used instead of AtTIR1, which was used in previous experiments, due to increased fertility of this stock. Embryos were collected for 1h and treated with 5mM auxin or PBS as a negative control for an additional 6h. This generated 6-7h AEL embryos, which corresponds to the peak accumulation of repressive H3K9me3 marks in w^{1118} embryos at *roo* insertions (see chapter 5.3.7).

I next investigated the abundance of H3K9me3 signal at the consensus sequence of transposons that showed specific accumulation of repressive marks at insertions during embryogenesis as well as strong maternal deposition of targeting piRNAs (*roo*, *297*). As a control, I further investigated H3K9me3 occupancy at transposons that showed no accumulation of H3K9me3 and lacked deposition of targeting piRNAs despite high transcriptional activity in w^{1118} embryos, as shown previously (*mdg1*, *412*).

H3K9me3 signal was highly enriched over the consensus sequence of *roo* in untreated embryos, as expected from w^{1118} experiments (see Figure 5.14, 5.19 A). However, treatment with 5mM auxin had a severe impact on H3K9me3 occupancy. Signal density decreased significantly by 46% ($t = 4.2318$, $df = 2.2051$, p-value = 0.04334, Welch Two Sample t-test) in treated embryos compared to the PBS control (Figure 5.19 B). However, the signal density over the *297* consensus sequence was only mildly affected with a non-significant reduction of signal between treated and control embryos of 17% ($t = 1.0229$, $df = 3.9824$, p-value = 0.3644) (Figure 5.19 A-B). H3K9me3 deposition at *297* insertions peaked early during embryogenesis (2-4h AEL) suggesting that *297* is an early target of maternally deposited Piwi. Therefore, the weak impact on H3K9me3 is likely due to the timing of auxin treatment. Embryos were collected for 1h and degradation of maternally deposited Piwi required a further

~30min, therefore, deposition of H3K9me3 could have occurred before sufficient Piwi degradation was achieved.

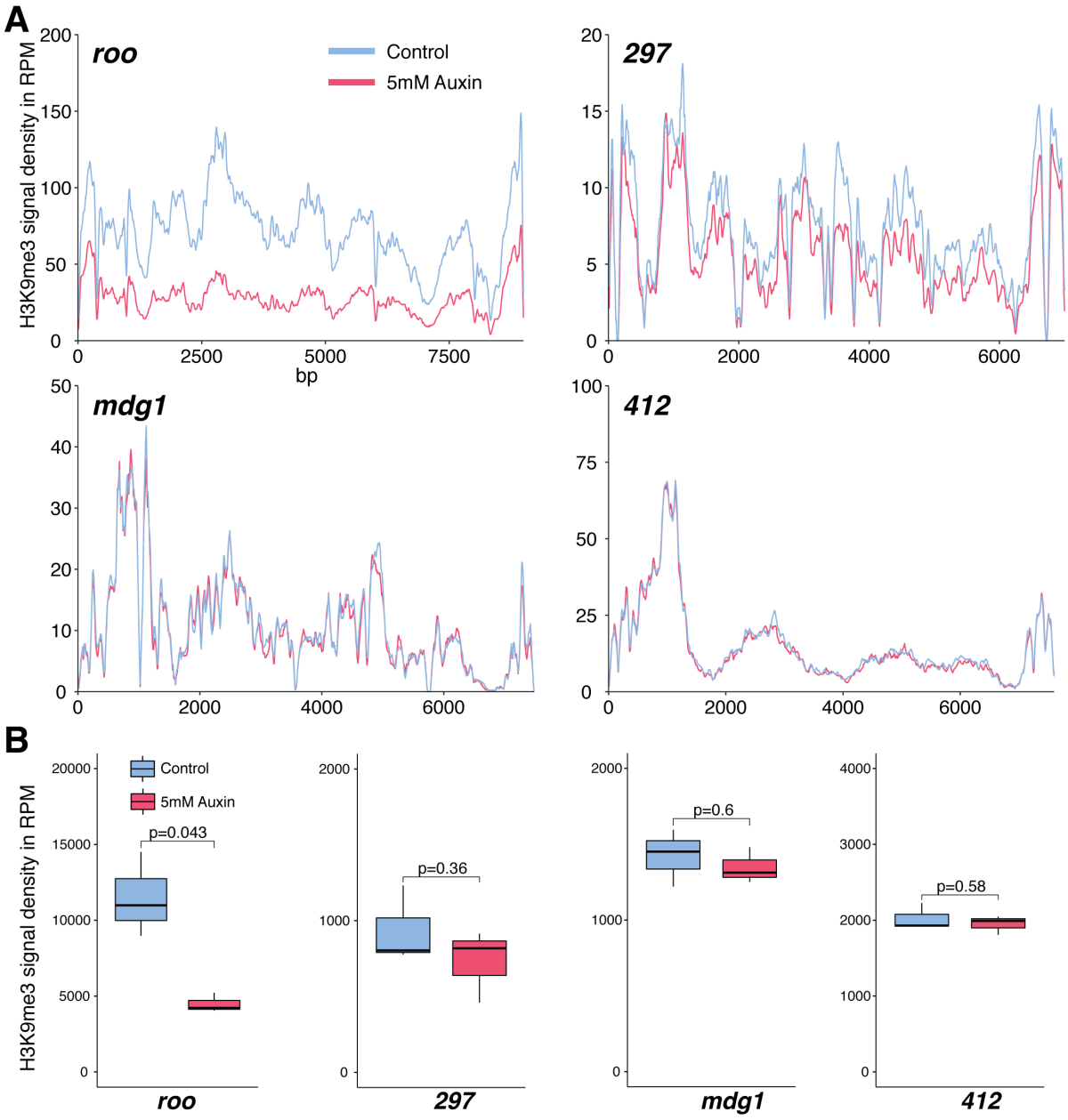

Figure 5.19: Maternally deposited Piwi induces epigenetic changes at targeted transposons

A) Density plot of H3K9me3 ChIP-Seq data over the indicated transposon consensus sequence for 6-7h GFP-AID-Piwi; OsTIR1 embryos treated with 5mM auxin or PBS (Control). Signal density in reads per million (RPM). Signal from concatenated biological replicates (n=3). **B**) Quantification of collapsed signal intensities from A) for individual replicates. Statistical significance is depicted as indicated p-value (Welch two sample t-test).

Transposons with low targeting potential of maternally deposited piRNAs did not show changes in H3K9me3 signal upon auxin treatment. Both, *mdg1* and *412* H3K9me3 levels remained stable with non-significant changes < 4% (t = 0.57769, df = 3.3606, p-value = 0.5999) and < 3% (t = 0.60564, df = 3.6235, p-value = 0.5806, Welch two sample t-test) in treated versus untreated embryos respectively (Figure 5.119 A-B).

Lastly, I examined the impact of depletion of maternally deposited Piwi on embryo survival. Embryos were treated as described above either with auxin or PBS as negative control. Auxin treated embryos showed a reduced hatching rate compared to PBS treated embryos (Figure 5.20). However, further experiments are necessary to confirm this initial observation.

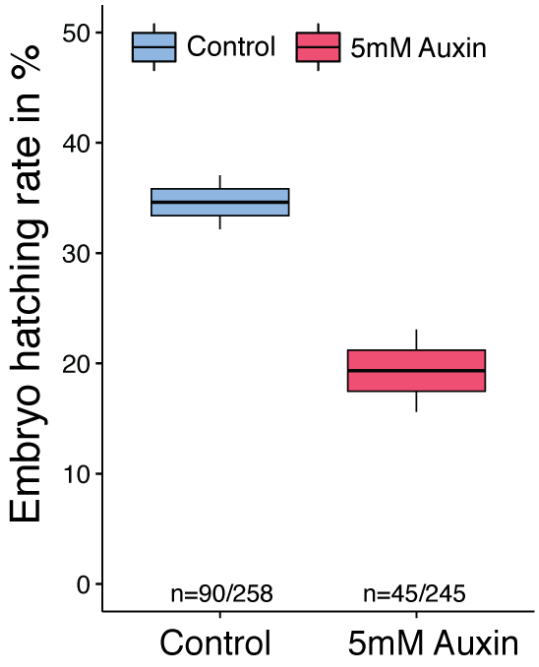

Figure 5.20: Degradation of maternally deposited Piwi affects viability of embryos

Hatching rate of GFP-AID-Piwi; OsTIR1 embryos treated with PBS (Control) or 5mM auxin for 2h. Boxplots show percentage of treated embryos that hatched into larvae. n indicates number of hatched embryos/ total counted embryos from 2 biological replicates.

5.4 Conclusion

The work presented in this chapter revealed a pivotal role of the piRNA pathway in transposon silencing during *Drosophila* embryogenesis.

Piwi protein is maternally inherited through the oocyte and localises to defined nuclear foci in embryonic cells of somatic origin and germ line progenitors. In addition, components of the piRNA pathway-dependent TGS machinery including the PICTS complex and downstream effector proteins, are maternally deposited or expressed in early embryos. Therefore, piRNA-guided TGS could occur during embryogenesis.

Transposons that are targeted by maternal piRNA-Piwi complexes are strongly expressed during the first hours of development and localise to somatic cells. Piwi engages early transposon transcripts likely guided by Piwi-bound piRNAs. Transposon insertions targeted by maternal piRNAs accumulate repressive H3K9me3 marks. The signal intensity at transposons shows highly dynamic changes during embryogenesis. While H3K9me3 signal peaked shortly after transposon peak expression, H3K9me3 was lost in embryos during later stages, strongly correlating with the disappearance of maternally deposited Piwi. This suggests that some epigenetic repressive chromatin states are maintained in a Piwi-dependent manner during embryogenesis.

Auxin-induced degradation of maternally inherited Piwi during early embryogenesis led to upregulation of the highly targeted transposon *roo* and subsequent loss of H3K9me3 marks at transposon insertions, indicating a direct relationship between maternal deposition of piRNA-Piwi complexes and transcriptional gene silencing of transposons in the developing embryo. Maternally deposited piRNAs, therefore represent a carrier of transgenerational epigenetic inherited information required for transposon silencing during early development of *Drosophila* embryos.

6 Discussion and Outlook

The piRNA pathway has been characterised as a crucial mechanism for small RNA-based transposon regulation in animal gonads. The work presented in this thesis has greatly contributed to our understanding of the mechanisms involved in transcriptional gene silencing and has revealed a function for the piRNA pathway in transposon silencing during *Drosophila* embryogenesis.

6.1 piRNA-guided TGS is dependent on the PICTS complex

piRNA pathway-induced transposon silencing is dependent on Piwi and piRNAs able to target nascent transposon transcripts. Recent studies associated the gonad-specific protein Panx with induction of TGS (Sienski et al., 2015; Yu et al., 2015). However, the mechanism linking small RNA-mediated target recognition with silencing events has remained unknown.

Our data demonstrated that the PICTS complex, comprised of Panx, Nxf2 and Nxt1, is essential for co-transcriptional gene silencing. Loss of either Nxf2 or Panx, resulted in deregulation of transposons *in vitro* and *in vivo* (Figure 3.1, 3.3). Additionally, epigenetic marks at transposon loci such as H3K9me3 were strongly decreased upon loss of PICTS, while permissive histone marks such as H3K4me2 increased. Appearance of permissive marks correlated with transcriptional output of affected transposons (Figure 3.2, 3.3-3.6).

The UBA domain of Nxf2 interacts with the carboxy-terminal region of Panx. The amino-terminal region of Panx does not contain any identifiable domains (Figure 3.7, 3.9). However, our data suggests the presence of a nuclear localisation sequence (NLS) at the amino-terminus of Panx. Deletion of this region had no effect on the assembly of the PICTS complex but prevented the import into the nucleus, thus confirming the dependency of the complex on the amino-terminal region of Panx for its nuclear localisation (Figure 3.8). Unsurprisingly, Panx-ΔN was unsuccessful in rescuing transposon silencing in Panx depleted OSCs likely due to improper

localisation. However, even forced nuclear import of Panx-ΔN did not rescue the silencing ability. This suggests that PICTS complex-dependent transposon silencing heavily depends on the amino-terminal region of Panx (Figure 3.7, 3.8). Indeed, we found that the silencing ability of the PICTS complex resides in this unstructured region (Figure 3.15).

We further revealed that binding of Nxt1 to the PICTS complex is necessary for its function but not for the interaction between Panx and Nxf2. Mutations in the NTF2 domain of Nxf2 abolished binding of Nxt1 to Nxf2, while its interaction with Panx was unaffected. Again, transposon silencing in rescue experiments was unsuccessful with Nxf2 unable to bind Nxt1 (Figure 3.10, 3.11). This data indicates that binding of Nxf2 and Panx alone is not sufficient to induce piRNA-dependent transcriptional silencing. Nxt1 binding is indispensable for the function of the PICTS complex.

The PICTS complex changes the epigenetic state of genomic regions in close proximity to active transposon insertions (Figure 3.2, 3.4-3.6), suggesting a direct or indirect recruitment of histone methyltransferases such as Egg as previously reported, thereby causing heterochromatin formation and transcriptional repression (Rangan et al., 2011; Sienski et al., 2015; Yu et al., 2015). Tethering individual components of the PICTS complex to RNA was sufficient to induce epigenetic changes at our sensor sequence (Figure 3.12-3.14). All three components were able to induce strong silencing of the RNA tethering reporter and showed high deposition of repressive chromatin marks (Figure 3.14). While tethering of a single component of the complex was successful in silencing the reporter, it is likely that our tagged components are incorporated into functional PICTS complexes with endogenous proteins. Therefore, silencing is likely to be carried out by properly assembled PICTS complexes. Tethering the individual PICTS components to DNA showed similar results. However, while Panx tethering was able to induce strong silencing and epigenetic changes, Nxf2 showed only a moderate effect and Nxt1 was unable to silence our reporter (Figure 3.12, 3.13). I also showed that tethering the amino-terminus of Panx directly to DNA was sufficient to induce silencing (Figure 3.15). This result is consistent with the failure of the PICTS complex lacking the amino-terminus of Panx to rescue *mdg1* silencing in our OSC assay (Figure 3.7), thereby linking both observations.

While I was able to identify the silencing domain of the PICTS complex, the exact mechanism of the amino-terminal silencing ability remains unknown. A direct interaction between Panx and Egg has not been reported, but some evidence suggests that the SUMO ligase Su(var)2-10 interacts with both Panx and Egg (Ninova et al., 2020a; Ninova et al., 2020b), therefore giving a possible explanation for Panx-dependent recruitment of downstream effectors and subsequent heterochromatin formation. However, further research is needed to confirm the functional relationship between association of those proteins and piRNA pathway-dependent heterochromatin formation.

Depletion of Nxf2 or Panx had a profound impact on the localisation of some transposon mRNAs. Cells depleted of either factor showed high levels of retained *mdg1* transcripts in the nucleus. RNA-FISH experiments revealed distinct nuclear foci indicating strong accumulation of transcripts at specific genomic locations. (Figure 3.16). One explanation is that Piwi holds nascent transposon mRNA in place and recruits the PICTS complex to the location of transcription thus shutting down expression by orchestrating epigenetic changes to histone states. Loss of the PICTS complex could result in the accumulation of transcripts due to a failure in TGS induction. Another possibility is that Piwi could trap transcripts at sites of transcription, thereby preventing their nuclear export. The nuclear *mdg1* RNA foci detected in PICTS-depleted cells might be a result of phase separation caused by Piwi-mediated RNA retention resembling other Argonaute-associated RNA-protein condensates such as P granules in *C. elegans* (Seydoux, 2018). Thereby, Piwi could serve as a fail-safe mechanism for transposon silencing in cells lacking PICTS. Removal of Piwi, however, resulted in the release of *mdg1* transcripts and export to the cytoplasm (Figure 3.16). Loss of Piwi likely prevents the recognition of transposon transcripts by the piRNA pathway machinery entirely and instead allows canonical export of transposon mRNA to the cytoplasm by Nxf1-Nxt1. This could also explain our observation of increased cell mortality for Piwi knockdowns in comparison to depletion of PICTS components. Piwi loss is accompanied by high cell toxicity, while depletion of the piRNA pathway-specific components of the PICTS complex, Panx and Nxf2, only has a moderate impact on cell mortality (data not shown). This could be due to

Piwi reducing export and translation of transposon mRNA in PICTS depleted cells by preventing recognition by the canonical export machinery, thereby maintaining genome integrity (Figure 6.1).

However, nuclear retention of transposon transcripts was only observed for *mdg1*, while *gypsy* transcripts showed only little change in localisation upon depletion of Piwi, Nxf2 or Panx. This indicates that localisation regulation mechanisms might be specific to some transposon families and are not a universal feature of all piRNA pathway-controlled transposons.

Our study supports a model in which piRNA-Piwi complexes scan nascent transcripts and recognise active transposons by sequence complementarity with the Piwi-bound piRNA. The precise recruitment mechanism of the PICTS complex remains elusive. Previous studies, together with our data, showed some association of Panx and Nxf2 with Piwi (Batki et al., 2019; Fabry et al., 2019; Sienski et al., 2015; Yu et al., 2015). However, we were unable to validate this interaction in our IP mass spectrometry experiment (Fabry et al., 2019). One possible reason for this is that it is likely that only a small fraction of the PICTS complex interacts with Piwi at any given time. Therefore, mass spectrometry assays might not be sensitive enough to detect the transient interaction of the PICTS complex with Piwi. Following the recruitment of the PICTS complex, the downstream silencing machinery is likely to induce transcriptional gene silencing by deposition of H3K9me3 marks and subsequent heterochromatin formation (Figure 6.1).

The piRNA pathway co-opted Nxf2 and Nxt1 for transposon control. Nxf2 belongs to the nuclear export factor (NXF) family that includes Nxf1, which is a central component of the canonical mRNA export machinery and functions together with its co-factor Nxt1 (Fribourg et al., 2001; Herold et al., 2001; Herold et al., 2000). Interestingly, the gonad-specific protein Nxf3 has recently been described to be involved in piRNA cluster transcript export (ElMaghraby et al., 2019; Kneuss et al., 2019), illustrating another example of co-option of NXF family members by the piRNA pathway. The function of the testis-specific Nxf4 has yet to be revealed. Tissue-specific expression and diversity of NXF family members seems to be conserved in mammals (Yang et al., 2001), potentially suggesting a conserved function in

transposon control. However, more research is needed to understand the function of different NXF family members in other animals.

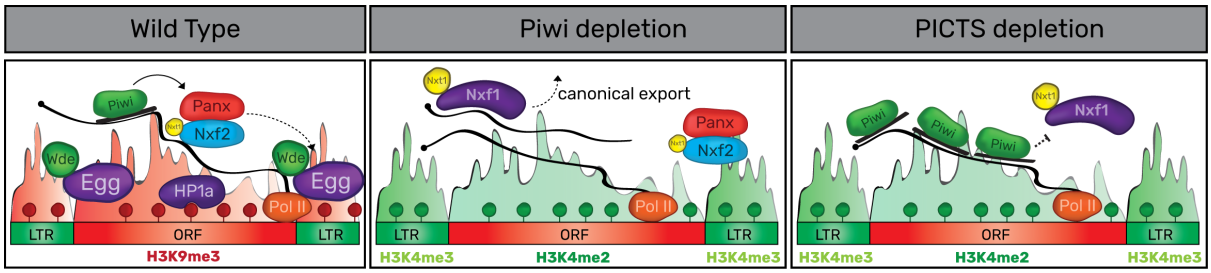

Figure 6.1: The PICTS complex silences transposons by deposition of repressive chromatin marks

Model for piRNA-guided co-transcriptional gene silencing. **Wild Type**: Piwi guided by piRNAs scans the genome for nascent transposon transcripts. Target engagement results in conformation change of Piwi and recruitment of the PICTS complex. The PICTS complex recruits the general silencing machinery of the cell, thus shutting down transcription. **Piwi depletion**: Transposon transcripts are not recognised by Piwi and are canonically exported by the Nxf1/Nxt1 complex. **PICTS depletion**: Transposon transcripts are identified by Piwi and trapped at loci of transcription. Transposon expression is not silenced due to the absence of the PICTS complex but transcripts are protected from canonical export.

6.2 Approaches for the degradation of maternally deposited proteins

The piRNA pathway has been studied extensively in *Drosophila* ovaries. However, Piwi shows strong maternal deposition in embryos (Brennecke et al., 2008), suggesting a function for the piRNA pathway during early development. This initial observation has been made over 10 years ago. However, due to the maternal deposition of Piwi and the limited tools available to inactivate Piwi's function in early embryos, studying the role of the piRNA pathway during early *Drosophila* embryogenesis experimentally has been very challenging in the past. Targeting Piwi or its downstream factors by traditional genetic methods in ovaries or embryos is not feasible. Disturbing the piRNA pathway at transcriptional level in ovaries leads to

sterility. Additionally, maternal deposition of proteins reduces the efficiency of transcriptional approaches in embryos greatly.

The methods described in chapter 3 provide successful alternative approaches to study the function of the piRNA pathway during embryogenesis. Firstly, I delocalised Piwi or Nxf2 form the nucleus thereby preventing transcriptional gene silencing. Secondly, I induced degradation of maternally deposited Piwi protein.

Changing the location of piRNA pathway proteins has the advantage that target proteins are not removed from the embryo. Therefore, this approach, in theory, is reversible. However, genetically encoded traps such as JabbaTrap make it difficult to return tethered proteins to their original location as long as the protein trap is in place. Additionally, the protein trap has to be expressed from the zygotic genome as maternal deposition risks disturbing the piRNA pathway during oogenesis. While there is limited conditional control of those genetic approaches, there are also advantages. Continued expression of protein traps ensures permanent relocalisation of target proteins, therefore ensuring indefinite deactivation of the piRNA pathway.

Trapping of proteins using JabbaTrap was effective *in vitro* (Figure 4.2). Nuclear proteins were efficiently tethered to lipid droplets in OSCs following co-expression of GFP-tagged targets and JabbaTrap. However, the precise dynamics of this process remain elusive since both GFP-tagged target and JabbaTrap were expressed simultaneously. *In vivo* application of the JabbaTrap tool led to delocalisation of GFP-tagged Piwi during oogenesis (Figure 4.3). Similarly, trapping of maternally deposited Piwi protein using JabbaTrap was effective in later stage embryos (Figure 4.4). The delayed onset of zygotic transcription during embryogenesis, however, limits the biological questions that can be addressed using this method. With promoters I used, JabbaTrap was not expressed during early embryogenesis (< 5h AEL). Therefore, the impact of Piwi during this time of development cannot be assessed. However, maternally inherited Piwi is highly abundant in somatic nuclei for several hours after egg laying and effects of Piwi delocalisation in later stages of embryogenesis can be examined using this approach. Using promoters of genes expressed earlier during embryogenesis could potentially increase the temporal resolution required to study early development. Microinjection

of JabbaTrap mRNA in early embryos has been used to study effects of Egg during early embryogenesis (Seller et al., 2019). This technique could be used to trap Piwi at embryogenesis stages even before ZGA, potentially overcoming the temporal limitation of this method.

Delocalisation approaches such as optogenetic tools are highly conditional for protein re-localisation and can be switched on or off with light. Both CRY2 and LEXY require only the addition of a tag at the protein of interest without the requirement of expressing further effector proteins. While CRY2-mCherry-Piwi lost nuclear localisation already in control conditions (Figure 4.5), modifying Nxf2 with a carboxy-terminal mCherry-LEXY tag had no adverse impact on protein function as evaluated by rescue experiments and localisation studies (Figure 4.8-4.10).

Illumination-induced delocalisation enables very precise spatio-temporal control. I was able to delocalise LEXY-tagged proteins within 30sec of light treatment and, in theory, even single cell illumination is feasible. This approach could facilitate not only functional protein studies in the entire embryo at desired stages, but also allow location-specific examinations. Optogenetic tools could therefore be utilised to dissect the impact of maternally deposited Piwi or downstream effectors in somatic and pole cells individually by selectively illuminating sections of an embryo. While the dynamics of de- and re-localisation are nearly instantaneous using optogenetics, these tools require continuous illumination of embryos over long periods of time. This could result in phototoxicity and alter normal embryogenesis. Further experiments are required to establish ideal illumination conditions while not affecting embryo development.

Precise illumination of embryos with methods such as light sheet and spinning disk microscopy yields the lowest phototoxicity and high spatial resolution (Icha et al., 2017). However, illuminating hundreds or thousands of embryos required for experiments such as epigenetic profiling by ChIP-Seq is not currently feasible. Therefore, consistently and reproducibly treating sufficient numbers of embryos for comprehensive downstream experiments is currently the bottleneck of this technology. This could, however, change with the development of low-input methods. For instance, commercially available kits for RNA-Seq are already available requiring as little as a

single embryo as input material as well as single-cell sequencing (scRNA-Seq) approaches.

While protein delocalisation requires the constant expression of protein traps or illumination, protein degradation is irreversible. This makes it a particularly useful technique for studying the function of maternally deposited piRNA-Piwi complexes. Zygotic Piwi expression in embryos is restricted to pole cells (chapter 5.3.1). Since Piwi localising to somatic nuclei is entirely of maternal origin, degradation of the finite pool of maternally deposited Piwi is irreversible and cannot be replenished by zygotic transcription. Again, some degradation systems requiring zygotic expression such as deGradFP are only partially useful in this context (Caussinus et al., 2011). Maternally deposited Piwi is only degraded upon expression of deGradFP and this restricts the temporal resolution greatly (Figure 4.12-4.13). There is also no rapid conditional activation or deactivation of degradation, further restricting its use.

Conditional degradation systems such as AID combine temporal resolution with the irreversible removal of maternal proteins. Both GFP-AID-Piwi and TIR1 can be safely maternally deposited without the risk of disturbing the piRNA pathway during oogenesis or embryogenesis in the absence of auxin (Figure 5.18). Additionally, the kinetics of degradation (< 25min) allow reasonable temporal control (Figure 4.19). This enables the study of Piwi's function for various stages of development.

6.3 Transgenerationally inherited piRNAs silence transposons during *Drosophila* embryogenesis

Extensive research over the last decade has contributed greatly to our understanding of the piRNA pathway in maintaining genome integrity. Transcriptional gene silencing mediated by the PICTS complex is at the core of this pathway (Batki et al., 2019; Fabry et al., 2019; Murano et al., 2019; Sienski et al., 2015; Yu et al., 2015; Zhao et al., 2019). piRNA-dependent epigenetic changes at transposon insertions are believed to lead to silencing of active transposons, thereby preventing transposition events. Our current understanding of the piRNA pathway, however, is mostly limited to *Drosophila* gonads and *in vitro* systems. Applying the AID approach described

above enabled me to finally examine the function of the piRNA pathway during other developmental stages.

Maternal deposition of Piwi in complex with piRNAs in embryos implied an additional function for the piRNA pathway during embryogenesis (Brennecke et al., 2008). I showed that not only Piwi is highly abundant during early development, but also its downstream effector complex PICTS (Figure 5.1, 5.4). General silencing factors required for PICTS-dependent induction of transcriptional gene silencing such as Egg and HP1a were highly expressed in early stages of embryogenesis, as reported (Seller et al., 2019; Yuan and O'Farrell, 2016) (Figure 5.5). Piwi localisation was not limited to germ cell progenitors, but primarily accumulated in somatic nuclei for the first 12h of embryogenesis (Figure 5.1). Piwi was not zygotically expressed in either pole cells or somatic cells initially but was clearly detectable in germ cell progenitors during late embryogenesis. Interestingly, Piwi of zygotic origin did not localise to nuclei of late stage embryos, suggesting that piRNA biogenesis is not yet functional during this developmental time point (Figure 5.2). This could be a consequence of immature cluster definition at this stage. Piwi has been shown to be involved in establishing clusters in adult flies during embryogenesis (Akkouche et al., 2017; Le Thomas et al., 2014). This data suggests that initiation of cluster re-definition or piRNA biogenesis occurs later on during development and might gradually increase throughout adulthood. Indeed, sterile females of dysgenic crosses can regain fertility when aged suggesting a dynamic adaptation of piRNA biogenesis throughout life (Khurana et al., 2011). Additionally, this data could imply a more prominent role of maternally deposited piRNAs in complex with Aub in the re-establishing of clusters. However, further investigation of these processes is necessary to fully understand the function of maternal inheritance on germ cell development.

Summarising these findings, a defined window of potential Piwi-dependent silencing for genomic targets emerged during the first half of embryogenesis. I further showed that transposons were highly expressed in early embryos. The highest transcribed transposon, *roo*, contributed more than 1% towards the entire embryonic transcriptome during its expression peak, thereby illustrating the need for transposon regulation (Figure 5.6).

roo transcripts localised to somatic cells positive for nuclear Piwi further indicating a function of the piRNA pathway in transposon regulation. Interestingly, maternal deposition of Piwi and *roo* expression in early embryos was conserved in a closely related species (Figures 5.7-5.8). The expression patterns and maternal deposition were almost identical in *D. simulans* compared to *D. melanogaster* embryos. Expression of *roo* was restricted to somatic cells, especially cell lineages giving rise to the adult mesoderm. Previous research suggested that *roo* is partially dependent on the endogenous transcription factors twist (twi) and snail (sna), which are highly expressed in embryogenic mesoderm (Bronner et al., 1995). However, *roo* expression was not detected in future germ cells. While maternal deposition of Piwi and *roo* expression was conserved in *D. melanogaster* and *D. simulans*, it remains to be explored whether this is a common feature in other Drosophilidae.

Transposition events have to occur in the genome of germ cells in order to be inherited by the next generation. This raises the question, why *roo* expression is restricted to somatic tissue during embryogenesis. The evolutionary interplay between the piRNA pathway and transposons has been often described as an arms race (Parhad and Theurkauf, 2019). Components of the piRNA pathway evolve rapidly under positive selection. However, transposons adapt by changing their genomic sequence to evade small RNA-based identification or by attempting to evade the piRNA pathway entirely. For instance, the *gypsy* and *ZAM* family of transposons are expressed primarily in somatic follicle cells surrounding the germ line during oogenesis. Virus-like particles produced in follicle cells are able to infect the oocyte, thereby inducing transposition events inherited by the next generation (Kim et al., 1994; Leblanc et al., 2000). The piRNA pathway likely adapted by partially extending its expression to *gypsy/ZAM*-expressing somatic follicle cells and establishing a functional response towards transposons by transcriptional gene silencing. Like *gypsy* and *ZAM*, it is possible that *roo* highjacked embryonic transcription factors such as twist and snail to evade silencing during oogenesis by the highly active piRNA pathway in gonads. In fact, *de novo* transposon insertion identification revealed *roo* as the most abundant TE in w^{1118}, further indicating the successful strategy of *roo* transposition (chapter 5.3.7).

We might be observing an ongoing war between *roo* and the piRNA pathway in early embryos. I was able to show that *roo* and other transposons produce protein output (Figure 5.10). *roo* and *gypsy* are special in their protein-coding capability compared to other retrotransposons. Besides coding for retrovirus-like proteins such as gag and pol, both additionally code for an envelope (env) protein that is absent from most other transposons (Kim et al., 1994). This suggests that *roo*, like *gypsy*, might be able to form virus-like particles that can infect other cells such as germline progenitors in early embryogenesis. The inheritance of the piRNA pathway in embryos could therefore be an attempt of the organism to protect embryos from *roo* inversion of the genome. However, further research is needed to understand the transposition mechanism of *roo* and to probe whether *roo* can actually infect other cells.

I found several indications that the piRNA pathway is able to actively target *roo* insertions. Maternally deposited piRNAs are produced in nurse cells during oogenesis. While *roo* is not regulated by the piRNA pathway in gonads (Sienski et al., 2012), over 15% of maternally inherited piRNAs target *roo* (Figure 5.11). Therefore, major capacities of the biogenesis machinery during oogenesis are employed to produce piRNAs exclusively for use during embryogenesis. This suggests that nurse cells are vital in mass-producing epigenetic information inherited by the next generation. Inheritance of small RNAs, therefore, enables the piRNA pathway to identify potential threats to genome integrity in the next generation during embryogenesis. Indeed, I was able to identify early targets of maternally deposited Piwi guided by antisense piRNAs. Maternal Piwi bound transposon transcripts, including *297* and *roo*, indicating a directed response towards early transposon transcription (Figure 5.12). Additionally, I was able to identify TGS signatures at transposon insertions with high resolution. Importantly, repressive chromatin marks accumulated exclusively at transposon insertions highly targeted by maternally deposited piRNAs and transcribed during embryogenesis but were absent at transposons lacking antisense piRNA deposition (Figure 5.14, 5.15).

Despite maternal deposition of piRNAs targeting *roo* and the presence of the piRNA pathway during embryogenesis, high levels of *roo* transcripts were detectable at certain embryonic stages (Figure 5.6-5-8). However, induced degradation of

maternal Piwi led to deregulation of *roo* during earlier stages of embryogenesis prior to its wild-type peak expression (Figure 5.16). The piRNA pathway might therefore supress transposon expression especially early on in development when germ cells are much more accessible and vulnerable for transposons to infect by delaying the expression peak of *roo*. Additionally, transposition events in somatic cells could have a deleterious impact on embryo development. This hypothesis is further supported by the high abundance of maternal Piwi in pre-gastrulation embryos (<3h AEL) compared to later stages (Figure 5.1).

Surprisingly, degradation of maternal Piwi affected genes during MZT (Figure 5.17). Maternally deposited mRNAs were cleared faster and zygotically expressed genes were activated prematurely in embryos depleted of Piwi compared to untreated controls. This suggests that maternal Piwi might have an additional function on regulating more fundamental processes during embryogenesis. However, more research is needed to better understand this function.

The piRNA pathway silences transposons by two mechanisms. Post-transcriptional gene silencing (PTGS) in the cytoplasm, mediated by Ago3 and Aub, and transcriptional gene silencing relying on nuclear Piwi. While Aub is maternally inherited and primarily localises to germ cell progenitors in embryos, Ago3 appears absent (Brennecke et al., 2008). This indicates that piRNA pathway-mediated silencing in somatic cells of embryos is likely mediated by TGS rather than PTGS and resembles therefore the pathway in somatic follicle cells. Indeed, I was able to show that maternal Piwi leads to the deposition of repressive chromatin marks in the form of H3K9me3 at transposon insertions (Figure 5.14-5.16). Auxin-induced degradation of Piwi resulted in a severe loss of H3K9me3 at *roo* sequences (Figure 5.19), thereby supporting the hypothesis that maternal Piwi regulates transposon expression by TGS.

This study uncovered a novel function for the piRNA pathway in transposon silencing during embryogenesis. Our data suggests that in somatic cells early transposon transcripts are bound by maternally inherited Piwi that is guided by antisense piRNAs. TGS is induced at targeted transposon loci likely by the PICTS complex and the general silencing machinery. Previous studies speculated that the

maintenance of silenced epigenetic states at transposon insertions is Piwi-independent and maintained throughout adult life (Gu and Elgin, 2013). My data presented in this thesis contradicts this view. Piwi protein was present throughout the first half of embryogenesis but was absent from somatic nuclei in late stage embryos. Similarly, H3K9me3 marks at transposon insertions regulated by the piRNA pathway were lost in late stage embryos once Piwi expression was confined to germ cells (Figure 5.14-5.15). This suggests a Piwi-dependent mode of heterochromatin maintenance at targeted loci likely by Piwi engaging leaked transposon transcripts and subsequent recruitment of the PICTS complex to associated chromatin (Figure 6.2).

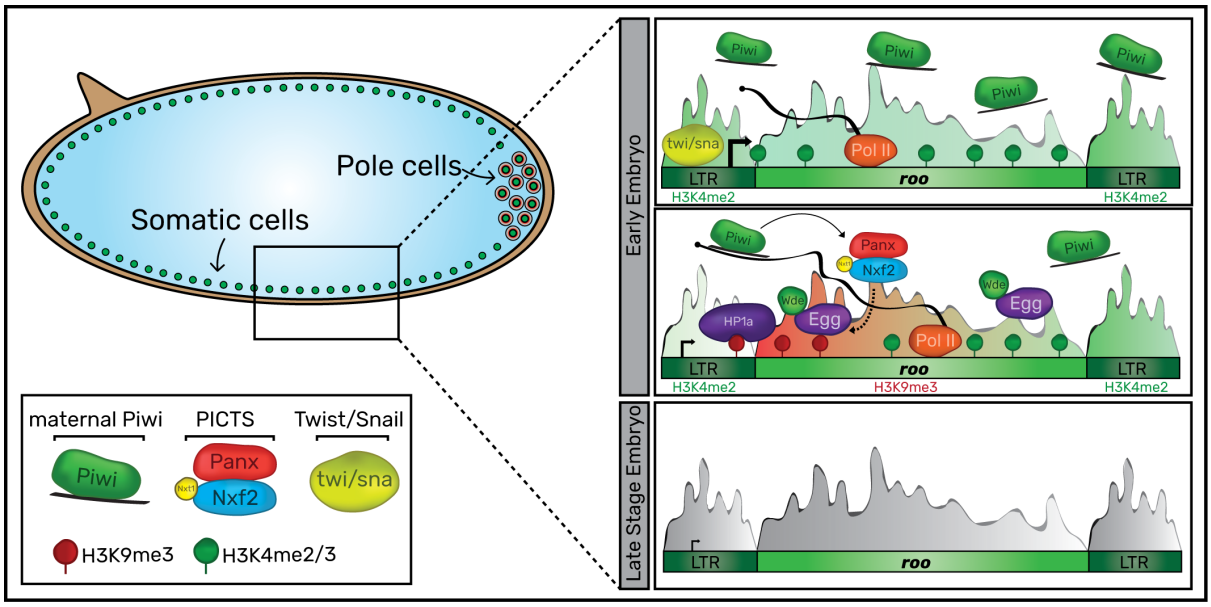

Figure 6.2: Transcriptional gene silencing in somatic cells during *Drosophila* embryogenesis

Non-gonadal somatic cells express transposons such as *roo* during early embryogenesis putatively driven by the endogenous transcription factors twist (twi) and snail (sna). Maternally deposited Piwi in complex with targeting piRNAs engage zygotic nascent transposon transcripts. Piwi recruits the PICTS complex comprised of Panx, Nxf2 and Nxt1. The PICTS complex instruments the general silencing machinery, including Eggless (Egg), which leads to the deposition of repressive chromatin marks such as H3K9me3. HP1a binds H3K9me3 and converts the formally active transposon locus into densely packed heterochromatin thus shutting down transcription. While Piwi is degraded in late stage embryos and H3K9me3 marks are lost, *roo* is not re-expressed, likely due to the lack of expression of the co-opted transcription factors.

This thesis improved our understanding of transcriptional gene silencing of transposons during oogenesis. However, it remains to be explored how the recruitment of the PICTS complex results in the recruitment of the general silencing machinery and subsequent conversion of transcriptionally active target loci into repressive heterochromatin. I was able to overcome technical limitations that prevented the exploration of the piRNA pathway during embryogenesis. Further improvements to the approaches presented above could prove invaluable for the functional characterisation of other maternally deposited proteins and greatly enhance our understanding of transgenerational epigenetic inheritance. I was able to reveal one function of the piRNA pathway in transposon silencing during embryogenesis. However, my data suggests control of further developmental processes such as MZT by maternal Piwi. More research is necessary to dissect the precise function the piRNA pathway plays during this sensitive moment of development.

7 References

Aboobaker, A.A., Tomancak, P., Patel, N., Rubin, G.M., and Lai, E.C. (2005). Drosophila microRNAs exhibit diverse spatial expression patterns during embryonic development. Proc Natl Acad Sci U S A *102*, 18017-18022.

Agrawal, A., Eastman, Q.M., and Schatz, D.G. (1998). Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. Nature *394*, 744-751.

Akkouche, A., Mugat, B., Barckmann, B., Varela-Chavez, C., Li, B., Raffel, R., Pelisson, A., and Chambeyron, S. (2017). Piwi Is Required during Drosophila Embryogenesis to License Dual-Strand piRNA Clusters for Transposon Repression in Adult Ovaries. Mol Cell *66*, 411-419 e414.

Allshire, R.C., and Madhani, H.D. (2018). Ten principles of heterochromatin formation and function. Nat Rev Mol Cell Biol *19*, 229-244.

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.

Ardehali, M.B., Mei, A., Zobeck, K.L., Caron, M., Lis, J.T., and Kusch, T. (2011). Drosophila Set1 is the major histone H3 lysine 4 trimethyltransferase with role in transcription. EMBO J *30*, 2817-2828.

Ashe, A., Sapetschnig, A., Weick, E.M., Mitchell, J., Bagijn, M.P., Cording, A.C., Doebley, A.L., Goldstein, L.D., Lehrbach, N.J., Le Pen, J.*, et al.* (2012). piRNAs can trigger a multigenerational epigenetic memory in the germline of C. elegans. Cell *150*, 88-99.

Baek, D., Villen, J., Shin, C., Camargo, F.D., Gygi, S.P., and Bartel, D.P. (2008). The impact of microRNAs on protein output. Nature *455*, 64-71.

Bagijn, M.P., Goldstein, L.D., Sapetschnig, A., Weick, E.M., Bouasker, S., Lehrbach, N.J., Simard, M.J., and Miska, E.A. (2012). Function, targets, and evolution of Caenorhabditis elegans piRNAs. Science *337*, 574-578.

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.

Barron, M.G., Fiston-Lavier, A.S., Petrov, D.A., and Gonzalez, J. (2014). Population genomics of transposable elements in Drosophila. Annu Rev Genet *48*, 561-581.

Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell *116*, 281-297.

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

Bate, M., and Arias, A.M. (1993). The Development of Drosophila melanogaster. Cold Spring Harbor Laboratory Press *I*.

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.

Batut, P., Dobin, A., Plessy, C., Carninci, P., and Gingeras, T.R. (2013). High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression. Genome Res *23*, 169-180.

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.

Ben-Hattar, J., and Jiricny, J. (1988). Methylation of single CpG dinucleotides within a promoter element of the Herpes simplex virus tk gene reduces its transcription in vivo. Gene *65*, 219-227.

Benachenhou, F., Sperber, G.O., Bongcam-Rudloff, E., Andersson, G., Boeke, J.D., and Blomberg, J. (2013). Conserved structure and inferred evolutionary history of long terminal repeats (LTRs). Mob DNA *4*, 5.

Bence, M., Jankovics, F., Lukacsovich, T., and Erdelyi, M. (2017). Combining the auxin-inducible degradation system with CRISPR/Cas9-based genome editing for the conditional depletion of endogenous Drosophila melanogaster proteins. FEBS J *284*, 1056-1069.

Berger, F., and Twell, D. (2011). Germline specification and function in plants. Annu Rev Plant Biol *62*, 461-484.

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.

Beuchle, D., Struhl, G., and Muller, J. (2001). Polycomb group proteins and heritable silencing of Drosophila Hox genes. Development *128*, 993-1004.

Bingham, P.M., Kidwell, M.G., and Rubin, G.M. (1982). The molecular basis of P-M hybrid dysgenesis: the role of the P element, a P-strain-specific transposon family. Cell *29*, 995-1004.

Biscotti, M.A., Canapa, A., Forconi, M., Olmo, E., and Barucca, M. (2015). Transcription of tandemly repetitive DNA: functional roles. Chromosome Res *23*, 463-477.

Black, L., Petruk, S., Fenstermaker, T.K., Hodgson, J.W., Caplan, J.L., Brock, H.W., and Mazo, A. (2016). Chromatin proteins and RNA are associated with DNA during all phases of mitosis. Cell Discov *2*, 16038.

Bleykasten-Grosshans, C., Friedrich, A., and Schacherer, J. (2013). Genome-wide analysis of intraspecific transposon diversity in yeast. BMC Genomics *14*, 399.

Boland, A., Huntzinger, E., Schmidt, S., Izaurralde, E., and Weichenrieder, O. (2011). Crystal structure of the MID-PIWI lobe of a eukaryotic Argonaute protein. Proc Natl Acad Sci U S A *108*, 10466-10471.

Borsos, M., and Torres-Padilla, M.E. (2016). Building up the nucleus: nuclear organization in the establishment of totipotency and pluripotency during mammalian development. Genes Dev *30*, 611-621.

Boskovic, A., and Rando, O.J. (2018). Transgenerational Epigenetic Inheritance. Annu Rev Genet *52*, 21-41.

Bownes, M. (1975). A photographic study of development in the living embryo of Drosophila melanogaster. J Embryol Exp Morphol *33*, 789-801.

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.

Braun, J.E., Huntzinger, E., Fauser, M., and Izaurralde, E. (2011). GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. Mol Cell *44*, 120-133.

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.

Brink, R.A. (1956). A Genetic Change Associated with the R Locus in Maize Which Is Directed and Potentially Reversible. Genetics *41*, 872-889.

Brink, R.A., and Weyers, W.H. (1957). Invariable Genetic Change in Maize Plants Heterozygous for Marbled Aleurone. Proc Natl Acad Sci U S A *43*, 1053-1060.

Bronner, G., Taubert, H., and Jackle, H. (1995). Mesoderm-specific B104 expression in the Drosophila embryo is mediated by internal cis-acting elements of the transposon. Chromosoma *103*, 669-675.

Brower-Toland, B., Findley, S.D., Jiang, L., Liu, L., Yin, H., Dus, M., Zhou, P., Elgin, S.C., and Lin, H. (2007). Drosophila PIWI associates with chromatin and interacts directly with HP1a. Genes Dev *21*, 2300-2311.

Bucheton, A., Paro, R., Sang, H.M., Pelisson, A., and Finnegan, D.J. (1984). The molecular basis of I-R hybrid dysgenesis in Drosophila melanogaster: identification, cloning, and properties of the I factor. Cell *38*, 153-163.

Buckley, B.A., Burkhart, K.B., Gu, S.G., Spracklin, G., Kershner, A., Fritz, H., Kimble, J., Fire, A., and Kennedy, S. (2012). A nuclear Argonaute promotes multigenerational epigenetic inheritance and germline immortality. Nature *489*, 447-451.

Buhler, M., Verdel, A., and Moazed, D. (2006). Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. Cell *125*, 873-886.

Bundo, M., Toyoshima, M., Okada, Y., Akamatsu, W., Ueda, J., Nemoto-Miyauchi, T., Sunaga, F., Toritsuka, M., Ikawa, D., Kakita, A.*, et al.* (2014). Increased l1 retrotransposition in the neuronal genome in schizophrenia. Neuron *81*, 306-313.

Calarco, J.P., Borges, F., Donoghue, M.T., Van Ex, F., Jullien, P.E., Lopes, T., Gardner, R., Berger, F., Feijo, J.A., Becker, J.D.*, et al.* (2012). Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. Cell *151*, 194-205.

Capuano, F., Mulleder, M., Kok, R., Blom, H.J., and Ralser, M. (2014). Cytosine DNA methylation is found in Drosophila melanogaster but absent in Saccharomyces cerevisiae, Schizosaccharomyces pombe, and other yeast species. Anal Chem *86*, 3697-3702.

Caussinus, E., Kanca, O., and Affolter, M. (2011). Fluorescent fusion protein knockout mediated by anti-GFP nanobody. Nat Struct Mol Biol *19*, 117-121.

Chan, C.S., Rastelli, L., and Pirrotta, V. (1994). A Polycomb response element in the Ubx gene that determines an epigenetically inherited state of repression. EMBO J *13*, 2553-2564.

Chang, H.H.Y., Pannunzio, N.R., Adachi, N., and Lieber, M.R. (2017). Non-homologous DNA end joining and alternative pathways to double-strand break repair. Nat Rev Mol Cell Biol *18*, 495-506.

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.

Chari, S., Wilky, H., Govindan, J., and Amodeo, A.A. (2019). Histone concentration regulates the cell cycle and transcription in early development. Development *146*.

Chekulaeva, M., Mathys, H., Zipprich, J.T., Attig, J., Colic, M., Parker, R., and Filipowicz, W. (2011). miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. Nat Struct Mol Biol *18*, 1218-1226.

Chen, C.Y., Zheng, D., Xia, Z., and Shyu, A.B. (2009). Ago-TNRC6 triggers microRNA-mediated decay by promoting two deadenylation steps. Nat Struct Mol Biol *16*, 1160-1166.

Chen, Y., Pane, A., and Schupbach, T. (2007). Cutoff and aubergine mutations result in retrotransposon upregulation and checkpoint activation in Drosophila. Curr Biol *17*, 637-642.

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.

Choi, H.M., Beck, V.A., and Pierce, N.A. (2014). Next-generation in situ hybridization chain reaction: higher gain, lower cost, greater durability. ACS Nano *8*, 4284-4294.

Christie, M., Boland, A., Huntzinger, E., Weichenrieder, O., and Izaurralde, E. (2013). Structure of the PAN3 pseudokinase reveals the basis for interactions with the PAN2 deadenylase and the GW182 proteins. Mol Cell *51*, 360-373.

Chuma, S., and Nakano, T. (2013). piRNA and spermatogenesis in mice. Philos Trans R Soc Lond B Biol Sci *368*, 20110338.

Ciabrelli, F., Comoglio, F., Fellous, S., Bonev, B., Ninova, M., Szabo, Q., Xuereb, A., Klopp, C., Aravin, A., Paro, R*., et al.* (2017). Stable Polycomb-dependent transgenerational inheritance of chromatin states in Drosophila. Nat Genet *49*, 876-886.

Ciccia, A., and Elledge, S.J. (2010). The DNA damage response: making it safe to play with knives. Mol Cell *40*, 179-204.

Cinalli, R.M., Rangan, P., and Lehmann, R. (2008). Germ cells are forever. Cell *132*, 559-562.

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

Cox, D.N., Chao, A., Baker, J., Chang, L., Qiao, D., and Lin, H. (1998). A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. Genes Dev *12*, 3715-3727.

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.

Cubbon, A., Ivancic-Bace, I., and Bolt, E.L. (2018). CRISPR-Cas immunity, DNA repair and genome stability. Biosci Rep *38*.

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.

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., and Pirrotta, V. (2002). Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell *111*, 185-196.

de Koning, A.P., Gu, W., Castoe, T.A., Batzer, M.A., and Pollock, D.D. (2011). Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet *7*, e1002384.

de la Chaux, N., and Wagner, A. (2009). Evolutionary dynamics of the LTR retrotransposons roo and rooA inferred from twelve complete Drosophila genomes. BMC Evol Biol *9*, 205.

de Vanssay, A., Bouge, A.L., Boivin, A., Hermant, C., Teysset, L., Delmarre, V., Antoniewski, C., and Ronsseray, S. (2012). Paramutation in Drosophila linked to emergence of a piRNA-producing locus. Nature *490*, 112-115.

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., 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.

Derrien, T., Estelle, J., Marco Sola, S., Knowles, D.G., Raineri, E., Guigo, R., and Ribeca, P. (2012). Fast computation and applications of genome mappability. PLoS One *7*, e30377.

Ding, D., and Lipshitz, H.D. (1994). Spatially regulated expression of retrovirus-like transposons during Drosophila melanogaster embryogenesis. Genet Res *64*, 167-181.

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.

Dooner, H.K., Robbins, T.P., and Jorgensen, R.A. (1991). Genetic and developmental control of anthocyanin biosynthesis. Annu Rev Genet *25*, 173-199.

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.

Eichhorn, S.W., Subtelny, A.O., Kronja, I., Kwasnieski, J.C., Orr-Weaver, T.L., and Bartel, D.P. (2016). mRNA poly(A)-tail changes specified by deadenylation broadly reshape translation in Drosophila oocytes and early embryos. Elife *5*.

Eisen, T.J., Eichhorn, S.W., Subtelny, A.O., and Bartel, D.P. (2020). MicroRNAs Cause Accelerated Decay of Short-Tailed Target mRNAs. Mol Cell *77*, 775-785 e778.

Ekwall, K., Olsson, T., Turner, B.M., Cranston, G., and Allshire, R.C. (1997). Transient inhibition of histone deacetylation alters the structural and functional imprint at fission yeast centromeres. Cell *91*, 1021-1032.

Elgin, S.C., and Reuter, G. (2013). Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila. Cold Spring Harb Perspect Biol *5*, a017780.

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.

Eyre-Walker, A., and Keightley, P.D. (2007). The distribution of fitness effects of new mutations. Nat Rev Genet *8*, 610-618.

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*.

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.

Finnegan, D.J. (1989). Eukaryotic transposable elements and genome evolution. Trends Genet *5*, 103-107.

Finnegan, D.J. (2012). Retrotransposons. Curr Biol *22*, R432-437.

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.

Frank, F., Sonenberg, N., and Nagar, B. (2010). Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. Nature *465*, 818-822.

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.

Fukunaga, R., Han, B.W., Hung, J.H., Xu, J., Weng, Z., and Zamore, P.D. (2012). Dicer partner proteins tune the length of mature miRNAs in flies and mammals. Cell *151*, 533-546.

Galiana-Arnoux, D., Dostert, C., Schneemann, A., Hoffmann, J.A., and Imler, J.L. (2006). Essential function in vivo for Dicer-2 in host defense against RNA viruses in drosophila. Nat Immunol *7*, 590-597.

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.

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.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods *6*, 343-345.

Golbabapour, S., Majid, N.A., Hassandarvish, P., Hajrezaie, M., Abdulla, M.A., and Hadi, A.H. (2013). Gene silencing and Polycomb group proteins: an overview of their structure, mechanisms and phylogenetics. OMICS *17*, 283-296.

Goodier, J.L., and Kazazian, H.H., Jr. (2008). Retrotransposons revisited: the restraint and rehabilitation of parasites. Cell *135*, 23-35.

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.

Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W.*, et al.* (2011). The developmental transcriptome of Drosophila melanogaster. Nature *471*, 473-479.

Greenberg, M.V.C., and Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. Nat Rev Mol Cell Biol *20*, 590-607.

Gu, T., and Elgin, S.C. (2013). Maternal depletion of Piwi, a component of the RNAi system, impacts heterochromatin formation in Drosophila. PLoS Genet *9*, e1003780.

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.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835-840.

Hall, I.M., Shankaranarayana, G.D., Noma, K., Ayoub, N., Cohen, A., and Grewal, S.I. (2002). Establishment and maintenance of a heterochromatin domain. Science *297*, 2232-2237.

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.

Han, J., Lee, Y., Yeom, K.H., Kim, Y.K., Jin, H., and Kim, V.N. (2004). The Drosha-DGCR8 complex in primary microRNA processing. Genes Dev *18*, 3016-3027.

Han, J.S. (2010). Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions. Mob DNA *1*, 15.

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.

Hartig, J.V., Esslinger, S., Bottcher, R., Saito, K., and Forstemann, K. (2009). Endo-siRNAs depend on a new isoform of loquacious and target artificially introduced, high-copy sequences. EMBO J *28*, 2932-2944.

Hartig, J.V., and Forstemann, K. (2011). Loqs-PD and R2D2 define independent pathways for RISC generation in Drosophila. Nucleic Acids Res *39*, 3836-3851.

Hayashi, K., de Sousa Lopes, S.M., and Surani, M.A. (2007). Germ cell specification in mice. Science *316*, 394-396.

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.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A.*, et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet *39*, 311-318.

Hendrickson, D.G., Hogan, D.J., McCullough, H.L., Myers, J.W., Herschlag, D., Ferrell, J.E., and Brown, P.O. (2009). Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. PLoS Biol *7*, e1000238.

Hermant, C., Boivin, A., Teysset, L., Delmarre, V., Asif-Laidin, A., van den Beek, M., Antoniewski, C., and Ronsseray, S. (2015). Paramutation in Drosophila Requires Both Nuclear and Cytoplasmic Actors of the piRNA Pathway and Induces Cis-spreading of piRNA Production. Genetics *201*, 1381-1396.

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.

Hollick, J.B. (2017). Paramutation and related phenomena in diverse species. Nat Rev Genet *18*, 5-23.

Horner, V.L., and Wolfner, M.F. (2008). Transitioning from egg to embryo: triggers and mechanisms of egg activation. Dev Dyn *237*, 527-544.

Horwich, M.D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., and Zamore, P.D. (2007). The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. Curr Biol *17*, 1265-1272.

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

Hu, K. (2019). On Mammalian Totipotency: What Is the Molecular Underpinning for the Totipotency of Zygote? Stem Cells Dev *28*, 897-906.

Huang, A., Amourda, C., Zhang, S., Tolwinski, N.S., and Saunders, T.E. (2017). Decoding temporal interpretation of the morphogen Bicoid in the early Drosophila embryo. Elife *6*.

Hug, C.B., Grimaldi, A.G., Kruse, K., and Vaquerizas, J.M. (2017). Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. Cell *169*, 216-228 e219.

Hutvagner, G., and Simard, M.J. (2008). Argonaute proteins: key players in RNA silencing. Nat Rev Mol Cell Biol *9*, 22-32.

Hyun, K., Jeon, J., Park, K., and Kim, J. (2017). Writing, erasing and reading histone lysine methylations. Exp Mol Med *49*, e324.

Icha, J., Weber, M., Waters, J.C., and Norden, C. (2017). Phototoxicity in live fluorescence microscopy, and how to avoid it. Bioessays *39*.

Iguchi-Ariga, S.M., and Schaffner, W. (1989). CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. Genes Dev *3*, 612-619.

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., 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.

Jackson, S.P., and Bartek, J. (2009). The DNA-damage response in human biology and disease. Nature *461*, 1071-1078.

Jang, J.K., Messina, L., Erdman, M.B., Arbel, T., and Hawley, R.S. (1995). Induction of metaphase arrest in Drosophila oocytes by chiasma-based kinetochore tension. Science *268*, 1917-1919.

Jayaprakash, A.D., Jabado, O., Brown, B.D., and Sachidanandam, R. (2011). Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. Nucleic Acids Res *39*, e141.

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.

Jullien, P.E., Susaki, D., Yelagandula, R., Higashiyama, T., and Berger, F. (2012). DNA methylation dynamics during sexual reproduction in Arabidopsis thaliana. Curr Biol *22*, 1825-1830.

Kaneuchi, T., Sartain, C.V., Takeo, S., Horner, V.L., Buehner, N.A., Aigaki, T., and Wolfner, M.F. (2015). Calcium waves occur as Drosophila oocytes activate. Proc Natl Acad Sci U S A *112*, 791-796.

Kapitonov, V.V., and Jurka, J. (2005). RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. PLoS Biol *3*, e181.

Kawaoka, S., Hayashi, N., Suzuki, Y., Abe, H., Sugano, S., Tomari, Y., Shimada, T., and Katsuma, S. (2009). The Bombyx ovary-derived cell line endogenously expresses PIWI/PIWI-interacting RNA complexes. RNA *15*, 1258-1264.

Kennedy, M.J., Hughes, R.M., Peteya, L.A., Schwartz, J.W., Ehlers, M.D., and Tucker, C.L. (2010). Rapid blue-light-mediated induction of protein interactions in living cells. Nat Methods *7*, 973-975.

Kerkow, D.E., Carmel, A.B., Menichelli, E., Ambrus, G., Hills, R.D., Jr., Gerace, L., and Williamson, J.R. (2012). The structure of the NXF2/NXT1 heterodimeric complex reveals the combined specificity and versatility of the NTF2-like fold. J Mol Biol *415*, 649-665.

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

Khairy, K., Lemon, W.C., Amat, F., and Keller, P.J. (2015). Light sheet-based imaging and analysis of early embryogenesis in the fruit fly. Methods Mol Biol *1189*, 79-97.

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

Kidwell, M.G., Kidwell, J.F., and Sved, J.A. (1977). Hybrid Dysgenesis in DROSOPHILA MELANOGASTER: A Syndrome of Aberrant Traits Including Mutation, Sterility and Male Recombination. Genetics *86*, 813-833.

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, T., and Buratowski, S. (2009). Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5' transcribed regions. Cell *137*, 259-272.

Kim, Y.K., and Kim, V.N. (2007). Processing of intronic microRNAs. EMBO J *26*, 775-783.

King, R.C. (1970). Ovarian development in Drosophila melanogaster / [by] Robert C. King (New York: New York : Academic Press, 1970.).

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

Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Khurana, J.S., Zhang, F., Schultz, N., Koppetsch, B.S., Nowosielska, A.*, et al.* (2009). The Drosophila HP1 homologue Rhino is required for transposon silencing and piRNA production by dual strand clusters. Cell.

Klenov, M.S., Sokolova, O.A., Yakushev, E.Y., Stolyarenko, A.D., Mikhaleva, E.A., Lavrov, S.A., and Gvozdev, V.A. (2011). Separation of stem cell maintenance and transposon silencing functions of Piwi protein. Proc Natl Acad Sci U S A *108*, 18760-18765.

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.

Koch, C.M., Honemann-Capito, M., Egger-Adam, D., and Wodarz, A. (2009). Windei, the Drosophila homolog of mAM/MCAF1, is an essential cofactor of the H3K9 methyl transferase dSETDB1/Eggless in germ line development. PLoS Genet *5*, e1000644.

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

Kramer, J.M., Kochinke, K., Oortveld, M.A., Marks, H., Kramer, D., de Jong, E.K., Asztalos, Z., Westwood, J.T., Stunnenberg, H.G., Sokolowski, M.B.*, et al.* (2011). Epigenetic regulation of learning and memory by Drosophila EHMT/G9a. PLoS Biol *9*, e1000569.

Krug, L., Chatterjee, N., Borges-Monroy, R., Hearn, S., Liao, W.W., Morrill, K., Prazak, L., Rozhkov, N., Theodorou, D., Hammell, M.*, et al.* (2017). Retrotransposon activation contributes to neurodegeneration in a Drosophila TDP-43 model of ALS. PLoS Genet *13*, e1006635.

Kuntz, S.G., and Eisen, M.B. (2014). Drosophila embryogenesis scales uniformly across temperature in developmentally diverse species. PLoS Genet *10*, e1004293.

Kwasnieski, J.C., Orr-Weaver, T.L., and Bartel, D.P. (2019). Early genome activation in Drosophila is extensive with an initial tendency for aborted transcripts and retained introns. Genome Res *29*, 1188-1197.

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.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W.*, et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Langie, S.A., Koppen, G., Desaulniers, D., Al-Mulla, F., Al-Temaimi, R., Amedei, A., Azqueta, A., Bisson, W.H., Brown, D.G., Brunborg, G.*, et al.* (2015). Causes of genome instability: the effect of low dose chemical exposures in modern society. Carcinogenesis *36 Suppl 1*, S61-88.

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. (2012). mRNA localization and translational control in Drosophila oogenesis. Cold Spring Harb Perspect Biol *4*.

Lau, N.C., Robine, N., Martin, R., Chung, W.J., Niki, Y., Berezikov, E., and Lai, E.C. (2009). Abundant primary piRNAs, endo-siRNAs, and microRNAs in a Drosophila ovary cell line. Genome Res *19*, 1776-1785.

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., 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.

Leatherman, J.L., Levin, L., Boero, J., and Jongens, T.A. (2002). germ cell-less acts to repress transcription during the establishment of the Drosophila germ cell lineage. Curr Biol *12*, 1681-1685.

Leblanc, P., Desset, S., Giorgi, F., Taddei, A.R., Fausto, A.M., Mazzini, M., Dastugue, B., and Vaury, C. (2000). Life cycle of an endogenous retrovirus, ZAM, in Drosophila melanogaster. J Virol *74*, 10658-10669.

Lechner, M.S., Schultz, D.C., Negorev, D., Maul, G.G., and Rauscher, F.J., 3rd (2005). The mammalian heterochromatin protein 1 binds diverse nuclear proteins through a common motif that targets the chromoshadow domain. Biochem Biophys Res Commun *331*, 929-937.

Lee, H.C., Gu, W., Shirayama, M., Youngman, E., Conte, D., Jr., and Mello, C.C. (2012). C. elegans piRNAs mediate the genome-wide surveillance of germline transcripts. Cell *150*, 78-87.

Lee, M.G., Wynder, C., Cooch, N., and Shiekhattar, R. (2005). An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. Nature *437*, 432-435.

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, 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, 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., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.

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, Q., Gent, J.I., Zynda, G., Song, J., Makarevitch, I., Hirsch, C.D., Hirsch, C.N., Dawe, R.K., Madzima, T.F., McGinnis, K.M.*, et al.* (2015). RNA-directed DNA methylation enforces boundaries between heterochromatin and euchromatin in the maize genome. Proc Natl Acad Sci U S A *112*, 14728-14733.

Li, X.Y., Harrison, M.M., Villalta, J.E., Kaplan, T., and Eisen, M.B. (2014). Establishment of regions of genomic activity during the Drosophila maternal to zygotic transition. Elife *3*.

Li, Z., Thiel, K., Thul, P.J., Beller, M., Kuhnlein, R.P., and Welte, M.A. (2012). Lipid droplets control the maternal histone supply of Drosophila embryos. Curr Biol *22*, 2104-2113.

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.

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.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Luger, K., and Hansen, J.C. (2005). Nucleosome and chromatin fiber dynamics. Curr Opin Struct Biol *15*, 188-196.

Lund, E., and Dahlberg, J.E. (2006). Substrate selectivity of exportin 5 and Dicer in the biogenesis of microRNAs. Cold Spring Harb Symp Quant Biol *71*, 59-66.

Luteijn, M.J., van Bergeijk, P., Kaaij, L.J., Almeida, M.V., Roovers, E.F., Berezikov, E., and Ketting, R.F. (2012). Extremely stable Piwi-induced gene silencing in Caenorhabditis elegans. EMBO J *31*, 3422-3430.

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.

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.

Marques, J.T., Kim, K., Wu, P.H., Alleyne, T.M., Jafari, N., and Carthew, R.W. (2010). Loqs and R2D2 act sequentially in the siRNA pathway in Drosophila. Nat Struct Mol Biol *17*, 24-30.

McClintock, B. (1950). The origin and behavior of mutable loci in maize. Proc Natl Acad Sci U S A *36*, 344-355.

McCullers, T.J., and Steiniger, M. (2017). Transposable elements in Drosophila. Mob Genet Elements *7*, 1-18.

McDaniel, S.L., Gibson, T.J., Schulz, K.N., Fernandez Garcia, M., Nevil, M., Jain, S.U., Lewis, P.W., Zaret, K.S., and Harrison, M.M. (2019). Continued Activity of the Pioneer Factor Zelda Is Required to Drive Zygotic Genome Activation. Mol Cell *74*, 185-195 e184.

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

Megosh, H.B., Cox, D.N., Campbell, C., and Lin, H. (2006). The role of PIWI and the miRNA machinery in Drosophila germline determination. Curr Biol *16*, 1884-1894.

Miyoshi, K., Tsukumo, H., Nagami, T., Siomi, H., and Siomi, M.C. (2005). Slicer function of Drosophila Argonautes and its involvement in RISC formation. Genes Dev *19*, 2837-2848.

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.

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.

Munoz-Lopez, M., and Garcia-Perez, J.L. (2010). DNA transposons: nature and applications in genomics. Curr Genomics *11*, 115-128.

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.

Nefedova, L., and Kim, A. (2017). Mechanisms of LTR-Retroelement Transposition: Lessons from Drosophila melanogaster. Viruses *9*.

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. (2020a). Su(var)2-10 and the SUMO Pathway Link piRNA-Guided Target Recognition to Chromatin Silencing. Mol Cell *77*, 556-570 e556.

Ninova, M., Godneeva, B., Chen, Y.A., Luo, Y., Prakash, S.J., Jankovics, F., Erdelyi, M., Aravin, A.A., and Fejes Toth, K. (2020b). The SUMO Ligase Su(var)2-10 Controls Hetero- and

Euchromatic Gene Expression via Establishing H3K9 Trimethylation and Negative Feedback Regulation. Mol Cell *77*, 571-585 e574.

Niopek, D., Wehler, P., Roensch, J., Eils, R., and Di Ventura, B. (2016). Optogenetic control of nuclear protein export. Nat Commun *7*, 10624.

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.

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., Ishizuka, A., Siomi, H., and Siomi, M.C. (2004). Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. Genes Dev *18*, 1655-1666.

Olena, A.F., and Patton, J.G. (2010). Genomic organization of microRNAs. J Cell Physiol *222*, 540-545.

Olina, A.V., Kulbachinskiy, A.V., Aravin, A.A., and Esyunina, D.M. (2018). Argonaute Proteins and Mechanisms of RNA Interference in Eukaryotes and Prokaryotes. Biochemistry (Mosc) *83*, 483-497.

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.

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.

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.

Parhad, S.S., and Theurkauf, W.E. (2019). Rapid evolution and conserved function of the piRNA pathway. Open Biol *9*, 180181.

Park, H., Kim, N.Y., Lee, S., Kim, N., Kim, J., and Heo, W.D. (2017). Optogenetic protein clustering through fluorescent protein tagging and extension of CRY2. Nat Commun *8*, 30.

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.

Peng, H., Shi, J., Zhang, Y., Zhang, H., Liao, S., Li, W., Lei, L., Han, C., Ning, L., Cao, Y.*, et al.* (2012). A novel class of tRNA-derived small RNAs extremely enriched in mature mouse sperm. Cell Res *22*, 1609-1612.

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.

Proudfoot, N.J., Furger, A., and Dye, M.J. (2002). Integrating mRNA processing with transcription. Cell *108*, 501-512.

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.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res *42*, W187-191.

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.

Raychaudhuri, N., Dubruille, R., Orsi, G.A., Bagheri, H.C., Loppin, B., and Lehner, C.F. (2012). Transgenerational propagation and quantitative maintenance of paternal centromeres depends on Cid/Cenp-A presence in Drosophila sperm. PLoS Biol *10*, e1001434.

Richards, E.J., and Elgin, S.C. (2002). Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. Cell *108*, 489-500.

Robinett, C.C., Straight, A., Li, G., Willhelm, C., Sudlow, G., Murray, A., and Belmont, A.S. (1996). In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. J Cell Biol *135*, 1685-1700.

Robinson, D.N., Cant, K., and Cooley, L. (1994). Morphogenesis of Drosophila ovarian ring canals. Development *120*, 2015-2025.

Rodriguez-Martin, B., Alvarez, E.G., Baez-Ortega, A., Zamora, J., Supek, F., Demeulemeester, J., Santamarina, M., Ju, Y.S., Temes, J., Garcia-Souto, D.*, et al.* (2020). Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. Nat Genet *52*, 306-319.

Rouget, C., Papin, C., Boureux, A., Meunier, A.C., Franco, B., Robine, N., Lai, E.C., Pelisson, A., and Simonelig, M. (2010). Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. Nature *467*, 1128-1132.

Rudolph, T., Yonezawa, M., Lein, S., Heidrich, K., Kubicek, S., Schafer, C., Phalke, S., Walther, M., Schmidt, A., Jenuwein, T.*, et al.* (2007). Heterochromatin formation in Drosophila is initiated through active removal of H3K4 methylation by the LSD1 homolog SU(VAR)3-3. Mol Cell *26*, 103-115.

Saerens, D., Pellis, M., Loris, R., Pardon, E., Dumoulin, M., Matagne, A., Wyns, L., Muyldermans, S., and Conrath, K. (2005). Identification of a universal VHH framework to graft non-canonical antigen-binding loops of camel single-domain antibodies. J Mol Biol *352*, 597-607.

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.

Saito, K., Nishida, K.M., Mori, T., Kawamura, Y., Miyoshi, K., Nagami, T., Siomi, H., and Siomi, M.C. (2006). Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. Genes Dev *20*, 2214-2222.

Saito, K., Sakaguchi, Y., Suzuki, T., Suzuki, T., Siomi, H., and Siomi, M.C. (2007). Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. Genes Dev *21*, 1603-1608.

Saksouk, N., Simboeck, E., and Dejardin, J. (2015). Constitutive heterochromatin formation and transcription in mammals. Epigenetics Chromatin *8*, 3.

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.

Schmidt, D., Wilson, M.D., Spyrou, C., Brown, G.D., Hadfield, J., and Odom, D.T. (2009). ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. Methods *48*, 240-248.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A.*, et al.* (2009). The B73 maize genome: complexity, diversity, and dynamics. Science *326*, 1112-1115.

Seller, C.A., Cho, C.Y., and O'Farrell, P.H. (2019). Rapid embryonic cell cycles defer the establishment of heterochromatin by Eggless/SetDB1 in Drosophila. Genes Dev *33*, 403-417.

Semotok, J.L., Cooperstock, R.L., Pinder, B.D., Vari, H.K., Lipshitz, H.D., and Smibert, C.A. (2005). Smaug recruits the CCR4/POP2/NOT deadenylase complex to trigger maternal transcript localization in the early Drosophila embryo. Curr Biol *15*, 284-294.

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.

Seum, C., Bontron, S., Reo, E., Delattre, M., and Spierer, P. (2007a). Drosophila G9a is a nonessential gene. Genetics *177*, 1955-1957.

Seum, C., Reo, E., Peng, H., Rauscher, F.J., 3rd, Spierer, P., and Bontron, S. (2007b). Drosophila SETDB1 is required for chromosome 4 silencing. PLoS Genet *3*, e76.

Seydoux, G. (2018). The P Granules of C. elegans: A Genetic Model for the Study of RNA-Protein Condensates. J Mol Biol *430*, 4702-4710.

Seydoux, G., and Braun, R.E. (2006). Pathway to totipotency: lessons from germ cells. Cell *127*, 891-904.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941-953.

Shirayama, M., Seth, M., Lee, H.C., Gu, W., Ishidate, T., Conte, D., Jr., and Mello, C.C. (2012). piRNAs initiate an epigenetic memory of nonself RNA in the C. elegans germline. Cell *150*, 65-77.

Shukla, R., Upton, K.R., Munoz-Lopez, M., Gerhardt, D.J., Fisher, M.E., Nguyen, T., Brennan, P.M., Baillie, J.K., Collino, A., Ghisletti, S.*, et al.* (2013). Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. Cell *153*, 101-111.

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.

Singer, B., and Kusmierek, J.T. (1982). Chemical mutagenesis. Annu Rev Biochem *51*, 655-693.

Sinzelle, L., Izsvak, Z., and Ivics, Z. (2009). Molecular domestication of transposable elements: from detrimental parasites to useful host genes. Cell Mol Life Sci *66*, 1073-1093.

Soares, L.M., He, P.C., Chun, Y., Suh, H., Kim, T., and Buratowski, S. (2017). Determinants of Histone H3K4 Methylation Patterns. Mol Cell *68*, 773-785 e776.

Song, J.J., Smith, S.K., Hannon, G.J., and Joshua-Tor, L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. Science *305*, 1434-1437.

Stabell, M., Eskeland, R., Bjorkmo, M., Larsson, J., Aalen, R.B., Imhof, A., and Lambertsson, A. (2006). The Drosophila G9a gene encodes a multi-catalytic histone methyltransferase required for normal development. Nucleic Acids Res *34*, 4609-4621.

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.

Straight, A.F., Belmont, A.S., Robinett, C.C., and Murray, A.W. (1996). GFP tagging of budding yeast chromosomes reveals that protein-protein interactions can mediate sister chromatid cohesion. Curr Biol *6*, 1599-1608.

Stratton, M.R., Campbell, P.J., and Futreal, P.A. (2009). The cancer genome. Nature *458*, 719-724.

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.

Sultana, T., Zamborlini, A., Cristofari, G., and Lesage, P. (2017). Integration site selection by retroviruses and transposable elements in eukaryotes. Nat Rev Genet *18*, 292-308.

Suyama, M., Doerks, T., Braun, I.C., Sattler, M., Izaurralde, E., and Bork, P. (2000). Prediction of structural domains of TAP reveals details of its interaction with p15 and nucleoporins. EMBO Rep *1*, 53-58.

Symington, L.S., and Gautier, J. (2011). Double-strand break end resection and repair pathway choice. Annu Rev Genet *45*, 247-271.

Tadros, W., and Lipshitz, H.D. (2009). The maternal-to-zygotic transition: a play in two acts. Development *136*, 3033-3042.

Teixeira, F.K., Heredia, F., Sarazin, A., Roudier, F., Boccara, M., Ciaudo, C., Cruaud, C., Poulain, J., Berdasco, M., Fraga, M.F.*, et al.* (2009). A role for RNAi in the selective correction of DNA methylation defects. Science *323*, 1600-1604.

Teixeira, F.K., Okuniewska, M., Malone, C.D., Coux, R.X., Rio, D.C., and Lehmann, R. (2017). piRNA-mediated regulation of transposon alternative splicing in the soma and germ line. Nature *552*, 268-272.

Thompson, P.J., Macfarlan, T.S., and Lorincz, M.C. (2016). Long Terminal Repeats: From Parasitic Elements to Building Blocks of the Transcriptional Regulatory Repertoire. Mol Cell *62*, 766-776.

Trost, M., Blattner, A.C., and Lehner, C.F. (2016). Regulated protein depletion by the auxin-inducible degradation system in Drosophila melanogaster. Fly (Austin) *10*, 35-46.

Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G., and Reuter, G. (1994). The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. EMBO J *13*, 3822-3831.

Tuschl, T., Zamore, P.D., Lehmann, R., Bartel, D.P., and Sharp, P.A. (1999). Targeted mRNA degradation by double-stranded RNA in vitro. Genes Dev *13*, 3191-3197.

Tzeng, T.Y., Lee, C.H., Chan, L.W., and Shen, C.K. (2007). Epigenetic regulation of the Drosophila chromosome 4 by the histone H3K9 methyltransferase dSETDB1. Proc Natl Acad Sci U S A *104*, 12691-12696.

Ugarkovic, D. (2005). Functional elements residing within satellite DNAs. EMBO Rep *6*, 1035-1039.

van der Krol, A.R., Mur, L.A., Beld, M., Mol, J.N., and Stuitje, A.R. (1990). Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. Plant Cell *2*, 291-299.

Van Doren, M., Williamson, A.L., and Lehmann, R. (1998). Regulation of zygotic gene expression in Drosophila primordial germ cells. Curr Biol *8*, 243-246.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., and Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science *297*, 1833-1837.

Von Stetina, J.R., and Orr-Weaver, T.L. (2011). Developmental control of oocyte maturation and egg activation in metazoan models. Cold Spring Harb Perspect Biol *3*, a005553.

Wang, H., An, W., Cao, R., Xia, L., Erdjument-Bromage, H., Chatton, B., Tempst, P., Roeder, R.G., and Zhang, Y. (2003). mAM facilitates conversion by ESET of dimethyl to trimethyl lysine 9 of histone H3 to cause transcriptional repression. Mol Cell *12*, 475-487.

Wang, L., Dou, K., Moon, S., Tan, F.J., and Zhang, Z.Z. (2018). Hijacking Oogenesis Enables Massive Propagation of LINE and Retroviral Transposons. Cell *174*, 1082-1094 e1012.

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.

Watt, F., and Molloy, P.L. (1988). Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. Genes Dev *2*, 1136-1143.

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.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wienholds, E., and Plasterk, R.H. (2005). MicroRNA function in animal development. FEBS Lett *579*, 5911-5922.

Wilkinson, A.C., Nakauchi, H., and Gottgens, B. (2017). Mammalian Transcription Factor Networks: Recent Advances in Interrogating Biological Complexity. Cell Syst *5*, 319-331.

Wilson, R.C., and Doudna, J.A. (2013). Molecular mechanisms of RNA interference. Annu Rev Biophys *42*, 217-239.

Wostemeyer, J., and Kreibich, A. (2002). Repetitive DNA elements in fungi (Mycota): impact on genomic architecture and evolution. Curr Genet *41*, 189-198.

Wu, J., Yamauchi, T., and Izpisua Belmonte, J.C. (2016). An overview of mammalian pluripotency. Development *143*, 1644-1648.

Yamada, T., Fischle, W., Sugiyama, T., Allis, C.D., and Grewal, S.I. (2005). The nucleation and maintenance of heterochromatin by a histone deacetylase in fission yeast. Mol Cell *20*, 173-185.

Yamaguchi, S., Oe, A., Nishida, K.M., Yamashita, K., Kajiya, A., Hirano, S., Matsumoto, N., Dohmae, N., Ishitani, R., Saito, K.*, et al.* (2020). Crystal structure of Drosophila Piwi. Nat Commun *11*, 858.

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.

Yesbolatova, A., Natsume, T., Hayashi, K.I., and Kanemaki, M.T. (2019). Generation of conditional auxin-inducible degron (AID) cells and tight control of degron-fused proteins using the degradation inhibitor auxinole. Methods *164-165*, 73-80.

Yu, H., Moss, B.L., Jang, S.S., Prigge, M., Klavins, E., Nemhauser, J.L., and Estelle, M. (2013). Mutations in the TIR1 auxin receptor that increase affinity for auxin/indole-3-acetic acid proteins result in auxin hypersensitivity. Plant Physiol *162*, 295-303.

Yu, R., Wang, X., and Moazed, D. (2018). Epigenetic inheritance mediated by coupling of RNAi and histone H3K9 methylation. Nature *558*, 615-619.

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.

Yuan, K., and O'Farrell, P.H. (2016). TALE-light imaging reveals maternally guided, H3K9me2/3-independent emergence of functional heterochromatin in Drosophila embryos. Genes Dev *30*, 579-593.

Yuan, S., Schuster, A., Tang, C., Yu, T., Ortogero, N., Bao, J., Zheng, H., and Yan, W. (2016). Sperm-borne miRNAs and endo-siRNAs are important for fertilization and preimplantation embryonic development. Development *143*, 635-647.

Zaessinger, S., Busseau, I., and Simonelig, M. (2006). Oskar allows nanos mRNA translation in Drosophila embryos by preventing its deadenylation by Smaug/CCR4. Development *133*, 4573-4583.

Zalokar, M. (1976). Autoradiographic study of protein and RNA formation during early development of Drosophila eggs. Dev Biol *49*, 425-437.

Zenk, F., Loeser, E., Schiavo, R., Kilpert, F., Bogdanovic, O., and Iovino, N. (2017). Germ line-inherited H3K27me3 restricts enhancer function during maternal-to-zygotic transition. Science *357*, 212-216.

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, H., Lang, Z., and Zhu, J.K. (2018). Dynamics and function of DNA methylation in plants. Nat Rev Mol Cell Biol *19*, 489-506.

Zhang, L., Ward, J.D., Cheng, Z., and Dernburg, A.F. (2015). The auxin-inducible degradation (AID) system enables versatile conditional protein depletion in C. elegans. Development *142*, 4374-4384.

Zhang, Z., Wang, J., Schultz, N., Zhang, F., Parhad, S.S., Tu, S., Vreven, T., Zamore, P.D., Weng, Z., and Theurkauf, W.E. (2014). 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.

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.

Zhuang, J., Wang, J., Theurkauf, W., and Weng, Z. (2014). TEMP: a computational method for analyzing transposable element polymorphism in populations. Nucleic Acids Res *42*, 6826-6838.

8 Analysis Scripts

Code 8.1: Example script used for RNA-Seq analysis

```
#!/bin/bash

#SBATCH -n 1 # one CPU
#SBATCH -N 1 # on one node
#SBATCH -t 0-2:00 # Running time of 2hr
#SBATCH --mem 20000 # Memory request
#SBATCH --mail-type=ALL
#SBATCH --mail-user=martin.fabry91@googlemail.com

source ~/software/MyVE/bin/activate


echo 'Start' ${1}
filename=${1%%.fq}
echo 'Start' $filename;
fastx_trimmer -Q33 -f 2 -l 49 -i $1 -o $filename.trimmed;
echo 'Trimmer done ' $filename;
#2 mismatches per read, add --alignIntronMax 1 to suppress spliced reads,
add --alignEndsType EndToEnd to account for difficult to map indels (makes
it more relatable to bowtie)
STAR --genomeDir
/Users/fabry01/genomes/drosophila/dm6/UCSC/dm6/Sequence/STAR-
Index/te_fused_clean_3 --outSAMtype BAM SortedByCoordinate --
outFilterMultimapNmax 1000 --outMultimapperOrder Random --outSAMmultNmax 1
--alignIntronMax 100 --alignEndsType EndToEnd --outFilterMismatchNoverLmax
0.04 --outReadsUnmapped Fastx --readFilesIn $filename.trimmed --runThreadN
6 --outFileNamePrefix $filename.te.;
echo 'STAR dm6 done' $filename;
samtools index $filename.te.Aligned.sortedByCoord.out.bam;
echo 'Index bai done' $filename;
STAR --genomeDir
/Users/fabry01/genomes/drosophila/dm6/UCSC/dm6/Sequence/STAR-Index/ --
outSAMtype BAM SortedByCoordinate --outFilterMultimapNmax 1000 --
outMultimapperOrder Random --outSAMmultNmax 1 --alignIntronMax 100 --
alignEndsType EndToEnd --outFilterMismatchNoverLmax 0.04 --readFilesIn
$filename.te.Unmapped.out.mate1 --runThreadN 6 --outFileNamePrefix
$filename.dm6.;
samtools index $filename.dm6.Aligned.sortedByCoord.out.bam;
echo 'Index bai done' $filename;

wait
```

```
# 16 = antisense, 0 = sense transcripts mapping to TE genome
# Use sorted bam file as input
# Generates list with read number for each TE

#Split te reads into sense (0) and antisense (16)

samtools view -f 0x10 -b $filename.te.Aligned.sortedByCoord.out.bam >
$filename.te.0.bam;
samtools index $filename.te.0.bam;
samtools idxstats $filename.te.0.bam | cut -f 1,3 >
$filename.te.0.chrom_reads.txt;


samtools view -F 0x10 -b $filename.te.Aligned.sortedByCoord.out.bam >
$filename.te.16.bam;
samtools index $filename.te.16.bam;
samtools idxstats $filename.te.16.bam | cut -f 1,3 >
$filename.te.16.chrom_reads.txt;

#Split dm6 reads into sense (0) and antisense (16)

samtools view -f 0x10 -b $filename.dm6.Aligned.sortedByCoord.out.bam >
$filename.dm6.0.bam;
samtools index $filename.dm6.0.bam;
samtools idxstats $filename.dm6.0.bam | cut -f 1,3 >
$filename.dm6.0.chrom_reads.txt;


samtools view -F 0x10 -b $filename.dm6.Aligned.sortedByCoord.out.bam >
$filename.dm6.16.bam;
samtools index $filename.dm6.16.bam;
samtools idxstats $filename.dm6.16.bam | cut -f 1,3 >
$filename.dm6.16.chrom_reads.txt;


wait

# Normalisation with scaling factor!!!

echo 'Start Normalization'

#Calculate total reads
var1=$(samtools view -c -F 260 $filename.dm6.Aligned.sortedByCoord.out.bam)

var2=$(samtools view -c -F 260 $filename.te.Aligned.sortedByCoord.out.bam)

var3=$((var1+var2))
```

```
#calculate reads for condition (te.0) / total reads

var4=$(samtools view -c -F 260 $filename.te.0.bam)

var5=$(bc <<<"scale=10; $var4 / $var3")

echo $var5 > $filename.te.0.scaling.factor.txt

bamCoverage -b $filename.te.0.bam -bs 10 -p 4 --scaleFactor $var5 --
normalizeUsing CPM -o $filename.te.0.cpm.norm.bw

wait

#calculate reads for condition (te.16) / total reads

var4=$(samtools view -c -F 260 $filename.te.16.bam)

var5=$(bc <<<"scale=10; $var4 / $var3")

echo $var5 > $filename.te.16.scaling.factor.txt

bamCoverage -b $filename.te.16.bam -bs 10 -p 4 --scaleFactor $var5 --
normalizeUsing CPM -o $filename.te.16.cpm.norm.bw

wait

#calculate reads for condition (dm6.0) / total reads

var4=$(samtools view -c -F 260 $filename.dm6.0.bam)

var5=$(bc <<<"scale=10; $var4 / $var3")

echo $var5 > $filename.dm6.0.scaling.factor.txt

bamCoverage -b $filename.dm6.0.bam -bs 10 -p 4 --scaleFactor $var5 --
normalizeUsing CPM -o $filename.dm6.0.cpm.norm.bw

wait

#calculate reads for condition (dm6.16) / total reads

var4=$(samtools view -c -F 260 $filename.dm6.16.bam)

var5=$(bc <<<"scale=10; $var4 / $var3")

echo $var5 > $filename.dm6.16.scaling.factor.txt
```

```
bamCoverage -b $filename.dm6.16.bam -bs 10 -p 4 --scaleFactor $var5 --
normalizeUsing CPM -o $filename.dm6.16.cpm.norm.bw


wait

wait

#htseq on dm6.Ali

htseq-count -s reverse -f bam -i gene_name
$filename.dm6.Aligned.sortedByCoord.out.bam
/Users/fabry01/genomes/drosophila/dm6/UCSC/dm6/Sequence/GTF-
files/ensembl_dm6_v2_rrna_dep.gtf > $filename.count.htseq

exit
```

Code 8.2: Example R script used to analysis differential expression for knockdowns

```
#####RNA-seq analysis for emf33 comb libraries (OSC)
library(ggplot2)
library(ggrepel)
library(stringr)
#RNA-seq for TE only

setwd("~/Dropbox (hannonlab)/Sequencing
Data/Roo/rnaseq/emf85_86/new_april20/count")

te_class <- read.table("~/Dropbox (hannonlab)/Sequencing
Data/deeptools_annotation/te_list_gl_vs_soma_merged.txt", as.is = TRUE)

###Cond1 is Aux, Cond2 +Aux

reads_cond1 <- read.table("emf86_3.te.0.chrom_reads.txt", as.is = TRUE)
reads_cond2 <- read.table("emf86_4.te.0.chrom_reads.txt", as.is = TRUE)

scale_f1 <- read.table("emf86_3.te.0.scaling.factor.txt", as.is = TRUE)
scale_f2 <- read.table("emf86_4.te.0.scaling.factor.txt", as.is = TRUE)

te_reads <- cbind(reads_cond1[1:122,],reads_cond2[1:122,2])


#Calculate RPM
rownames(te_reads) <- te_class[,1]
#te_reads[,2:5] <- te_reads[,2:5]*(1000000/colSums(te_reads[,2:5]))
```

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

#Normalize to scaling factor (te/dm6 reads) !!!!!!!!!!!!Always set the
scaling factor!!!!!!!!!!!!!!!!!
te_reads[,2] <- te_reads[,2]*as.numeric(scale_f1)
te_reads[,3] <- te_reads[,3]*as.numeric(scale_f2)

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



########Calculate FC for tes###############

fc_te <- te_reads[,3]/te_reads[,2]
fc_te <- as.data.frame(fc_te)
fc_te <- cbind(te_reads[,1],fc_te)

########################################
#Log2 scaling
te_reads[,2:3] <- log2(te_reads[,2:3])

#edit rownames and add class column

test <- str_split_fixed(rownames(te_reads), "_", 2)
te_reads[,1] <- test[,1]
te_reads <- cbind(te_reads, test[,2])
colnames(te_reads) <- c("te", "cond1", "cond2", "class")



#RNA-seq for dm6

dm6_reads_cond1 <- read.table("emf86_3.count.htseq", as.is = TRUE)
dm6_reads_cond2 <- read.table("emf86_4.count.htseq", as.is = TRUE)

dm6_reads <- cbind(dm6_reads_cond1, dm6_reads_cond2[,2])
dm6_reads <- dm6_reads[1:17622,]
#Calculate rpm
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("cond1", "cond2")
#Removes all rows with at least one value < 1
dm6_reads_rpm <- dm6_reads_rpm[!rowSums(dm6_reads_rpm < 1),]
```

```
########Calculate FC for genes###############

fc_genes <- dm6_reads_rpm[,3]/dm6_reads_rpm[,2]
fc_genes <- as.data.frame(fc_genes)
fc_genes <- cbind(dm6_reads_rpm[,1],fc_genes)

########################################

dm6_reads_rpm_log2 <- log2(dm6_reads_rpm[2:3])


# Combine Genes and TE in one plot

colnames(te_reads) <- c("te", "cond1", "cond2", "class")
dm6_reads_rpm_log2 <-
cbind(rownames(dm6_reads_rpm_log2),dm6_reads_rpm_log2)
dm6_reads_rpm_log2 <- data.frame(dm6_reads_rpm_log2,"gene")
colnames(dm6_reads_rpm_log2) <- c("te", "cond1", "cond2", "class")
te_dm6_comb_reads_log2 <- rbind(dm6_reads_rpm_log2,te_reads)

###################Add columns for fc#######

#FC > 2
fc_genes <- as.matrix(fc_genes[,2])
fc_te <- as.matrix(fc_te[,2])
fc_all <- rbind(fc_genes,fc_te)
fc_all_pos <- fc_all > 2
te_dm6_comb_reads_log2 <- cbind(te_dm6_comb_reads_log2, fc_all_pos)

#FC > 0.5

fc_all_neg <- fc_all < 0.5
te_dm6_comb_reads_log2 <- cbind(te_dm6_comb_reads_log2,fc_all_neg)


colnames(te_dm6_comb_reads_log2) <- c("te", "cond1", "cond2",
"class","fc_+2","fc_-2")

####### Set all genes fc to false #####

te_dm6_comb_reads_log2[c(grep("gene", te_dm6_comb_reads_log2[,4])),5] <-
"TEST"
te_dm6_comb_reads_log2[c(grep("gene", te_dm6_comb_reads_log2[,4])),6] <-
"TEST"

# Calculate genes correlation
x <- te_dm6_comb_reads_log2[grep("gene", te_dm6_comb_reads_log2[,4]),2]
```

```
y <- te_dm6_comb_reads_log2[grep("gene", te_dm6_comb_reads_log2[,4]),3]
rsquarelm2 <-cor(x,y)

########################################

ggplot(te_dm6_comb_reads_log2, aes(x=cond1, y=cond2)) +
  geom_point(data=te_dm6_comb_reads_log2[1:7162,], aes(x=cond1, y=cond2),
colour="grey", pch=1) +

geom_point(data=te_dm6_comb_reads_log2[grep("TRUE",te_dm6_comb_reads_log2[,
5]),], aes(x=cond1, y=cond2), colour="dodgerblue1") +

geom_point(data=te_dm6_comb_reads_log2[grep("FALSE",te_dm6_comb_reads_log2[
,5]),], aes(x=cond1, y=cond2), colour="purple") +

geom_point(data=te_dm6_comb_reads_log2[grep("TRUE",te_dm6_comb_reads_log2[,
6]),], aes(x=cond1, y=cond2), colour="dodgerblue1") +

#geom_text_repel(data=te_dm6_comb_reads_log2[grep("TRUE",te_dm6_comb_reads_
log2[,5]),], aes(label=te,hjust=0, vjust=0)) +

#geom_text_repel(data=te_dm6_comb_reads_log2[grep("TRUE",te_dm6_comb_reads_
log2[,6]),], aes(label=te,hjust=0, vjust=0)) +

geom_text_repel(data=te_dm6_comb_reads_log2[grep("\\broo\\b",te_dm6_comb_re
ads_log2[,1]),], aes(label=te,hjust=0, vjust=0)) +
  geom_abline(intercept = 1, slope = 1, linetype="dotted", colour="red") +
  geom_abline(intercept = 0, slope = 1, linetype="dotted", colour="black")
+
  geom_abline(intercept = -1, slope = 1, linetype="dotted", colour="blue")
+
  scale_x_continuous(expand = c(0, 0)) +
  scale_y_continuous(expand = c(0, 0)) +
  coord_cartesian(ylim = c(0,15),xlim = c(0,15)) +
  ylab("log2 RPM Auxin") +
  xlab("log2 RPM Control") +
  theme_classic() +
  annotate("text", x = 11, y = 14, label = bquote(R^2 ==
.(round(rsquarelm2, 2))))  +
  theme(axis.text.x = element_text(size = 15, angle = 0, hjust = .5, vjust
= .5, face = "plain"),
        axis.text.y = element_text(size = 15, angle = 0, hjust = .5, vjust
= .5, face = "plain"),
        axis.title.x = element_text(size = 15, angle = 0, hjust = .5, vjust
= .5, face = "plain"),
        axis.title.y = element_text(size = 15, angle = 90, hjust = .5,
vjust = .5, face = "plain"))
```

```
#ggsave(filename="~/Dropbox (hannonlab)/Sequencing
Data/Roo/rnaseq/emf85_86/new_april20/analysis/emf86_3_vs4.svg", plot =
last_plot(), width=5, height=5)
```

Code 8.3: Example R script used for Deseq2 differential gene expression analysis

```
### Analysis of eMF73 auxin treated embryos

library("ggplot2", lib.loc="~/Library/R/3.6/library")
library("DESeq2")
library("pheatmap")
library(ggrepel)
library(stringr)
library(svglite)
library("ggpubr", lib.loc="~/Library/R/3.6/library")

directory <- "Dropbox (hannonlab)/Sequencing
Data/rna_seq/emf73/new_indi_april20/count/"


sampleFiles <- grep("te_dm6_count.txt",list.files(directory),value=TRUE)
sampleCondition <- as.character(replicate(3,c("ctrl", "auxin")))
sampleBatch <- as.character(c("a","a","b","b","c","c"))
sampleName <- gsub("_te_dm6.*", "", sampleFiles)
sampleTable <- data.frame(sampleName = sampleName,
                          fileName = sampleFiles,
                          condition = sampleCondition,
                          batch = sampleBatch)



ddsHTSeq <- DESeqDataSetFromHTSeqCount(sampleTable = sampleTable,
                                       directory = directory,
                                       design= ~ batch + condition)

ddsHTSeq$condition <- factor(ddsHTSeq$condition, levels =
c("ctrl","auxin"))

dds <- DESeq(ddsHTSeq)
res <- results(dds)

resLFC <- lfcShrink(dds, coef="condition_auxin_vs_ctrl", type="apeglm")
write.table(resLFC, "~/Desktop/test.txt", quote = F, sep = "\t")

resOrdered <- res[order(res$pvalue),]
```

```
plotMA(res, ylim=c(-2,2))
plotMA(resLFC, ylim=c(-2,2))

results <- as.data.frame(resLFC)
results$te <- grepl("FBgn", rownames(results))
results$gene <- str_split_fixed(rownames(results), "_", 2)[,2]
results[,1] <- log10(results[,1])


te <- ggplot(results, aes(x=baseMean, y=log2FoldChange)) +
  geom_point(data=results[which(results[,6]=='TRUE'& results[,5] > 0.05),],
aes(x=baseMean, y=log2FoldChange), colour="grey", alpha=1) +
  geom_point(data=results[which(results[,6]=='TRUE'& is.na(results[,5])),],
aes(x=baseMean, y=log2FoldChange), colour="grey", alpha=1) +
  geom_point(data=results[which(results[,6]=='TRUE'& results[,5] < 0.05 &
results[,2] < 1),], aes(x=baseMean, y=log2FoldChange), colour="purple",
size=2, alpha=0.5) +
  geom_point(data=results[which(results[,6]=='TRUE'& results[,5] < 0.05 &
results[,2] > 1),], aes(x=baseMean, y=log2FoldChange), colour="red",
size=2) +
  geom_text_repel(data=results[which(results[,6]=='TRUE'& results[,5] <
0.05 & results[,2] < 1),], aes(label=gene,hjust=0, vjust=0)) +
  geom_text_repel(data=results[which(results[,6]=='TRUE'& results[,5] <
0.05 & results[,2] > 1),], aes(label=gene,hjust=0, vjust=0)) +
  geom_abline(intercept = 1, slope = 0, linetype="dotted", colour="red") +
  geom_abline(intercept = 0, slope = 0, linetype="dotted", colour="black")
+
  geom_abline(intercept = -1, slope = 0, linetype="dotted", colour="blue")
+
  scale_x_continuous(expand = c(0, 0)) +
  scale_y_continuous(expand = c(0, 0)) +
  coord_cartesian(ylim = c(-1.6,1.6),xlim = c(1,5)) +
  ylab("log2 Fold Change auxin vs ctrl") +
  xlab("log10 mean expression") +
  theme_classic() +
  theme(axis.text.x = element_text(size = 15, angle = 0, hjust = .5, vjust
= .5, face = "plain"),
        axis.text.y = element_text(size = 15, angle = 0, hjust = .5, vjust
= .5, face = "plain"),
        axis.title.x = element_text(size = 15, angle = 0, hjust = .5, vjust
= .5, face = "plain"),
        axis.title.y = element_text(size = 15, angle = 90, hjust = .5,
vjust = .5, face = "plain"))

#ggsave(filename="~/Dropbox (hannonlab)/Sequencing
Data/rna_seq/emf73/new_indi_april20/analysis/emf73_te_dm6_deseq2_te.svg",
plot = last_plot(), width=5, height=5)
```

```
dm6 <-ggplot(results, aes(x=baseMean, y=log2FoldChange)) +
  geom_point(data=results[which(results[,6]=='FALSE'& results[,5] >
0.05),], aes(x=baseMean, y=log2FoldChange), colour="grey", alpha=0.1) +
  geom_point(data=results[which(results[,6]=='FALSE'&
is.na(results[,5])),], aes(x=baseMean, y=log2FoldChange), colour="grey",
alpha=0.1) +
  geom_point(data=results[which(results[,6]=='FALSE'& results[,5] <
0.05),], aes(x=baseMean, y=log2FoldChange), colour="purple", alpha=0.3) +
  geom_point(data=results[which(results[,6]=='FALSE'& results[,5] < 0.05 &
results[,2] < -1),], aes(x=baseMean, y=log2FoldChange), colour="blue") +
  geom_point(data=results[which(results[,6]=='FALSE'& results[,5] < 0.05 &
results[,2] > 1),], aes(x=baseMean, y=log2FoldChange), colour="red") +
  geom_abline(intercept = 1, slope = 0, linetype="dotted", colour="red") +
  geom_abline(intercept = 0, slope = 0, linetype="dotted", colour="black")
+
  geom_abline(intercept = -1, slope = 0, linetype="dotted", colour="blue")
+
  scale_x_continuous(expand = c(0, 0)) +
  scale_y_continuous(expand = c(0, 0)) +
  coord_cartesian(ylim = c(-2.5,2.5),xlim = c(1,5)) +
  ylab("log2 Fold Change auxin vs ctrl") +
  xlab("log10 mean expression") +
  theme_classic() +
  theme(axis.text.x = element_text(size = 15, angle = 0, hjust = .5, vjust
= .5, face = "plain"),
        axis.text.y = element_text(size = 15, angle = 0, hjust = .5, vjust
= .5, face = "plain"),
        axis.title.x = element_text(size = 15, angle = 0, hjust = .5, vjust
= .5, face = "plain"),
        axis.title.y = element_text(size = 15, angle = 90, hjust = .5,
vjust = .5, face = "plain"))

#ggsave(filename="~/Dropbox (hannonlab)/Sequencing
Data/rna_seq/emf73/new_indi_april20/analysis/emf73_te_dm6_deseq2_genes.svg"
, plot = last_plot(), width=5, height=5)




### Data sets used for defining mat or zygotic genes
```

```
htseq_0_2h <- read.table("~/Dropbox (hannonlab)/Sequencing
Data/Roo/rnaseq/emf22/comb/dm6/count/1_comb.count.htseq", as.is=T, nrows =
17622)
htseq_3_4h <- read.table("~/Dropbox (hannonlab)/Sequencing
Data/Roo/rnaseq/emf22/comb/dm6/count/4_comb.count.htseq", as.is=T, nrows =
17622)

htseq_0_2h[,2] <- as.numeric(htseq_0_2h[,2]*(1000000/sum(htseq_0_2h[,2])))
htseq_3_4h[,2] <- as.numeric(htseq_3_4h[,2]*(1000000/sum(htseq_3_4h[,2])))

mat_zyg <- cbind(htseq_0_2h,htseq_3_4h[,2])


#Removes all rows with sum < 2 rpm
mat_zyg <- mat_zyg[!rowSums(mat_zyg[,2:3]) < 2,]

### Define maternally deposited genes only (not zygotically expressed in 3-
4h embryos)
fold10 <- mat_zyg[,2]/mat_zyg[,3] >10

mat_dep_only <- cbind(mat_zyg[,1:2], mat_zyg[,3], fold10)

mat_dep_only <- mat_dep_only[grep("TRUE", mat_dep_only[,4]),1]



### Define zygotically expressed genes only (not maternally deposited but
expressed in 3-4h embryos)
fold0.1 <- mat_zyg[,2]/mat_zyg[,3] <0.2

zygote_only <- cbind(mat_zyg[,1:2], mat_zyg[,3], fold0.1)

zygote_only <- zygote_only[grep("TRUE", zygote_only[,4]),1]


results[,8] <- grepl(paste(mat_dep_only, collapse="\\b|\\b"),
rownames(results))
results[,9] <- grepl(paste(zygote_only, collapse="\\b|\\b"),
rownames(results))



te_only <- te_dm6_comb_reads_log2[-grep("gene",
te_dm6_comb_reads_log2[,4]),]
genes_only <- te_dm6_comb_reads_log2[grep("gene",
te_dm6_comb_reads_log2[,4]),]
```

```
colnames(results) <- c("baseMean", "log2FoldChange", "lfcSE", "pvalue",
"padj", "te", "gene", "mat", "zyg")

mat_zyg <- ggplot(results, aes(x=baseMean, y=log2FoldChange)) +
  geom_point(data=results[which(results[,6]=='FALSE'& results[,5] >
0.05),], aes(x=baseMean, y=log2FoldChange), colour="grey", alpha=0.1) +
  geom_point(data=results[which(results[,6]=='FALSE'&
is.na(results[,5])),], aes(x=baseMean, y=log2FoldChange), colour="grey",
alpha=0.1) +
  geom_point(data=results[which(results[,6]=='FALSE'& results[,5] <
0.05),], aes(x=baseMean, y=log2FoldChange), colour="purple", alpha=0.3) +
  geom_point(data=results[which(results[,6]=='FALSE'& results[,5] < 0.05 &
results[,2] < -1),], aes(x=baseMean, y=log2FoldChange), colour="blue") +
  geom_point(data=results[which(results[,6]=='FALSE'& results[,5] < 0.05 &
results[,2] > 1),], aes(x=baseMean, y=log2FoldChange), colour="red") +
  geom_point(data=results[which(results[,6]=='FALSE'& results[,5] < 0.05 &
results[,2] < -1 &results[,8]=='TRUE'),], aes(x=baseMean,
y=log2FoldChange), colour="magenta", size=2) +
  geom_point(data=results[which(results[,6]=='FALSE'& results[,5] < 0.05 &
results[,2] > 1 &results[,8]=='TRUE'),], aes(x=baseMean, y=log2FoldChange),
colour="magenta", size=2) +
  geom_point(data=results[which(results[,6]=='FALSE'& results[,5] < 0.05 &
results[,2] < -1 &results[,9]=='TRUE'),], aes(x=baseMean,
y=log2FoldChange), colour="orange", size=2) +
  geom_point(data=results[which(results[,6]=='FALSE'& results[,5] < 0.05 &
results[,2] > 1 &results[,9]=='TRUE'),], aes(x=baseMean, y=log2FoldChange),
colour="orange", size=2) +
  geom_abline(intercept = 1, slope = 0, linetype="dotted", colour="red") +
  geom_abline(intercept = 0, slope = 0, linetype="dotted", colour="black")
+
  geom_abline(intercept = -1, slope = 0, linetype="dotted", colour="blue")
+
  scale_x_continuous(expand = c(0, 0)) +
  scale_y_continuous(expand = c(0, 0)) +
  coord_cartesian(ylim = c(-2.5,2.5),xlim = c(1,5)) +
  ylab("log2 Fold Change auxin vs ctrl") +
  xlab("log10 mean expression") +
  theme_classic() +
  theme(axis.text.x = element_text(size = 15, angle = 0, hjust = .5, vjust
= .5, face = "plain"),
        axis.text.y = element_text(size = 15, angle = 0, hjust = .5, vjust
= .5, face = "plain"),
        axis.title.x = element_text(size = 15, angle = 0, hjust = .5, vjust
= .5, face = "plain"),
        axis.title.y = element_text(size = 15, angle = 90, hjust = .5,
vjust = .5, face = "plain"))
```

```
#ggsave(filename="~/Dropbox (hannonlab)/Sequencing
Data/rna_seq/emf73/new_indi_april20/analysis/emf73_te_dm6_deseq2_genes_mat_
zyg.svg", plot = last_plot(), width=5, height=5)


ggarrange(te, dm6, mat_zyg,
          labels = c("A", "B", "C"),
          ncol = 3, nrow = 1)

#ggsave(filename="~/Dropbox (hannonlab)/Sequencing
Data/rna_seq/emf73/new_indi_april20/analysis/emf73_te_dm6_deseq2_comb.gif",
plot = last_plot(), width=15, height=5)

te

#ggsave(filename="~/Dropbox (hannonlab)/Sequencing
Data/rna_seq/emf73/new_indi_april20/analysis/emf73_te_dm6_deseq2_te.svg",
plot = last_plot(), width=5, height=5)


ggarrange(dm6, mat_zyg,
          labels = c("A", "B"),
          ncol = 2, nrow = 1)

#ggsave(filename="~/Dropbox (hannonlab)/Sequencing
Data/rna_seq/emf73/new_indi_april20/analysis/emf73_te_dm6_deseq2_genes_all.
svg", plot = last_plot(), width=10, height=5)
```

Code 8.4: Example Shell script used to analyse fRIP-Seq

```
#!/bin/bash

#SBATCH -n 1 # one CPU
#SBATCH -N 1 # on one node
#SBATCH -t 0-2:00 # Running time of 2hr
#SBATCH --mem 20000 # Memory request
#SBATCH --mail-type=ALL
#SBATCH --mail-user=martin.fabry91@googlemail.com

source ~/software/MyVE/bin/activate

echo 'Start' ${1}
filename=${1%%.fq}
fastx_trimmer -Q33 -f 2 -l 49 -i ${1} -o $filename.trimmed;
# Align to crap list and pipe remaining reads in next alignment
```

```
STAR --genomeDir
/Users/fabry01/genomes/drosophila/dm6/UCSC/dm6/Sequence/STAR-Index/craplist
--outSAMtype BAM SortedByCoordinate --outFilterMultimapNmax 1000 --
outMultimapperOrder Random --outSAMmultNmax 1 --alignIntronMax 100 --
alignEndsType EndToEnd --outFilterMismatchNoverLmax 0.04 --outReadsUnmapped
Fastx --readFilesIn $filename.trimmed --runThreadN 4 --outFileNamePrefix
$filename.crap.;
echo 'Trimmer done ' $filename;
#2 mismatches per read, add --alignIntronMax 1 to suppress spliced reads,
add --alignEndsType EndToEnd to account for difficult to map indels (makes
it more relatable to bowtie)
STAR --genomeDir
/Users/fabry01/genomes/drosophila/dm6/UCSC/dm6/Sequence/STAR-
Index/te_fused_clean_3_satellites --outSAMtype BAM SortedByCoordinate --
outFilterMultimapNmax 1000 --outMultimapperOrder Random --outSAMmultNmax 1
--alignIntronMax 100 --alignEndsType EndToEnd --outFilterMismatchNoverLmax
0.04 --outReadsUnmapped Fastx --readFilesIn
$filename.crap.Unmapped.out.mate1 --runThreadN 4 --outFileNamePrefix
$filename.te.;
echo 'STAR dm6 done' $filename;
samtools index $filename.te.Aligned.sortedByCoord.out.bam;
echo 'Index bai done' $filename;
STAR --genomeDir
/Users/fabry01/genomes/drosophila/dm6/UCSC/dm6/Sequence/STAR-Index/ --
outSAMtype BAM SortedByCoordinate --outFilterMultimapNmax 1 --
outMultimapperOrder Random --outSAMmultNmax 1 --alignIntronMax 100 --
alignEndsType EndToEnd --outFilterMismatchNoverLmax 0.04 --readFilesIn
$filename.te.Unmapped.out.mate1 --runThreadN 4 --outFileNamePrefix
$filename.dm6.;
samtools index $filename.dm6.Aligned.sortedByCoord.out.bam;
echo 'Index bai done' $filename;

rm $filename.crap.Unmapped.out.mate1
rm $filename.te.Unmapped.out.mate1

wait

# 16 = antisense, 0 = sense transcripts mapping to TE genome
# Use sorted bam file as input
# Generates list with read number for each TE

#Split te reads into sense (0) and antisense (16)

samtools view -f 0x10 -b $filename.te.Aligned.sortedByCoord.out.bam >
$filename.te.0.bam;
samtools index $filename.te.0.bam;
samtools idxstats $filename.te.0.bam | cut -f 1,3 >
$filename.te.0.chrom_reads.txt;
```

```
samtools view -F 0x10 -b $filename.te.Aligned.sortedByCoord.out.bam >
$filename.te.16.bam;
samtools index $filename.te.16.bam;
samtools idxstats $filename.te.16.bam | cut -f 1,3 >
$filename.te.16.chrom_reads.txt;

#Split dm6 reads into sense (0) and antisense (16)

samtools view -f 0x10 -b $filename.dm6.Aligned.sortedByCoord.out.bam >
$filename.dm6.0.bam;
samtools index $filename.dm6.0.bam;
samtools idxstats $filename.dm6.0.bam | cut -f 1,3 >
$filename.dm6.0.chrom_reads.txt;


samtools view -F 0x10 -b $filename.dm6.Aligned.sortedByCoord.out.bam >
$filename.dm6.16.bam;
samtools index $filename.dm6.16.bam;
samtools idxstats $filename.dm6.16.bam | cut -f 1,3 >
$filename.dm6.16.chrom_reads.txt;


wait

# Normalisation with scaling factor!!!

echo 'Start Normalization'

#Calculate total reads
var1=$(samtools view -c -F 260 $filename.dm6.Aligned.sortedByCoord.out.bam)

var2=$(samtools view -c -F 260 $filename.te.Aligned.sortedByCoord.out.bam)

var3=$((var1+var2))


#calculate reads for condition (te.0) / total reads

var4=$(samtools view -c -F 260 $filename.te.0.bam)

var5=$(bc <<<"scale=10; $var4 / $var3")

echo $var5 > $filename.te.0.scaling.factor.txt

bamCoverage -b $filename.te.0.bam -bs 10 -p 4 --scaleFactor $var5 --
normalizeUsing CPM -o $filename.te.0.cpm.norm.bw
```

```
wait

#calculate reads for condition (te.16) / total reads

var4=$(samtools view -c -F 260 $filename.te.16.bam)

var5=$(bc <<<"scale=10; $var4 / $var3")

echo $var5 > $filename.te.16.scaling.factor.txt

bamCoverage -b $filename.te.16.bam -bs 10 -p 4 --scaleFactor $var5 --
normalizeUsing CPM -o $filename.te.16.cpm.norm.bw

wait

#calculate reads for condition (dm6.0) / total reads

var4=$(samtools view -c -F 260 $filename.dm6.0.bam)

var5=$(bc <<<"scale=10; $var4 / $var3")

echo $var5 > $filename.dm6.0.scaling.factor.txt

bamCoverage -b $filename.dm6.0.bam -bs 10 -p 4 --scaleFactor $var5 --
normalizeUsing CPM -o $filename.dm6.0.cpm.norm.bw

wait

#calculate reads for condition (dm6.16) / total reads

var4=$(samtools view -c -F 260 $filename.dm6.16.bam)

var5=$(bc <<<"scale=10; $var4 / $var3")

echo $var5 > $filename.dm6.16.scaling.factor.txt

bamCoverage -b $filename.dm6.16.bam -bs 10 -p 4 --scaleFactor $var5 --
normalizeUsing CPM -o $filename.dm6.16.cpm.norm.bw


wait

wait

#htseq on dm6.Ali
```

```
htseq-count -s yes -f bam -i gene_name
$filename.dm6.Aligned.sortedByCoord.out.bam
/Users/fabry01/genomes/drosophila/dm6/UCSC/dm6/Sequence/GTF-
files/ensembl_dm6_v2_rrna_dep.gtf > $filename.count.htseq
exit
```

Code 8.5: Example Shell script used for small RNA-Seq analysis

```
#!/bin/bash

# Analysis for Clust1, alignment to TE list only

source ~/software/MyVE/bin/activate


for i in *.fq; do (
        filename=${i%%.fq}
        echo 'Start' $filename;
        fastx_clipper -a AGATCGGAAGAGCACACGTCT -l 15 -c -Q33 -i $i -o
$filename.clipped;
        seqtk trimfq -b 4 -e 4 $filename.clipped > $filename.trimmed;
        ) &
done

wait

echo 'Start Part 1'

for i in *.trimmed; do (
        filename=${i%%.trimmed}
        #2 mismatches per read, add --alignIntronMax 1 to suppress spliced
reads, add --alignEndsType EndToEnd to account for difficult to map indels
(makes it more relatable to bowtie)
        STAR --outSAMunmapped Within --genomeDir
/Users/fabry01/genomes/drosophila/dm6/UCSC/dm6/Sequence/STAR-
Index/te_fused_clean_3 --limitBAMsortRAM 1098266374 --outSAMtype BAM
SortedByCoordinate --outFilterMultimapNmax 1000000000 --outMultimapperOrder
Random --outSAMmultNmax 1 --alignIntronMax 100 --alignEndsType EndToEnd --
outFilterMismatchNoverLmax 0.04 --readFilesIn $filename.trimmed --
runThreadN 2 --outFileNamePrefix $filename.;
        echo 'STAR dm6 done' $filename;
        samtools index $filename.Aligned.sortedByCoord.out.bam;
        samtools view -hf 4 $filename.Aligned.sortedByCoord.out.bam >
$filename.unmap.sam;
        samtools view -S -b $filename.unmap.sam > $filename.unmap.bam;
        bamToFastq -i $filename.unmap.bam -fq $filename.unmap.fq;
```

```
        STAR --genomeDir
/Users/fabry01/genomes/drosophila/dm6/UCSC/dm6/Sequence/STAR-Index/ --
outSAMtype BAM SortedByCoordinate --outFilterMultimapNmax 100000 --
outMultimapperOrder Random --outSAMmultNmax 1 --alignIntronMax 100 --
alignEndsType EndToEnd --outFilterMismatchNoverLmax 0.04 --readFilesIn
$filename.unmap.fq --runThreadN 2 --outFileNamePrefix $filename.dm6.multi.;
        samtools index $filename.dm6.multi.Aligned.sortedByCoord.out.bam;
        echo 'Index bai done' $filename;
        ) &
done

wait



for i in *.Aligned.sortedByCoord.out.bam; do (
        filename=${i%%.Aligned.sortedByCoord.out.bam}
        samtools view -h $i |awk '(length($10) > 20) && (length($10) < 22)
|| $1 ~ /^@/' | samtools view -bS - >
$filename.21bp.Aligned.sortedByCoord.out.bam
        samtools view -h $i |awk '(length($10) > 22) && (length($10) < 30)
|| $1 ~ /^@/' | samtools view -bS - >
$filename.23_29bp.Aligned.sortedByCoord.out.bam
        ) &
done

wait

# the flag in samtools view removes all reads not associated with a chr or
TE, that mean all read counts printed can be used immediately

for i in *.Aligned.sortedByCoord.out.bam; do (
        filename=${i%%.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;
        echo $filename.0.te.chrom_reads.txt;
        cut -f 2 $filename.0.te.chrom_reads.txt | paste -sd+ - | bc;
        ) &
done

wait

for i in *.Aligned.sortedByCoord.out.bam; do (
        filename=${i%%.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;
        echo $filename.16.te.chrom_reads.txt;
        cut -f 2 $filename.16.te.chrom_reads.txt | paste -sd+ - | bc;
        ) &
done

wait

exit
```

Code 8.6: Example Shell script used to analyse ChIP-Seq data

```
#!/bin/bash

#SBATCH -n 1 # one CPU
#SBATCH -N 1 # on one node
#SBATCH -t 0-2:00 # Running time of 2hr
#SBATCH --mem 20000 # Memory request
#SBATCH --mail-type=ALL
#SBATCH --mail-user=martin.fabry91@googlemail.com

source ~/software/MyVE/bin/activate


echo 'Start' ${1}
filename=${1%%.fq}
echo 'Start' $filename;
fastx_trimmer -Q33 -f 2 -l 49 -i $1 -o $filename.trimmed;
echo 'Trimmer done ' $filename;
#2 mismatches per read, add --alignIntronMax 1 to suppress spliced reads,
add --alignEndsType EndToEnd to account for difficult to map indels (makes
it more relatable to bowtie)
STAR --genomeDir
/Users/fabry01/genomes/drosophila/dm6/UCSC/dm6/Sequence/STAR-
Index/te_fused_clean_3 --outSAMtype BAM SortedByCoordinate --
outFilterMultimapNmax 1000 --outMultimapperOrder Random --outSAMmultNmax 1
--alignIntronMax 100 --alignEndsType EndToEnd --outFilterMismatchNoverLmax
0.04 --outReadsUnmapped Fastx --readFilesIn $filename.trimmed --runThreadN
6 --outFileNamePrefix $filename.te.;
echo 'STAR dm6 done' $filename;
samtools index $filename.te.Aligned.sortedByCoord.out.bam;
echo 'Index bai done' $filename;
STAR --genomeDir
/Users/fabry01/genomes/drosophila/dm6/UCSC/dm6/Sequence/STAR-Index/ --
outSAMtype BAM SortedByCoordinate --outFilterMultimapNmax 1000 --
```

```
outMultimapperOrder Random --outSAMmultNmax 1 --alignIntronMax 100 --
alignEndsType EndToEnd --outFilterMismatchNoverLmax 0.04 --readFilesIn
$filename.te.Unmapped.out.mate1 --runThreadN 6 --outFileNamePrefix
$filename.dm6.;
samtools index $filename.dm6.Aligned.sortedByCoord.out.bam;
echo 'Index bai done' $filename;

wait

# Normalisation with scaling factor!!!

echo 'Start Normalization'

#Calculate total reads
var1=$(samtools view -c -F 260 $filename.dm6.Aligned.sortedByCoord.out.bam)

var2=$(samtools view -c -F 260 $filename.te.Aligned.sortedByCoord.out.bam)

var3=$((var1+var2))


#calculate reads for condition (te) / total reads

var4=$(samtools view -c -F 260 $filename.te.Aligned.sortedByCoord.out.bam)

var5=$(bc <<<"scale=10; $var4 / $var3")

echo $var5 > $filename.te.scaling.factor.txt

bamCoverage -b $filename.te.Aligned.sortedByCoord.out.bam -bs 10 -p 4 --
scaleFactor $var5 --normalizeUsing CPM -o $filename.te.scpm.norm.bw

wait



exit
```

Code 8.7: Shell script used to generate metaplot matrix for ChIP-Seq data

```
#!/bin/bash

source /Users/fabry01/software/MyVE/bin/activate

bw_files=$(find . -type f -name "*bw")

computeMatrix scale-regions -S $bw_files  -R
/Users/fabry01/annotation/roo.bed -o out.txt.gz --outFileNameMatrix
out_matrix.txt -m 9000 -bs 10

sed '1,3d' out_matrix.txt > matrix_values.txt
sed -n '1,2p' out_matrix.txt > temp.txt
echo $bw_files | cat - temp.txt matrix_values.txt > matrix_output_r.txt
sed 's/#//'g matrix_output_r.txt > matrix_output2_r.txt
rm matrix_values.txt
rm temp.txt
rm matrix_output_r.txt
rm out_matrix.txt
rm out.txt.gz

wait

Rscript /Users/fabry01/scripts/plot_deeptools.R

exit
```

Code 8.8: R Script used to generate metaplots for ChIP-Seq data

```
library(ggplot2)
library(RColorBrewer)
library(reshape2)

test <- read.table("matrix_output2_r.txt", header=F, as.is = TRUE, fill
=TRUE)
test <- as.matrix(test)

#Calculates the number of non-overlapping bins
x = as.numeric(gsub("body:", "", test[3,3]))/as.numeric(gsub("size:", "",
test[3,5]))

data = test[1,grep(".bw", test[1,])]
seq_length <-c(1:length(data))

empt <- matrix(0, ncol = x, nrow = length(data))
```

```
test <- as.matrix(as.numeric(as.character(test[4,])))
for (val in seq_length) {
  empt[val,] <- test[(1+x*(val-1)):(val*x),]
}
rownames(empt) <- data

coverage <- cbind(c(1:x), t(empt))
coverage[,1] <- matrix(nrow= x, ncol =1, c(seq(10,(x*10), by=10)))
coverage <- as.data.frame(coverage)

plot <- melt(coverage, id = c("V1"))

ggplot(plot, aes(x=V1, y=value, color=variable, group=variable))+
geom_line()

ggsave(filename="test_mdg1.pdf", plot = last_plot(), width=10, height=8)
```

Code 8.9: Fiji macro used for cytoplasmic and nuclear signal counting

```
/*
 * This macro analyzes RNA FISH images. It is fully interactive and can
work in two modes. The first one is manual, and for
 * each image the analysis settings are to be input. I suggest using it if
small tweaks have to be added from image to image.
 * The second one instead creates a file with the settings to be used and
uses them until the file is deleted. To find the file
 * after it has been created, run the command:
 */

//print(getDirectory("macros"));

/*
 * It is in there and it's called settings.txt. Just delete it and restart
the macro to create a new one.
 * Alessandro Passera, August 2018.
 */

/*
 * About the edge detection. This macro implements a difference of gaussian
(DoG, or mexican hat) filter, which is suitable
 * for both the detection of dots and edges: let's call them "features". If
your feature of interest is 10 pixel
 * of radius, and there's other irrelevant things in the picture, both
smaller and bigger, blurring the picture
 * with a gaussian filter of, say, 12 pixels of sigma will make anything
smaller unrecognizable. That's both
```

```
 * your features of interest and the smaller ones. The bigger ones will be
quite unchanged. A sigma of, say,
 * 8 pixels will, instead, blur only the smaller features, leaving intact
the big ones and the interesting ones.
 * If now we subtract the "larger sigma filtered" one from the other, the
small features (blurred more or less
 * equally in both) will go away, same for the big ones (left more or less
unchanged in both). This will leave
 * only the features between 8 and 12 pixels. The downside of this is you
have to know, more or less, the size
 * of the features. Also, the two sigmas should be as close as possible, so
that the two pictures are blurred
 * as similarly as possible and thus small and large features cancel out as
much as possible. However, this
 * isn't a big problem, since it's quite flexible: only a rough idea is
necessary, since using even quite different
 * sigmas still gives excellent results. So, don't be afraid to just
measure a couple of border widths and go
 * with the average. A better alternative is to drw a line across the
border, fit a gaussian distribution to the plot
 * profile and use the standard deviation of that distribution, but it
isn't really needed.
 * I personally used 7 and 3, since the bright edge of the membrane
staining was usually around
 * 8-9 micrometers, which is around 10 pixels, so 5 is in the 7-3 range.
Just for reference, I tested other,
 * albeit less performing, filters: laplacian, gradient magnitude (both
Robert's cross gradient and Sobel), the simple
 * "Edges" LUT, a more refined filter based on fuzzy set theory, and a
maximum filtered - minimum filtered approach.
 */

/*
 * About the median filter: this filter can take out features, and is
therefore to be used VERY carefully.
 * Specifically, I didn't really know what mask each "radius" value
corresponded too, so I made some tests.
 * A radius from 1 to 2 (not included) corresponds to a 3x3 mask. This
takes out anything equal to or smaller than 4 pixel.
 * A radius of 2,5 means a 5x5 mask: that takes out anything smaller than
12 pixels. That's definitely too much.
 * I couldn't be sure of radius=2. However, a weighted median filter with a
4x4 mask centered on the pixel and
 * weighing the pixels on the border half as much as the ones inside works.
I don't know if that's the rule.
 * Knowing this, I still believe that the radius=2 filter gives the
cleanest results. If you think it's too
```

```
 * big, put it to 1. A good test could be looking at a plot frequency over
area of your dots.
 */

setBatchMode(true);

Dialog.create("Hi! Do you want to process a lot of images all in the same
way?");
Dialog.addChoice("", newArray("No", "Yes"));
Dialog.show();
lifechoice = Dialog.getChoice();

if(lifechoice == "No"){
        Dialog.create("You need at least nuclei and signal for this to
work. Which channel do you have your nuclei in?");
        Dialog.addNumber("Nuclei channel:", 1);
        Dialog.show();
        nucleichannel = Dialog.getNumber();

        Dialog.create("Do you have a cytoplasmic staining? Or do you want
to use the nuclei to infer the cell areas?");
        Dialog.addChoice("Cell area from:", newArray("nuclei",
"cytoplasm"));
        Dialog.show();
        cytochoice = Dialog.getChoice();

        if(cytochoice == "cytoplasm"){
                Dialog.create("Then I need the channel it is in.");
                Dialog.addNumber("Cytoplasms channel", 3);
                Dialog.show();
                cytochannel = Dialog.getNumber();
        }

        Dialog.create("Do you have a nuclear membrane staining? Or do you
want to use DAPI?");
        Dialog.addChoice("Nucleus from:", newArray("DAPI", "nuclear
membrane"));
        Dialog.show();
        nucleuschoice = Dialog.getChoice();

        if(nucleuschoice == "nuclear membrane"){
                Dialog.create("Then I need the channel it is in.");
                Dialog.addNumber("Nuclear membrane channel", 3);
                Dialog.show();
                nucmembchannel = Dialog.getNumber();
```

```
                Dialog.create("Input a number bigger than half the width of
the nuclear membrane in your staining. If it were 0.9 micron, 0.66 is fine
:)");
                Dialog.addNumber("", 0.66);
                Dialog.show();
                firstnucleus = Dialog.getNumber();

                Dialog.create("Input a number smaller than half the width
of the nuclear membrane in your staining. If it were 0.9 micron, 0.29 is
fine :)");
                Dialog.addNumber("", 0.29);
                Dialog.show();
                secondnucleus = Dialog.getNumber();
        }

        Dialog.create("What channel is the signal to measure in?");
        Dialog.addNumber("Signal channel:", 2);
        Dialog.show();
        signalchannel = Dialog.getNumber();

        Dialog.create("Finally, do you want your signal to be considered as
dots? Or just as brightness?");
        Dialog.addChoice("Signal as:", newArray("dots", "just signal"));
        Dialog.show();
        signalchoice = Dialog.getChoice();
} else {
        if(File.exists(getDirectory("macros")+"settings.txt") == 0){
                runMacro("ladsetup.ijm");
        } else {
                settings =
File.openAsString(getDirectory("macros")+"settings.txt");
                lines = split(settings,"\n");
                mainline = split(lines[0], ",");

                nucleichannel = mainline[0];
                signalchannel = mainline[1];
                signalchoice = mainline[2];
                cytochoice = mainline[3];
                nucleuschoice = mainline[4];
                if(nucleuschoice == "nuclear membrane"){
                        secondline = split(lines[1], ",");

                        nucmembchannel = secondline[0];
                        firstnucleus = secondline[1];
                        secondnucleus = secondline[2];
                }
        }
}
```

```
//If something is in the ROI manager, it is deleted

if(roiManager("Count") != 0){
        roiManager("Deselect");
        roiManager("Delete");
        roiManager("Show None");
        roiManager("Show All");
        roiManager("Show None");
}

//These are just useful things that will be used later

width = getWidth();
height = getHeight();
titleOrig = getTitle();
run("Duplicate...","duplicate");
titleDupl = getTitle();
run("Set Measurements...", "area mean redirect=[titleOrig] decimal=3");
run("Split Channels");
setForegroundColor(255, 255, 255);

////////////////////////////////////////////////////Let's start!
////////////////////////////////////////////////////////////////
////////////////////////////////////////////////Nuclear
outlines////////////////////////////////////////////////////////////////
if(nucleuschoice == "nuclear membrane"){
        selectWindow("C"+nucmembchannel+"-"+titleDupl);rename("1");
        run("Duplicate...", " ");rename("2");

        selectWindow("1");
        run("Gaussian Blur...", "sigma="+firstnucleus+" scaled");
        selectWindow("2");
        run("Gaussian Blur...", "sigma="+secondnucleus+" scaled");
        imageCalculator("Subtract create", "2", "1");
        rename("DoG");
        selectWindow("1");close();
        selectWindow("2");close();

        selectWindow("DoG");
        setAutoThreshold("Triangle dark");
        run("Convert to Mask");
        run("Skeletonize");
/*
 * Skeletonize takes the middle line of the membrane selection. This was
done because I thought that, since antibodies should
 * bind to the outside of the nucleus, and the fluorophores would stick
even more outside, this could counteract it bringing
```

```
 * the border a bit more inside. However, if you would prefer to use the
outline rather than the middle line, comment out the
 * above line with // in front and comment out the following code taking
out this symbols /*.
 */
        /*
        setForegroundColor(0, 0, 0);
        run("Analyze Particles...", "add");
        AOOOO = roiManager("Count");
        for(i=0; i<AOOOO; i++){
                roiManager("Select", i);
                roiManager("Fill");
        }

        run("Outline");
         */

        run("Analyze Particles...", "add");
        AOOOO = roiManager("Count");

        //The next part takes out open lines

        for(i=0; i<AOOOO; i++){
                roiManager("Select", i);
                roiManager("Measure");
                if(getResult("Mean", i) == 255){
                        roiManager("Select", i);
                        roiManager("Fill");
                }
        }

        run("Clear Results");
        roiManager("Deselect");
        roiManager("Delete");

        run("Analyze Particles...", "add");
        AOOOO = roiManager("Count");
        setForegroundColor(0, 0, 0);

//The next part is needed to make outlines with, like, two bubbles inside,
become just one big bubble

        for(i=0; i<AOOOO; i++){
                roiManager("Select", i);
                roiManager("Fill");
        }

        run("Outline");
```

```
        roiManager("Deselect");
        roiManager("Delete");

        run("Analyze Particles...", "add");
        AOOOO = roiManager("Count");
        setForegroundColor(255, 255, 255);

//The next part excludes features too small to be nuclei. Often small
circles with long appendages are created, and this
//helps in taking them out. The area can be altered though.

        for(i=0; i<AOOOO; i++){
                roiManager("Select", i);
                roiManager("Measure");
                if(getResult("Area", i)<2){
                        roiManager("Fill");
                }
        }

        run("Analyze Particles...", "add");
        AOOOO = roiManager("Count");

        for(i=0; i<AOOOO; i++){
                roiManager("Select", i);
                roiManager("Measure");
                if(getResult("Mean", i) == 255){
                        roiManager("Select", i);
                        roiManager("Fill");
                }
        }

        run("Clear Results");
        roiManager("Deselect");
        roiManager("Delete");
        roiManager("Show None");
        roiManager("Show All");
        roiManager("Show None");

} else if(nucleuschoice == "DAPI"){
        selectWindow("C"+nucleichannel+"-"+titleDupl);
        run("Duplicate...", " ");rename("DoG");
        run("Gaussian Blur...", "sigma=0.19 scaled");
        setAutoThreshold("Mean dark");
        run("Convert to Mask");
}
```

```
////////////////////////////////////////////////////////Cytoplasms/////////
////////////////////////////////////////////

/*
 * About the watershed: if you wish to run it (maybe the cells were too
cluttered and the nuclei overlap), I strongly
 * advise also adding a size-filtering step afterwards, to counteract the
possible over-segmentation due to watershed.
 * The specific size depends on the size of your average nucleus, of
course. This is the code, just add it after the watershed.

run("Analyze Particles...", "add");
AOOOO = roiManager("Count");
setForegroundColor(255, 255, 255);

for(i=0; i<AOOOO; i++){
        roiManager("Select", i);
        roiManager("Measure");
        if(getResult("Area", i)<2){
                roiManager("Fill");
        }
}
 */

if(cytochoice == "nuclei"){

        selectWindow("C"+nucleichannel+"-"+titleDupl);
        run("Duplicate...", " ");rename("cells");
        run("Gaussian Blur...", "sigma=0.38 scaled");
        setAutoThreshold("IsoData dark");
        run("Convert to Mask");
        //run("Watershed");
        run("Voronoi");
        setThreshold(0, 0);
        run("Convert to Mask");
        run("Analyze Particles...", "add");
        AOOOO = roiManager("Count");

} else if(cytochoice == "cytoplasm"){

        selectWindow("C"+nucleichannel+"-"+titleDupl);
        rename("lines");
        run("Gaussian Blur...", "sigma=0.38 scaled");
        setAutoThreshold("Otsu dark");
        run("Convert to Mask");
        run("Watershed");

        run("Analyze Particles...", "add");
```

```
AOOOO = roiManager("Count");
setForegroundColor(255, 255, 255);

for(i=0; i<AOOOO; i++){
        roiManager("Select", i);
        roiManager("Measure");
        if(getResult("Area", i)<2){
                roiManager("Fill");
        }
}

roiManager("Deselect");
roiManager("Delete");
roiManager("Show None");
roiManager("Show All");
roiManager("Show None");

run("Voronoi");
setThreshold(0, 0);
run("Convert to Mask");
run("Divide...", "value=255.000");

selectWindow("C2-"+titleDupl);
rename("cyto");
run("Gaussian Blur...", "sigma=0.19 scaled");
setAutoThreshold("MinError dark");
run("Convert to Mask");
run("Analyze Particles...", "size=0-20 add");

AOOOO = roiManager("Count");
setForegroundColor(255, 255, 255);

for(i=0; i<AOOOO; i++){
        roiManager("Select", i);
        roiManager("Fill");
}

run("Clear Results");
selectWindow("Results");run("Close");
roiManager("Deselect");
roiManager("Delete");
roiManager("Show None");
roiManager("Show All");
roiManager("Show None");

imageCalculator("Multiply create", "cyto","lines");
rename("cells");
```

```
        run("Analyze Particles...", "add");
        AOOOO = roiManager("Count");
}

////////////////////////////////////////////////////////Signal//////////////
////////////////////////////////////////////////

if(signalchoice == "dots"){

        selectWindow("C"+signalchannel+"-"+titleDupl);
        run("Duplicate...", " ");rename("dots");
        run("Median...", "radius=0.19 scaled");
        setAutoThreshold("Default dark");
        run("Convert to Mask");
        roiManager("Deselect");
        roiManager("Delete");

        print("/,outsidefraction,cell");
        run("Set Measurements...", "area redirect=[titleOrig] decimal=3");

        for(i=0; i<AOOOO; i++){//For each cell area
                selectWindow("DoG");
                run("Duplicate...", " ");rename("DoG2");
                selectWindow("dots");
                run("Duplicate...", " ");rename("dots2");
                selectWindow("cells");
                run("Analyze Particles...", "add");//Which are recognized
here
                selectWindow("DoG2");
                makeRectangle(0, 0, width, height);
                roiManager("Add");
                roiManager("Select", newArray(i,AOOOO));
                roiManager("XOR");
                roiManager("Add");
                roiManager("Select", AOOOO+1);
                roiManager("Fill");//We delete everything outside of it in
the outline channel
                selectWindow("dots2");
                setForegroundColor(255,255,255);
                roiManager("Select", AOOOO+1);
                roiManager("Fill");//And in the dots signal, leaving only
what's inside (all done on copies)
                roiManager("Deselect");
                roiManager("Delete");
                roiManager("Show None");
                roiManager("Show All");
                roiManager("Show None");
                selectWindow("DoG2");
```

```
                run("Analyze Particles...", "add");

                if(roiManager("Count")==1){//If there is one and one only
nuclear outline in that area, we proceed
                        roiManager("Select", 0);
                        roiManager("Measure");
                        nucleus = getResult("Area", 0);//We measure the
nuclear area
                        run("Clear Results");
                        roiManager("Deselect");
                        roiManager("Show None");
                        roiManager("Show All");
                        roiManager("Show None");
                        selectWindow("dots2");
                        run("Analyze Particles...", "add");//And then we
recognize each dot
                        AOOOO2 = roiManager("Count");

                        if(AOOOO2>1){//If there is at least one dot
                                for(j=1; j<AOOOO2; j++){//For each one
                                roiManager("Select", j);
                                roiManager("Measure");//We measure it
                                roiManager("Select", newArray(0,j));
                                roiManager("Combine");//And then we combine
it with the nucleus, and measure this area too
                                roiManager("Add");//The difference is the
area of the dot that's outside
                                roiManager("Select", AOOOO2);
                                roiManager("Measure");
                                y = getResult("Area",0);
                                x = ((getResult("Area",1)-nucleus)/y);
                                print(y+","+x+","+i);//Then we print dot
area, outside fraction and current cell in a CSV line
                                roiManager("Select", AOOOO2);
                                roiManager("Delete");
                                run("Clear Results");
                                }
                        }
                        roiManager("Deselect");
                        roiManager("Delete");
                        selectWindow("DoG2");close();
                        selectWindow("dots2");close();
                } else if(roiManager("Count")>1){//Nothing happens if
there's more than one path in the outline channel
                        roiManager("Deselect");
                        roiManager("Delete");
                        run("Clear Results");
                        selectWindow("DoG2");close();
```

```
                        selectWindow("dots2");close();
                } else {//Nothing happens if there's nothing in the outline
channel either
                        run("Clear Results");
                        selectWindow("DoG2");close();
                        selectWindow("dots2");close();
                }
        }

        selectWindow("Results");run("Close");
        selectWindow("dots");close();
        selectWindow("cells");close();
        selectWindow("DoG");
        run("Analyze Particles...", "add");
        close();
        selectWindow(titleOrig);
        roiManager("Show All");
        setBatchMode(false);

//In the end, save the log as a CSV file  by right clicking on it and
you're done. You can see the original with the nuclear
//membranes superimposed to check if it's alright.

} else if (signalchoice == "just signal"){

        selectWindow("C"+signalchannel+"-"+titleDupl);
        run("Duplicate...", " ");rename("dots");

        print("integrnuclearsignal,integrtotalsignal,nuclеararea,cellarea,c
ell");
        run("Set Measurements...", "area mean integrated
redirect=[titleOrig] decimal=3");
        run("8-bit");
        for (i=0; i<50; i++) {
                run("Replace value", "pattern=i replacement=0");
                }
        for(i=0; i<AOOOO; i++){//For each cell area
                selectWindow("cells");
                run("Analyze Particles...", "add");//That is recognized
here
                selectWindow("DoG");
                roiManager("Select", i);
                run("Analyze Particles...", "add");

                if(roiManager("Count")==AOOOO + 1){//If there is one and
only one valid nuclear envelope inside
                        selectWindow("dots");
```

```
                        roiManager("Select", A0000);//This is the single
nucleus
                        roiManager("Measure");
                        nucleus = getResult("Area", 0);
                        integrnuclearsignal = getResult("IntDen", 0);
                        run("Clear Results");
                        roiManager("Select", i);//This is the cell
                        roiManager("Measure");
                        cell = getResult("Area", 0);
                        integrtotalsignal = getResult("IntDen", 0);
                        run("Clear Results");

// In this case, the output is the integrated signal inside the nucleus,
the integrated signal in the whole cell (not in the
// cytoplasm only!!), the area of the nucleus that had signal in it (not
the whole nuclear area!!) and the cellular area that
// had signal in it, plus an identification for the cell.


        print(integrnuclearsignal+","+integrtotalsignal+","+nucleus+","+cel
l+","+i);

                        roiManager("Deselect");
                        roiManager("Delete");
                        roiManager("Show None");
                        roiManager("Show All");
                        roiManager("Show None");
                        run("Clear Results");
                } else if(roiManager("Count")>1){//Again, nothing happens
if there's more than one nucleus
                        roiManager("Deselect");
                        roiManager("Delete");
                        run("Clear Results");
                } else {//Or if there is none
                        run("Clear Results");
                }
        }

        selectWindow("Results");run("Close");
        selectWindow("dots");close();
        selectWindow("cells");close();
        selectWindow("DoG");
        run("Analyze Particles...", "add");
        close();
        selectWindow(titleOrig);
        roiManager("Show All");
        setBatchMode(false);
}
```

```
//In the end, save the log as a CSV file by right clicking on it and you're
done. You can see the original with the nuclear
//membranes superimposed to check if it's alright.
```

Code 8.10: R script used to analyse and display FISH signal

```
setwd("~/Dropbox
(hannonlab)/PhD_fabry/Projects/Nxf2/FISH/emf56_2/analysis/gypsy_genomic/nxf
2")

files = list.files(pattern="*.csv")
nxf2 = do.call(rbind, lapply(files, function(x) read.csv(x,
stringsAsFactors = FALSE)))
nxf2[,5] <- c("nxf2")

setwd("~/Dropbox
(hannonlab)/PhD_fabry/Projects/Nxf2/FISH/emf56_2/analysis/gypsy_genomic/pan
x")

files = list.files(pattern="*.csv")
panx = do.call(rbind, lapply(files, function(x) read.csv(x,
stringsAsFactors = FALSE)))
panx[,5] <- c("panx")

setwd("~/Dropbox
(hannonlab)/PhD_fabry/Projects/Nxf2/FISH/emf56_2/analysis/gypsy_genomic/piw
i")

files = list.files(pattern="*.csv")
piwi = do.call(rbind, lapply(files, function(x) read.csv(x,
stringsAsFactors = FALSE)))
piwi[,5] <- c("piwi")

setwd("~/Dropbox
(hannonlab)/PhD_fabry/Projects/Nxf2/FISH/emf56_2/analysis/gypsy_genomic/wde
")

files = list.files(pattern="*.csv")
wde = do.call(rbind, lapply(files, function(x) read.csv(x, stringsAsFactors
= FALSE)))
wde[,5] <- c("wde")
```

```
setwd("~/Dropbox
(hannonlab)/PhD_fabry/Projects/Nxf2/FISH/emf56_2/analysis/gypsy_genomic/gfp
")

files = list.files(pattern="*.csv")
gfp = do.call(rbind, lapply(files, function(x) read.csv(x, stringsAsFactors
= FALSE)))
gfp[,5] <- c("gfp")

setwd("~/Dropbox
(hannonlab)/PhD_fabry/Projects/Nxf2/FISH/emf56_2/analysis/gypsy_genomic/mae
l")

files = list.files(pattern="*.csv")
mael = do.call(rbind, lapply(files, function(x) read.csv(x,
stringsAsFactors = FALSE)))
mael[,5] <- c("mael")

setwd("~/Dropbox
(hannonlab)/PhD_fabry/Projects/Nxf2/FISH/emf56_2/analysis/gypsy_genomic/zuc
")

files = list.files(pattern="*.csv")
zuc = do.call(rbind, lapply(files, function(x) read.csv(x, stringsAsFactors
= FALSE)))
zuc[,5] <- c("zuc")

nxf2_fil <- nxf2[(nxf2$integrtotalsignal > 300) & nxf2$nuclearearea > 10 &
nxf2$cellarea > 50,]
piwi_fil <- piwi[(piwi$integrtotalsignal > 300) & piwi$nuclearearea > 10 &
piwi$cellarea > 50,]
panx_fil <- panx[(panx$integrtotalsignal > 300) & panx$nuclearearea > 10 &
panx$cellarea > 50,]
wde_fil <- wde[(wde$integrtotalsignal > 300) & wde$nuclearearea > 10 &
wde$cellarea > 50,]
mael_fil <- mael[(mael$integrtotalsignal > 300) & mael$nuclearearea > 10 &
mael$cellarea > 50,]
zuc_fil <- zuc[(zuc$integrtotalsignal > 300) & zuc$nuclearearea > 10 &
zuc$cellarea > 50,]
gfp_fil <- gfp[(gfp$integrtotalsignal > 300) & gfp$nuclearearea > 10 &
gfp$cellarea > 50,]

median(nxf2_fil[,1]/(nxf2_fil[,2]-nxf2_fil[,1]))
median(piwi_fil[,1]/(piwi_fil[,2]-piwi_fil[,1]))
median(panx_fil[,1]/(panx_fil[,2]-panx_fil[,1]))
median(zuc_fil[,1]/(zuc_fil[,2]-zuc_fil[,1]))
median(mael_fil[,1]/(mael_fil[,2]-mael_fil[,1]))
median(wde_fil[,1]/(wde_fil[,2]-wde_fil[,1]))
```

```
median(gfp_fil[,1]/(gfp_fil[,2]-gfp_fil[,1]))



nxf2_fil[,1]/(nxf2_fil[,2]-nxf2_fil[,1])

mean(nxf2_fil[,1])
mean(nxf2_fil[,2]-nxf2_fil[,1])

mean(piwi_fil[,1])
mean(piwi_fil[,2]-piwi_fil[,1])

mean(panx_fil[,1])
mean(panx_fil[,2]-panx_fil[,1])

mean(wde_fil[,1])
mean(wde_fil[,2]-wde_fil[,1])

mean(mael_fil[,1])
mean(mael_fil[,2]-mael_fil[,1])

mean(gfp_fil[,1])
mean(gfp_fil[,2]-gfp_fil[,1])

mean(zuc_fil[,1])
mean(zuc_fil[,2]-zuc_fil[,1])

print("Hello")

median(nxf2_fil[,1])
median(nxf2_fil[,2]-nxf2_fil[,1])

median(piwi_fil[,1])
median(piwi_fil[,2]-piwi_fil[,1])

median(panx_fil[,1])
median(panx_fil[,2]-panx_fil[,1])

median(wde_fil[,1])
median(wde_fil[,2]-wde_fil[,1])

median(mael_fil[,1])
median(mael_fil[,2]-mael_fil[,1])

median(gfp_fil[,1])
median(gfp_fil[,2]-gfp_fil[,1])

median(zuc_fil[,1])
median(zuc_fil[,2]-zuc_fil[,1])
```

```
test <- rbind(nxf2_fil,panx_fil,mael_fil,gfp_fil,wde_fil,zuc_fil,piwi_fil)


ggplot(test) +
  geom_jitter(position=position_jitter(0.3), colour="blue",aes(x=cell,
y=test$integrnuclearsignal))+
  geom_jitter(position=position_jitter(0.3), colour="red",aes(x=cell, y=-
(integrtotalsignal-integrnuclearsignal)))+
  ylim(-5000, 5000) +
  theme_bw()

ggplot(test) +
  geom_boxplot(colour="blue",aes(x=cell, y=integrnuclearsignal))+
  geom_boxplot(colour="red",aes(x=cell, y=-(integrtotalsignal-
integrnuclearsignal)))+
  ylim(-5000, 5000) +
  theme_bw()

ggplot(test) +
  geom_boxplot(colour="blue",aes(x=cell, y=integrtotalsignal))+
  ylim(-5000, 5000) +
  theme_bw()
```