Dissection of piRNA precursor biogenesis
and
the epigenetic mechanisms of the piRNA pathway

DISSERTATION ZUR ERLANGUNG DES DOKTORGRADES
DER NATURWISSENSCHAFTEN (DR. RER. NAT.)
DER FAKULTÄT FÜR BIOLOGIE UND VORKLINISCHE MEDIZIN
DER UNIVERSITÄT REGENSBURG
vorgelegt von

EVELYN STUWE

aus

SCHWETZINGEN

im Jahr 2015

Das Promotionsgesuch wurde eingereicht am:
13.02.2015

Die Arbeit wurde angeleitet von:
PROF. GUNTER MEISTER

Das Promotionskolloquium fand statt am:
24.07.2015

Zusammensetzung des Prüfungsausschusses:
Vorsitz: PD Achim Griesenbeck
1. Prüfer: Prof. Gunter Meister
2. Prüfer: Prof. Gernot Längst
3. Prüfer: Prof. Klaus Grasser
Vertretung: PD Attila Nemeth

Unterschrift:

EVELYN STUWE

Contents

1 Summary **1**

2 Introduction **3**
2.1 Small RNAs in RNA silencing . . . 3
2.2 The discovery of small RNA silencing . . . 4
2.3 Argonaute proteins . . . 6
2.4 miRNAs . . . 8
2.4.1 miRNA biogenesis . . . 9
2.4.2 miRNA function . . . 13
2.4.2.1 Post-transcriptional gene silencing in the cytoplasm . . . 13
2.5 siRNAs . . . 14
2.5.1 siRNA biogenesis . . . 14
2.5.2 siRNA function . . . 14
2.5.2.1 siRNAs in plants . . . 15
2.5.2.2 Transcriptional gene silencing in the nucleus . . . 15
2.6 piRNAs . . . 17
2.6.1 piRNA biogenesis . . . 18
2.6.1.1 piRNA cluster architecture and transcription . . . 18
2.6.1.2 Biogenesis of primary piRNAs . . . 20
2.6.2 Nuclear and cytoplasmic function of the piRNA pathway . . . 24
2.6.2.1 The role of piRNAs in regulating chromatin structure . . . 24
2.6.2.2 Ping-pong: the cytoplasmic function of the piRNA pathway 28
2.7 piRNAs are trans-generational epigenetic factors . . . 29
2.7.1 Introduction into epigenetic mechanisms . . . 29
2.7.2 Trans-generational effects of piRNAs in *Drosophila* . . . 31
2.7.3 Hybrid dysgenesis - an old genetic phenomenon caused by piRNAs 33

3 Results **39**
3.1 Characterization of piRNA precursors . . . 39
3.1.1 Objective . . . 39
3.1.2 Introduction into piRNA precursors . . . 39

3.1.3 Cloning and sequencing of 5' capped RNAs . . . 41
3.1.4 Single stranded piRNA cluster transcripts are capped and transcribed from one defined start site . . . 42
3.1.5 Capped piRNA precursors from double stranded clusters originate from internal transcription initiation . . . 46
3.1.6 piRNA precursors are 5' mono-phosphorylated . . . 53
3.1.7 Approach for piRNA precursor detection . . . 53
3.1.8 piRNA precursors are 5' mono-phosphorylated long RNAs . . . 54
3.1.9 Identification of primary piRNA biogenesis factors involved in first cleavage events . . . 57
3.1.10 The investigated factors affect precursor levels, but not the nucleotide bias . . . 61
3.1.11 Efforts towards the elucidation of piRNA precursors to be co-transcriptionally processed . . . 64
3.2 Transgenerational inheritance of piRNAs converts piRNA targets to piRNA sources . . . 68
3.2.1 Objective . . . 68
3.2.2 Conversion of an inactive to an active piRNA cluster is accompanied by acquisition of the H3K9me3 mark . . . 68
3.2.3 A paternally transmitted piRNA cluster does not recover piRNA production . . . 72
3.2.4 The multiple effects of maternally inherited piRNAs . . . 76
3.2.4.1 Maternally inherited piRNAs are necessary for piRNA production in the next generation . . . 76
3.2.4.2 Inherited piRNAs induce primary piRNA biogenesis at target loci . . . 80
3.2.4.3 piRNA targets are enriched in H3K9me3, Rhino and Cutoff 82
3.2.4.4 Read-through transcription on piRNA targets is increased 84
3.3 Transcriptional regulation of piRNA clusters . . . 86
3.3.1 Objective . . . 86
3.3.1.1 Rhino and Cutoff promote piRNA cluster transcription . 87
3.3.1.2 Cutoff is an anti-terminator and protects 5' monophosphate ends . . . 92

4 Discussion **97**
4.1 Tracing piRNA precursors . . . 97
4.1.1 Uni-stranded cluster transcripts are mRNA doppelganger . . . 98
4.1.2 Double stranded clusters do not have defined transcription start sites 99
4.1.3 piRNA precursors are processed in intermediate steps . . . 100
4.1.4 Zucchini's role in piRNA precursor biogenesis . . . 101
4.2 Trans-generational effect of piRNAs on source and target loci . . . 103
4.2.1 maternally inherited piRNAs ensure piRNA biogenesis in progeny 103
4.2.2 Maternally inherited piRNAs induce ping-pong amplification . . . 104
4.2.3 piRNA clusters rely on inherited piRNAs to stay active . . . 104
4.2.4 Maternally inherited piRNAs act as an epigenetic switch . . . 106
4.2.5 H3K9me3 attracts piRNA biogenesis factors . . . 107

4.2.6 Cutoff acts as an anti-terminator . . . 107
4.2.7 Cutoff protects free 5' monoposhate ends . . . 108

5 Conclusions **111**

6 Methods and Materials **113**
6.1 Standard prcedures . . . 113
6.1.1 RNA isolation . . . 113
6.1.1.1 Classic Phenol/Chloroform Extraction . . . 113
6.1.1.2 RNA extraction using Ribozol Reagent . . . 113
6.1.2 Quantification of nucleic acids . . . 114
6.1.2.1 Spectrophotometric DNA and RNA quantification . . . 114
6.1.2.2 Fluorometric DNA and RNA quantification . . . 114
6.1.3 PCR . . . 115
6.1.4 DNase treatement . . . 115
6.1.5 Reverse transcription . . . 116
6.1.6 Quantitative PCR (qPCR) . . . 116
6.1.7 Strand specific RT and qPCR for detection of read through transcription . . . 117
6.1.8 DEPC treatment of water . . . 117
6.1.9 T7 RNA transcirption . . . 117
6.2 Specific Experimental Methods . . . 119
6.2.1 Chromatin Immunoprecipitation - ChIP . . . 119
6.2.2 Nuclear Run-On / GRO-Seq . . . 119
6.2.2.1 Preparations . . . 119
6.2.2.2 Nuclei isolation . . . 120
6.2.2.3 nuclear run on reaction . . . 120
6.2.2.4 Immunoprecipitation of run-on RNA . . . 120
6.2.3 5' capped RNA cloning . . . 121
6.2.3.1 Removal of RNAs shorter than 200 nt . . . 121
6.2.3.2 CIP treatment . . . 122
6.2.3.3 TAP treatment . . . 122
6.2.3.4 5' adapter ligation . . . 122
6.2.3.5 Introduction of the 3' adapter . . . 123
6.2.3.6 Linear PCR . . . 123
6.2.3.7 Final library amplification . . . 124
6.2.4 5' mono-phosphate RNA cloning . . . 125
6.2.5 Nascent transcript isolation . . . 125
6.2.5.1 Nuclei isolation . . . 125
6.2.5.2 Chromatin precipitation and nascent RNA isolation . . . 125
6.2.6 Oligo(dT) beads assay to determine polyA containing RNAs . . . 126
6.2.7 Bioinformatic analysis of RNA sequencing data . . . 126
6.2.7.1 RNA mapping . . . 126
6.3 Methods related to work with *Drosophila* . . . 127
6.3.1 LacZ staining in *Drosophila* ovaries . . . 127
6.3.2 Fly crosses and stocks . . . 127

6.4 Materials 133

7 Acknowledgement **141**

8 Appendix **145**
8.1 Publications related to this thesis 145
8.2 Curriculum Vitae 146

List of Figures

2.1 Schematic view of an Argonaute protein . . . 7
2.2 miRNA and siRNA biogenesis pathways . . . 10
2.3 piRNA clusters in flies and mouse . . . 19
2.4 piRNA biogenesis pathways . . . 22
2.5 Schematic of Hybrid Dysgenesis . . . 32
2.6 Transgeneational effect of piRNAs . . . 34
2.7 Multi-generation experiment by Bucheton . . . 36

3.1 Scheme of cloning capped RNAs . . . 43
3.2 5'Cap RNA annotations . . . 44
3.3 5'Cap RNA genic signal . . . 45
3.4 RPKMS for 5' capped RNA libraries . . . 46
3.5 UCSC tracks for flamenco . . . 47
3.6 UCSC tracks for cluster 20A . . . 48
3.7 UCSC tracks for cluster 42AB . . . 49
3.8 UCSC tracks for cluster 80EF . . . 50
3.9 UCSC tracks for clusters 38C.1 and 38C.2 . . . 51
3.10 UCSC tracks for cluster flamenco . . . 52
3.11 Distribution of reads on 5'P and 5'OH spikes . . . 55
3.12 Gene ontology analysis for 5'P libraries . . . 56
3.13 RPKMS for 5'P RNA libraries . . . 57
3.14 Nucelotide distribution in piRNA precursors from wild type *Drosophia* . . . 58
3.15 qPCRs in sh knockdowns . . . 60
3.16 total RNA sequencing in sh knockdown lines . . . 62
3.17 piRNA precursor abundance in sh knockdown lines . . . 63
3.18 5'nucleotide distribution for piRNA precursors . . . 64
3.19 scheme of assays to assess purity of nascent transcripts . . . 66
3.20 PAT-Assays . . . 67
3.21 T1-BX2 scheme . . . 70
3.22 T1-BX2 H3K9me3 ChIP . . . 71
3.23 T1-BX2 nuclear run-on . . . 71

3.24 Schematic overview of crosses to follow a paternally transmitted piRNA producing locus . . . 73
3.25 T1-BX2 LacZ silencing . . . 75
3.26 H3K9me3 levels on the LacZ transgene after paternal transmission and in subsequent generations . . . 76
3.27 Schematic of P1152/BC69 fly system . . . 78
3.28 P1152 and BC69 piRNA production . . . 79
3.29 BC69 unique region spanning piRNAs . . . 81
3.30 P1152 targeted endogenous loci . . . 83
3.31 H3K9me3, Rhino and Cutoff ChIPs upon maternal deposition . . . 84
3.32 Read through transcription upon maternal deposition . . . 85
3.33 Schematic overview over nuclear run-on procedure . . . 88
3.34 GroSeq signal in shRhi, shCuff and control . . . 89
3.35 GroSeq signal in piRNA clusters . . . 90
3.36 GroSeq density in double-stranded clusters . . . 91
3.37 GroSeq density in single-stranded clusters . . . 91
3.38 Cutoff structure model . . . 93
3.39 piRNA precursor levels in shCuff, shRat1 and shWhite . . . 95
3.40 Cuff acts as anti-cleavage factor . . . 96

4.1 Model of uni-stranded and double-stranded cluster transcription . . . 100
4.2 Model for effect of Zucchini knockdown . . . 102
4.3 Model for effects of maternally inherited piRNAs . . . 110

6.1 Strand specific RT for read-throug-PCRs . . . 118

List of Tables

3.1 piRNA cluster coordinates 40
3.2 Read numbers and mapping statistics of 5'Cap RNA libraries 42
3.3 Sequencing statistics of two 5'P RNA replicates 56
3.4 Gro-Seq sequencing statistics 88

6.1 *Drosophila* strains 129
6.2 List of buffers 134
6.3 List of antibodies 137
6.4 List of enzymes 137
6.5 List of equipment 137
6.6 List of reagents 138
6.7 List of reagents 139
6.8 List of kits 139

CHAPTER *1*

SUMMARY

RNA interference (RNAi) is a cellular mechanism where small RNAs (20 to 35 nucleotides in length) pair with proteins to build effector complexes guiding them to RNA targets through homologous base pairing.

The piRNA pathway is a germ line specific RNAi pathway that protects the gonads of animals from harmful genetic sequences, termed transposable elements (TEs). TEs are viral DNA remnants that can autonomously mobilize within genomes and cause damage through insertion into new genomic loci and double-strand breaks in DNA. TE activity in germline cells ultimately leads to sterility and its suppression through the piRNA pathway is therefore essential for propagation. While the piRNA pathway shares common characteristics with the two major RNAi pathways (miRNAs and siRNAs), it additionally exhibits several unique features, particularly during biogenesis. piRNAs originate from large non-genic regions, so called piRNA clusters, or from mRNAs of transposable elements. Neither of these RNA transcripts contain the requisite structural features capable of triggering small RNA production, seen in both siRNAs and miRNAs. Consequently, piRNAs are produced in a Dicer independent manner. Therefore, it is not yet understood, how target transcript are recognized by the piRNA pathway.

Furthermore, the extent of piRNA functionality *in vivo* are not well defined, as they seem to act on multiple levels to protect the germline genome from the propagation of transposable elements. piRNAs lead to direct cleavage of TE mRNAs, but also seem to change the chromatin landscape of regions expressing RNA targets. Additionally, the effect of piRNAs reaches into the next generation through a transgenerational epigenetic mechanism.

This study focuses on the characterization of the early stages in piRNA biogenesis, starting at their transcription through the first processing steps. The data presented suggest that the two different piRNA cluster types, uni-stranded and double-stranded clusters, differ greatly in the properties of their transcription. Uni-stranded clusters exhibit a defined transcription start site, giving rise to 5' m^7G-capped piRNA precursor molecules that resemble canonical mRNAs. Whereas, double-stranded clusters do not show a distinct start site and can be transcribed through internal initiation or read through transcription from neighboring genes. The work furthermore reveals the existence of piRNA intermediates: they are over 200 nucleotides long, which contain a 5' monophosphorylated terminus and already exhibit the nucleotide biases similar to mature piRNAs.

After processing of piRNA precursors into mature piRNAs, their effects can reach into subsequent generations. To investigate the trans-generational aspect of piRNAs, I utilized transgenic fly models. The results reveal that piRNAs are epigenetic factors influencing a plethora of mechanisms in the piRNA pathway: ensuring a strong piRNA production in the next generation through maintenance of both piRNA production pathways (primary and secondary piRNA production). Furthermore, they maintain the characteristic structural chromatin features of piRNA clusters in the next generation. However, most strikingly piRNAs are capable of transforming regions expressing targeted transcripts into loci producing primary piRNAs. Therefore, the latter result informs a revision to the current definition of piRNA sources and their subsequent targets. piRNA sources and targets might not be strictly distinct, but any piRNA target acquires piRNA cluster signatures and starts producing piRNAs.

The discovery that inherited piRNAs transform a piRNA target into a source, allowed the investigation of the molecular changes at loci targeted by the piRNA pathway. These newly established piRNA clusters acquire a specific chromatin configuration consisting of the histone mark H3K9me3 and two piRNA protein factors Rhino and Cutoff. Together, these factors correlate with increased transcription of their targeted loci. Cutoff appears to be the key factor in piRNA production, through both promotion of read-through transcription beyond stop signals and protection of RNA termini against exonucleolytic degradation.

Overall, the presented data contribute to an understanding of the molecular mechanisms in piRNA biogenesis, from the initial processing steps through the trans-generational effect of piRNAs. It suggests that the piRNA system is organized as a giant feed-forward loop, where piRNAs themselves reach into the next generation to ensures reliable propagation of the system and effective silencing of deleterious TE sequences.

CHAPTER 2

INTRODUCTION

2.1 Small RNAs in RNA silencing

Since the discovery of the first small RNAs in the early 1990 (Lee et al., 1993) and the later realization that those small RNAs play a pivotal role in gene expression and genome stability, research on small RNA silencing has been an ever expanding universe (reviewed by (Ghildiyal and Zamore, 2009)). It has been less then 20 years ago, when Fire and Mello discovered that double-stranded RNA molecules can induce gene silencing. This phenomenon was termed RNA interference (RNAi) and the elucidation of its mechanisms, as well as its implementation as a molecular tool in everyday research, have changed the understanding of gene regulation. Small RNA silencing mechanisms can be found in plants, eucaryotes and even some procaryotes and they take part in a variety of cellular processes. They regulate and shape gene expression, cell growth and differentiation, they maintain genome stability and defend cells against invading viruses and mobile genetic elements. The mechanisms of their actions influence chromatin structure, chromosome segregation, transcription, RNA processing, RNA stability and translation (Ghildiyal and Zamore, 2009; Castel and Martienssen, 2013; Meister, 2013). At the core of this pathway a small RNA of 20-30 nucleotide length pairs with a protein of the Argonaute family to form the effector of RNA silencing: the RNA-induced silencing complex (RISC). In this complex the small RNA provides sequence specificity through complementary base pairing with a specific target, whereas the Argonaute exerts the function. Small RNAs are divided into several classes, depending on their origin, length, mode of targeting and type of Argonaute protein they associate with. Three major classes of small RNAs were

defined to date: siRNAs, miRNAs and piRNAs. They all share common features of RNA silencing mechanisms, but every class has distinct characteristics.

2.2 The discovery of small RNA silencing

Our understanding of small RNA silencing emerged from the efforts of many different research groups in seemingly unrelated fields. Several phenomena of gene expression regulation through RNA had been described in plants and animals, before some hallmark discoveries linked those observations to our current understanding of small RNA silencing mechanisms.

One of the first observations of small RNA silencing was made in plants. Napoli and coworkers (Napoli et al., 1990) described, how introduction of an extra copy of the chalcone synthase transgene to enhance pigmentation of petunia flowers led to a surprising effect. Instead of enhancing the color of the flowers through increased production of the coloring pigment, the flowers showed variegated colouring, or were white altogether, suggesting the silencing of the endogenous locus. In the same year, van der Krol and coworkers (van der Krol et al., 1990), also investigating the chalcone synthase pathway, introduced an antisense chalcone synthase gene into petunia and observed silencing of the endogenous gene. Even though RNA interference with an antisense RNA had been used before to manipulate gene expression in cells (Izant and Weintraub, 1984), the common assumption was, that an antisense RNA molecule directly inhibits targeted mRNAs through hybridization. Through dissection of the requirements for the silencing effect, though, van der Krol and colleagues concluded that the mechanism of antisense inhibition reaches beyond RNA duplex formation of sense and antisense RNA.

Several other labs observed that plants, infected with a replicating RNA virus, were able to suppress symptoms of a viral infection and at the same time exhibited silencing of endogenous genes with sequence homology to viral RNA(Dougherty et al., 1994; Angell and Baulcombe, 1997; Ruiz et al., 1998). An observation similar to the transgene silencing in petunia, made by Napoli and colleagues, was made in the filamentous fungus *Neurospora crassa*, where the introduction of an exogenous sequence led to the decreased expression of the corresponding edogenous gene (Romano and Macino, 1992). Those observations were termed 'quelling' in fungus, and post-transcriptional gene silencing (PTGS) in plants.

Only a few years after the first reports of transgene silencing in plants, animals and fungi, Andy Fire and Craig Mello conducted hallmark experiments in *Caenorhabditis elegans*, which would pave the way to understanding the principles underlying the observed phenomena (Fire et al., 1998). They reported that double-stranded RNA leads to inher-

itable gene silencing in *Caenorhabditis elegans* and that this silencing effect of double-stranded RNAs was orders of magnitudes more potent than purified single stranded RNA. It became apparent that something about the double-strandedness of the RNA leads to a "catalytic or amplification component in the interference process" (Fire et al., 1998). The phenomenon that a double stranded RNA was able to induce silencing of any homologous sequence was named RNA interference (RNAi).

Originally, the term RNAi was used to describe silencing through siRNAs, the class of small non-coding RNAs discovered by Fire and Mello. With our growing understanding about mechanistic details of different RNA silencing pathways, however, clear distinctions between different pathways became less apparent. They are partially overlapping and converging, and thus the term "RNAi" is now broadly used for silencing through small RNAs.

The discovery of RNAi changed and advanced our understanding of fundamental cellular principles and it became an invaluable tool in basic and applied research. It was rewarded with a Nobel Prize for Fire and Mello in 2006.

At the same time of the first phenomenological reports of transgene silencing in plants and fungi, the first small RNA, the miRNA lin-4, was discovered in 1993 by Victor Ambros and Gary Ruvkun (Lee et al., 1993). They found a transcript from the *lin-4* locus to be a negative regulator of the protein LIN-14. Cloning and investigation of this locus showed that it was not coding for a protein, but produces two small RNA species, a 61 nucleotide (nt) long *lin-4*L and a 22 nt long *lin-4*S molecule. The sequence of this *lin-4* transcript is highly conserved between different *Caenorhabditis elegans* species and has a 10 nt long block of complementarity to a repeated sequence element in the 3' untranslated region (UTR) of lin-14 mRNA. This publication presents hallmarks in miRNA mechanisms. It described the miRNA precursor *lin-4*L and suggested its stem-loop structure. Furthermore, it described the first miRNA *lin-4*S and proposed that it regulates its targets through complementary binding in its 3'UTR. The 10 nt block of complementarity is today known as 'the seed' of miRNAs.

The second essential component of RNA silencing are Argonaute proteins. They were first described in plants and *Drosophila melanogaster* (Lin and Spradling, 1997; Cox et al., 1998; Bohmert et al., 1998) and their name stems from the phenotype of AGO1 mutants in *Arabidopsis thaliana*. The unusual appearance of mutant plants reminded the researchers of a squid, and thus they named the gene after the pelagic octopus, *Agonauta argo* (Bohmert et al., 1998). Those first reports on Argonaute connected it to a function in stem cell differentiation and growth. Later, genetic and biochemical studies on RNAi established them as the second essential component of small RNA mediated silencing. They were found in all eucaryotes with an RNAi system, as well as in some archaea and

bacteria (Makarova et al., 2009).

Even though the first structural insights into the Argonaute architecture originated from procaryotic Argonautes, their procaryotic function still remains unclear. While RNAi is the crucial defense mechanism of many organisms against invasive genetic elements, procaryotes protect their genome through the analogous, but not homologous CRISPR-associated system. Only recently some reports suggest that the function of bacterial Argonautes might be a second line of defense against phages and plasmids, additionally to the CRISPR system (Carmell et al., 2002; Olovnikov et al., 2013; Swarts et al., 2014; Vogel, 2014).

2.3 Argonaute proteins

Proteins of the Argonaute family are found in plants, animals and fungi (Hutvagner and Simard, 2008; Meister, 2013). As the central effector proteins of RNA silencing they are diversified in certain aspects, but their core domains remain conserved throughout evolution. Argonautes can be divided into three subfamilies, the Ago, WAGO and the Piwi subfamily. Argonautes of the Ago clade are expressed ubiquitously and are involved in various aspects of RNA silencing pathways such as miRNA or siRNA mediated silencing; WAGOs are worm specific Argonautes. Piwi Argonautes are specific to animals and their expression is mainly restricted to germline cells. They were first described as factors for germline maintenance in the *Drosophila* germline (Lin and Spradling, 1997; Cox et al., 1998). Since the mutation of the germline specific Argonautes leads to small gonads due to derepression of transposable elements, they were dubbed "P-Element Induced Whimpy Testis". The name PIWI persisted for the germline specific Argonautes, and one of the main protein domains of all Argonautes is called PIWI.

Throughout evolution, Argonaute proteins underwent significant gene duplications and functional diversification. In yeast, for example, *Saccharomyces pombe* carries only one Argonaute protein, which covers all RNAi related functions from PTGS to heterochromatin silencing (Sigova et al., 2004). *Saccharomyces cervisiae* lost its Argonaute protein with the accompanied RNAi machinery altogether. To the other extreme, the *Caenorhabditis elegans* genome encodes 27 distinct Argonautes, where specific proteins conduct distinct aspects of RNA silencing (Buck and Blaxter, 2013). 18 of the 27 Argonautes in *Caenorhabditis elegans* are worm specific, and they are thus called WAGOs. In between those extremes, plants have ten Argonautes (Mallory and Vaucheret, 2010), humans have four Argonautes and four Piwis and *Drosophila melanogaster* has two Argonautes and three Piwis.

The basic common architecture of Argonaute proteins is their division into four do-

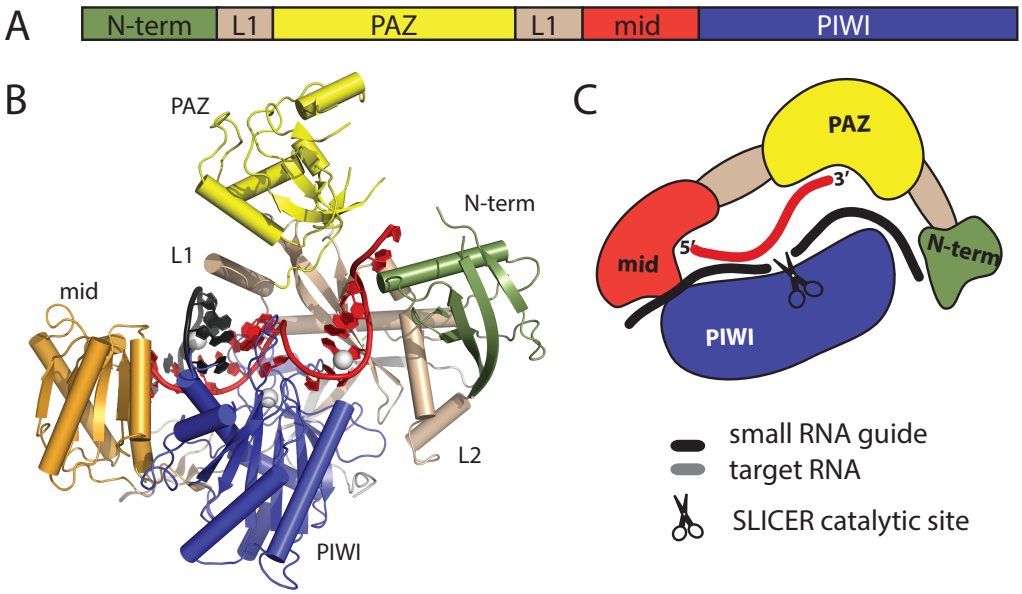

Figure 2.1: **Architecture of an Argonaute protein.** (A) Domains of Argonaute proteins. (B) Structure of the human Argonaute hAgo2 in complex with a micro RNA (PDB code 4F3T). (C) Schematic representation of an Argonaute protein in a cartoon. Domain colors are green for the N-terminal domain, yellow for the PAZ domain, red for the mid domain, blue for the Piwi domain and wheat for the two linker domains L1 and L2. The small RNA guide (red) is anchored in the PAZ domain with its 3' and and in the MID domain with its 5' end. In case of catalytically active Argonautes, slicing of the target (black) is facilitated by the Piwi domain.

mains: the N domain, the PAZ (Piwi-Argonaute-Zwille) domain, the MID (middle domain) and the PIWI (P-element-induced wimpy testes) domain. Crystal structures of procaryotic (archea and eubacteria), yeast and other eucaryotic Argonautes, provided insights into the overall architecture and molecular mechanisms of their functions (Song et al., 2004; Parker et al., 2005; Ma et al., 2005; Wang et al., 2009, 2008; Schirle and MacRae, 2012; Nakanishi et al., 2012; Elkayam et al., 2012)).

The PAZ domain harbors the anchor point for the 3' end of the small RNA. It provides a binding pocket, which specifically recognizes the single stranded 3' overhang of a double stranded RNA molecule. Those 3' overhangs are the characteristic product of upstream processing by the RNAseIII endonucleases Drosha and Dicer. The MID domain anchors the 5' end of the small RNA (Boland et al., 2010; Frank et al., 2010) and in some cases, as seen for the hAgo2 MID domain, it seems to have a preference for certain nucleotides, leading to a bias for 5' nucleotides of Argnaute bound small RNAs. The PIWI domain has an RNase-H-like fold. RNase-H endonucleases generally cleave RNAs guided by a DNA template. In some Argonautes, the endonucleolytic activity of the PIWI domain is conserved and those Argonautes are commonly referred to as "Slicers". Slicers use the

small RNA as a guide and cleave the targeted RNA, leaving a phosphate at the 5' end. Only a few Argonautes are cleavage-competent. In *Arabidopsis thaliana* only two (Ago1 and Ago4) out of the ten Argonautes are slicers (Baumberger and Baulcombe, 2005; Qi et al., 2005, 2006), in *Drosophila melanogaster* both Argonautes (Ago1 and Ago2), and all three Piwis are cleavage-competent. In human, only Ago2 is an active Slicer, while Ago1, 3 and 4 are not (Meister et al., 2004; Liu et al., 2004). Studies on those human Argonautes revealed the structural requirements for an Argonaute to be an active slicer. Sequence alignments and later structural studies revealed that the catalytic center of the hAgo2 PIWI domain contains a DEDH tetrade, which is mutated in the other, non-slicing Argonautes. This catalytic center, however, is not the only requirement to make an Aronaute an active slicer. The close investigation of the PIWI domain showed that besides the DEDH tetrade, there are other crucial sequence requirements for the PIWI domain to be active. Most likely, those regions are responsible for the proper orientation of the catalytic center. Domain swap, mutagenesis and DNA shuffling experiments revealed that also the N domain of an Argonaute needs to meet certain sequence requirements in order to make an Argonaute a Slicer (Hauptmann et al., 2013; Faehnle et al., 2013; Schurmann et al., 2013; Hauptmann et al., 2014).

2.4 miRNAs

Until now, miRNAs were found in the plant and animal branches of the tree of eukaryotic organisms. They are 21-25 nt long and originate from double-stranded RNA molecules. The precursor RNAs for miRNA production (pri-miRNAs) are usually transcribed by RNA Polymerase II (Ghildiyal and Zamore, 2009) and they originate from intergenic regions, 3' UTRs or introns. Pri-miRNAs contain a 5' Cap and a poly-adenylate tail (Kim, 2005). The defining feature for miRNA production is a stem-loop structure, where the stem is 33 nt long and consists of imperfectly paired bases (Bartel, 2004). One pri-miRNA transcript can contain only one, or a cluster of several distinct miRNAs. In animals, maturation of the pri-miRNA into a 20-25 nt miRNA is facilitated sequentially in the nucleus and in the cytoplasm by two RNAse III endonucleases with their double-stranded RNA binding domain (dsRBD) partner proteins. In plants, this two-step process occurs entirely in the nucleus, carried out by only one RNAse III endonuclease (Zhu, 2008).

2.4.1 miRNA biogenesis

In animals, the first step of miRNA processing takes place in the nucleus. Initially, the dsRBD DGCR8 (in mammals) or Pasha (in flies), recognizes the stem-loop structure and attracts its partner RNAse III endonuclease Drosha. In this complex, which is called Micorprocessor, DGCR/Pasha positions Drosha precisely at the stem of the stem-loop and Drosha catalyzes the excision of the RNA stem-loop (Gregory et al., 2006). The resulting product is called pre-miRNA and it has a two-nucleotide overhang at its 3' end and a 5' phosphate group. The 3' end of this product will not be further trimmed, and thus Drosha defines the 3' end of miRNAs.

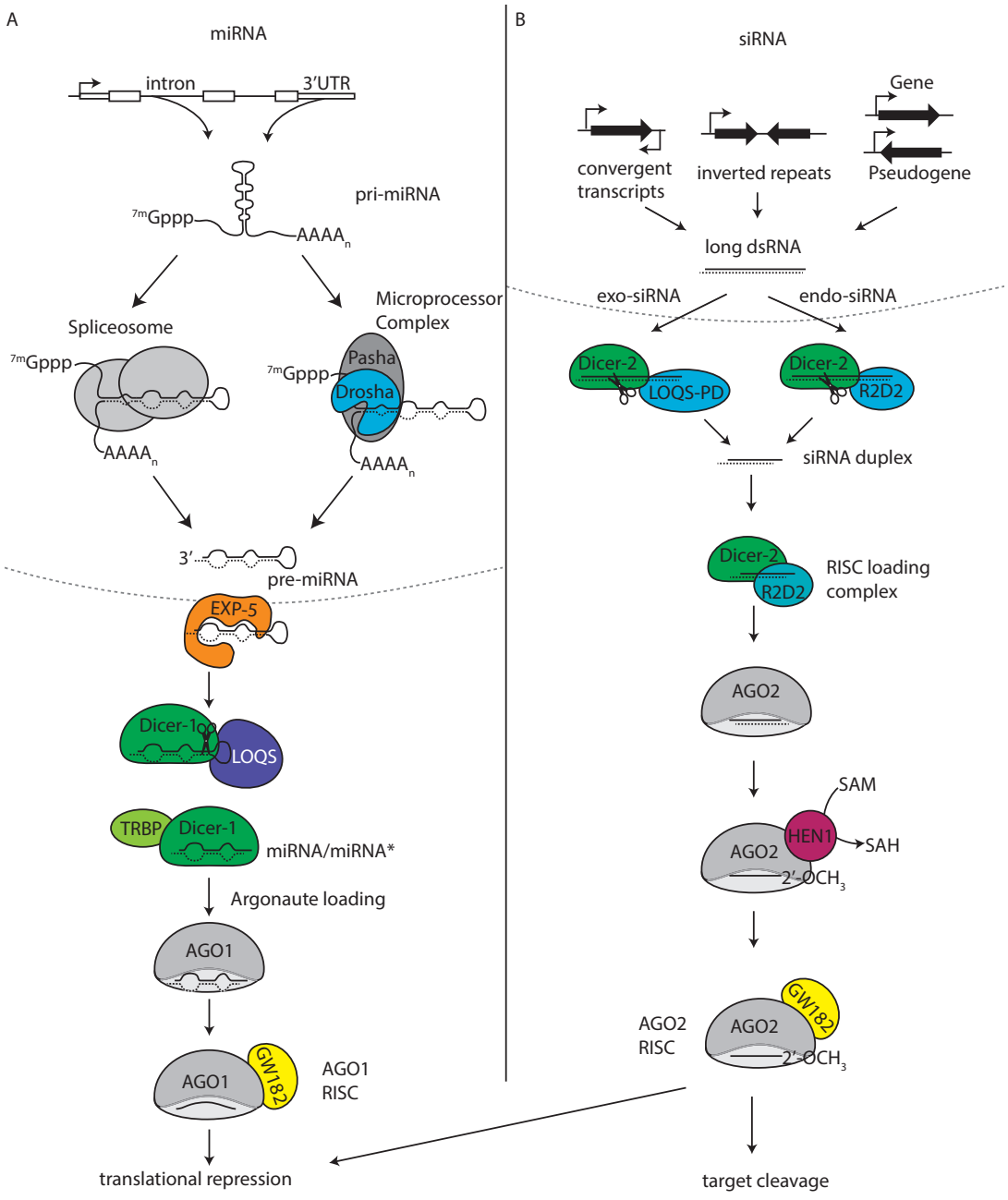

Figure 2.2:
miRNA and siRNA biogenesis pathways (partially adapted from (Ghildiyal and Zamore, 2009). *[Legend see next page]*

Figure 2.2: miRNA and siRNA biogenesis pathways. *[The figure is on the previous page]* Biogenesis pathways for miRNAs and siRNAs exemplified with *Drosophila melanogster* proteins. (A) miRNAs are transcribed from intergenic regions, 3' UTRs or introns into primary miRNAs (pri-miRNA). The stem-loop structure is recognized by Pasa and cleaved by Drosha. Together, they build the Microprocessor complex. An alternative route is the excission of the stem-loop through the pre-mRNA splicing pathway. The pre-miRNA is exported into the cytoplasm through Exportin-5 and loaded into Dicer-1, assisted by the double-stranded RNA binding protein LOQS. The miRNA/miRNA* duplex is then loaded into AGO1, aided by TRBP and possibly other co-factors. After passenger strand degradation and association with the cofactor GW182, the guide miRNA leads AGO1 RISC to its targets.
(B) siRNAs are derived from long, double-stranded precursor molecules, which get processed into siRNA dublexes by Dicer-2. To detect a dsRNA as a substrate, Dicer-2 relies on a double-stranded RNA binding protein. In *Drosophila*, those are Loqs-PD for exo-siRNAs and R2D2 for endo-siRNAs. Dicer together with the processed siRNA dublex and the respective double-strand RNA binding protein forms the RISC loading complex and transfer the dublex into AGO2. The passenger strand is degraded and the guide strand is methylated on its 3' terminus through HEN1 to increase stability. Association with GW182 forms the AGO2 RISC, which is guided to its targets by the siRNA.

In worms, flies and mammals, a Microprocessor independent route has been described for the nuclear processing of so-called mirtrons. In the case of mirtrons, the excision of the stem-loop is facilitated by the nuclear pre-mRNA splicing pathway (Okamura et al., 2007; Ruby et al., 2007; Berezikov et al., 2007; Babiarz et al., 2008). In animals, pre-miRNAs are exported into the cytoplasm through Exportin-5 (Yi et al., 2003; Bohnsack et al., 2004; Lund and Dahlberg, 2006). The next cut to define the 5' end of the mature miRNA is performed in the cytoplasm by Dicer, assisted by its dsRBD partner protein TRBP (in mammals) or Loquacious (LOQS) (in flies). Dicer binds the 3' end of the pre-miRNA with its PAZ-domain, which is a common feature of Dicer and Argonautes. Cleavage is facilitated through its RNAse III catalytic site. The catalytic site is positioned two helical turns (or 22 bp) away from the PAZ domain and consists of a cleft, formed by an intramolecular dimer involving the RNAse III domains. Each RNAse III domain cleaves one of the two RNA strands, which results in an RNA duplex with a 2 nt 3 overhang. It is thus the distance of Dicer's PAZ domain to its catalytic site, which makes it act as a molecular ruler for miRNAs. The two resulting strands are called miRNA and miRNA*.

However, to this general principle of miRNA biogenesis, there are some exceptions. In zebrafish and mouse, a specific miRNA, miR-451, has been described to be processed independently of Dicer by Ago2 (Cifuentes et al., 2010; Cheloufi et al., 2010). A class of miRNAs derived from small nucleolar RNAs, seem to be processed in a Dicer dependent, but DGCR8/Drosha independent way. (Ender et al., 2008)

In plants, the processing of pri-miRNAs into miRNA-miRNA* duplexes is not divided into a nuclear and a cytoplasmic phase, but takes place entirely in the nucleus. There, Dcl1 (Dicer-like protein), together with its dsRBD partner HYL1 and the C2H2 Zn-finger protein SE, performs the cleavages solo (Kurihara and Watanabe, 2004; Dong et al., 2008). Another distinction between miRNA biogenesis in plants and animals is the 2'-O-methylation of the 3' ends of plant miRNAs by HEN1. (Park et al., 2002; Yu et al., 2005; Yang et al., 2006). This modification is believed to enhance stability of miRNAs by preventing 3' uridinylation, which acts as a signal for degradation (Li et al., 2005).

One of the two strands from a miRNA- miRNA* duplex guides the Argonaute to its targets. This chosen strand is called "guide", whereas the other strand will ultimately be degraded and is called "passenger". Which of the strands is chosen to be the guide, depends on the relative thermodynamic stability of the duplex - in most cases the 5' end is chosen from the end with the less stable base-pairing (Khvorova et al., 2003; Schwarz et al., 2003).

The Argonaute together with its small RNA ultimately forms the RNA-induced Si-

lencing Complex (RISC). Loading of a miRNA from Dicer into an Argonaute depends on many accessory co-factors, which furthermore differ between organisms.

The current model is that a RISC loading complex consists of Dicer, TRBP and Argonaute. After degradation of the passenger strand, the loaded Argonaute together with GW182 forms the mature RISC. Functional studies show that GW182 is necessary and sufficient to induce Ago mediated gene silencing in humans, *Caenorhabditis elegans* and *Drosophila* (Liu et al., 2005; Ding and Han, 2007; Eulalio et al., 2008; Carthew and Sontheimer, 2009; Ding and Grosshans, 2009).

2.4.2 miRNA function

In the miRISC complex each function is clearly assigned to one component. The miRNA acts as a sequence specific guide, while the Argonaute protein provides the function. Gene silencing through small RNAs interferes with gene expression on different levels. A nuclear pathway silences on a transcriptional level (transcriptional gene silencing, TGS), whereas a cytoplasmic pathway targets mRNAs in the cytoplasm (post-transcriptional gene silencing PTGS). The cytoplasmic PTGS through miRNAs will be described subsequently, while principles of the nuclear TGS will be addressed in section 2.5.2.2.

2.4.2.1 Post-transcriptional gene silencing in the cytoplasm

Targeted mRNAs undergo three possible fates, depending on the type of Argonaute protein and the degree of complementarity to the miRNA. They can be either directly cleaved, channeled into a degradation pathway, or be translationally suppressed.

Cleavage of the target is generally induced by perfect or almost perfect complementarity between miRNA and target. High degree of complementarity is mainly observed in plant miRNAs, and this mode of action is restricted to RISC complexes containing a Slicer.

The two other processes - degradation and translation inhibition - are initiated through imperfect miRNA-target complementarity and can be performed by any Argonaute protein. They seem to be the prevalent mode of regulation in animals.

The details of non-cleavage mediated silencing is intensively studied and several, mutually not exclusive mechanisms are proposed (Gu and Kay, 2010). Detection of the miRNA target sites in the 3' UTR of an mRNA can lead to deadenylation and degradation of the target. Alternatively, initiation of translation can be inhibited through prevention of 5' m^7G-cap recognition or through prevention of ribosome assembly. Some data suggest that even after translation initiation, protein synthesis can be repressed by slowed-down elongation or by dissociation of the ribosome. One key player in translation-inhibition is

GW182 (Ding and Han, 2007; Zekri et al., 2009), which interacts with Argonaute proteins, as well as with the poly-A binding protein (PABP). It localizes RISC comlpexes into specific cytoplasmic compartments, P-bodies, which are centers for mRNA regulation and degradation.

On the other hand, Argonaute proteins show specific interaction with 5' m^7G-caps and thus could compete with eIF4E, the cap binding factor necessary for translation initiation. Since a proper interaction between the 5' cap and the polyA tail of an mRNA is required for translation initiation, it is conceivable that miRNA mediated translational repression is a consequence of interference with those basic processes.

2.5 siRNAs

siRNAs are produced from linear, perfectly base paired precursor molecules and they can be subdivided into two groups, depending on the origin of the precursor. If the long doublestranded RNA (dsRNA) originates from an exogenous source, like in an experimental laboratory setup, or a virus, they are called exo-siRNAs. If the dsRNA stems from endogenous sources, such as transposable elements, repetitive sequences, pseudogenes or convergent transcription, they are called endo-siRNAs (Golden et al., 2008).

2.5.1 siRNA biogenesis

Like miRNAs, siRNAs are produced from double-stranded precursor RNA by Dicer, leaving a 5' phosphorylated end and a 3' OH end with a 2 nt overhang. In mammals and *Caenorhabditis elegans*, one Dicer produces miRNAs as well as siRNAs. Other organisms have designated Dicers, for instance in *Drosophila* Dicer-1 makes miRNAs, whereas Dicer-2 produces siRNAs. To detect a dsRNA as a substrate and channel it into the siRNA pathway, Dicer relies on double-strand binding proteins (dsRBPs). In *Caenorhabditis elegans* this is RDE-4, in *Drosophila* R2D2 and Loqs-PD. (Liu et al., 2003; Parker et al., 2006; Mirkovic-Hosle and Forstemann, 2014; Hartig et al., 2009; Hartig and Forstemann, 2011). In human cells, Dicer can associate with two different dsRBPs, protein activator of PKR (PACT) and trans-activation response RNA-binding protein (TRBP). (Lee et al., 2013). Loading and maturation of the siRISC is in many parts convergent with the miRNA pathway and relies on mainly the same components (see figure 2.2).

2.5.2 siRNA function

siRNAs are distinguished from miRNAs based on their origin, but also through their perfect complementarity to their targets. Perfect base pairing between small RNA and

target triggers direct cleavage of the target, as opposed to a non-cleavage dependent repression. Perfect complementarity is especially observed in plants, where the siRNA pathway is a powerful defense against viral infections. It appears that the absence of a protein-based immune system in plants expanded the breadth of siRNAi. In most animals, the mechanisms between siRNA and miRNA mediated silencing are converging. In plants and worms, however, the response to long dsRNAs differs from most other organisms. They possess a system, in which they can amplify the silencing.

2.5.2.1 siRNAs in plants

Exogenous dsRNAs trigger processing into siRNAs through Dicer. Those so called primary siRNAs lead RISC to the target, which is cleaved by the Argonaute. Plants and worms encode an RNA dependent RNA polymerase (RdRP), which uses this cleaved transcript as a template and synthesizes a second strand, resulting in a long dsRNA. The plant, which is widely used for a model organism to investigate plant RNAi, is *Arabidopsis thaliana*. *Arabidopsis thaliana* expresses 4 different Dicers and one of them is specialized to process the products of RdRP into secondary siRNAs (Bologna and Voinnet, 2014). The resulting siRNA can once more be loaded into Argonautes and direct nuclear or cytoplasmic silencing. In *Caenorhabditis elegans* the mechanism to produce secondary siRNAs is sligthly different. Detection of the target RNA by the Argonaute RDE-1 attracts the RdRP, which produces small RNAs from the target template.

2.5.2.2 Transcriptional gene silencing in the nucleus

Nuclear RNA silencing mechanisms have first been described in yeast, ciliates and plants, where they have been found to modify the chromatin structure of targeted genomic regions (Castel and Martienssen, 2013). Animals also seem to have nuclear RNA silencing pathways, but mechanistic insights are still sparse.

The fission yeast *Saccharomyces pombe* only has one Argonaute protein, Ago1, and its main function appears to be nuclear - formation of heterochromatin on centromers (Moazed et al., 2006). Those regions consist of repetitive sequences, which are bidirectionally transcribed into long dsRNAs. Those dsRNA molecules are diced into siRNAs and loaded into Ago1 to form the RITS complex (RNA-induced transcriptional silencing). The siRNA brings RITS to nascent transcripts of the centromeric region, where it ensures the maintenance of heterochromatin through different mechanisms. RITS attracts the Histone-methyl transferase Clr4, which methylates histone H3 lysine 9. This chromatin mark attracts the HP1 homolog Swi6, which establishes heterochromatin at this locus. At the same time, RITS recruits the RNA-dependent RNA polymerase Rdp1, which am-

plifies the siRNA signal through further generation of dsRNA from this locus. (Motamedi et al., 2004) Further, genetic studies showed that Ago1, Dicer and Rdp1 are crucial for heterochromatin formation (Volpe et al., 2002), that Clr4 is required for siRNA production (Noma et al., 2004) and that Swi6 is required for Rdp1 localization. This whole system is organized in a feed-forward loop, where siRNA production, RITS localization and H3K9 methylation are reinforcing each other.

In *Arabidopsis thaliana* the establishment of pericentromeric heterochromatin also depends on Argonaute proteins and small RNAs. Yet, the platform for heterochromatin formation does not rely on histone tail modification, but rather on DNA methylation (Matzke et al., 2009). The centromeric repetitive regions are transcribed by the plant specific RNA polymerase IV (Pontier et al., 2005; Kanno et al., 2005) and the resulting transcripts are templates for the RNA dependent polymerase RDR2, which produces the dsRNA for siRNA production through Dicer. Those siRNAs are loaded into Ago4 in the cytoplasm and are re-imported into the nucleus, where they recognize nascent transcripts from the second plant-specific RNA polymerase V (Pontier et al., 2005; Kanno et al., 2005). Upon detection, the PolV transcribed region gets repressed through DNA methylation by the DNA methyltransferase DRM2.

The two mechanisms of RNAi mediated pericentromeric heterochromatin formation necessitate an intricate regulation: heterochromatin is presumably silenced, but still needs to be transcribed in order to be established. The nuclear role of RNAi in fission yeast and *Arabidopsis thaliana* seems to be mainly focused on establishment of a structural pericentromeric heterochromatin.

In animal somatic cells evidence for nuclear RNAi is getting denser. The molecular mechanism, however, is far less understood (Ohrt et al., 2012; Schraivogel and Meister, 2014). Argonautes have been reported to shuttle between the cytoplasm and the nucleus (Weinmann et al., 2009) and Dicer, too, can localize into the nucleus (Doyle et al., 2013). For both proteins the nuclear levels seem to be tightly regulated. Other reports suggest the presence of a complete nuclear RNAi machinery containing all components that are essential in cytoplasmic RNAi. (Gagnon et al., 2014). In analogy to *Saccharomyces pombe* and *Arabidopsis thaliana*, mammalian RNAi also was implicated in heterochromatin formation of structural heterochromatin on satellites, different other repeat-rich loci and transposable elements (Peng and Karpen, 2007; Fagegaltier et al., 2009; Deshpande et al., 2005). One report suggested a direct interaction of Ago1 and Dicer2 with chromatin and RNA polymerase II in somatic *Drosophila* cells (Kavi and Birchler, 2009). Another study showed a direct association of Dicer with Polymerase II on actively transcribed genes and suggests transcriptional silencing (White et al., 2014). To understand

the full extend of nuclear RNAi functions in mammals, further studies will be required.

2.6 piRNAs

The piRNA pathway is an RNA silencing pathway, which is specialized in repression of transposble elements. piRNAs associate with the Piwi-clade Argonautes, which is specific to metazoans and show gonad-specific expression (Houwing et al., 2007a; Lin and Spradling, 1997; Kuramochi-Miyagawa et al., 2001, 2004; Cox et al., 1998; Carmell et al., 2002).

The domain structure of Piwi Argonautes is very similar to canonical Argonautes, with the MID domain anchoring the 5' end of a piRNA, and the PAZ domain lodging the 3' end. In *Drosophila* all three Piwi proteins have slicer activity. Slicer deficient Piwi, however, does not result in any phenotype (Sienski et al., 2012; Darricarrere et al., 2013), whereas Aub and Ago 3 absolutely require their slicer activity. What distinguishes Piwi clade Argonautes from other Argonautes, is the presence of Arginine-rich motifs near their N-termini. These residues are post-translationally dimethylated and arginine methylation allows Piwi proteins to interact with Tudor family proteins, which are key components of the piRNA pathway (Vagin et al., 2009; Nishida et al., 2009; Kirino et al., 2009; Liu et al., 2010).

The eponymous Tudor domain binds methylated arginines on Piwi proteins. Since many Tudor proteins often have several Tudor domains, they can act as a scaffold for the formation of higher-order complexes and possibly coordination of piRNA biogenesis and function (Mathioudakis et al., 2012; Huang et al., 2011).

In germ cells of *Drosophila melanogaster* and in mouse testes, recognition of arginine-methylated effector proteins seems to compartmentalize the piRNA pathway into distinct perinuclear granules, which have been described by cell biologists as "nuage" (for their clowd-like morphology, 'nuage' in french means 'cloud'). They are similar to P-bodies and harbor multiple components of the piRNA pathway (Brennecke et al., 2007; Aravin et al., 2008, 2009; Malone et al., 2009)

With a typical size between 24 and 30 nucleotides piRNAs are longer than siRNAs and miRNAs. They are generated either from RNA transcripts of active TE copies or from transcripts originating from specialized loci in the genome, called piRNA clusters. Clusters are intergenic regions harboring defunct remnants of TEs, forming the basis of germline defense against TE propagation (Brennecke et al., 2007; Aravin et al., 2007, 2008). piRNAs that are generated from piRNA clusters are mostly antisense to TE mRNA sequences and thus can guide the piRISC to mRNAs of active TEs through complementary base pairing.

The piRNA pathway has often been compared to an adaptive immune system, as it conveys memory of previous transposon invasions by storing TE sequence information in piRNA clusters. By amplifying piRNAs that are complementary to the active transposon sequence, the piRNA pathway can respond quickly and specifically to acute TE activation. In addition to targeted cleavage of TE mRNAs within the cytoplasm, Piwi proteins also function on the chromatin level to silence TE transcription. Silencing is mediated through formation of repressive chromatin on TE copies (Rozhkov et al., 2013; Sienski et al., 2012; Le Thomas et al., 2013).

2.6.1 piRNA biogenesis

2.6.1.1 piRNA cluster architecture and transcription

piRNA populations are highly complex: deep sequencing of piRNAs from mouse and fly revealed millions of individual, distinct piRNA molecules. Neither piRNAs, nor their precursor sequences show any structural motif or sequence bias except for a preference for Uracil as the first 5' nucleotide (1U bias) of the piRNA. However, when mapped to the genome, this highly diverse population of piRNAs falls into a few discrete genomic loci, called piRNA clusters (Ro et al., 2007; Brennecke et al., 2007; Lau et al., 2006; Grivna et al., 2006; Aravin et al., 2007; Houwing et al., 2007b; Aravin et al., 2006). Clusters are up to 200 kilobases long, mostly located in pericentromeric and subtelomeric heterochromatin and are highly enriched in transposable element sequences (Brennecke et al., 2007). Upon invasion of a new transposon the TE eventually jumps into one of the clusters leaving a memory of the invasion. Once a transposable element leaves a trace in a cluster, it can be targeted by the piRNA machinery, which will limit further propagation of the TE. TEs insert into different genomic sites in a non-random fashion (Huang et al., 2012; Yamanaka et al., 2014). Studies on the P-element transposon established that those elements preferably jump into the immediate 5' region of genes or into 5' exons (Spradling et al., 2011; Yamanaka et al., 2014). Another transposable element, piggyBac, was also shown to preferably insert near promoter regions (Thibault et al., 2004; Yamanaka et al., 2014). Hence, TEs seem to require an open chromatin structure, which is characteristic for actively transcribed regions (Yamanaka et al., 2014).

In fly ovaries, piRNA clusters are expressed in two cell types: cells of germline origin that include the developing oocyte and associated nurse cells, and somatic support cells called follicular cells. Interestingly, the structure of clusters differs depending on where they are expressed: Germline clusters are transcribed bi-directionally, generating both sense and antisense piRNA in relation to the transposon mRNA. Conversely, somatic clusters appear to be transcribed uni-directionally, producing mostly piRNA that are

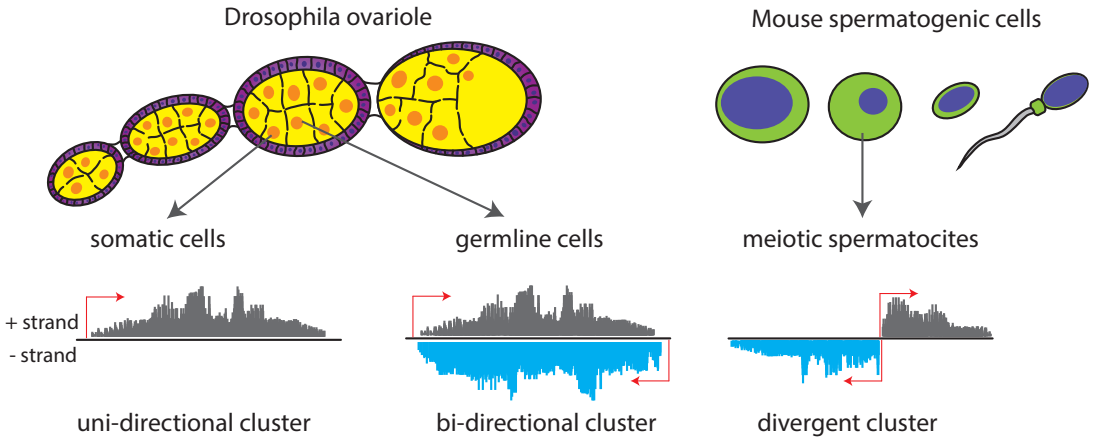

Figure 2.3: **piRNA clusters *Drosophila* and mouse.** In the *Drosophila* ovaiole germline cells (in yellow) are surrounded by a layer of somatic cells (called follicular cells, in pink). The somatic cells express uni-stranded piRNA clusters, in germline cells the majority of piRNA clusters is transcribed in a bi-directional manner, giving rise to piRNAs mapping to both strands. In meiotic spermatocites of mouse testes, most piRNA clusters are transcribed divergently from a central promoter region.

antisense to TE coding regions (Brennecke et al., 2007) (see figure 2.3). Overall, the cellular source of piRNA cluster transcripts destines how they are processed.

In mouse testes, spermatogenic cells have two kinds of piRNA clusters: one class is transcribed during embryonic development, and like in *Drosophila melanogaster*, piRNAs derived from these clusters defend the germline against transposable elements. A second type of piRNA cluster is expressed in spermatogenic cells of adolescent mice during the first division of meiosis. These are called pachytene clusters, since piRNAs derived from these clusters are highly abundant in the pachytene stage of meiosis. Pachytene piRNAs are not enriched in transposon sequences and to date their function is unknown. An interesting feature of many pachytene clusters is that they are transcribed divergently (i.e. in both directions) from a central promoter (figure 2.3) (Aravin et al., 2006; Houwing et al., 2007b) .

To date, the transcriptional regulation of piRNA clusters in flies and in embryonic stages of mouse spermatogenesis remains elusive. Chromatin immunoprecipitation (ChIP) analysis on histone modifications in the silkworm ovary-derived BmN4 cell line (this cell line contains a functional piRNA pathway and is used as an analogous model for the *Drosophila* piRNA pathway) revealed that piRNA clusters display features of euchromatin. They are enriched for histone modifications associated with transcriptional activity such as H3K4 di- and tri-methylation in addition to H3K9 acetylation. At the same time, they are devoid of repressive histone H3K9 di- and tri-methyl marks (Kawaoka et al., 2013). In ChIP performed on *Drosophila* ovaries, however, H3K9me3 marks are

present on clusters and transposon loci (Rangan et al., 2011). The heterochromatin protein (HP1) homolog of *Drosophila*, Rhino, seems to bind those histone marks on piRNA clusters and is essential for piRNA cluster transcription (Klattenhoff et al., 2009). Still, it remains unresolved if these chromatin marks are a cause or consequence of cluster transcription.

In summary, data about piRNA clusters in *Drosophila* suggests that they are similar to pericentromeric heterochromatin in *Saccharomyces pombe*, where a specific chromatin configuration is required for generation of small RNAs.

2.6.1.2 Biogenesis of primary piRNAs

piRNAs in flies are 23-25nt in length, whereas in mouse they are slightly longer, averaging 25-28nt. Similar to other small RNA classes, piRNAs are processed from longer precursors. In contrast to miRNAs and siRNAs, however, the precursors are single stranded transcripts without obvious hairpin structures and they are produced in a Dicer independent pathway. It is thought that piRNA biogenesis begins with the endonucleolytic cleavage of the long precursor transcript, generating shorter piRNA precursors. This cleavage event possibly specifies the 5' end of the future piRNA (figure 2.4). Genetic screens and structural studies suggest that the endonuclease, Zucchini (Zuc), might conduct this first cleavage of piRNA precursors (Voigt et al., 2012; Nishimasu et al., 2012; Ipsaro et al., 2012; Olivieri et al., 2010; Haase et al., 2010; Muerdter et al., 2013; Handler et al., 2013; Czech et al., 2013). Endonucleolytic cleavage of ssRNA in vitro through Zucchini results in a 5' phosphate end and a 3' hydroxyl end. (Nishimasu et al., 2012; Ipsaro et al., 2012). Since mature piRNAs possess a 5' phosphate, Zucchini seems a good candidate for this cleavage step. Nishimasu and colleagues investigated the cleavage activity of Zucchini on heterogeneous and poly(U) ssRNA (Nishimasu et al., 2012) and found that it exhibits no strong bias towards any nucleotides. It is therefore unclear, how the 1U bias of primary piRNAs is generated. A bias for the first nucleotide in Argonaute bound small RNAs has been observed in many systems. miRNAs, for example, exhibit a strong bias for Uridine at the 5' end in *Caenorhabditis elegans* (Lau et al., 2001) and *Drosophila* (Czech et al., 2009; Ghildiyal et al., 2010). siRNAs in flies and in plants tend to have a 1C bias (Lee et al., 2010; Ghildiyal et al., 2008; Kawamura et al., 2008). Structural studies on the MID domain of human and *Arabidopsis thaliana* Argonautes showed that a so-called "specificity loop" seems to be a determinant for 5' nucleotide specificity (Frank et al., 2010, 2012) for miRNAs and siRNAs and it is suggested that this bias is detected during loading of the double-stranded Dicer product. The aminoacid sequence of the nucleotide specificity loop of Piwi Argonautes is different from hAGO2 and AtAGO1, and still all those Argonautes exhibit a 1U selectivity. This suggested that the specificity

loop interacts with the 5' nucleotide through hydrogen bonds and not through specific aminoacid side chains. Studies in *Drosophila melanogaster* embryo lysate showed that the 5' nucleotide of miRNAs and siRNAs seems to be sensed and monitored during RISC loading and unwinding of the passenger strand (Kawamata et al., 2011). After cleavage of the piRNA cluster transcript, the 5' end of piRNA precursors gets loaded into a Piwi protein. In flies, primary piRNA are loaded into two of the three Piwi proteins, Piwi and Aubergine (Aub). The factors that constitute the piRNA loading machinery are currently unknown, however several studies have identified the proteins Shutdown (Shu), Vreteno (Vret), Brother-of-Yb (BoYb) and Sister-of-Yb (SoYb) as components involved in loading of Piwi and Aub with primary piRNA (Olivieri et al., 2010; Handler et al., 2013). Loading of Piwi seems to require some additional factors including Armitage (Armi) and Zucchini (Zuc) and in somatic ovary cells, the Tudor domain protein, Yb (Szakmary et al., 2009; Olivieri et al., 2010).

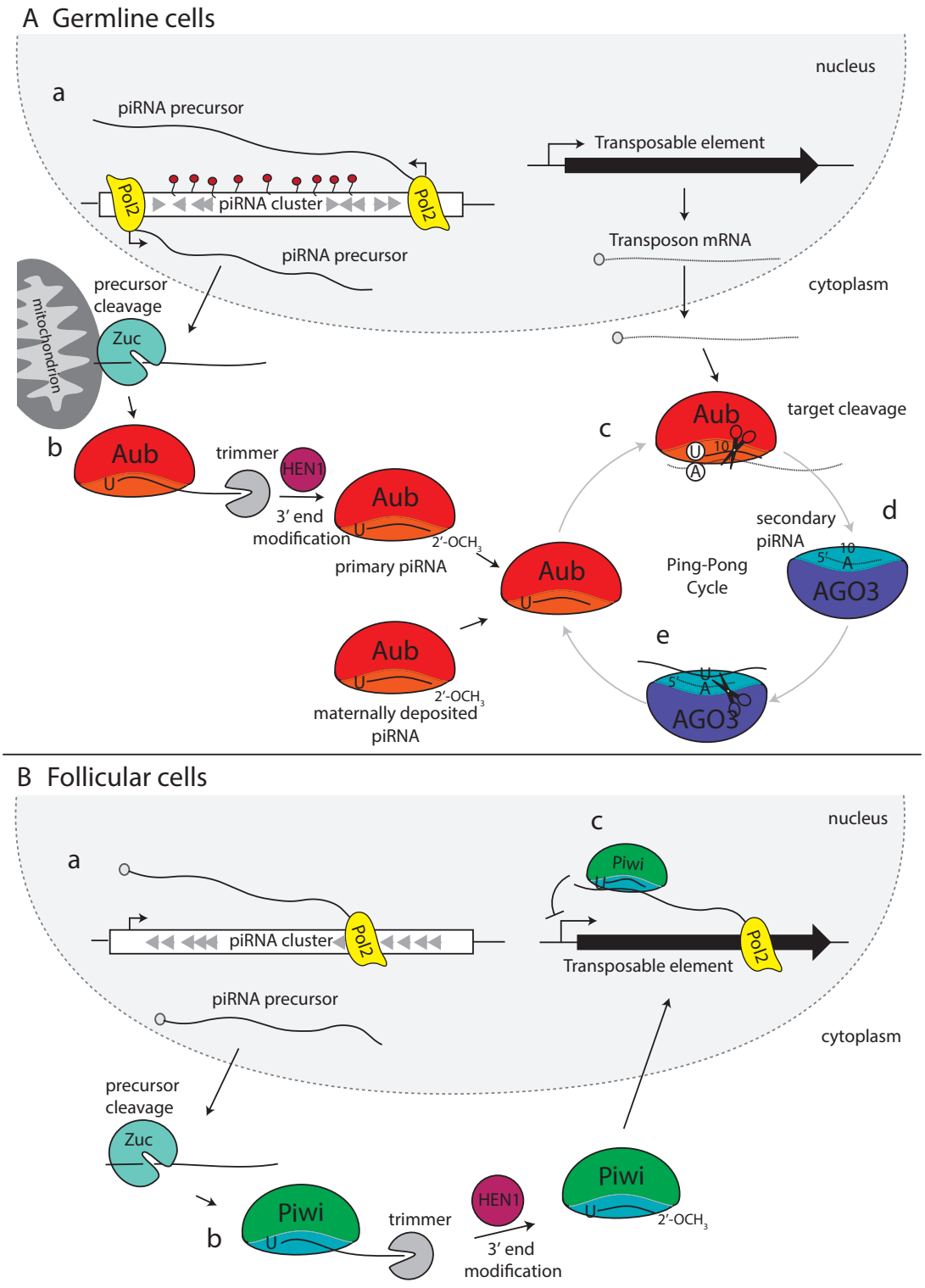

Figure 2.4:
piRNA biogenesis pathways in germline cells and somatic follicular cells of the *Drosophila* ovary. *[Legend see next page]*

Figure 2.4: piRNA biogenesis pathways in germline cells and somatic follicular cells of the *Drosophila* ovary. *[The figure is on the previous page]*
Biogenesis of piRNA in flies begins in the nucleus and initiates the ping-ping cycle in the cytoplasm. (A) piRNA pathway in germline cells. a) Precursor transcripts originate from bi-directionall piRNA clusters, containing TE fragments in both orientations (grey arrows). They are exported from the nucleus into the cytoplasm and processed into smaller fragments, possibly by the mitochondrial-associated endonuclease Zucchini. b) Aub binds precursor piRNA fragments with a preference for fragments that contain a 5 terminal uracil (1U). The 3' end of piRNA precurosrs is generated when an unknown 3- 5 exonuclease trims the precursor piRNA fragment bound to Aub and Hen1 catalyzes the 2-O methylation. This cluster derived piRNA is called primary piRNA. c) When mRNA of active transposons is exported from the nucleus, Aub loaded with a piRNA complementary to this element is guided to the transcript and cleaves it. d) The cleaved transcript is loaded into Ago3, anchored with its 5' end, followed by trimming and 2' O-methylation of its 3' end, to generate a mature secondary piRNA. e) Ago3 loaded with the secondary piRNA is then believed to target and cleave newly generated precursor transcripts that are loaded into Aub to initiate a new round of ping-pong. (B) piRNA pathway in somatic follicular cells. a) Uni-stranded clusters containing TE fragments mainly in antisense orientation are transcribed into piRNA precursor molecules and exported into the cytoplasm, where they are cleaved through Zucchini. b) Piwi incorporates the piRNA precursor with a prefrernce for 1U, which gets trimmed and methylated at its 3' end. The Piwi-piRNA complex re-enters the nucleus and finds transcriptionally active TEs through complementarity to the nascent transcript. There it silences the TE on a transcriptional level.

While Dicer determines the size of miRNAs and siRNAs, piRNAs are sized according to the Argonaute, they associate with. Once loaded into Piwi proteins, the piRNA precursor is trimmed at its 3' end by an unknown 3' - 5' exonuclease. This nuclease is tentatively named "Trimmer". The final length of the piRNA is determined by the piRNA binding pocket of the Piwi protein: Piwi associated piRNAs are 25 nt long, whereas piRNAs associated with AUB or AGO3 are 24 nt or 23 nt long, respectively. After trimming, piRNAs acquire a characteristic 2' methyl modification at their 3' end that is introduced by the piRNA methyl-transferase Hen-1 and leads to stabilization of the small RNA (Kawaoka et al., 2011). Overall, while the exact biochemistry and the involved factors are still not fully known, piRNA biogenesis consists of multiple consecutive steps from transcription of precursors, through precursor cleavage and loading, to trimming and 2' O-methylation. Defects in any of these steps leads to diminished piRNAs, upregulation of transposons and sterility.

2.6.2 Nuclear and cytoplasmic function of the piRNA pathway

Transposons are selfish genetic elements that can be classified into DNA or retrotransposons based on the mechanism by which they transpose in the genome. DNA transposons mobilize by a 'cut and paste' mechanism that is accomplished with the help of an encoded transposase protein. Retrotransposons propagate by reverse transcription of an RNA intermediate. The required enzymes are encoded by the transposable elements. Another class of TE is the semi-autonomous sequences, such as Alu repeats, that require the protein machinery of other transposons for their replication. In either case the transposable element needs to be transcribed as an mRNA in order to produce the enzymes necessary for its transposition. Accordingly, transposition can be regulated on multiple levels, from the efficiency of transcription, to transcript stability and rate of translation of required factors, or by interfering with the process of re-integration into the genome. The piRNA pathway seems to target two steps that are required for all transposons: In the nucleus, piRNAs are implicated in the regulation of chromatin structure that can affect transcription of targeted loci. In the cytoplasm, the piRNA pathway directly targets and destroys RNAs of transposons that escaped transcriptional silencing.

2.6.2.1 The role of piRNAs in regulating chromatin structure

Transcriptional regulation of specific loci is influenced by the overall chromatin architecture. In mammals this regulation is achieved on two levels: through methylation of DNA and through modifications of histone proteins or even incorporation of special histone variants. In *Drosophila melanogaster*, histone modifications and histone variants

are mainly responsible for defining the properties of chromatin. In both flies and mice, one of the three Piwi proteins localizes to the nucleus, suggesting a nuclear role for the piRNA pathway (Cox et al., 1998; Aravin et al., 2008). Such a mechanism is further strengthened by the observation that in flies Piwi shows a banding pattern on polytene chromosomes of nurse cells, suggesting a direct interaction with chromatin (Le Thomas et al., 2013).

Nuclear piRNA effects in mouse

In mouse, DNA methylation plays an essential role in imprinting and in silencing of transposable elements (TEs). Male germ cells are established in a narrow developmental window during embryonic germ cell development, following global genome de-methylation. This is a critical point in the fate of spermatogenic cells, since the failure to methylate sequences of retrotransposons scattered throughout the genome leads to their activation and subsequent meiotic failure and sterility (Bourc'his and Bestor, 2004). Several lines of evidence indicate that piRNAs are responsible for directing DNA methylation to its genomic targets. First, several proteins involved in the piRNA pathway have been described to localize to the nucleus: Miwi2, Tdrd9, and Mael (Soper et al., 2008; Shoji et al., 2009; Aravin et al., 2008, 2009). While Tdrd9 and Mael are expressed throughout male germ cell development, Miwi2 is expressed in a narrow developmental window exactly at the same time of *de novo* DNA methylation in the embryonic spermatocyte. Additionally, genetic data implicate piRNAs in establishing *de novo* DNA methylation patterns in the mouse male germline on regulatory regions of TEs (Kuramochi-Miyagawa et al., 2008; Carmell et al., 2007; Aravin et al., 2008). Dnmt3L is an accessory factor for *de novo* DNA methyltransferases Dnmt3a and Dnmt3b, and is essential for *de novo* DNA methylation. The defects observed in Dnmt3L knock-out animals are strikingly similar to those observed in animals deficient in many piRNA pathway components, including the two Piwi proteins, MILI and MIWI2: all result in meiotic arrest of spermatogenesis and apoptosis of germ cells. Furthermore, methylation patterns were not reestablished on retrotransposon sequences in mice lacking Mili or Miwi2 (Kuramochi-Miyagawa et al., 2008), whereas piRNAs are still produced in Dnmt3L mutants (Aravin et al., 2008).

A second mouse Piwi protein, MILI is concomitantly expressed in the embryonic stage of spermatogenesis. While it localizes to the cytoplasm, its expression is necessary for piRNA loading and nuclear localization of MIWI2. The phenotype of Mili mutants is similar to that of Miwi2 suggesting that piRNAs serve as sequence-specific guides that direct Miwi2 to sites where DNA methylation is established (Kuramochi-Miyagawa et al., 2001, 2004, 2008; Aravin et al., 2008).

Indeed, piRNA directed DNA methylation is one mode of regulation for transposable

elements. A study on DNA methylation patterns in wild-type and Mili mutant mice found that young, still very active transposable elements that show similarities to active genes, seem to evade a general wave of DNA methylation during spermatogenesis (Molaro et al., 2014). Another study showed a direct link between the piRNA pathway and H3K9 tri-methylation of transposable elements (Pezic et al., 2014). It revealed that the promoter regions of certain active transposable elements are targeted for H3K9 tri-methylation in a Miwi2 dependent manner. How DNA methylation and H3K9 methylation are orchestrated and if they are interdependent, remains to be established.

Nuclear piRNA effects in flies

In contrast to mammals, DNA methylation in *Drosophila* does not play a major role in the regulation of chromatin structure and gene expression. In fruit flies it is believed that the piRNA pathway induces TE repression through deposition of specific histone marks at target loci. Multiple studies have shown that the piRNA pathway transcriptionally represses a subset of transposons. In cell culture experiments, deletion of Piwis nuclear localization signal leads to failure of transposon silencing, even though Piwi proteins are loaded with piRNAs in the cytoplasm (Saito et al., 2009). Similarly, in flies the deletion of the N-terminus of Piwi leads to its delocalization from the nucleus, a slight increase in active histone marks (di-methylation of H3K79 and H3K4) and a decrease of repressive marks (di-/tri-methylation of H3K9) over several transposons (Klenov et al., 2007). Additionally, an increase in transcription and accumulation of several telomeric retrotransposons was observed in *Drosophila* ovaries upon mutation of piRNA pathway components (Chambeyron et al., 2008; Shpiz et al., 2009, 2011). The telomeric transposons, which showed increased transcription, showed slight changes in their chromatin marks (Shpiz et al., 2011). These observations all indirectly hinted towards an involvement of Piwi in transcriptional regulation of transposable elements. Finally, three genome-wide studies confirmed this assumption: knockdown of Piwi leads to transcriptional derepression of a significant fraction of TEs both in cell culture and in ovaries as assessed by Pol II ChIP-seq and by GRO-seq (Rozhkov et al., 2013; Sienski et al., 2012; Le Thomas et al., 2013).

While it is conceivable that transcriptional silencing of transposable elements is achieved through establishment of a repressive chromatin state, a correlation between repressive chromatin marks and silencing has not yet been shown directly. Likewise, factors involved in mediating Piwi's repressive function are not known. Recently, two factors have been proposed to be involved in inducing transcriptional silencing. Maelstom (Mael) is a conserved factor that has been implicated in TE silencing in both mouse and flies (Soper et al., 2008; Aravin et al., 2009; Sienski et al., 2012). Knockdown of Mael in cell culture

of ovarian somatic *Drosophila* cells leads to increased Pol II occupancy over TEs similar to changes observed upon Piwi knockdown (Sienski et al., 2012). The same study showed that in ovarian somatic cell-culture Piwi silences a significant number of novel TE insertions through tri-methylation of H3K9 (Sienski et al., 2012). The silencing mark does not only cover the site of the novel TE insertion, but spreads up to 15 kb downstream in the direction of transcription. This observation indicates that transcription plays a role in establishing repressive chromatin. The spreading of repressive marks upon insertion of a transposable element can even lead to repression of proximal protein coding genes. In fly, however, similar new insertions in proximity to protein coding genes were not observed (Le Thomas et al., 2013). It is conceivable that while in cultured cells the silencing of nearby genes is tolerated and can be detected, in a living organism there is a strong selective pressure against insertions that would compromise functional gene expression.

The piRNA sequences bound to Piwi identify the sequences that are targeted for Piwi-mediated repression. Additionally, loading Piwi with artificial sequences like for example mapping to the lacZ transgene leads to silencing of lacZ-expressing loci *in vivo*. While the mechanism of target recognition remains unresolved, it is likely that piRNAs recognize nascent transcript targets through sequence complementarity. This is supported by the observation that Piwi knockdown results in decreased H3K9 tri-methylation of the genomic environment of active transposons, while untranscribed fragments of transposable elements throughout the genome remained unaffected (Sienski et al., 2012).

Transcriptional repression by Piwi occurs in both germ and follicular cells of *Drosophila* ovaries. Follicular cells are the somatic support cells in fly ovaries, and while they are not germline cells, they do express a uniquely tailored version of the piRNA pathway. Out of the three Piwi proteins in *Drosophila*, only Piwi is expressed in follicular cells. In these cells it is specifically loaded with piRNA generated from the unidirectional cluster flamenco (Malone et al., 2009). The flamenco cluster appears to primarily target TEs of the Gypsy family, whose expression can lead to the formation of viral particles that might infect the germline (Song et al., 1997). Therefore, transcriptional repression of Gypsy loci by Piwi in follicular cells is crucial for germline survival.

Transcriptional gene silencing seems to be a paradox mechanism: silenced regions need to produce small RNAs in order to be silenced. In case of piRNA directed transposable element repression, a constant level of piRNAs could provide the needed small RNAs and induce transcricptional silencing in trans. Some data from this thesis, however, propose a second, not mutually exclusive mechanism. It is conceivable that H3K9 tri-methylation on target loci recruits the piRNA pathway componenets Rhino and Cutoff and turns the target into a "mini-cluster": its transcription might be reduced, but is not abolished and the residual transcript is processed into primary piRNAs. Thus, each target turns into a

source of piRNAs and a low level of residual transcription provides the neccesary supply of substrate for piRNA biogenesis.

2.6.2.2 Ping-pong: the cytoplasmic function of the piRNA pathway

The beauty and power of the piRNA pathway lies in its ability to store information of previous TE invasions in piRNA clusters and to quickly and specifically respond to TE activation by distinctively amplifying sequences complementary to active elements. The latter is achieved through the interplay of two cytoplasmic Piwi proteins in *Drosophila*, Aubergine (Aub) and Argonaute-3 (Ago3) in a process termed the ping-pong cycle. As these two proteins are only expressed in the germline, the ping-pong cycle is also restricted to germ cells (Brennecke et al., 2007).

In *Drosophila*, Aub and Ago3 act as partners in the defense against active transposable elements, with Aub binding to piRNAs mainly coming from piRNA cluster transcripts and Ago3 enriched in piRNAs from transposon mRNAs. At the beginning of the cycle, Aub loads with primary piRNA sequences that are derived from piRNA clusters and are antisense to TE transcripts, or alternatively are already present from a maternally deposited pool (see figure 2.4). Those Aub-bound primary piRNAs have a bias for Uridine at their first 5' position and they guide Aub to scavenge the cytoplasm for complementary transcripts. When a TE mRNA with sequence complementarity to an Aub-bound piRNA is identified, Aub cleaves the target RNA ten nucleotides downstream of the piRNA 5' end. This cleavage of transposon mRNA serves two purposes: first, it destroys the transposon transcript and second, it generates the 5' end of a new piRNA precursor. This new precursor gets incorporated into Ago3 and is processed into a so-called secondary piRNA. Due to the fact that Ago3-bound piRNA is sense to the transposable element, this piRNA can promote the production of a new round of cluster-derived piRNA by recognizing complementary cluster transcripts (which then get cleaved and incorporated into Aub). As this pathway uses mRNAs of TEs as substrates, the ping-pong cycle shapes a specific response against active TEs. Aub-bound piRNAs have a strong bias for uracil at their 5' ends, and piRNAs that are loaded into Ago3 have a bias for Adenosine at position 10 (10A bias) and are complementary to the Aub-bound piRNAs over a 10 nucleotide stretch. The relationship between Aub piRNA and Ago3 piRNA is termed the ping-pong signature. It was for the longest believed to be caused by target cleavage between positions 10 and 11 of the guide, which would result in an Adenine at position 10 of the target. A recent publication, however, is challenging this model (Wang et al., 2014). Wang and colleagues show that Aub prefers a 10A target irrespective of the first nucleotide of its guide. Multiple structural analyses suggested that the first guide nucleotide is not available for pairing with the target and thus this result provides an

explanation for the 10A bias.

The precise mechanism of how Aub and Ago3 are coordinated in the ping-pong cycle is not yet fully understood. Protein components that are implicated in the ping-pong cycle include not only Aub and Ago3, but also the Tudor domain-containing proteins Qin, Krimper (Krimp), and Spindle-E (SpnE) in addition to the essential germline helicase, Vasa. SpnE and Vasa are both putative DEAD-box RNA helicases, and Qin contains a RING E3 ligase domain with unknown function (Malone et al., 2009; Li et al., 2009; Anand and Kai, 2012; Zhang et al., 2011). While the exact role of these additional proteins is not known, mutation of any of these components leads to aberrations in the ping-pong cycle and sterility.

2.7 piRNAs are trans-generational epigenetic factors

2.7.1 Introduction into epigenetic mechanisms

Parts of this section have been published in (Stuwe et al., 2014)

The cytoplasmic role of the piRNA pathway includes deposition of H3K9me3 on target loci, which is a classical epigenetic signal. Epigenetics generally describes the long-term maintenance of distinct gene expression profiles in cells despite their same genetic composition (DNA sequence). Two major modes of epigenetic regulation have been proposed: self-maintaining networks of transcription factors and chromatin modifications. Expression of particular sets of transcription factors that form gene regulatory networks can generate distinct patterns of gene expression that can be stable over long time periods and impact the identity of the cell. This was demonstrated by the observation that inducing expression of four transcription factors (the so-called Yamanaka factors) is sufficient to trigger a stable switch in cellular identity and to generate iPS cells from differentiated cells (Wernig et al., 2007; Yu et al., 2007; Nakagawa et al., 2008; Okita et al., 2007; Takahashi and Yamanaka, 2006; Takahashi et al., 2007)

A second mechanism of cell identity determination relies on maintenance of distinct chromatin states. Chromatin is generally defined as DNA in tight complex with histone and non-histone proteins. DNA itself can be marked by covalent modifications, for example through cytosine methylation in CpG dinucleotides in mammals (Hackett and Surani, 2013). Multiple residues in histone proteins undergo posttranscriptional modifications, like for example methylation, acetylation and phosphorylation. Early work demonstrated that DNA methylation as well as certain histone modifications correlate with the transcriptional status of genes (Jones and Takai, 2001; Kouzarides, 2007). Specifically, loci enriched in methylation of Lysine 4 of the histone H3 tail (H3K4me) are associated ac-

tively transcribed genes, whilst other marks such as methylation of Lysine 9 (H3K9me2/3) accumulate over repressed genes. These observations lead to the hypothesis that maintenance of a particular pattern of chromatin mark distribution provides the physical basis for an epigenetic memory of gene expression.

However, it is important to note that profiling of chromatin modifications only reveal correlations between a specific chromatin mark and the transcriptional status of genes. It can neither prove that a particular chromatin mark directly influences transcription nor that the mark is stably maintained through cellular divisions, two properties necessary to provide true epigenetic signal. In fact, it is likely that certain chromatin modifications are the consequence rather than the cause of transcriptional state.

The observation of a phenomenon called Position-Effect Variegation (PEV) more than 80 years ago (Muller, 1930) was the first clue that chromatin structure can influence gene expression of neighboring loci and be transmitted through cellular divisions. Chromosomal rearrangements in *Drosophila melanogaster* that placed the gene responsible for red eye pigmentation in the vicinity of heterochromatin resulted in flies that have intermingled patches of red and white cells in their eyes. Patches of cells without pigmentation (white color) indicate that the vicinity to heterochromatin causes repression of the gene, presumably by spreading of the repressive chromatin structure from heterochromatin into the gene environment. Importantly, large patches of white and red color suggest that the on- or off-state of gene expression was randomly established in a few cells in the developing eye and then maintained in the progeny of these cells through mitotic cell divisions.

The maintenance of a distinct cell identity shows how epigenetic signals can be stable throughout mitotic cell divisions within the same organism. There are, however, examples for transmission of epigenetic signals not only through mitosis from one cell to its progeny, but from one generation to the next. These cases of trans-generational epigenetic phenomena are characterized by inheritance of a distinct phenotype that is not encoded by a genetic source (DNA sequence). One such example is the fur coloration in the Agouti mouse strain (Argeson et al., 1996; Morgan et al., 1999). In this inbred strain the animals are genetically identical and still the fur color of individual animals can vary. The study shows that the differences in fur color depend on the DNA methylation status of a retrotransposon inserted close to the promoter of the Agouti gene, which is involved in determining fur color. The DNA methylation status of the Agouti locus is inherited from the mother and supports the role of DNA methylation as a trans-generational epigenetic mark. In other cases of trans-generational epigenetic inheritance the underlying mechanisms are not understood. In most animals the parental chromatin marks are erased and newly established after the haploid genomes fuse in the zygote. One possible explanation

for trans-generational inheritance of epigenetic marks would be the incomplete erasure of parental chromatin marks, which would allow re-establishment of expression patterns in the progeny that are similar to those in the parent.

Though trans-generational epigenetic inheritance has been described in just a few cases, as the Agouti example illustrates, they can provide invaluable insights into the mechanisms that likely operate to maintain cellular identities.

The next section will describe how piRNAs can be viewed as epigenetic factors that establish a cellular memory that is maintained through cell divisions as well as from one generation to the next.

2.7.2 Trans-generational effects of piRNAs in *Drosophila*

The sequence content of piRNA clusters constitutes the genetic component of the piRNA pathway. The piRNA pathway, however, contains a second, epigenetic component that was revealed in studies with transgenic as well as with wild type flies. piRNAs produced from a lacZ transgene insertion into a telomeric piRNA cluster can silence a lacZ reporter in trans. Unexpectedly, silencing is only observed if the lacZ-generating cluster is maternally inherited, while no repression is seen if the transgene is inherited from the father, despite the fact that the progeny are genetically identical (Josse et al., 2007). This result indicates that the presence of the piRNA cluster in the genome is not sufficient to induce silencing and that an additional epigenetic signal that is transmitted through the maternal germline is required for repression.

Silencing of native transposons in the progeny also depends on the direction in which two fly strains are crossed; this phenomenon was called hybrid dysgenesis (see figure 2.5).

It is observed in crosses between two fly strains where a particular transposable element is present in the genome of one but not the other strain. In the progeny of crosses between two such strains, the TE is activated if it was inherited from the father leading to sterile, dysgenic progeny, while the TE remains repressed and the progeny fertile if the TE is inherited from the mother (Bucheton, 1979; Bingham et al., 1982; Rubin et al., 1982) (figure 2.5). More recent studies revealed that derepression of the TE in a dysgenic cross correlates with the absence of cognate piRNAs targeting this element. This implicates a failure of piRNA-mediated silencing as a primary reason for element activation (Brennecke et al., 2008; Khurana et al., 2011; Grentzinger et al., 2012). Similar to the crosses with the lacZ transgene described above, the progeny in the two crosses are genetically identical. Therefore, the transgenerational epigenetic signal that is required to mount the silencing response in the next generation are the piRNAs themselves. Indeed, Piwi proteins loaded with piRNAs are deposited from the maternal germline into the oocyte and are present in the early embryo before the start of zygotic transcription

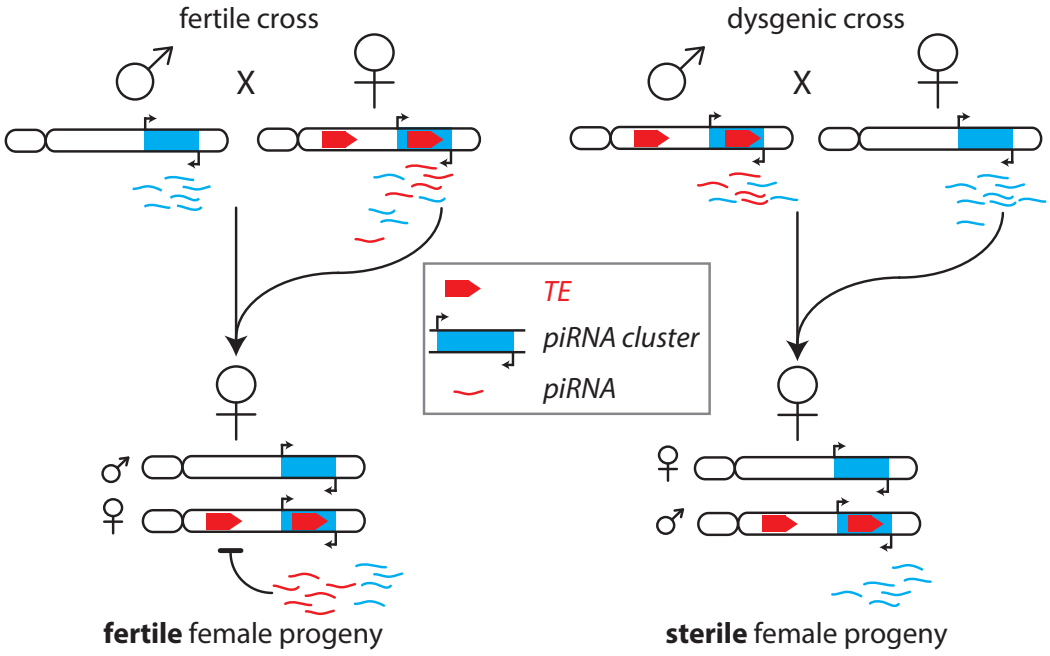

Figure 2.5: **Scheme of hybrid dysgenesis.** The presence of a potentially active transposable element (TE) in the genome correlates with expression of piRNAs targeting this element derived from piRNA clusters that contain the TE sequence. piRNAs targeting the element are transmitted from the maternal germline into the embryo and silence the TE in the progeny, keeping it fertile. If the transposable element is present in the paternal, but not the maternal genome, no piRNAs targeting the TE are deposited into the embryo. The presence of the TE sequences in the paternally inherited piRNA cluster is not sufficient to protect the progeny against sterility.

(Brennecke et al., 2008). The maternal deposition of piRNAs correlates with the ability of the adult progeny to repress cognate sequences in its ovaries (Brennecke et al., 2008). In the dysgenic cross no piRNAs against the particular TE are transmitted to the embryos as the mother lacks both the element and the cognate piRNAs, since the father does not transmit piRNAs.

How can piRNAs inherited from the previous generation ensure repression of targets in the germ cells of the progeny? It is unlikely that the piRNA molecules deposited into the oocyte are able to endure until adult stage and be sufficient to mediate repression. Therefore, piRNAs inherited from the previous generation must somehow trigger production of new cognate piRNAs in the next generation. The proof that maternally provided piRNAs are indeed able to activate production of piRNAs from cognate regions in the next generation was found in experiments with transgenic piRNA-generating loci (de Vanssay et al., 2012). Ronsseray and colleagues found that the exposure of a naïve transgenic locus that does not produce piRNAs to cognate piRNA species leads to generation of piRNAs from the locus. After initial activation, the property to generate piRNAs is inherited if the activated locus is transmitted through the maternal germline, further supporting the idea that maternally deposited piRNAs are required to maintain piRNA production. Thus, at least in *Drosophila*, piRNAs can provide an epigenetic signal that is transmitted between generations. How those piRNAs achieve a boost of piRNAs in the next generation was subject of parts of this thesis.

2.7.3 Hybrid dysgenesis - an old genetic phenomenon caused by piRNAs

piRNAs have only been discovered less than a decade ago and they are relatively young members of the repertoire of small RNAs. However, their discovery and an understaning of their workings can explain the mysterious phenomenon of hybrid dysgenesis, which has already been described more than 35 years ago. The female progeny of crosses between certain flies are sterile if the parental cross was performed in one direction, but they are not sterile, if the cross was performed in the opposite direction.(Kidwell et al., 1977; Bucheton, 1979, 1990) In case of the dysgenic cross - that is, when the progeny is sterile - certain transposons were highly upregulated. Since the progeny of those two recipocal crosses are gentically identical, sterility must be caused or prevented through a factor, which is not encoded in the genome. Today we know that dysgenesis occurs if a transposable element is present in the father and not in the mother and that sterility in the progeny is prevented through maternal piRNAs.

During oogenesis, piRNAs are defending the germline cells against transposable ele-

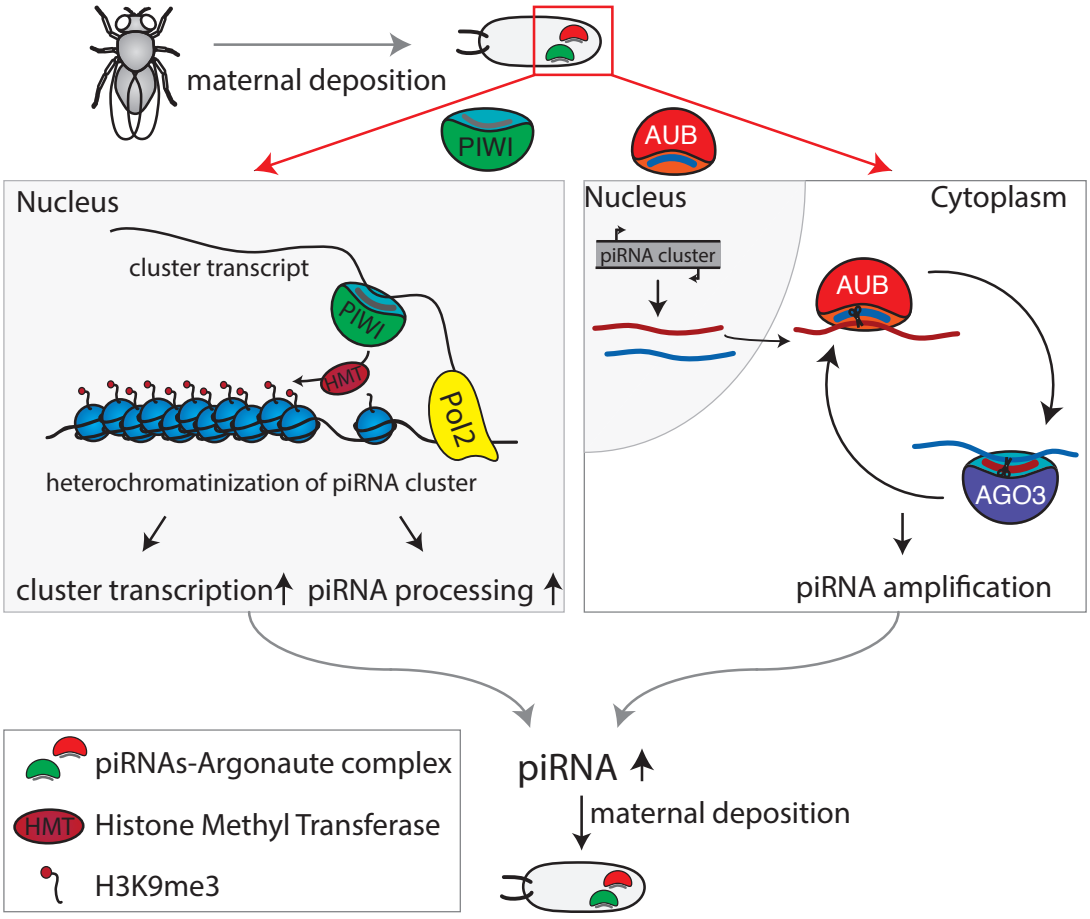

Figure 2.6: **Possible mechanisms of trans-generational effects of piRNAs.** Maternally transmitted piRNAs can lead to stable piRNA production in the next generation by two mechanisms. In the nucleus piRNAs in complex with PIWI could ensure piRNA precursor production by inducing establishment of a chromatin state over cognate piRNA clusters that permits piRNA production. Such a chromatin environment can either enhance transcription of the locus or result in channeling of the cluster transcript into the piRNA processing machinery, or both. In the cytoplasm the maternally deposited piRNAs, associated with AUB, could initiate cleavage of piRNA cluster transcripts in the ping-pong cycle.

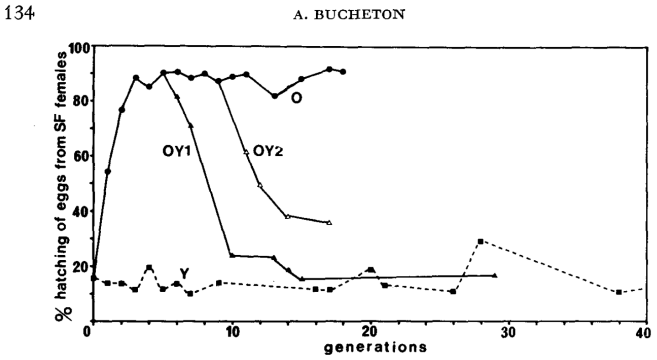

FIGURE 1.—Cumulative and reversible effects induced by aging. The graphs represent the change of reactivity level of various replicate stocks. Each point corresponds to the average of 15 measurements. ●, ▲, △ (solid lines), ■ (dotted line) correspond to the O, OY1, OY2, Y stocks, respectively (see text).

Figure 2.7: **Multi-generation experiment by Bucheton.** Figure from (Bucheton, 1979) describing how ageing mothers in a dysgenic cross leads to recovery of fertility. "Y" (dotted line) describes fertility of female flies from the "young stock", in which flies of each generation were collected from eggs laid by females two to four days old. The fertility rate ranges between 10 and 20%. The "O" line (black line, closed circles) are flies from a stock where females of each generation were collected from eggs laid by females aged between 45 and 55 days. For this stock the fertility increases within a few generations to a stably high level. When flies from the old stock were submitted to the same breeding condition as the "Y" stock, their fertility dropped to levels comparable to the "Y" strains within a few generations.

ments to ensure production of egg cells. But this is not the end of the piRNAs reach. piRNAs in complex with Aubergine are deposited into the oocyte and end up in the progeny's primordial germ cells. Therefore, the mother's repertoire of piRNAs decides, which transposable elements the progeny will be protected against. If the paternal genome contains a specific transposable element, which the mother was never exposed to, the embryos are lacking maternally deposited piRNAs against it and subsequently, the element is not repressed. Maternally inherited piRNAs are therefore necessary to mount an effective silencing response in the next generation and prevent hybrid dysgenesis (Brennecke et al., 2008)

The paternal genome, however, contains not only the specific transposable element, but also piRNA clusters containing sequences targeting the paternally inherited element. It has been noted that ageing of dysgenic females leads to recovery from sterility in a stochastic manner: some flies start laying eggs, which can mature to adult flies (Bucheton, 1979).

This recovery seems to be due to *de novo* piRNA production through processing of transcripts from paternally inherited clusters (Khurana et al., 2011). The flies seem to accumulate enough piRNAs to partially recover from sterility. Those piRNAs are again

maternally transmitted and in subsequent crosses where the mothers are always aged, fertility can reach levels almost as high as in the non-dysgenic cross. If a female fly from such subsequently aged crosses is not aged before mating, however, its progeny will be sterile again. This illustrates the limited power of the paternal piRNA cluster without maternally inherited piRNAs (Bucheton, 1979) (figure 2.7). Even though the cluster is hard-wiring the genetic footprint of a transposable invasion, it is not able to convey immunity against this element.

The presence of maternally inherited piRNAs is necessary to keep the piRNA system effective. There are two possible and not necessarily mutually exclusive scenarios, how piRNAs maintain piRNA production in the progeny on a sufficient level. On one hand, they can fuel the ping-pong cycle through initiation of direct transcript cleavage. On the other hand, they could trigger production of piRNAs through "keeping a piRNA cluster active". Indeed, experiments by Ronsseray and collegues (de Vanssay et al., 2012) demonstrated that maternally provided piRNAs are able to turn on zygotic piRNA clusters. They used two transgenic fly strains containing an insertion of 7 LacZ genes in the same euchromatic location, where one of those two strains produces piRNAs from this locus, the other does not. The only known difference between the two strains seems to be rearrangements at distant chromosomal sites, introduced through X-ray treatment. Strikingly, maternal transmission of piRNAs from this artificial cluster can switch the formerly unproductive locus to produce piRNAs. For this conversion the two alleles do not have to interact with each other; elegant crosses demonstrated that solely the maternal piRNAs suffice to convert the locus. This phenomenon is reminiscent of a similar effect described in plants as paramutation (Brzeski and Brzeska, 2011). There too, alleles can stably convert each other in a small RNA dependent manner.

In natural *Drosophila* strains the piRNA pathway acts on a diverse set of transposable elements and the repetitive and multicopy nature of those elements renders it extremely challenging to distinguish source from target. The T1/BX2 system provides a tool to investigate the nature of a piRNA cluster in a simplified way, since it can be "switched on" through paramutation in only one fly cross.

CHAPTER *3*

RESULTS

3.1 Characterization of piRNA precursors

3.1.1 Objective

During biogenesis of miRNAs and siRNAs, the double-stranded nature of their precursors seems to be one of the main triggers to identify them for small RNA processing. piRNAs, however, are derived from long, single-stranded precursors that lack any apparent structure- or sequence-motifs. The aim of this project was to identify properties of piRNA precursors from their transcription to first processing steps. The questions I wanted to answer were: Do piRNA precursor molecules possess a 5' m^7G-cap structure? Do they originate from one defined transcription start site or from multiple ones? And do processing intermediates exist in analogy to pri-mRNAs or pre-miRNAs?

To answer those questions, I employed high-throughput sequencing of specific RNA molecules from *Drosophila melanogaster* ovaries.

After establishing that piRNA intermediates indeed exist, I sought to determine factors involved in their production. Therefore, I profiled piRNA precursors in ovaries of *Drosophila melanogaster* after knock-down of candidate factors.

3.1.2 Introduction into piRNA precursors

The primary source of piRNAs are piRNA clusters, which are composed of remnants and fragments of transposable elements. In *Drosophila* ovaries, there are two major types

of piRNA clusters: uni-stranded and double-stranded; their cytological coordinates are listed in Table 3.1

The biggest uni-stranded cluster is located on the X chromosome and it is called flamenco. TE remnants in flamenco are oriented antisense to the direction of transcription and subsequently, flamenco derived piRNAs are predominantly antisense to TEs. Flamenco is expressed in the somatic follicular cells of *Drosophila* ovaries (Lau et al., 2009; Malone et al., 2009). The second big uni-stranded cluster is located upstream of flamenco and is called "20A", it is expressed in somatic as well as in germline cells.

The second class of piRNA clusters, double-stranded clusters, are mainly expressed in germline cells. Their task is - just like somatic piRNA clusters - production of piRNAs for transposable element repression.

The four biggest double-stranded clusters 42AB, 38C (two clusters) and 80EF are mainly expressed in germline cells.

Table 3.1: piRNA cluster coordinates

Name	**Chromosome**	**start**	**stop**
Flamenco	chrX	21502935	21844576
42AB	chr2R	2144349	2386719
80EF	chr3L	23281510	23310943
38C.1	chr2L	20148652	20223262
38C.2	chr2L	20104577	20116331
Xup	chrX	21392175	21431907

piRNA clusters are transcribed by the canonical polymerase for mRNA transcription, RNA polymerase II (PolII) (Li et al., 2013; Mohn et al., 2014; Zhang et al., 2014). Even though piRNA clusters in *Drosophila melanogaster* are transcribed by PolII, they are distinct from canonical gene coding regions. Until now, their transcriptional regulation is enigmatic: no conserved promoter regions have been found yet and several studies suggest that different clusters are transcribed in different ways: Transcription of flamenco seems to be initiated from one distinct promoter region (Brennecke et al., 2007; Mevel-Ninio et al., 2007), whereas 42AB might be transcribed through internal initiation (Pane et al., 2011). Generally, there are several different ways how transcription of piRNA clusters could be initiated: either from (1) flanking genic promoters, (2) not-yet defined piRNA cluster promoters, (3) from internal promoters, like e.g. the remaining TE promoter sequences or (4) piRNAs are transcribed as individual transcriptional units like in nematodes (Gu et al., 2012).

After transcription from a cluster, piRNA precursor molecules need to be further processed. The putative site of piRNA biogenesis is the perinuclear nuage structure,

containing the main components of the piRNA pathway. It is, however, not clear if there are intermediate nuclear processing steps and in which form piRNA cluster transcripts arrive at this cytoplasmic site.

To characterize the first piRNA biogenesis steps - transcription and the first cleavage events of precursors - I employed an RNA sequencing approach.

I first cloned and sequenced 5' cap containing RNAs from *Drosophila* ovaries investigate, if piRNA cluster transcript are capped. If they were capped, mapping of those reads to the genome should reveal possible promoter regions or visualize transcription initiation from within piRNA clusters. This part proved to be challenging, since the signal to noise ratio is high in 5' cap RNA libraries and the determination of real signal is not straightforward. (In a parallel approach to find transcription start sites of piRNA clusters, a colleague in my laboratory, Adrien Le Thomas, performed Chromatin Immunoprecipitation (ChIP) on PolII.) Since mature piRNAs possess a monophosphorylated 5' end, the next question was if long piRNA precursors with corresponding 5' termini exist. Therefore, I cloned and sequenced 5' P RNAs from *Drosophila* ovary. After establishing that piRNA precursors are indeed 5' monophosphorylated, I performed a small screen in *Drosophila* to find potential piRNA biogenesis factors, which might be involved in the first cleavage events of piRNA cluster transcripts.

In parallel, I tested several methods of nascent transcript isolation with the goal to detect co-transcriptional intermediates of piRNA processing. If nascent transcripts already showed hallmark features of mature piRNAs (like a bias for 1U), this would suggest that the piRNA biogenesis is divided into nuclear and cytoplasmic steps.

3.1.3 Cloning and sequencing of 5' capped RNAs

The aim of the following experiments was to find out if piRNA precursors possess a 5' m^7G-cap like other Polymerase II transcripts. Strong signals of capped RNAs from piRNA clusters would additionally reveal transcription start sites of piRNA clusters. I therefore cloned and deep-sequenced 5' cap containing RNAs from *Drosophila* ovaries.

A common way of cloning capped transcripts is immunoprecipitation of RNA with an antibody directed against the 5' m^7G-cap, which is prone to high background signal due to unspecific binding of RNA to beads. Thus, I employed an enzymatic method to capture capped RNAs (detailed description in 6.2.3). The scheme of the library preparation was adapted from Gu and Mello (Gu et al., 2012). Briefly, enzymatic treatment of total RNA converts previously capped RNAs into substrates for 5' adapter ligation. After introduction of the 3' adapter through reverse transcription the samples were amplified and sequenced. Indeed, no final library was amplified in the negative control. Libraries

were cloned from RNA of wild type *Drosophila* ovaries as shown in figure 3.1.

Table 3.2 contains details of sequencing of two replicate libraries, figure 3.2 shows annotations of the mapped reads. Approximately 60% of all reads mapped to rRNA, which is surprising, since they are not capped and should have been de-phosphorylated in the CIP treatment step. Roughly 20% of all reads mapped to the genome, 67% of which were annotated as exons. Overall, the libraries are comparable and for further analysis replicate 2 was used.

Table 3.2: Read numbers and mapping statistics of 5'Cap RNA libraries

	5'Cap RNA rep.1	**5'Cap RNA rep.2**
Total number of reads	17,494,597	31,978,559
Reads left after trimmig of adapter	17,492,165	31,972,725
reads mapping to rRNA	12,102,496	20,217,176
reads not mapping to genome	2,391,795	4,691,848
reads uniquely mapping to genome	**2,997,874**	**7,063,624**

Visual inspection of the mapped reads on the UCSC genome browser shows a clear enrichment on the 5' end of genes (figure 3.3). This served as a first proof that the method is suitable to detect 5' caped RNAs.

Next, I calculated RPKMs (explained in 6.2.7) for the three categories transcription start sites (TSS), exons and piRNA clusters. To account for different expression levels in those categories, the 5' cap RNA library was additionally normalized to a total rRNA depleted RNA-seq library.

As seen in figure 3.4, the median signal is highest at TSS, as expected for capped RNAs. The signal from piRNA clusters is lower than at TSS, but higher than in exons. This implies that the signal of capped RNA from within piRNA clusters is higher than from regions, where no capped signal is expected. After establishing that the method worked, I visually inspected piRNA clusters in the UCSC genome browser, which are shown in figures 3.5, 3.6, 3.7, 3.9 and 3.8.

3.1.4 Single stranded piRNA cluster transcripts are capped and transcribed from one defined start site

One of the big uni-stranded piRNA cluster, 20A, is expressed in somatic as well as in germline cells. 20A is neighbored by a non-coding RNA CR45082 and the gene Cyp6t1, which are both divergently transcribed from the putative TSS of cluster 20A (figure 3.6). The upper panel of figure 3.6 shows a signal at the left side of the cluster, which in the lower panel is shown in higher magnification. While a small cap-RNA signal is visible for

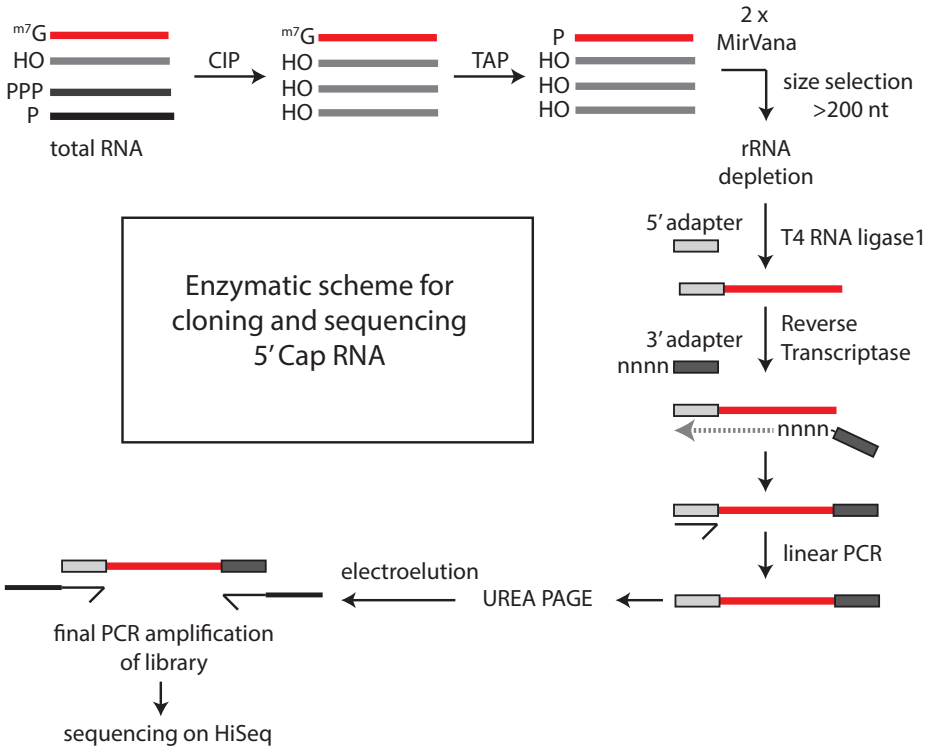

Figure 3.1: **Schematic overview of cloning procedure for 5' capped RNAs.** Total RNA was first treated with calf intestine phosphatase (CIP) to convert all 5' terminiy but the ones, which are protected by a 5' m^7G-cap into HO-groups. Subsequent treatment with Tobacco Acid Phosphatase removes the 5' m^7G-cap through hydrolizing the phosphoric acid anhydride bond in the triphosphate bridge of the cap structure. This leaves a 5'-monophosphate end. Two MirVana selections and one rRNA depletion step lead to material, which was used for 5' adapter ligation with an RNA ligase 1 specifically picking up only the 5'phosphorylated RNA molecules, which were previously capped. Subsequently, the 3' adapter was introduced through reverse transcription. A linear PCR enriched the sample for 5' ligated products and a subsequent denaturing Urea Polyacrylamide Gel cleaned the sample from unused adapters and PCR primer. The linear PCR product was size selected and electroeluted. Subsequent final PCR amplification lead to the material to be sequenced on an Illumina HiSeq 2500

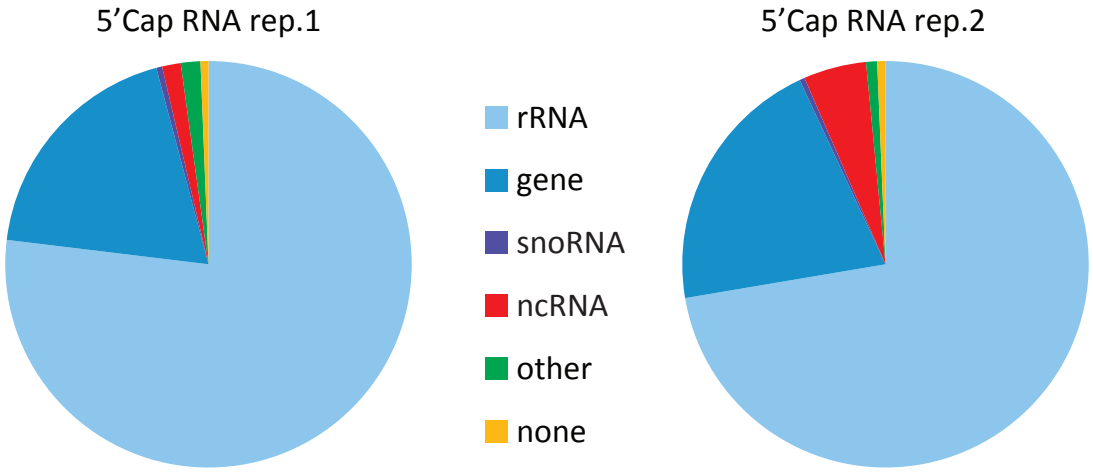

Figure 3.2: Gene ontology analysis of 5' capped RNA libraries

the gene Cyp6t1, a stronger signal is apparent upstream of this TSS. This signal does not overlap with the CR45082 transcript and thus seems to be a real capped RNA defining the transcription start site of cluster 20A.

As mentioned above, the uni-stranded, germline specific flamenco cluster was suspected to be transcribed from a single promoter, since P-element insertions in the 5' region of this cluster result in complete loss of piRNAs (Brennecke et al., 2007). The neighboring gene on the 5' end, Dip1, is transcribed divergently from flamenco. Dip1 shows a clear signal for a capped transcript, consistent with a strong PolII signal (figure 3.5). The putative TSS of flamenco, however, shows a very weak signal. While the PolII ChIP data suggests that flamenco is indeed transcribed from one defined promoter, the 5' cap RNA sequencing dataset can not clearly confirm this result. A recently published study from the Brennekce lab (Mohn et al., 2014) also contained a 5' cap data set from *Drosophila* ovaries, which I processed the same way like my data (unique mapping, zero mismatches) and compared tracks in the UCSC genome browser. Both data sets show a very comparable pattern of signals. Figure 3.10 shows that the gene next to flamenco, Dip1, is represented with 232 reads in the published 5' cap RNA library and with 144 reads in my 5' cap library. A 5' cap RNA signal at the beginning of flamenco is detected in the published dataset with 51 reads, in my dataset with 4 reads. Even though 4 reads represents a very low number, those sequences likely represent the real 5' ends of capped transcripts from the flamenco cluster.

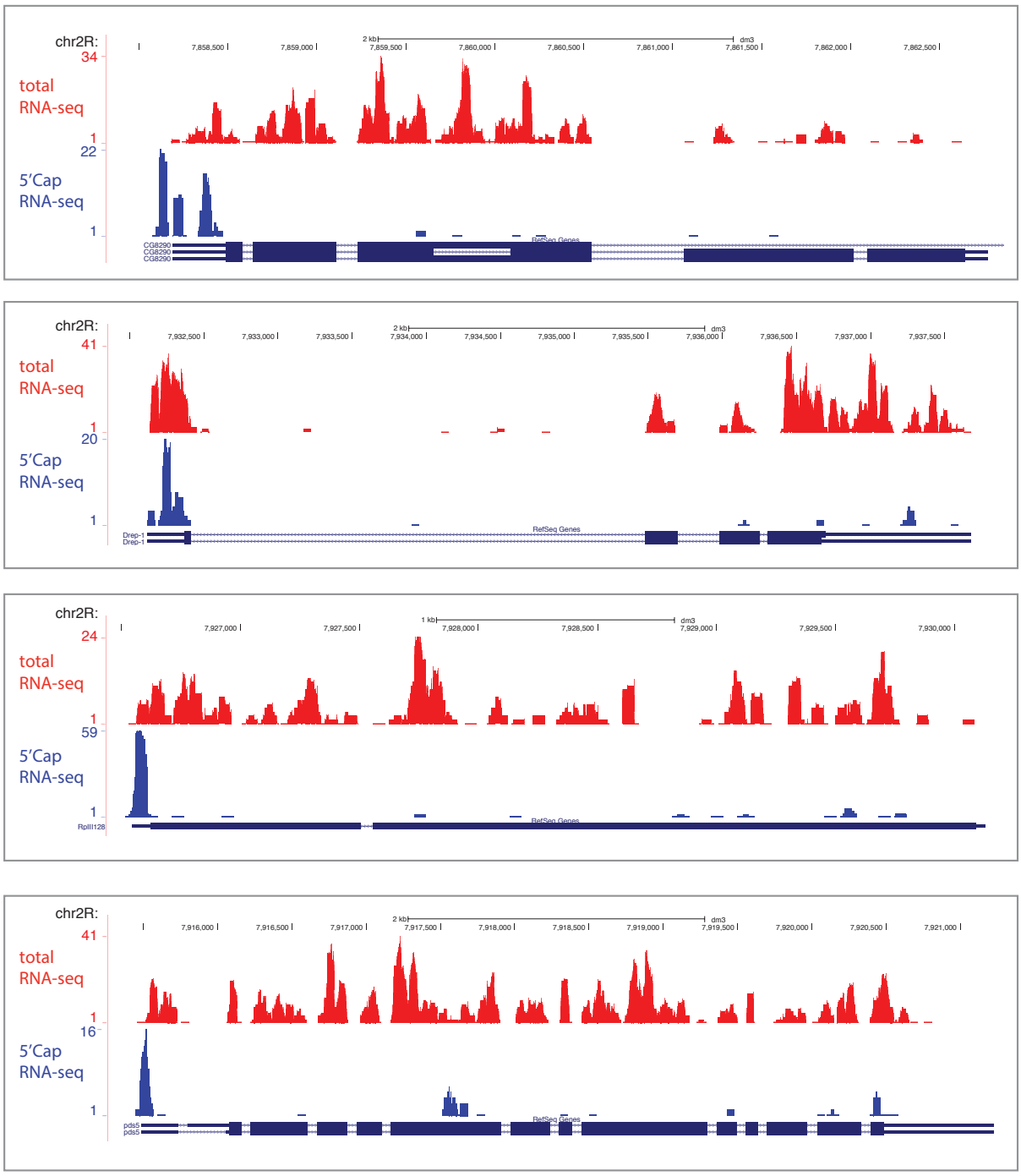

Figure 3.3: **5' Cap libraries detect start sites of mRNAs.** Upper tracks in red show uniquely mapped reads from a total RNA library. Blue tracks show uniquely mapped reads from the 5'Cap RNA rep.2 library. A clear enrichment is seen at transcription start sites of the example genes.

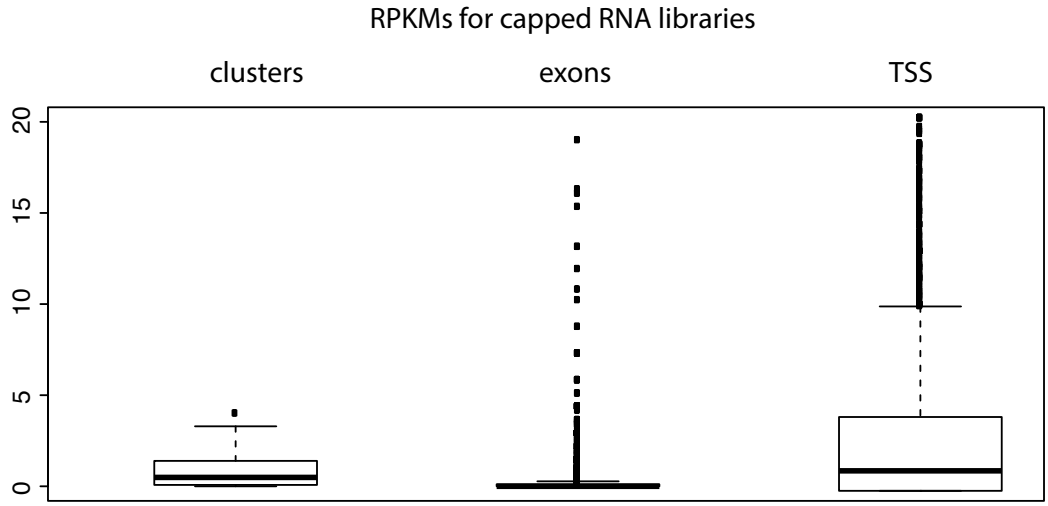

Figure 3.4: **5'Cap signal in three genomic features.** RPKMs calculated for the three genomic features piRNA clusters, exons and transcription start sites (TSS). Read numbers of the 5' cap RNA and a total RNA control library were counted per feature and then normalized for the total number of reads in the respective library and the length of the feature. RPKMs of the 5' capped RNA library was divided by RPKMs of the total RNA library. Boxplots were created in R. The range of the box comprises 25% to 75% of the values with the bar representing the median. Whiskers extend to 1.5 fold of the interquantile range and outliers are shown as dots.

3.1.5 Capped piRNA precursors from double stranded clusters originate from internal transcription initiation

The largest dual-stranded piRNA cluster 42AB is flanked by the gene Pld, which is transcribed in the same direction as the plus strand of the cluster. As seen in uni-stranded clusters, the signal for capped transcripts corresponds to a strong signal of PolII at the TSS of this neighboring gene. There is, however, no separate transcription start site for the cluster itself. Transcription initiated from Pld could explain the production of piRNAs from the plus strand. The next gene transcribed towards 42AB on the minus strand, however, is 139 kb downstream of the border of 42AB. Interestingly, 42AB shows signal for capped RNAs throughout the whole cluster that still has 50% of the intensity of the Pld gene.

The dual-stranded cluster 80EF is a particularly peculiar locus, since neither a prominent signal for capped RNAs, nor a peak for PolII can be seen at any side of this piRNA producing locus. The upper panel of figure 3.8 includes the two nearest genes, which are transcribed convergently into the direction of the cluster. The two genes alpha-Cat and CG32230 show a cap signal and a PolII signal, the gene nAcRalpha-80B, however, does not seem to be transcribed in ovaries. How the transcripts for piRNA production from

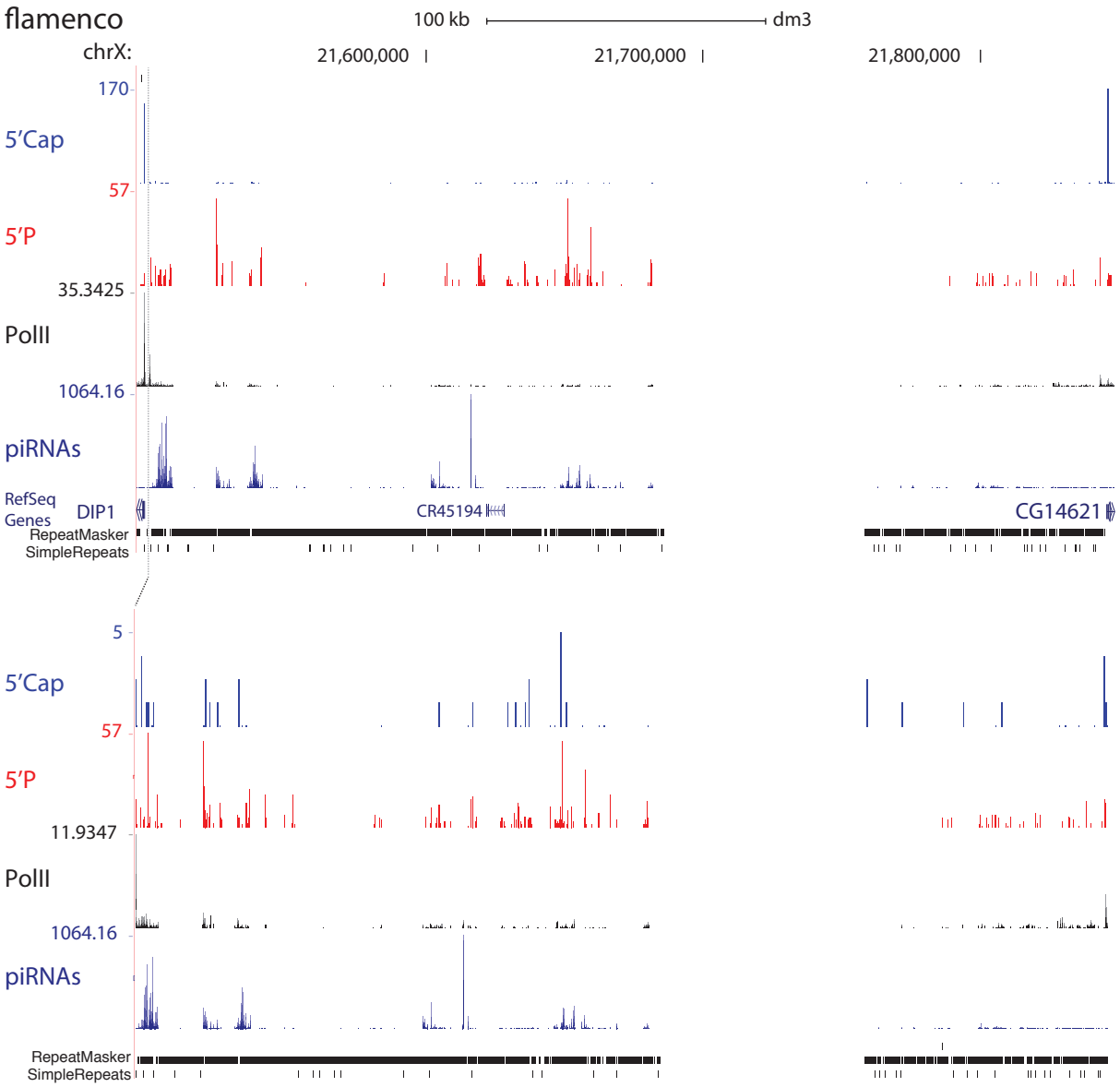

Figure 3.5: Shown are reads mapping to the flamenco cluster. 5' cap and 5' monophosphate tracks: unique reads mapping within the flamenco cluster coordinates. PolII ChIP track: signal intensity of Polymerase II ChIP, piRNA signal normalized to total library depth. (PolII and piRNA data from Adrien Le Thomas)

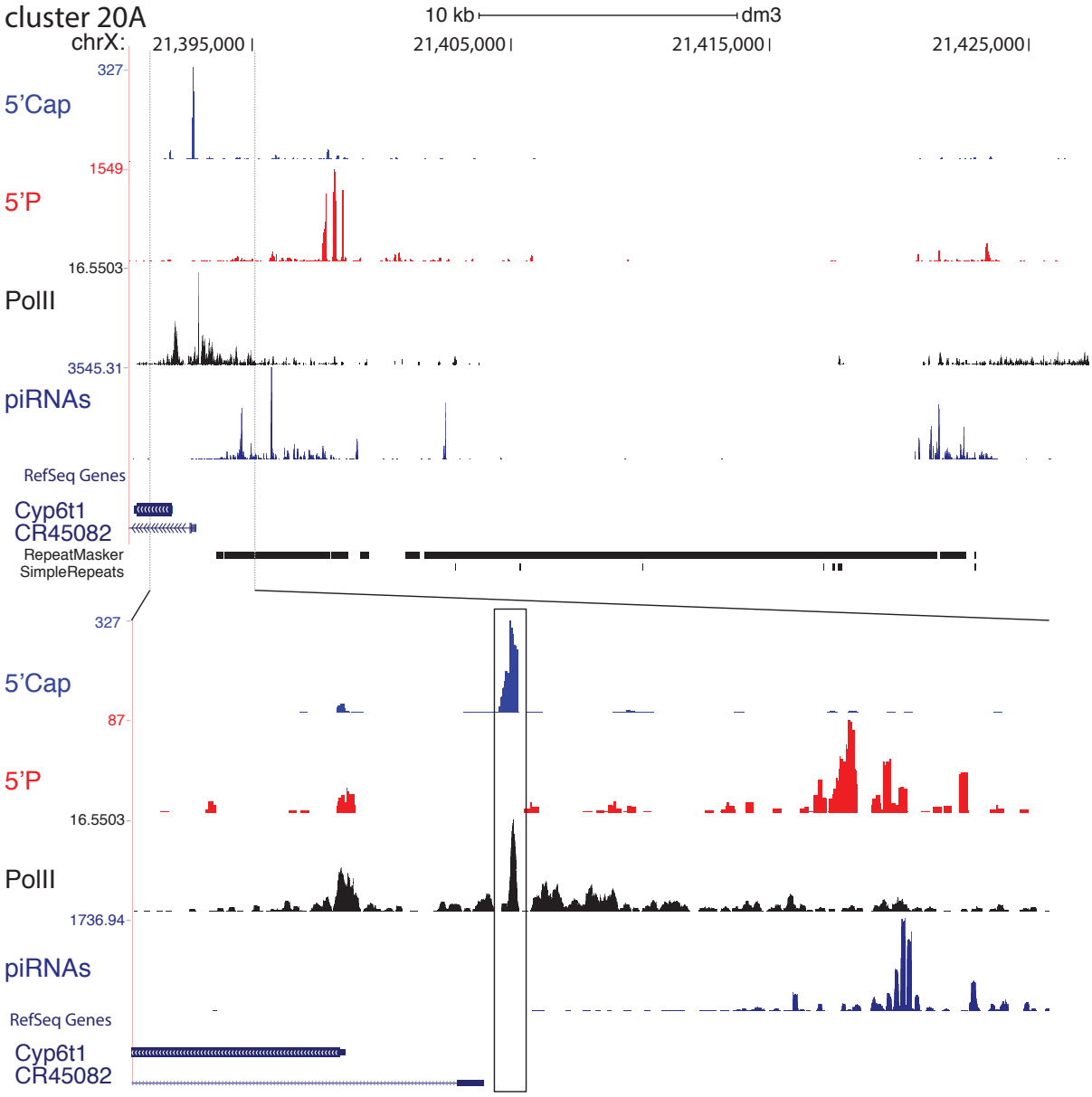

Figure 3.6: Shown are reads mapping to the uni-stranded cluster 20A, which is upstream of flamenco. 5' cap and 5' monophosphate tracks: unique reads mapping within the flamenco cluster coordinates. PolII ChIP track: signal intensity of Polymerase II ChIP, piRNA signal normalized to total library depth. The lower panel is a zoom into the putative promoter region. (PolII and piRNA data from Adrien Le Thomas)

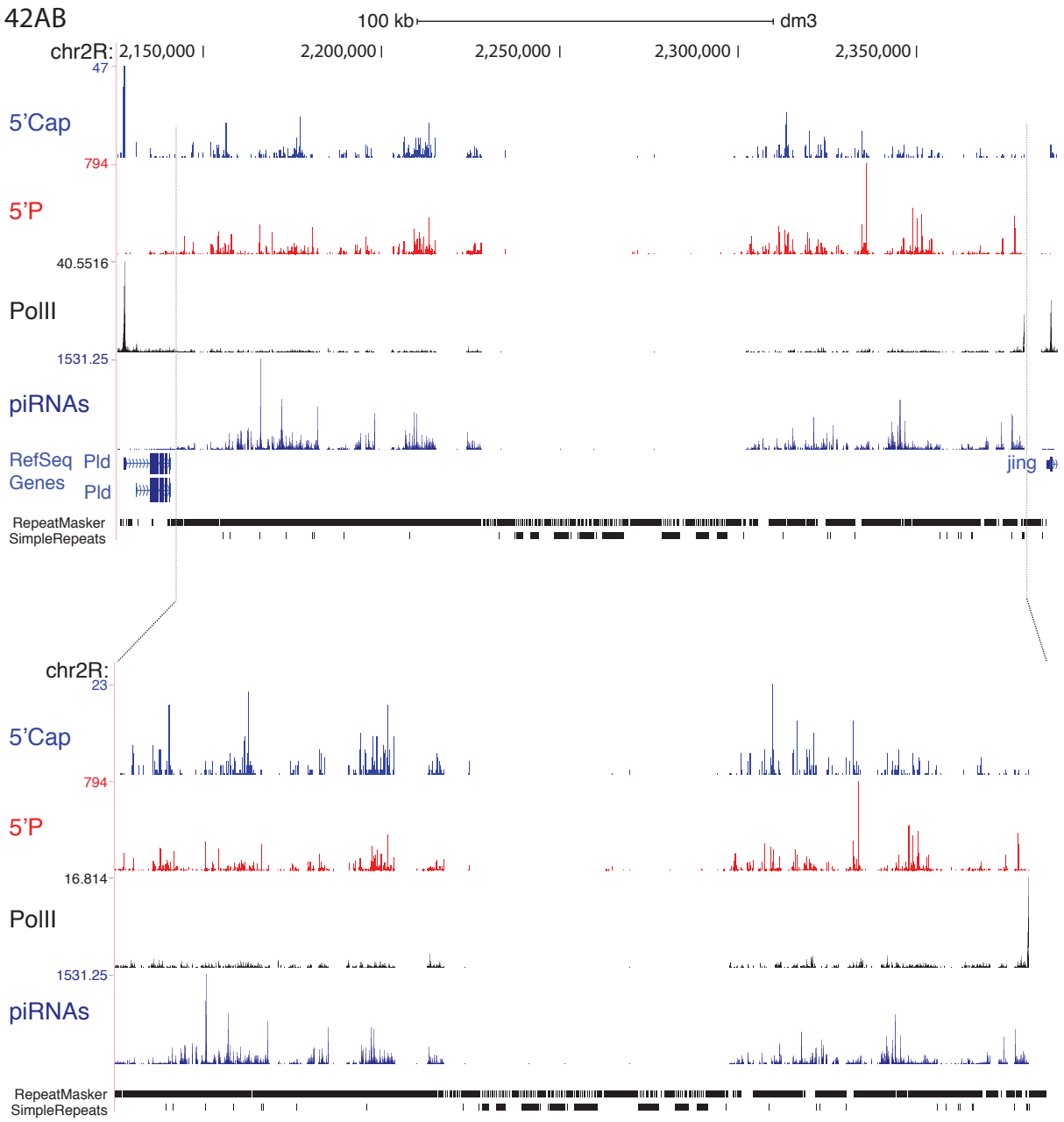

Figure 3.7: Shown are reads mapping to the double-stranded cluster 42AB on chromosome arm 2R. 5' cap and 5' monophosphate tracks: unique reads mapping within coordinates of cluster 42AB. PolII ChIP track: signal intensity of Polymerase II ChIP, piRNA signal normalized to total library depth. The upper panel contains signal from the neighbouring gene Pld, which is transcribed into the direction of the cluster. (PolII and piRNA data from Adrien Le Thomas)

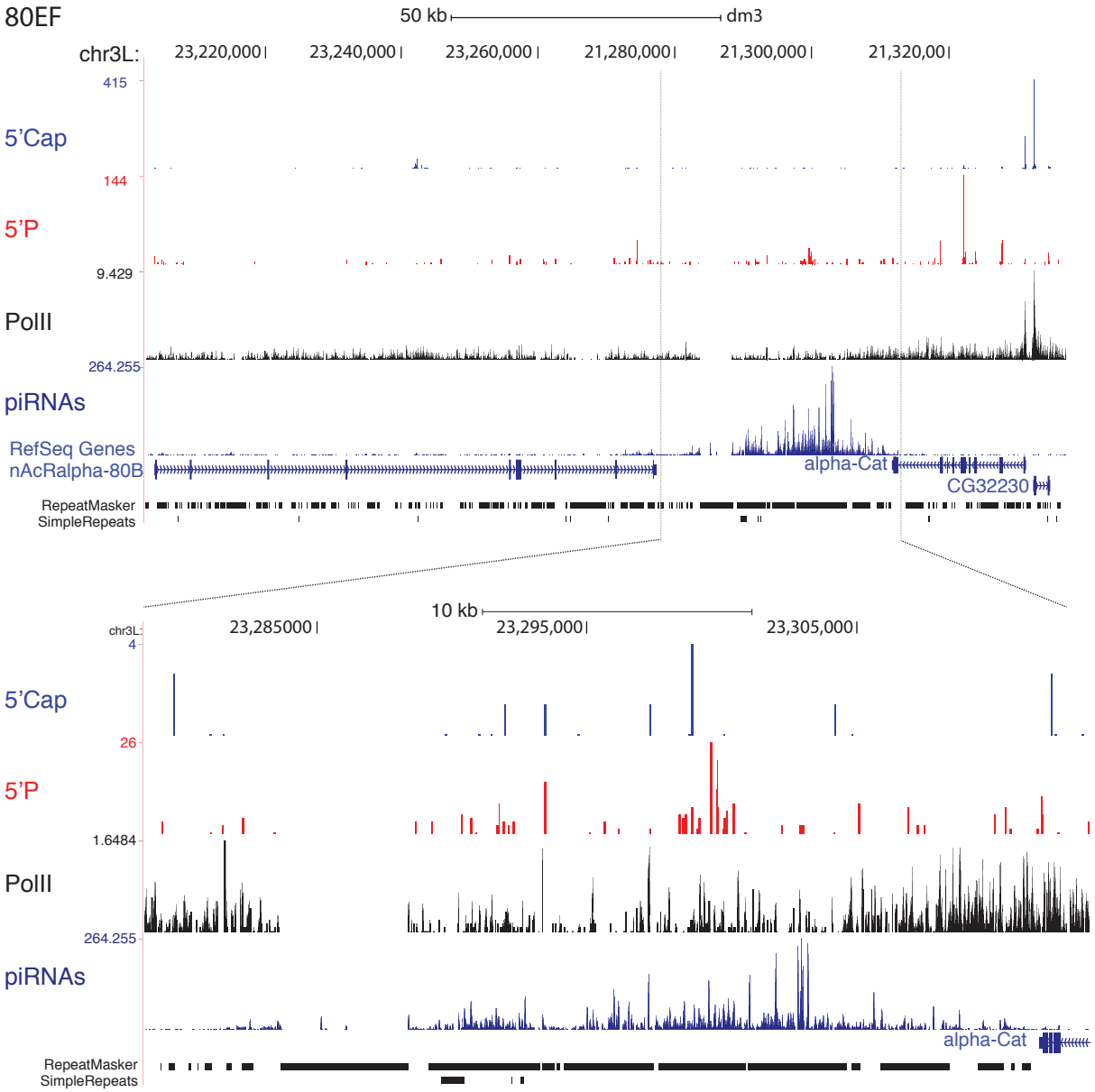

Figure 3.8: The upper panel shows a 100 kb window of chromosome 3L, which containes the cluster 80EF. The lower panel shows a zoom into the region of piRNA production. 5' cap and 5' monophosphate tracks: unique reads mapping within the flamenco cluster coordinates. PolII ChIP track: signal intensity of Polymerase II ChIP, piRNA signal normalized to total library depth. The two neighbouring genes nAcRalpha-80B and alpha-Cat are convergently transcribed into the piRNA producing region. (PolII and piRNA data from Adrien Le Thomas)

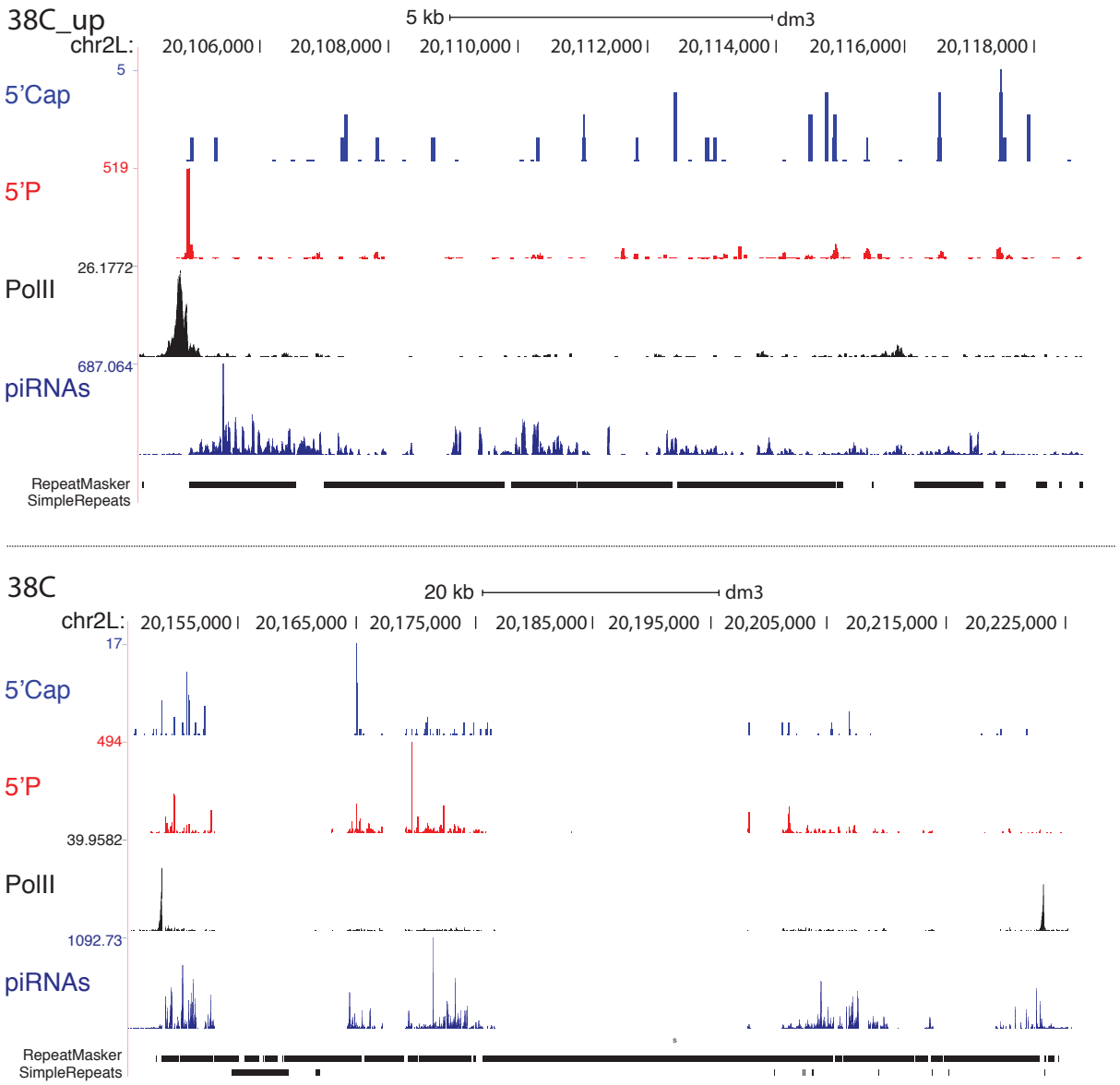

Figure 3.9: The upper panel shows reads mapping to the first piRNA cluster in the cytological band 38C of crhomosome 2L, the lower panel shows the second cluster in this chromosomal band. 5' cap and 5' monophosphate tracks: unique reads mapping within the flamenco cluster coordinates. PolII ChIP track: signal intensity of Polymerase II ChIP, piRNA signal normalized to total library depth. (PolII and piRNA data from Adrien Le Thomas)

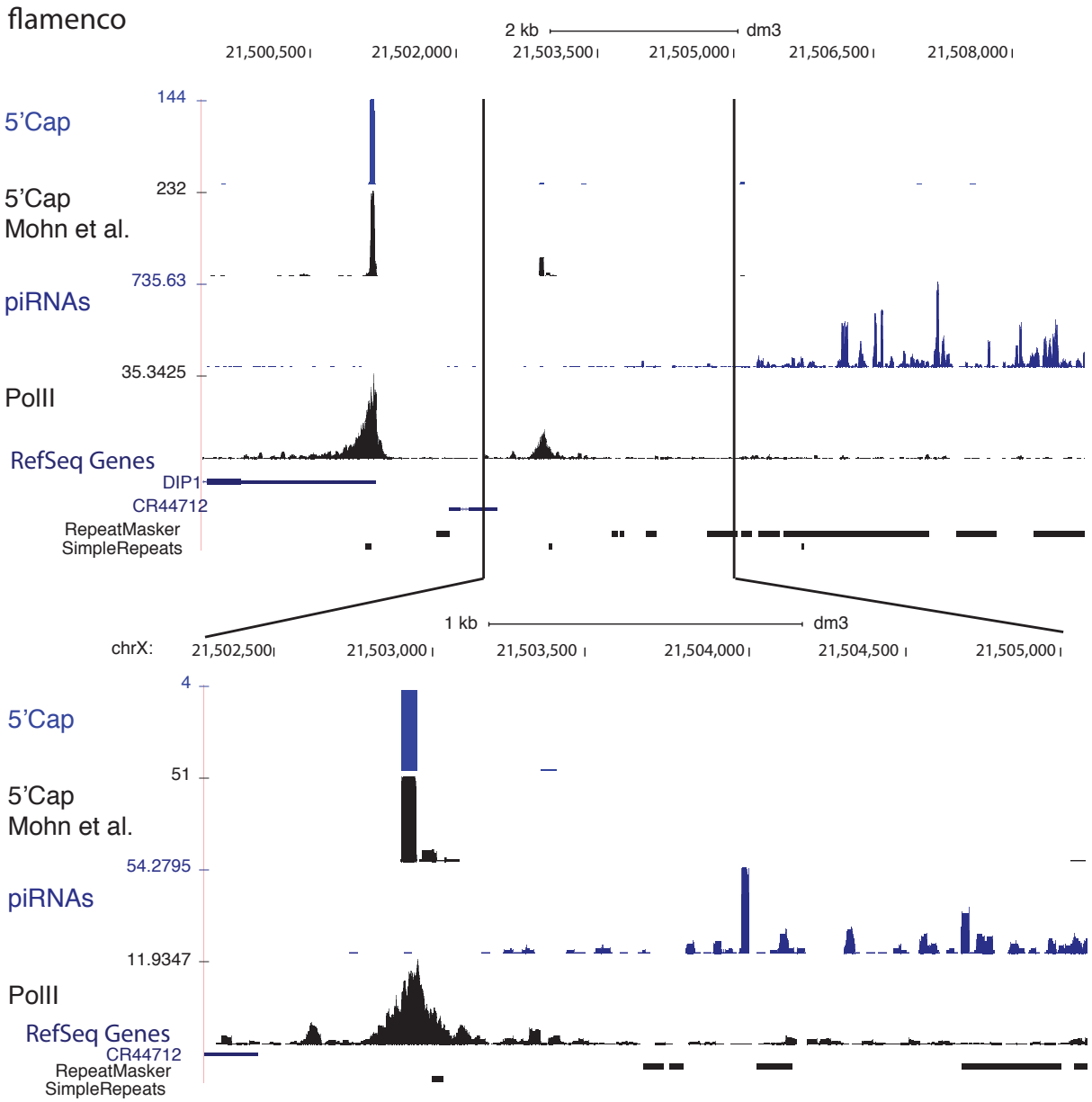

Figure 3.10: UCSC genome browser tracks for 5' capped RNA sequencing. The first track (blue) shows my data set, the second track (black) shows data from (Mohn et al., 2014) (GEO accession GSE55824, experiment GSM1346542, dataset SRX485201).

this locus are produced, can not be explained by the available datasets.

Based on the PolII ChIP dataset, cluster 38C was divided into two parts. The upper part (38C.1) has a clear PolII peak on one side, but no signal for capped RNA. 38C.2 has one PolII promoter on either side, however, there is no sharp signal for capped transcripts while some reads map within the cluster.

Overall, the 5' cap RNA sequencing experiments reveal that transcripts form single-stranded clusters indeed originate from defined transcription start sites and contain a 5' m^7G-cap structure. They resemble canonical mRNAs transcribed by Polymerase II. Double-stranded clusters also seem to give rise to 5' m^7G-capped RNAs, their exact start sites, however, could not be defined.

3.1.6 piRNA precursors are 5' mono-phosphorylated

The aim of the following experiments was to establish if intermediate piRNA precursor molecules exist. To this end, I cloned and deep-sequenced 5' monophosphate containing RNAs from *Drosophila* ovaries.

Maturation of piRNA precursors requires a number of cleavage events. At the beginning of the project very little was known about initial processing steps. It was clear that piRNAs are transcribed from long intergenic clusters and that mature piRNAs have mono-phosphorylated 5' termini. It was also known that piRNA production and maturation takes place in the cytoplasmic nuage. Intermediate forms of piRNA precursors, however, had not been described. Detecting intermediate long piRNA precursor molecules with a 5' end would establish that such precursor molecules actually exist. Therefore, 5' monophosphorylated RNAs from *Drosophila* ovaries were cloned and sequenced.

3.1.7 Approach for piRNA precursor detection

The first step in most conventional RNA library cloning protocols is the conversion of the RNA into cDNA, which subsequently is ligated to DNA adapters on both ends. Since this cloning approach detects RNAs irrespective of their 5' ends, I modified and adapted several cloning approaches to establish an RNA cloning method, which specifically detects 5' mono-phosphorylated RNA molecules.

Briefly, selection of 5' mono-phosphorylated RNAs takes place in the very first step, where a ssRNA adapter is ligated to the 5' end of the RNA, using an RNA ligase1. This enzyme ligates a 5' monophosphorylated nucleic acid to the 3' hydroxylated end of an acceptor nucleic acid. 5' capped, di- or tri-phoshphorylated, or hydroxylated RNAs will

not be ligated. Subsequently, the 3' adapter is introduced through reverse transcription with a DNA oligo containing a random hexamer on its 5' half and the adapter sequence in its 3' half. Since the goal was to detect piRNA precursor molecules, which should be longer than mature piRNAs, the first libraries were cut from an agarose gel after final PCR amplification to yield an insert size of 70-220 nucleotides. Sequencing of those first libraries showed that 96% of all reads are derived from ribosomal RNAs. Theoretically the 5' ends of rRNAs contain a tri-phosphate 5' terminus. They are, however, highly abundant and thus represent an overwhelming majority of RNAs cloned with this method. In order to deplete the RNA samples of those contaminating seuqences, a commercial rRNA depletion kit (Epicentre) was used in further experiments. To date there is no rRNA depletion kit available, which would be optimized for use with *Drosophila* material. Therefore, the commercial kit was supplemented with custom made oligo-probes in order to increase depletion of *Drosophila* specific rRNAs. Besides rRNA derived reads, a second large gorup of over-repressented sequences was apparent, which were nucleolar snoRNAs and spliceosomal U-RNAs. Since most snoRNAs and U-RNAs are smaller than 200 nt and since the aim of this study was to investigate long piRNA precursor molecules, those contaminants were removed from the sample by size selecting the initial material for RNAs longer than 200 nt. A third proportion of the sequenced libraries did not map to the genome, since they contained concatemers of very abundant short RNAs, like the 31 nt 2S rRNA. This problem was resolved after size selection of the initial RNA sample. As an internal control for the cloning procedure RNA samples were spiked with 2 heterologous RNAs. Those were synthesized in vitro through T7 transcription and their 5' ends were modified to contain either a mono-phosphate or a hydroxyl group. The two RNA spikes were added in equimolar amounts and to a final proportion of 1.6% of the total starting material. The 5' monophosphate RNA served as a positive control for the method, the 5' hydrohylated RNA should not be picked up with the cloning procedure. Mapping of the sequenced libraries to the respective spike sequences confirmed that the established protocol indeed predominantly clones the 5' monophosphorylated spike and does not unspecifically pick up signal from the 5' hydroxylated spike (see figure 3.11)

3.1.8 piRNA precursors are 5' mono-phosphorylated long RNAs

After establishing a protocol for cloning and sequencing 5' monophosphorylated RNAs, sequences were mapped to the genome using Bowtie with the flags v0 and m1, only allowing uniquely mapped reads with zero mismatches. Details of the sequencing statistics of two replicates are shown in table 3.3 and the gene annotations after mapping are shown in the graph of figure 3.12.

About 30% of reads mapped to rRNA, and roughly 45% mapped to the genome. I

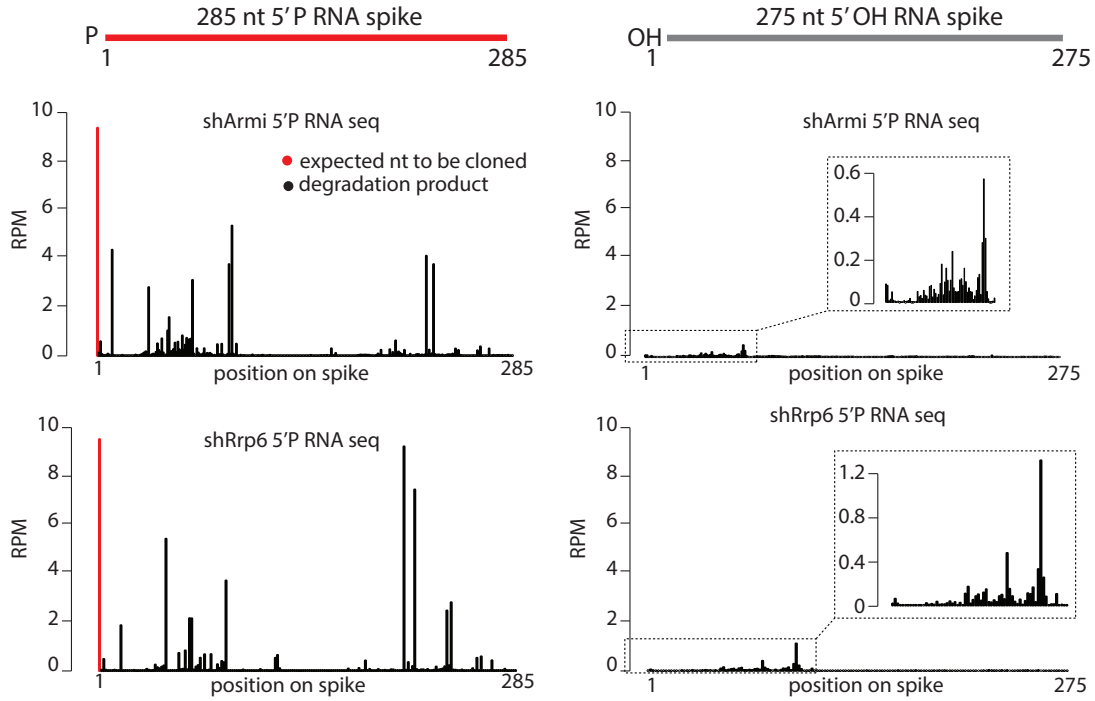

Figure 3.11: **The established RNA cloning protocol detects the 5' monophosphate spike and not the 5' hydroxylated spike.** Both RNA spikes were transcribed in vitro and their sequences do not map to the *Drosophila* genome. Samples were spiked with equimolar amounts of both spikes. Shown are reads normalized to total of mapped reads in the library (RPM) in two examplary libraries. Mapping of sequenced reads to the spike sequences shows a strong signal from the first nucleotide of the 5' monophosphorylated spike (left panel), but not from the first position of the 5' hydroxylated spike (right panel). Signal from within the spike sequence is likely to reflect degradation products containing a 5'P end. Since the experimental procedure to obtain the spike material contained one more enzymatic reaction and subsequent gel-purification for the 5' mono-phosphorylated spike molecules, degradation products are more abundant for this sequence.

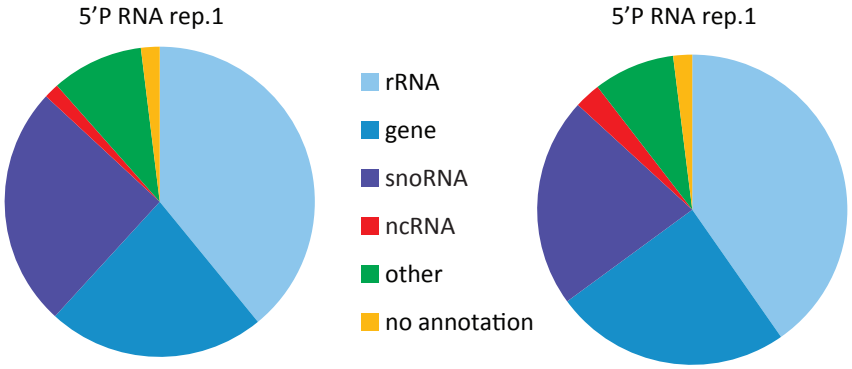

Figure 3.12: **Gene ontology analysis for 5' monophosphate RNA libraries.** Approximately 30% of all mapped reads alingn to rRNA sequences despite depletion for those molecules with a commercial kit. Roughly 25% of reads are annotated as genes, and another 25% as snoRNAs. In this gene ontology analysis piRNA clusters fall within the category "no annotation"

established if piRNA cluster reads are above signal from TSS or exons, in analogy to the analysis on 5' capped RNAs. RPKMs normalized to a total RNA library show that the median signal arising from piRNA clusters is dramatically higher than from exons or transcription start sites (TSS) (see figure 3.13). This confirms that piRNA precursor molecules are 5' monophosphorylated.

Table 3.3: Sequencing statistics of two 5'P RNA replicates

	5'P rep1	5'P rep2
Total number of reads	11,455,382	488,351
reads mapping to rRNA	3,426,443	148,395
reads not mapping to genome	2,989,038	131,898
reads mapping to genome	**5,039,901**	**208,056**

Figure 3.13 shows that despite the small percental contribution of piRNA cluster sequences in the 5'mono-phosphate libraries (see categorie "no annotation" in figure 3.12), piRNA clusters show a strong signal for those RNA species, when normalized to expression level. We therefore concluded that those RNA molecules are *bona fide* piRNA precursors. Figures 3.5, 3.6, 3.7, 3.9 and 3.8 show the distribution of piRNA precursors within clusters.

To uncover whether the piRNA precursor molecules already posses the piRNA signature of a Uridine at their first position, the nucleotide distribution over the 50 nt read lenght was blotted for unique sequences mapping to single-stranded piRNA clusters and double-stranded piRNA clusters. As a control the nucleotide distribution of all unique sequences was blotted.

Figure 3.14 shows the distribution of nucleotides along the piRNA precursors cloned in this experiment. A strong bias for Uridine is apparent in sequences from single- as well

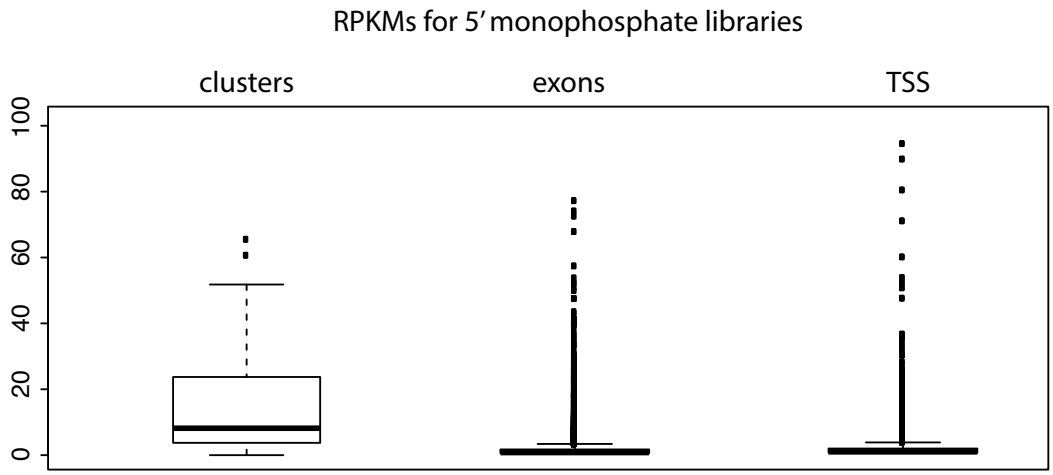

Figure 3.13: **5'P RNA signal in three genomic features.** piRNA clusters show stronger sigal for 5' mono-phosphorylated RNAs than transcription start sites (TSS) or exons. RPKMs of the 5'P library was normalized to a total RNA library. Boxplots were created in R. The range of the box comprises 25% to 75% of the values with the bar representing the median. Whiskers extend to 1.5 fold of the interquantile range and outliers are shown as dots.

as from double-stranded clusters. The double-stranded cluster reads show an additional bias for Adenosine at position 10. If all unique sequences of this library are investigated, no nucleotide bias is apparent.

Overall, those results show that piRNA precursors are indeed processed into long (larger than 200 nt) intermediates that already contain a 5' monophosphate terminus. Those precursors furthermore show the characteristic nucleotide biases of mature piRNAs: a 1U and a 10A enrichment.

3.1.9 Identification of primary piRNA biogenesis factors involved in first cleavage events

The detection of piRNA precursors enabled me to perform a small-scale screen for factors that are involved in their generation. Therefore, I knocked down candidate factors in *Drosophila melanogaster* ovaries and investigated the effect on piRNA precursors. After establishing an effect on steady state levels of piRNA cluster transcripts and transposable elements, I cloned and deep-sequenced 5' monophosphate piRNA precursors.

The candidate genes were knocked down through tissue specific expression of an RNA hairpin targeting the gene of interest with the bipartite UASP-Gal4 system. The hairpin is under the control of an Upstream Activating Seuqence (UAS) element, for which Gal4

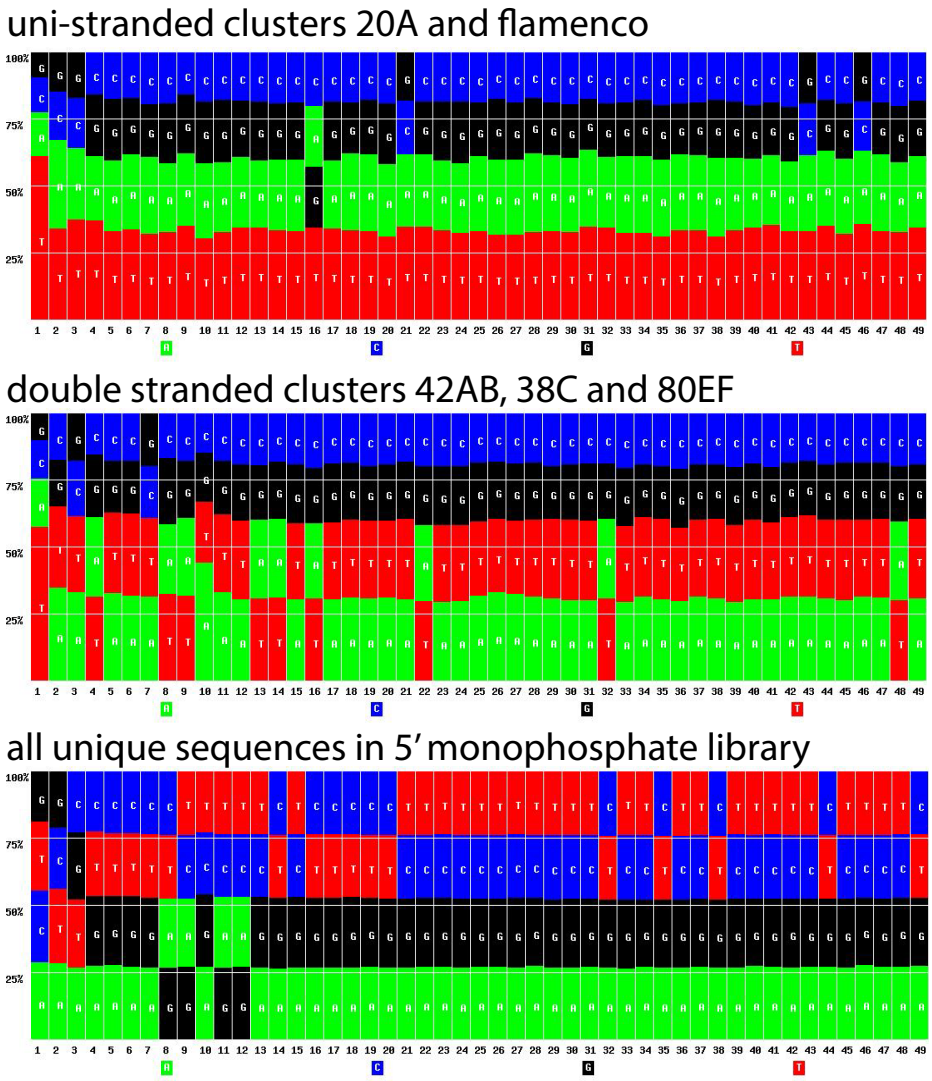

Figure 3.14: **Nucelotide distribution in piRNA precursors from wild type *Drosophila*.** Shown are the percentual distributions of each nucleotide along the 50 nt sequences for unique sequences mapping to either the two single stranded piRNA clusters flamenco and 20A (total of 1,223 sequences) or the double stranded clusters 42AB, 38C and 80EF (total of 3,682 sequences). The nucleotide distribution of all sequences of this library is shown in the lowest panel (74,770 sequences). This graphic was obtained from the database server of Ravi Sachidanandam

acts as a transcriptional activator. Crossing a fly containing the UAS-hairpin contruct to a fly, expressing Gal4 under the control of a tissue specific promoter, the progeny of this cross will express the hairpin in the desired tissue. In the case of this screen, expression of Gal4 was driven by the germline specific nanos promoter.

At the time of designing the experiments, 14 factors involved in primary piRNA biogenesis were described in two publications (Olivieri et al., 2010; Haase et al., 2010) and I selected some of those identified factors, as well as some new candidates, such as endonucleases or factors known to be involved in transcription termination. Since piRNA clusters are up to 100 kb long, one hypothesis was that they are not transcribed into one continuous, capped precursor, but rather that the first cleavage events could already take place co-transcriptionally. Thus, the nuclear factors were of special interest, since they would be candidates for such a co-transcriptional processing.

In the first step, the expression levels of transposable elements and piRNA cluster transcripts were established using quantitative PCR (qPCR). Accumulation of piRNA precursors and simultaneous upregulation of transposable elements would suggest an impairment in piRNA precursor processing.

The expression level of piRNA cluster transcripts was normalized to a control knockdown and the effects of the components was categorized into 5 levels: up, slightly up, no change, down and slightly down. The classification "no data" applied to samples, where the qPCR reaction failed, or replicas showed contradictory results (figure 3.15).

Previous screens utilized the upregulation of transposable elements to identify factors involved in the piRNA pathway. While many factors were identified that way, the drawback of this method is the positive scoring of any component, which plays a very general role in ovary metabolism. For example, some export factors were included in this first part of the screen. One of them, NXT1, showed a promising accumulation of piRNA precursor molecules and an upregulation of transposable elements. NXT1, however, was described as a general essential component of mRNA export in *Drosophila* ovary tissue (Caporilli et al., 2013). Thus, the effect on transposable elements and piRNA cluster transcripts could be a secondary effect caused by impairment of the export of an essential factor's mRNA.

Overall, most factors showed expected results based on published literature, while others were inconsistent with published data. Zucchini, for example, only scored in cluster 20A with an upregulation. Other cluster transcripts seemed decreased, wich contradicted the results of Haase and colleagues (Haase et al., 2010). Based on the qPCR results and literature, 7 genes were further selected for the second part of this screen, which aimed at quantifying piRNA precursors in those sh knockdown flies.

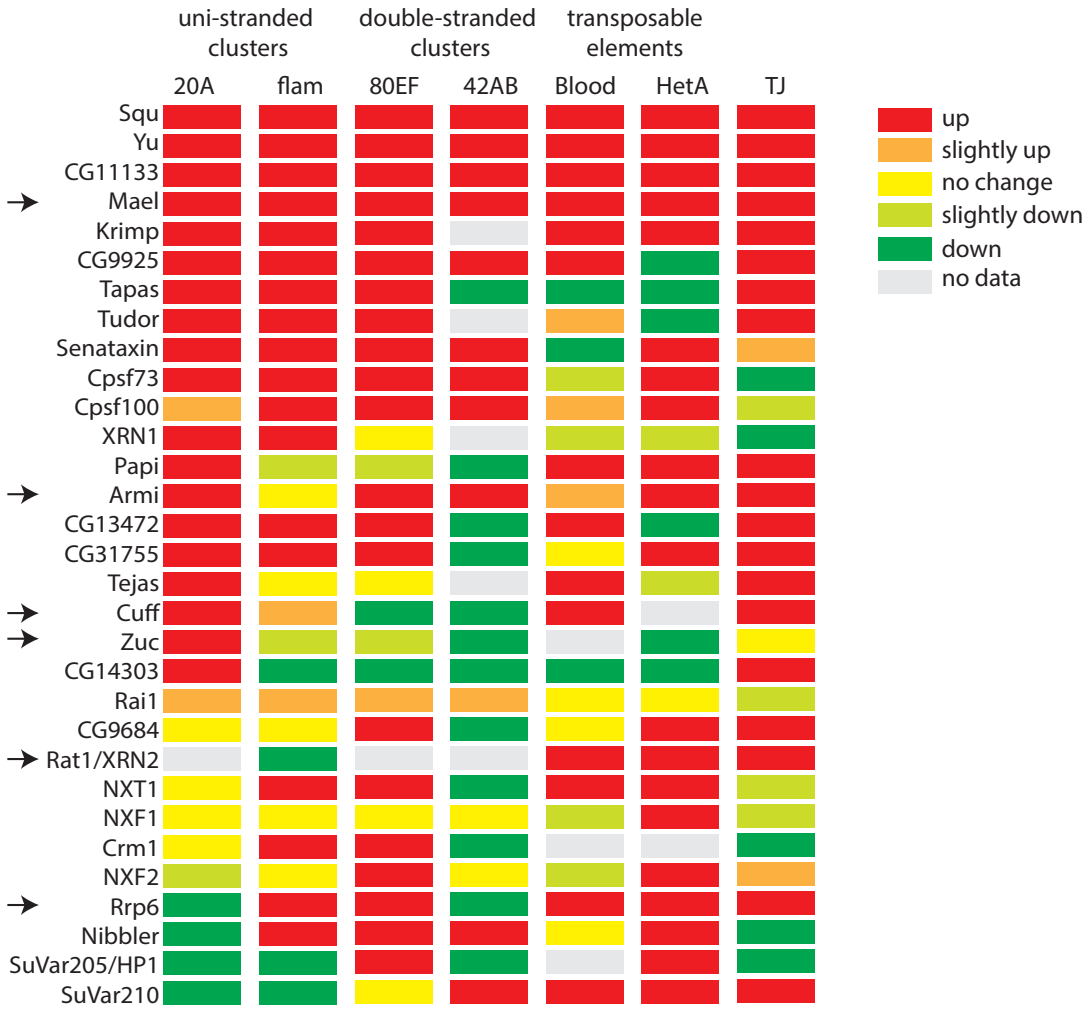

Figure 3.15: **Cluster and TE expression in *Drosophila* sh knockdown lines.** Expression levels of piRNA clusters and transposable elements in sh knockdown lines, measured by qPCR and normalized to shWhite control knockdown. The changes in transcript were categorized into the categories up (>3 fold up), slightly up (between 1.1 and 2.9 fold up), no change, slightly down (between 1.1 and 2.9 fold down) and down (>3 fold down). For "no data" the qPCR reactions failed or did not give conclusive results. sh lines with an arrow were further investigated with RNA cloning.

- Maelstrom is a nuclear piRNA factor and has been shown to be involved in the Piwi-induced silencing of targets (Sienski et al., 2012).
- Armi is a putative helicase and is implied in binding to piRNA intermediates (Saito et al., 2010) and maturation of RISC (Tomari et al., 2004).
- Cuff is a nuclear piRNA biogenesis factor (Chen et al., 2007; Pane et al., 2011).
- Zucchini is the candidate endonuclease for cleavage of precursor transcripts (Voigt et al., 2012; Nishimasu et al., 2012; Ipsaro et al., 2012)
- Rrp6 is a 3' - 5' exonuclease and a component of the nuclear exosome complex (Chlebowski et al., 2013).
- Rat1 is a nuclear 5' - 3' exonuclease involved in transcription termination (Dengl and Cramer, 2009).
- Rhino was added to the next step of this experiment, since it was shown to localize to piRNA clusters (Klattenhoff et al., 2009; Pane et al., 2011)

5' monophosphorylated RNA as well as total RNA libraries were cloned and sequenced from total ovary RNA of the above knockdown lines and one control. The expectation was that if a component is involved in first cleavage steps, its knockdown would result in the accumulation of total RNA with a simultaneous decrease in 5' monophosphorylated piRNA precursors.

Since a decreased level of mono-phosphorylated piRNA precursors could not be only caused by impaired processing, but also result from decreased transcription of piRNA clusters, the signals from 5' mono-phosphate libraries were normalized to total RNA levels.

RPKMs of the total RNA libraries can be seen in figure 3.16 and figure 3.17 shows the normalized 5' monophosphate data. None of the investigated components showed a very strong effect of total RNA accumulation and simultaneous decrease in precursor molecules.

3.1.10 The investigated factors affect precursor levels, but not the nucleotide bias

For analysis of the RNA sequencing data, reads mapping within the piRNA cluster coordinates were counted in the 5'P RNA and in the total RNA libraries. In order to account for differences in library depth the RPM was calculated by dividing numbers of

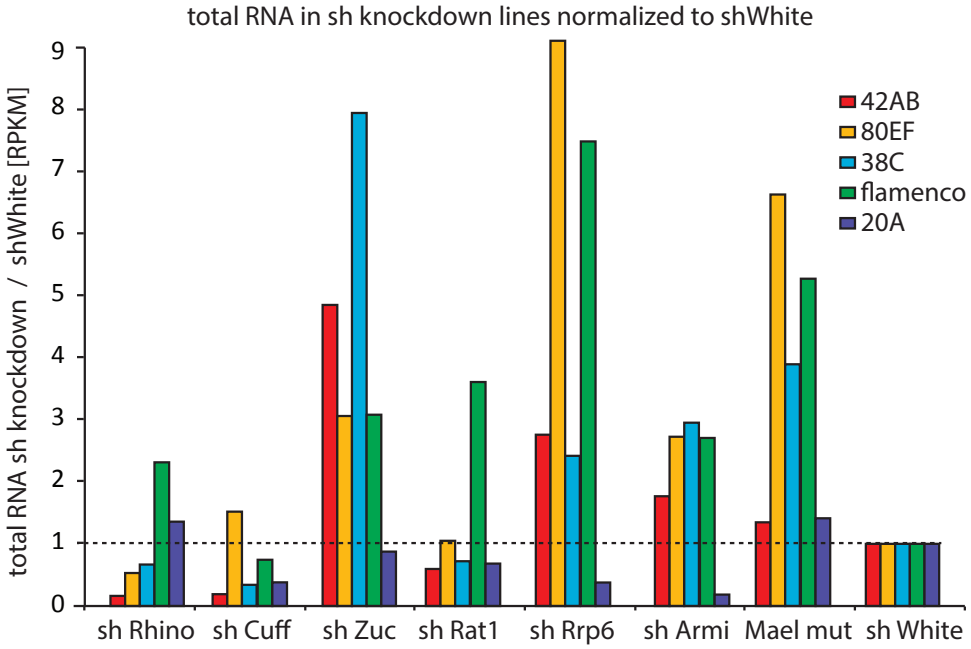

Figure 3.16: **Total RNA sequencing in sh knockdown lines.** Total RNA sequencing in the sh knockdown lines establishes the steady state levels of piRNA cluster transcripts. RPKMs were calculated by counting the number of mapped reads to the respective cluster and dividing by the total length of the cluster (in kb) and total number of mapped reads of the respective library (in million). Values for the different knockdown lines were normalized to the shWhite control through division.

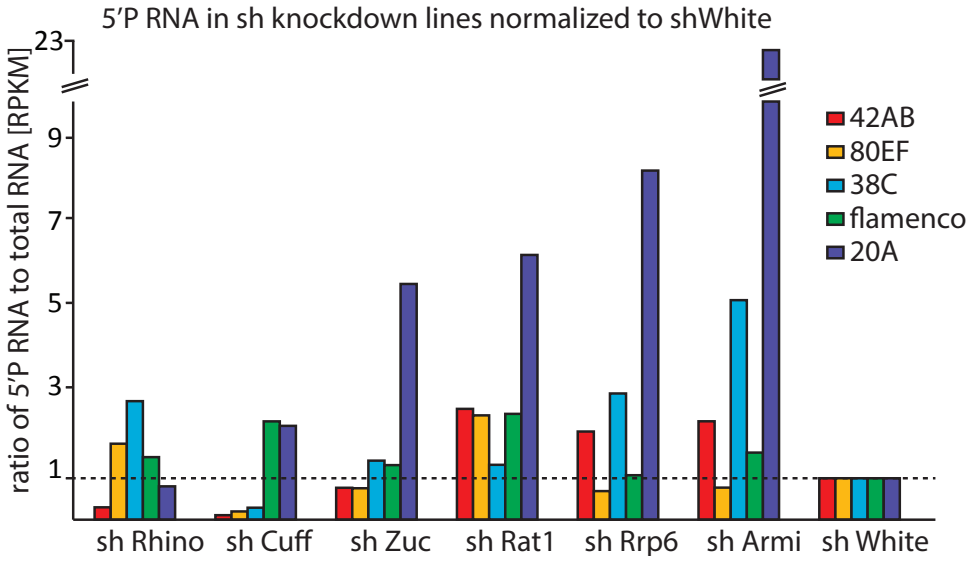

Figure 3.17: **piRNA precursor levels in sh knockdown libraries.** RPKMs were calculated by counting the number of mapped reads to the respective cluster and dividing by the total length of the cluster (in kb) and total number of mapped reads of the respective library (in million). Values for the different knockdown lines were normalized to the shWhite control and to total RNA levels.

reads uniquely mapping to the respective piRNA cluster by the number of total uniquely mapped reads in the library.

Figure 3.17 shows the reads of 5' monophosphate precursors normalized to expression level and a control knockdown.

Zucchini is a strong candidate for being the endonuclease performing the first cleavage step in piRNA biogenesis (Voigt et al., 2012; Nishimasu et al., 2012; Ipsaro et al., 2012; Olivieri et al., 2010; Muerdter et al., 2013; Handler et al., 2013; Czech et al., 2013) and qPCR data by Haase and colleagues (Haase et al., 2010) showed an accumulation of piRNA cluster transcripts. Indeed, total RNA sequencing confirmed this result: with the exception of transcripts from cluster 20A, all other piRNA cluster transcripts are elevated between 3 and 8 fold over levels in the shWhite control (figure 3.16). If the accumulation of piRNA cluster transcripts was due to the lack of an endonucleolytic processing through Zucchini, we expected to see a decrease of 5' monophosphate piRNA precursors. The 5' monophosphate RNA sequencing data, however does not show this expected result. The levels of processed precursor molecules are only slightly below the levels in the control for clusters 42AB and 80EF and are moderately to strongly increased in 38C, Flamenco and X-upstream.

To investigate the effect of the candidate proteins on the 1U bias of piRNA precursors, first nucleotides were counted in all reads mapping to piRNA clusters (figure 3.18). Overall, the 1 U nucleotide bias in piRNA precursors was around 50% and therefore weaker

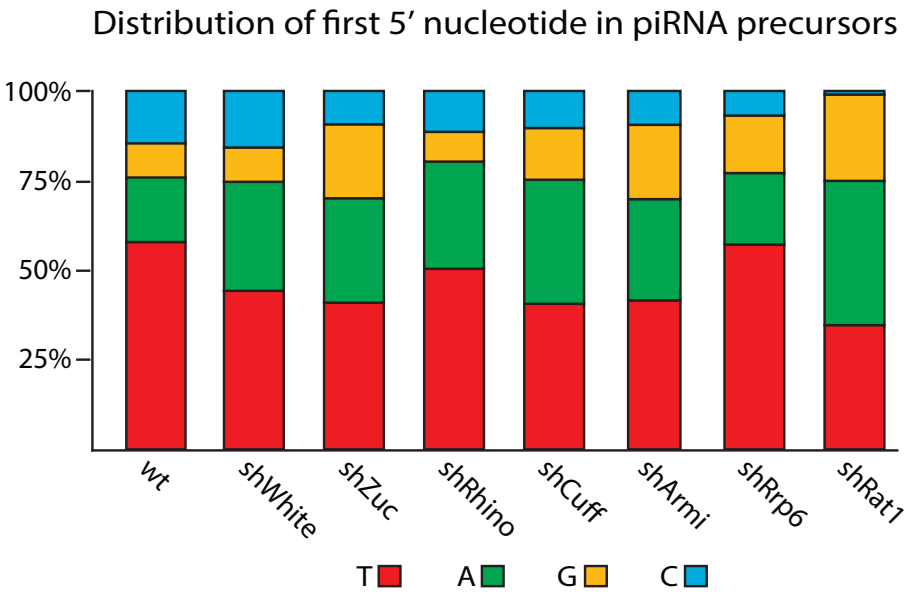

Figure 3.18: Distribution of the first nucleotide for all reads mapping to piRNA clusters

than the observed bias in piRNAs of approximately 80%. The wild type precursors show a slightly higher 1U bias than the shWhite library. This difference could be a measure of experimental variation of detection of the 1U bias. None of the investigated components shows a convincing change of the 1U bias.

Taken together, the effects of the investigated factors on piRNA precursors were not as strong as expected. No component showed an effect that would be convincing enough to identify it as a main (direct or indirect) factor for piRNA precursor biogenesis. The 5' monophosphate RNA deep sequencing data for Cutoff and Rat, however, were of great value for establishing a model for Cutoff's role in piRNA cluster transcription (see section 3.3.1.2).

3.1.11 Efforts towards the elucidation of piRNA precursors to be co-transcriptionally processed

Investigations in our laboratory on a *Drosophila* strain with an artificial sequence inserted into a piRNA cluster lead to a very interesting observation: while most genic piRNAs exclusively map to exons, piRNAs derived from the artificial sequence yielded some piRNAs mapping to intronic sense sequeneces (Muerdter et al., 2013). Since piRNA biogenesis factors localize to the cytoplasmic nuage, it was until then assumed that piRNA precursors are - just like genic PolII transcripts - fully spliced and then exported into the cytoplasm, where they are processed. The appearance of intronic piRNAs in some cases,

however, could have two reasons. Either, piRNA precursors are transcribed while RNA processing signals are ignored, leading to the retention of intronic seuqences within the precursor. Or there is a co-transcriptional processing step, which acts upstream of the splicing machinery, leading to precursor molecules that are already partially processed when exported into the cytoplasm. A recent report suggested that, according to the first hypothesis, splicing is impaired in piRNA cluster transcripts (Zhang et al., 2014).

At the starting point of the project, however, the second hyphothesis seemed equally likely. To test the second hypothesis, I performed a pilot experiment, where I isolated nuclear RNAs and performed 5' RACE for 5' monophosphorylated transcripts. In this preliminary experiment the nuclear fraction contained already cleaved piRNA cluster transcripts, which exhibited a slight 1U bias. To exclude that the observed intermediate piRNA precursors result from a contamination of the nuclear preparation with cytoplasmic RNA, I sought to employ a very stringent isolation of nascent transcripts and sequence 5' monophosphorylated RNAs. If nascent transcripts would already show signatures of processed piRNA precursors, this would ultimately demonstrate a co-transcriptional piRNA processing step. One challenge in nascent RNA isolation from *Drosophila* ovaries was to establish a protocol, which would yield enough material for library construction and at the same time be devoid of non-nascent RNA contaminations. The NET-seq method by Churchman and colleagues, for example, uses yeast cells in very high quantities, which are not obtainable from *Drosophila* ovaries (Churchman and Weissman, 2012). Other procedures require formaldehyde crosslinking followed by sonication to preserve DNA - RNA-Polymerase - nascent RNA interatcions during stringent immunoprecipitation conditions (Bittencourt and Auboeuf, 2012). Sonication, however potentially shears RNA and makes the detection of real 5' ends of nascent transcripts impossible.

I decided to employ a method that was first described by Wuarin and Schiebler to isolate nascent RNA transcripts from mouse liver cells (Wuarin and Schibler, 1994) and later modified by Khodor and colleagues for nascent transcript isolation from S2 cells (Khodor et al., 2011). I adopted this method for *Drosophila* ovaries. Briefly, nuclei were isolated and then fractionated into nucleoplasmic supernatant and a chromatin pellet. The chromatin pellet contained the ternary transcription complex consisting of template DNA, RNA polymerase and the nascent RNA transcript. The nascent transcripts were isolated through Phenol-Chloroform extraction.

The quality of the presumably nascent transcripts was assesed with two different assays (see figure 3.19).

The first "qPCR assay" relied on the assumption that the sample containing nascent transcripts should be enriched in the 5' portion of RNAs compared to the total RNA sample, containing mature mRNAs. Primer pairs for qPCR were designed against the 5'

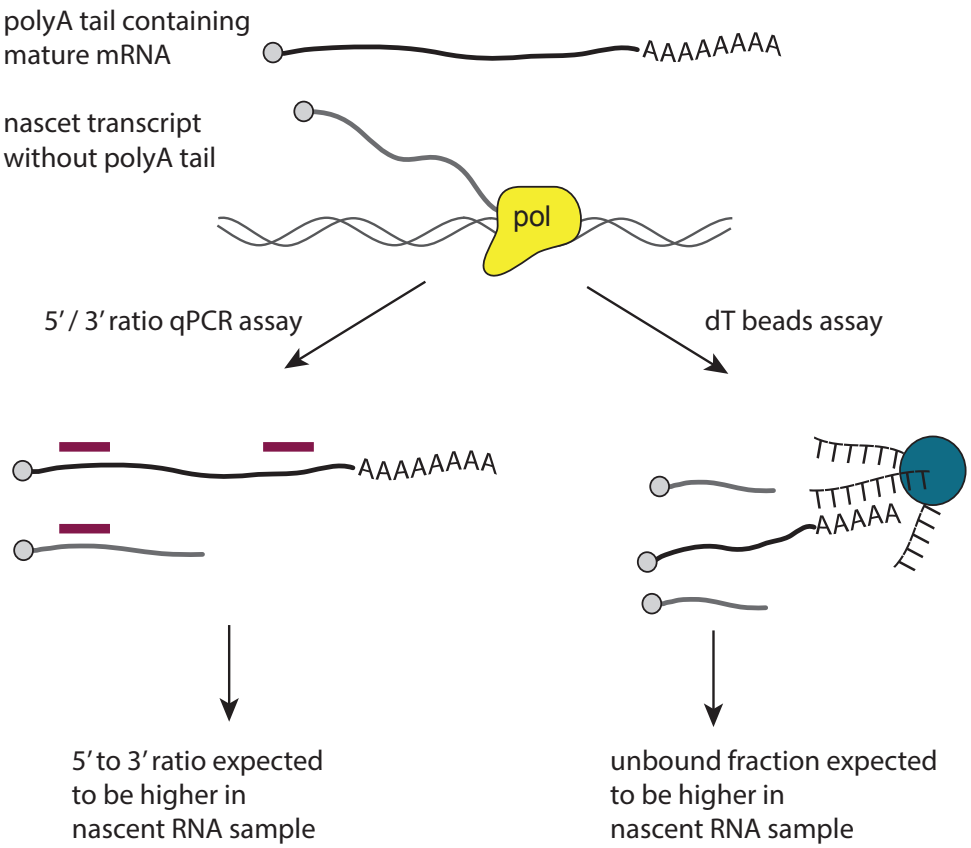

Figure 3.19: **Scheme of assays to assess purity of nascent transcripts.** Total RNA samples are enriched in full-length polyA tail containing mRNAs, whereas nascent transcripts are shorter and lack a polyA tail. Those two features are checked as a quality control of the nascent transcript isolation procedure.

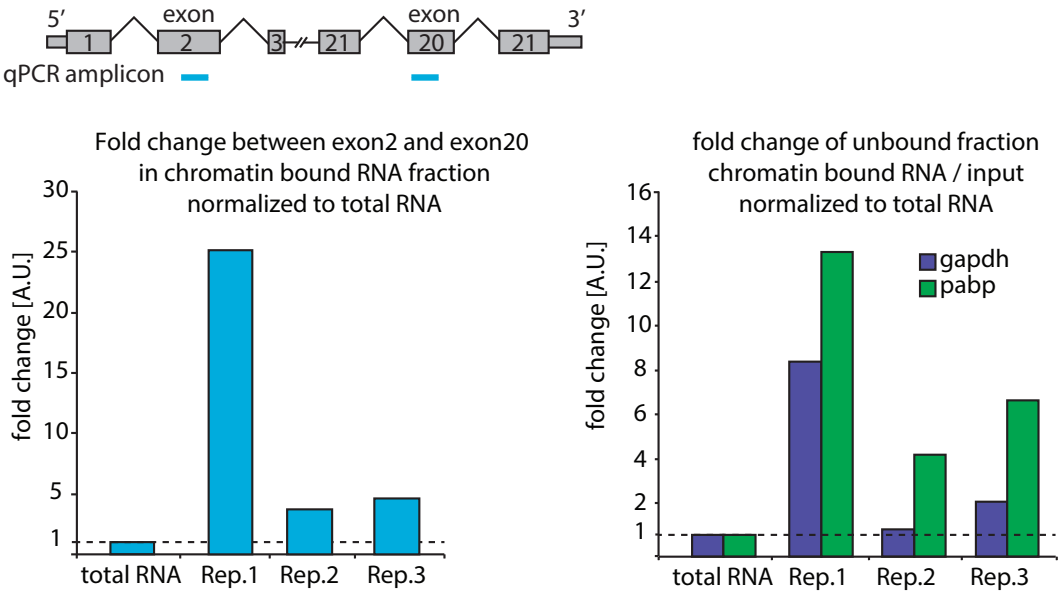

Figure 3.20: **PolyA assays in nascent transcript RNA preparations.** (A) assay to determine 5' to 3' ratio in nascent transcript sample compared to total RNA. (B) Fold change of gapdh and pabp transcripts in unbound fraction (no polyA tail) to input fraction. qPCR data is normalized to a total RNA sample

portion (Exon 2) and the 3' portion (Exon 20) of long RNAs. The ratio of 5' amplicons to 3' amplicons should be higher in the nascent transcript sample than in total RNA. The second assay tested for the presence of poly-adenytlated mRNAs in the nascent transcript sample. Since the polyA tail is only added to the 3' end of an RNA after cleavage of the nascent RNA chain, this feature distinguishes nascent from non-nascent RNAs and the presence of poly-adenylated RNAs is a measure for contaminations. The nascent transcript sample was split into two equal parts, one was kept as "input", the other was passed over magnetic beads coated with dT-oligonuclotides. The abundance of control transcripts was measured by qPCR in the input sample and the unbound sample and compared to a total RNA control. In the nascent-transcript sample most RNA should be present in the unbound fraction, since nascent transcripts do not have a poly-A tail yet. The results for three independent nascent RNA isolations (called rep1, rep2 and rep3) from wild type *Drosophila* ovaries are shown in figure 3.20. Overall, independent experiments did not shown consistent results.

After several rounds of modifications and optimizations of the nascent transcript isolation procedure it became apparent that very stringent wash conditions had to be applied in order to obtain samples enriched in nascent RNA. Therefore, the initial amount of material had to be significantly increased. The protocol was thus adopted for large scale nascent transcript isolation from suspension culture of S2 cells (somatic *Drosophila* cell culture system). In order to obtain sufficient material, the nascent transcript isolation

had to be performed on 200 ml of suspension culture. This quantity of cells was impossible to obtain from *Drosophila* ovaries through manual dissection. Therefore, the efforts to clone and sequence 5' mono-phosphorylated nascent transcripts from *Drosophila* ovaries were not further continued

3.2 Transgenerational inheritance of piRNAs converts piRNA targets to piRNA sources

3.2.1 Objective

Epigenetic effects of chromatin play a crucial role in many aspects of the piRNA pathway. piRNA producing clusters, on one hand, possess a specific chromatin configuration that enables them to produce piRNAs. Genomic regions targeted by piRNAs, on the other hand, also seem to acquire a specific chromatin environment leading to silencing. The aim of this part of my PhD thesis was to establish chromatin modifications on piRNA source and target loci and establish if they correlate with transcriptional activity and piRNA biogenesis.

A piRNA producing locus passed from one generation to the next, loses its activity if it is inherited through the paternal side of the parental cross. To investigate, if a paternally transmitted, inactivated piRNA cluster resumes piRNA production, I followed such a locus for multiple generations.

The analysis of piRNA sources and targets is generally hampered by the repetitive and multi-copy nature of transposable elements. Therefore, all described experiments were performed in transgenic *Drosophila melanogaster* strains producing piRNAs from heterologous sequences that allow for unambiguous mapping in the genome.

3.2.2 Conversion of an inactive to an active piRNA cluster is accompanied by acquisition of the H3K9me3 mark

To investigate chromatin changes on piRNA clusters, I used a transgenic *Drosophila* system. In this system, the same transgenic sequence inserted in the same genomic locus leads to piRNA biogenesis in one fly strain, but not in the other. Most interestingly, crosses between those two strains can convert the inactive locus into a piRNA producing locus. This conversion is also termed "paramutation". The system enabled me to investigate the changes on a locus after conversion into a piRNA producing cluster. I uncovered changes in H3K9 histone methylation, but not transcriptional activity when a locus is transformed into a piRNA cluster.

The fly system used in the following experiments was described by Ronsseray and colleagues and a schematic overview is shown in figure 3.21. Both strains, T1 and BX2 contain an identical number of tandem repeats of the P-lacZ transgene inserted in the same euchromatic locus in the middle of arm 2L. This genomic position does not give rise to piRNAs in other strains (Dorer and Henikoff, 1997; Ronsseray et al., 2001). In the T1 strain, the transgene gives rise to abundant piRNAs in germ cells, and these piRNAs are able to silence expression of another lacZ gene in trans. In contrast, no piRNAs are produced from the same transgene in BX2 (de Vanssay et al., 2012).

Since the two loci do not differ in sequence or in their immediate genomic environment, I performed Chromatin-immunoprecipitation (ChIP) on the H3K9me3 mark to find out, if their chromatin structure is different and thus might be the cause for piRNA production. Indeed, the "active" T1 strain shows a threefold enrichment of H3K9me3 over the transgene when compared to the "inactive" BX2 strain. This confirms the observation that piRNA production correlates with H3K9me3 occupancy (Rangan et al., 2011). To find out, if the exposure to maternally inherited piRNAs and the subsequent conversion of the inactive BX2 into the active BX2* locus is accompanied by an increase in H3K9me3, H3K9me3 ChIP was performed on the paramutating cross, where only piRNAs from T1, but not the T1 locus itself are transmitted into the progeny. After conversion, levels of H3K9me3 are increased almost to T1 levels, as seen in figure 3.22A, suggesting that exposure to transgenerationally inherited piRNAs leads to increased H3K9me3 levels on the BX2 locus after already one generation.

An obvious question about the conversion of an inactive into an active locus is the transcriptional activity in both states. Does exposure to piRNAs convert a transcriptionally inactive locus into an active one? Or is transcription unchanged and paramutation leads to channeling of transcripts into the piRNA processing machinery? deVanssay and colleagues performed qPCRs on the BX2 and BX2* transgene loci and did not observe any changes in steady state RNA levels. Since the transgene RNA is processed into piRNAs in the BX2* strain, but not in the BX2 strain, however, steady-state levels can not answer the question about the transcriptional activiy of the locus. I therefore employed nuclear run-on on ovaries of T1, BX2 and BX2* flies.

When measuring total RNA levels, transgene derived RNA is less abundant in the active T1 strain and BX2* strain, compared to the inactive BX2 (figure 3.23A). This result is counter-intuitive at first, but can be explained, since in the active T1 and BX2* strain the transcripts are processed into piRNAs and thus seem to be less abundant. Interestingly, the nuclear run-on data in figure 3.23B shows that transcription levels in BX2 and BX2* are slightly higher than in T1. But there is no difference between the inactive BX2 and the converted, active BX2* locus. Thus, in this system activation of

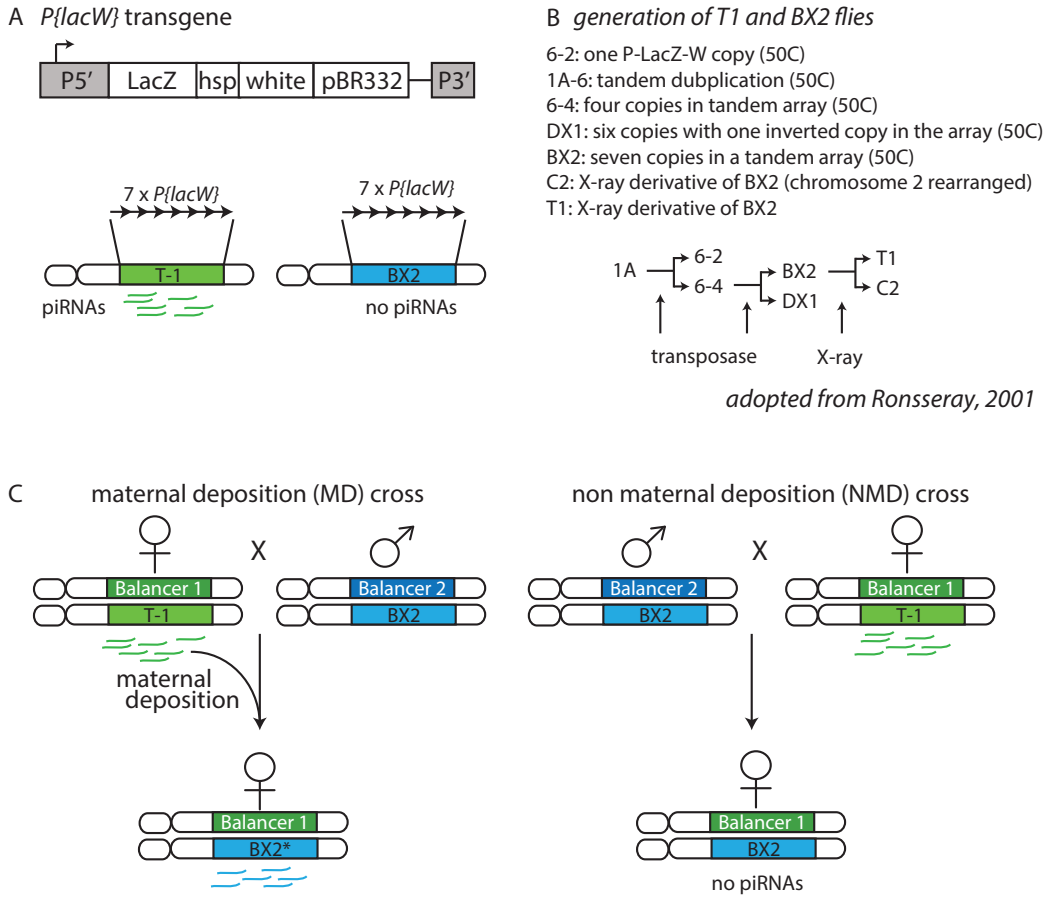

Figure 3.21: **Scheme of fly crosses between T-1 and BX2 flies.** (A) structure of the P{lacW} transgene containing an in-frame translational fusion of the *e.coli lacZ* gene to the second exon of the P-element transposase gene and a mini-white marker gene. P5' and P3' indicate the P-element sequences. (B) Generation of BX2 and T1 fly strains. Transposase-mediated mobilization of the initial 1A tandem repeat generated a single-copy cluster 6-2 and four-copy cluster 6-4. Additional mobilization generated the seven copy BX2 cluster. X-ray mutagenesis of BX2 males generated the two additional strains carrying the 7-copy clusters C2 nd T1 (de Vanssay et al., 2012). (C) In the paramutating maternal deposition (MD) cross the female progeny obtains the BX2 locus from the paternal side and a Balancer chromosome from the maternal side. The maternally deposited piRNAs induce piRNA production from BX2 (now called BX2*). In the reciprocal cross the BX2 chromosome comes from the maternal side and The balancer from the paternal, wich does not deposit the T1 derived piRNAs. The female progeny of this cross does not produce piRNAs from the BX2 locus.

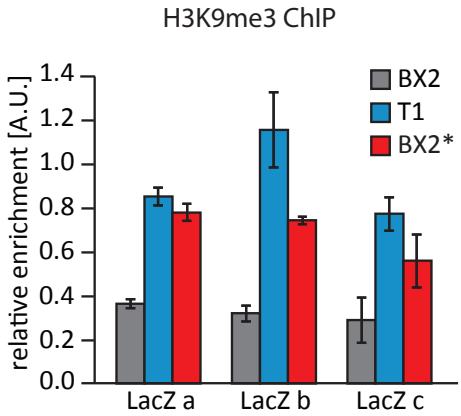

Figure 3.22: **H3K9me3 levels in T1, BX2 and BX2*.** H3K9me3 levels on the transgene locus are higher in the active strains T1 and BX2* than in the inactive BX2. Three primer pairs withing the LacZ transgene were used and enrichment was normalized to H3K9me3 signal on cluster 42AB, a region known to contain this histone modification. Values are the average between two biological replicates consisting of two technical replicates each. Standard deviation was calculated between the two biological replicates.

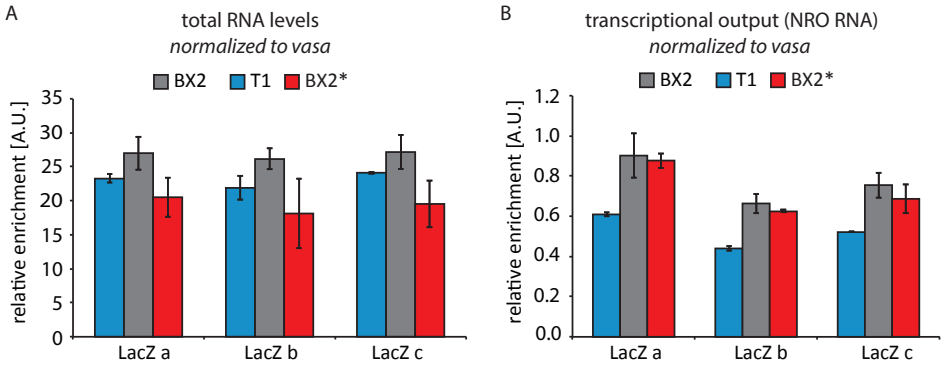

Figure 3.23: **Levels of transcripts and transcription of the original and the paramutated loci.** (A) qPCR on total RNA of T1, BX2 and BX2* total ovary RNA. Transcript levels are lower in the two active strains T-1 and BX2*. (B) qPCR on nuclear run-on RNA of T1, BX2 and BX2* total ovaries. The transcriptional activity in the BX2 and BX2* strains are comparable, thus paramutation does not lead to transcriptional activation of the paramutated locus. T1 shows lower transcriptional activity than BX2 and BX2*. Values are the average between two biological replicates consisting of two technical replicates each. Standard deviation was calculated between the two biological replicates.

BX2 into a piRNA producing locus does not lead to a detectable increase in transcription of the paramutated locus.

3.2.3 A paternally transmitted piRNA cluster does not recover piRNA production

In hybrid dysgenic fly crosses, female progeny are sterile, if they are exposed to a new transposable element from the paternal site. If the element was present in the mother, the progeny is fertile. Experiments following those crosses for multiple generations revealed that the piRNA response to a newly invading transposable element gets stronger with the age of the female fly and the generation time it has been present in a stock. In analogy to those experiments, I wanted to find out how a paternally transmitted piRNA cluster behaves over time and if ageing effects the process of piRNA biogenesis. Does the locus start producing piRNAs after several generations? Do aged mothers provide their progeny with more piRNAs than young mothers?

If the active T1 locus is transmitted into the next generation through the paternal side of a cross, it is not accompanied by piRNAs and thus loses silencing capacity of a LacZ reporter (de Vanssay et al., 2012). The ageing experiments in hybrid dysgenesis crosses by Bucheton and colleagues (Bucheton, 1979) showed that a paternally transmitted TE first leads to sterility in the female progeny. But as the flies age, they gain the ability to silence this newly acquired element (see figure 2.7 in section 2.7.3). If subsequent crosses are performed in a way that females are always aged before crossing, the progeny of the following generations recover fertility to almost normal levels. If a female progeny of "old mothers" in the previous generations, is not aged before mating to a wild type fly, the fertility is lost again in the next generations.

This illustrates that piRNA production slowly but steadily increases over time, moreover, if those piRNAs are then maternally inherited into the next generation, they can establish silencing of a just recently acquired transposable element. The ability of the locus to produce piRNAs without reinforcement, however, is not stable enough to protect against the transposable element. The mother still needs time to acquire enough protective piRNAs in order to ensure fertility in its female progeny.

In a cross of a male containing the active T1 locus to a wild type female, which does not contain any LacZ sequence, the transgenic T1 locus was transmitted into the next generation without its piRNAs. The female progeny of this cross was then aged for only 3 days or 3 weeks before mating to a wild type male. A schematic overview of the crossing procedure and the experiments is depicted in figure 3.24.

As a control ("reciprocal cross") a female T1 fly was crossed to a wild type male.

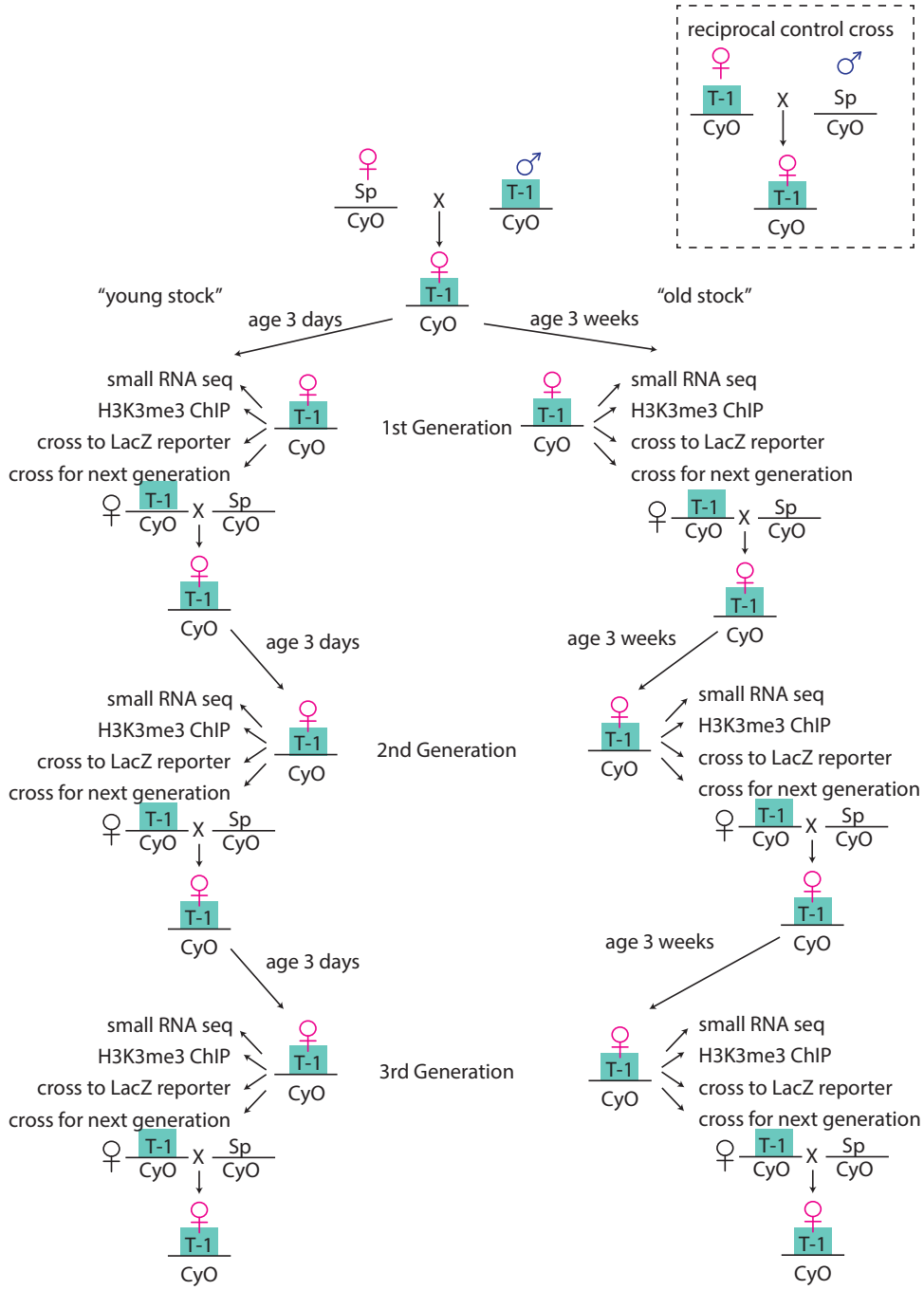

Figure 3.24: **Schematic overview of T1 transgenerational project.** The T1 locus was transmitted into the next generation (1. generation) through the paternal side. Therefore, a male fly containing the T1 locus was crossed to a double-balancer female. The female progeny of this cross was selected for the presence of the T1 locus and aged for 3 days (young stock) or 3 weeks (old stock) before they were crossed again to double-balancer male flies. From their sisters piRNAs were cloned, H3K9me3 ChIP was performed to measure the acquisition of this "piRNA cluster marker". Some were crossed to LacZ reporter flies to assess their potential to silence a reporter construct. Silencing of the reporter is a measure for piRNA production.

To see whether the T1 locus in the female progeny starts producing piRNAs, flies were crossed to a reporter fly containing a LacZ construct and ovaries of the progeny were treated with X-gal to reveal the expression or silencing of the LacZ gene. A decrease in LacZ expression in those reporter crosses indicates piRNA production and subsequent LacZ silencing in the maternal fly lines. To assess the capacity of the T1 locus to silence expression of the reporter, two different positive controls were performed. In the first positive control T1 females were directly crossed to the reporter fly. The female progeny of this cross showed nearly 100% silencing of LacZ (figure 3.25). In the second positive control the T1 construct was transmitted maternally for one generation before it was crossed to the reporter fly. This cross reveals the loss of silencing through one maternal transmission, which is the way the reporter crosses were performed in this experiment. 15% of ovarioles showed LacZ expression. As a negative control the reporter females were stained for LacZ expression. Nearly 80% of the reporter ovarioles show LacZ staining, which gives the maximum level of expression in this reporter system. When female flies of different generations were crossed to reporter males, LacZ silencing was not exceeding the negative control after 4 generations with young mothers, or after 3 generations with old mothers. This result suggests that the initially paternally transmitted T1 locus does not start producing piRNAs in the following generations investigated in this experiment.

At the same time the chromatin state of the T1 locus after paternal transmission were tested for H3K9me3 levels with ChIP qPCR (figure 3.26). Compared to the reciprocal cross where T1 was maternally transmitted, the H3K9me3 signal drops to approximately 50% in the first generation after paternal T1 transmission. Those flies were aged for either 3 days or 3 weeks before H3K9me3 levels were measured. Their female siblings were crossed into the next generation and female progeny was investigated after they were additionally aged 3 days or 3 weeks, respectively. The "young stock" were followed up to the 8th generation, the "old stock" were followed up to the 4th generation. To evaluate if H3K9me3 levels are increasing over time or generation, p-values comparing following generations were calculated: "1. generation 3 days" to "8. generation 3 days". This comparison asseses the effect of multiple generations within the "young flies" line. "1. generation 3 days" to "1. generation 3 weeks". This comparison assess the effect of aging on H3K9me3 levels. "1. generation 3 days" to "4. generation 3 weeks". This test compares the two most extreme cases between age and generation. Only one of the most extreme comparisons between the 1. generation 3 days and 4. generation 3 weeks, showed a significant difference in one qPCR reaction with a the two-sided t-tests resulted in a p-value 0.005. Therefore, consistent with the LacZ staining experiments, H3K9me3 over the transgenic locus does not recover after paternal transmission within the tested time frame of the trans-generational experiments.

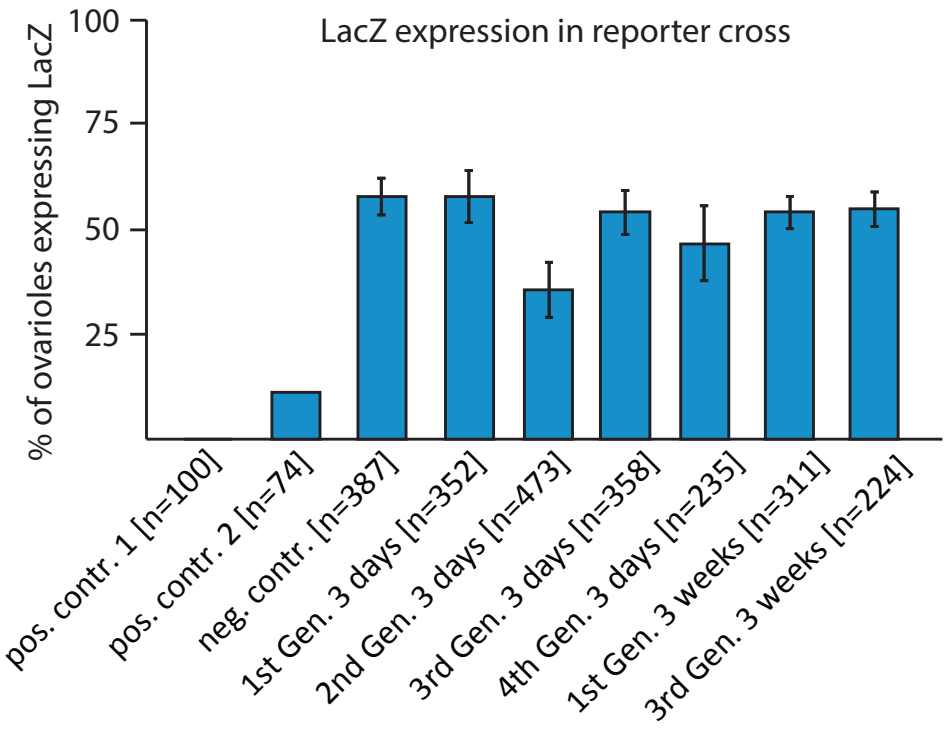

Figure 3.25: **Silencing capacity of females containing the initially paternally transmitted T1 construct.** Positive control 1: female T1 crossed to reporter; Positive control 2: female T1 crossed to wild type male, the progeny of this cross was then crossed to the reporter fly. Shown are the percentages of ovarioles, which were positively staining for LacZ expression. The negative control (reporter only) establishes the expression of un-silenced LacZ. Crossing to females of any generation did not lead to any increase in silencing. The numbers in brackets are the total numbers of counted ovarioles. Error bars represent the standard deviation between two technical replicates (staining of ovaries in two independent batches)

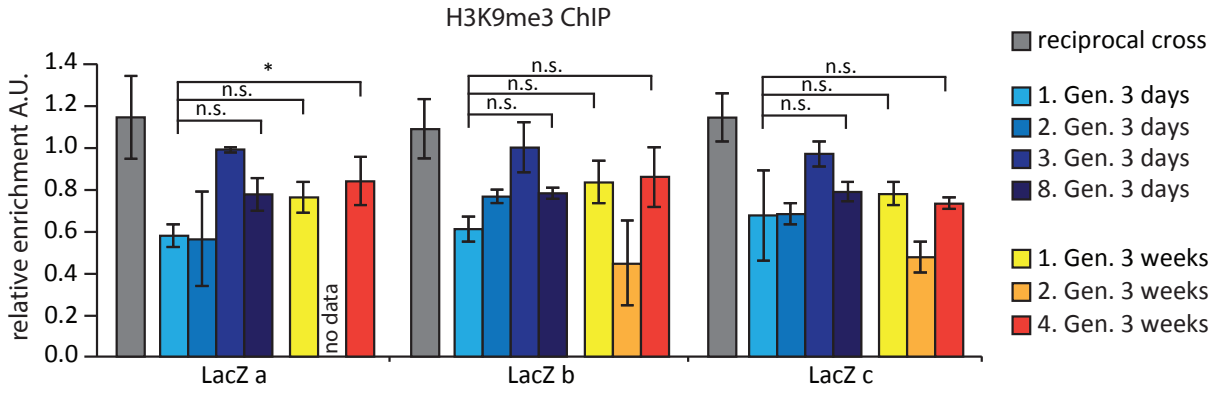

Figure 3.26: **H3K9me3 levels on the LacZ transgene after paternal transmission.** The reciprocal cross was performed between a T1 female and a wild type (double balancer) male fly. H3K9me3 enrichment was normalized to a region inside the cluster 42AB that is known to contain this histone modification. Error bars represent the standard deviation between three technical replicates. Significance was calculated with a two sided t-test, * = $p<0.05$, n.s. = $p>0.05$

In summary, those data show that the presence of maternally inherited piRNAs is pivotal for the production of heterologous piRNAs from the same locus in subsequent generations. The investigated piRNA cluster does not recover its activity during the time of this experiment.

3.2.4 The multiple effects of maternally inherited piRNAs

The results of the following sections demonstrate the breadth of effects that maternally deposited piRNAs impose on the piRNA system of the next generation. I found that they activate both piRNA biogenesis pathways - primary and secondary piRNA biogenesis - and that they establish a specific chromatin environment on targets as well as on piRNA sources themselves.

3.2.4.1 Maternally inherited piRNAs are necessary for piRNA production in the next generation

The previous experiments on T1/BX2 showed that maternally transmitted piRNAs correlate with increased H3K9 tri-methylation of the locus, which is converted into a piRNA producing locus. Reciprocally, an active locus loses its elevated H3K9me3 signal and becomes inactive if it is not exposed to or accompanied by homologous piRNAs as observed for the reciprocal cross in figure 3.26. The paramutated locus in this system consists of LacZ tandem repeats, which start producing piRNAs upon exposure to homologous piRNAs. This demonstrates that piRNAs themselves have the capacity to convert a target

locus into a piRNA producing cluster. To investigate if paramutation through transgenerational inheritance is generally required for piRNA production, I used an additional transgenic system with a single-copy unique sequence.

The flies used in this system were produced in screens of random P-element driven integration of sequences throughout the genome. One of the fly strains P{lArB}A171.1F1, also called P-1152, contains a single copy integration of the 18.3 kb P{lArB} transgene in a telomeric piRNA cluster on the X-chromosome (Wilson et al., 1989; Roche and Rio, 1998; Muerdter et al., 2012) (figure 3.27). P{lArB} consists of the complete sequence of the bacterial LacZ gene, the hsp70 3'UTR and fragments of the adh gene and the rosy gene. Since the insertion is located inside a piRNA cluster, this fly strain produces piRNAs from all of those exogenous sequences. Another fly strain, BC69, contains a very similar insertion, called P{A92}, consisting of the LacZ coding sequence, the hsp70 3'UTR and fragments of rosy. The P{A92} insertion, though, is located in a euchchromatic site on the second chromosome. BC69 flies express LacZ from the P{A92} transgene in their ovaries, which can be detected through staining in a X-galactosidase reaction. In a cross between P-1152 females and and BC69 males P-1152 transgene derived piRNAs lead to silencing of the LacZ in the female progeny. Those crosses are called "maternal deposition crosses" (MD) and this effect was described as trans-silencing effect (Ronsseray et al., 2001; Muerdter et al., 2012). In the reciprocal or "no maternal deposition cross" (NMD), the LacZ gene in the reporter is not silenced.

To profile levels of piRNAs in the MD and the NMD crosses, total piRNA libraries were cloned by Nikolay Rozhkow, who additionally defined piRNA populations bound to PIWI, AUB and AGO3. A large number of piRNAs mapping to the P1152 transgene can be detected in the parental P1152 strain, as reported before (de Vanssay et al., 2012; Muerdter et al., 2012); those piRNAs are present in Piwi, Aub and Ago3 (figure 3.28). In the BC69 strain practically no piRNAs mapping to the P1152 transgene can be found. The level of piRNAs mapping to the P1152 transgene drops twofold in the MD cross, which reflects the fact that only one allele of the transgene is present in the MD flies compared to the two copies in the parental P1152 strain. In contrast, the level of P1152 piRNAs decreases to approximately 10% in the NMD progeny. Since the MD progeny and NMD progeny are genetically identical, this confirms - as already seen in the T1/BX2 system - that maternally inherited piRNAs are essential to keep piRNA production up in the next generation.

One possible mechanism by which transgenerationally inherited piRNAs can fuel piRNA production in the next generation would be induction of ping-pong processing at piRNA clusters. Since germline clusters are generally transcribed from both strands, it is concievable that inherited piRNAs can target piRNA cluster transcripts through comple-

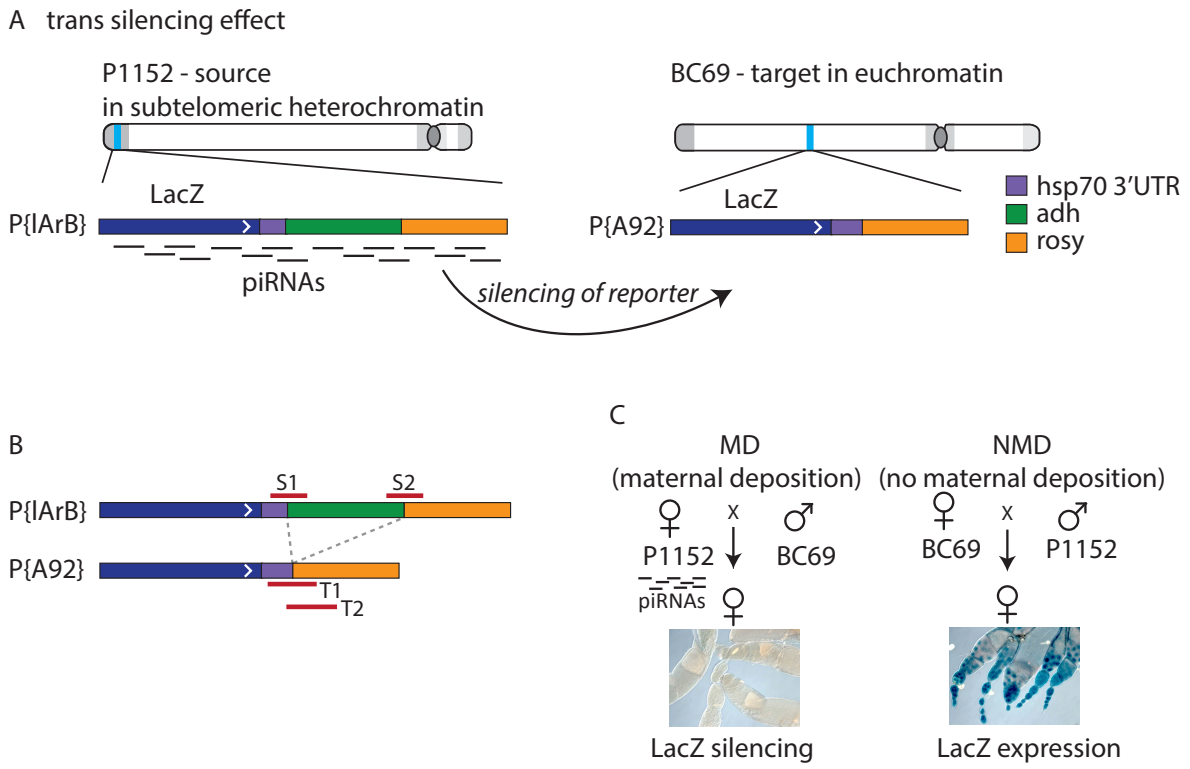

Figure 3.27: **Transgenic *Drosophila melanogaster* strains used for the study.** (A) The strain P1152 contains a single copy of the P{lArB} insertion at the telomeric X-TAS piRNA cluster on the X chromosome, which generates abundant piRNAs from both genomic strands (Josse et al., 2007; Muerdter et al., 2012). Strain BC69 contain a single copy of the P{A92} insertion at the euchromatic site 35B-C of chromosome 2 that does not generate piRNAs. (B) The P{lArB} and P{A92} transgenes are identical in sequence, except for the Adh region that is specific to the P{lArB} construct in P1152. The bars (S1-2 and T1-2) indicate sequences used for ChIP-qPCR to measure chromatin the structure of the P1152 (source) and the BC69 (target) transgenes. (C) BC69 females express lacZ from the transgene in the ovary, however, lacZ is repressed by lacZ piRNAs generated from the P1152 transgene if both transgenes are combined. The progeny designated as Maternal Deposition or MD inherited P1152-derived piRNAs from their mothers, the LacZ transgene is silenced. In contrast the progeny of the reciprocal cross, designated as No Maternal Deposition or NMD, did not inherit these piRNAs and LacZ is expressed.

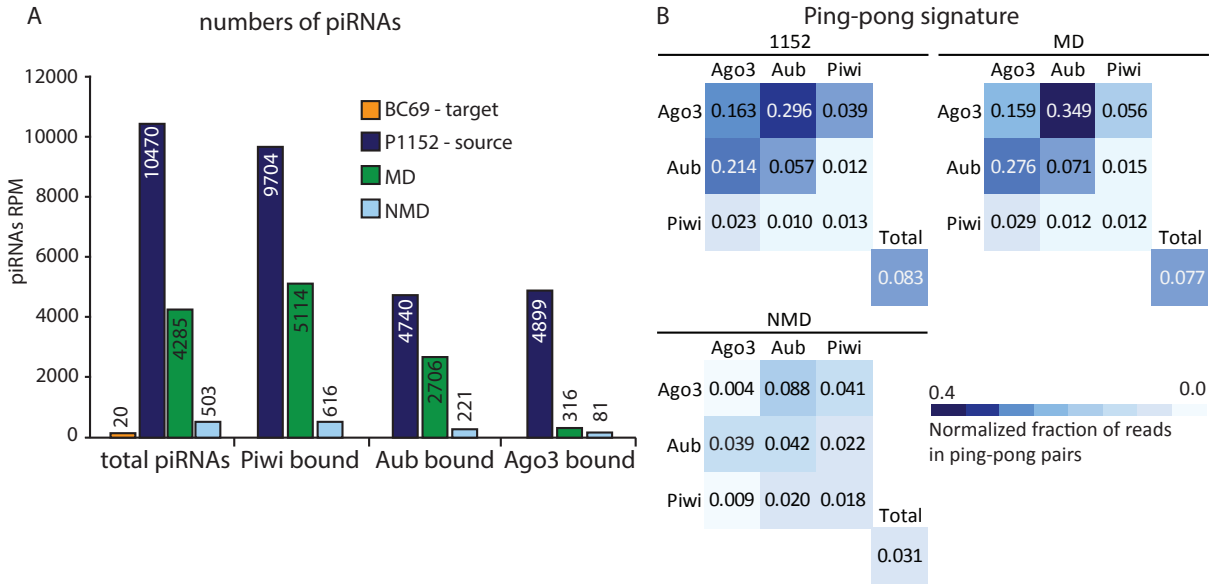

Figure 3.28: **piRNA production in P1152, BC69, MD and NMD crosses.** (A) piRNA production from the P1152 and the BC69 transgenes in the parental strains and the MD and NMD progenies. Shown are the numbers of piRNA reads normalized to library depth (RPM) mapped to the P{lArB} transgene in total RNA and in RNA purified from the PIWI, AUB and Ago3 complexes. Abundant piRNAs are present in the P1152 (piRNA source), but not in the BC69 (piRNA target) parental strain. piRNA levels in the MD progeny are 2-fold lower than in the P1152 strain reflecting the heterozygous presence of the source locus in the progeny. The piRNA level is 10-fold lower in the NMD compared to the MD progeny in the total cellular piRNA population and in the Argonaute complexes. (B) Analysis of ping-pong processing. An equal number of piRNA reads mapped to the P{lArB} transgene in each library was sampled 1000-times to calculate the fraction of reads in ping-pong pairs. Shown are the mean values of the normalized fraction of piRNAs that are in ping-pong pairs. Both the P1152 parental strain and the MD progeny have high level of AUB- and AGO3-bound piRNAs generated by ping-pong processing. The fraction of piRNA generated by ping-pong processing drops 4 to 7-fold in NMD progeny. PIWI-bound piRNAs are generated by primary biogenesis mechanism and accordingly do not have a strong-ping-pong signature. This bioinformatic analysis was performed by Georgi Marinov.

mentarity and initial processing. To quantify ping-pong processing in the parental strain, MD progeny and NMD progeny, we assessed the degree of the ping-pong characteristic feature of piRNA pairs with complementarity of the first 10 nucleotides from their 5' end. This analysis was performed by Georgi Marinov. To account for the different numbers of sequences in each library, the same number of reads was sampled in each library 1000 times. As expected (Brennecke et al., 2007; Li et al., 2009), a large fraction of piRNAs shows ping-pong signature in the parental P1152 strain and those piRNAs were found in complexes of AUB and AGO3. The MD progeny shows a degree of piRNAs produced through ping-pong processing similar to the parental strain. In the NMD progeny, ping-pong pairs were dramatically decreased up to sevenfold in the strongest ping-pong pair of AUB and AGO3, while the overall level of endogenous piRNAs is unaffected. Those results confirm that the high numbers of piRNAs in the MD progeny indeed are mainly a result of ping-pong processing.

Therefore, maternally inherited piRNAs maintain piRNA production in the female progeny through activation of the ping-pong cycle.

3.2.4.2 Inherited piRNAs induce primary piRNA biogenesis at target loci

When investigating the Argonaute bound piRNAs in MD and NMD crosses, one interesting observation suggested that not only ping-pong processing is fueled, but also primary biogenesis is enhanced upon maternal deposition of piRNAs. The levels of PIWI bound piRNAs is 8.3 fold higher in the presence of maternally deposited piRNAs (616 RPM compared to 5114 RPM, figure 3.28). The effect is comparable in total and Aub bound piRNAs. Piwi bound piRNAs do not show a ping-pong signature and thus are produced through primary piRNA biogenesis (Brennecke et al., 2007). Their presence in the MD progeny suggests that maternally inherited piRNAs do not only boost ping-pong processing, but that even primary piRNA biogenesis is enhanced.

I therefore investigated if primary piRNAs are generally produced from loci that are targeted by inherited piRNAs and found several examples, where this is the case.

The target insertion of BC69 lacks the adh fragment of P1152 and thus, the junction between the hsp70 3'UTR and rosy is unique to BC69. This made it possible to investigate how piRNA sources and targets are affected by transgenerationally inherited piRNAs. Mapping of piRNAs in MD and NMD crosses to the target construct revealed that the MD cross progeny contains piRNAs, which span the unique BC69 target sequence (figure 3.29). Since those piRNAs can not have been in the pool of maternally transmitted ones, they had to be produced *de novo* from the target sequence. Two out of nine distinct sequences were detected in the total cellular piRNA library, and seven were detected in the Piwi-bound population. None was present in the parental strains, in the NMD cross

		piRNAs from total ovaries RNA	piRNAs cloned after immunoprecipitation					
sequence	origin	Total-MD	Piwi-MD	Aub-MD	Ago3-MD	Piwi-NMD	Aub-NMD	Ago3-NMD
TGCAGCCAAGCTTCTTTTGATTGAGCT	plus strand	0	1	0	0	0	0	0
AATCAAAAGAAGCTTGGCTGCAGGTC	minus strand	0	1	0	0	0	0	0
GCAGCCAAGCTTCTTTTGATTGAGC	plus strand	0	1	0	0	0	0	0
AGCTCAATCAAAAGAAGCTTGGCT	minus strand	0	1	0	0	0	0	0
AATCAAAAGAAGCTTGGCTGCAGG	minus strand	1	0	0	0	0	0	0
GAAGCTTGGCTGCAGGTCGACGGAT	minus strand	0	1	0	0	0	0	0
CTGCAGCCAAGCTTCTTTTGATT	plus strand	0	1	0	0	0	0	0
GCTCAATCAAAAGAAGCTTGGC	minus strand	0	1	0	0	0	0	0
TCTTTTGATTGAGCTTATATTTTCTTG	plus strand	1	0	0	0	0	0	0
total number of sequences		2	7	0	0	0	0	0

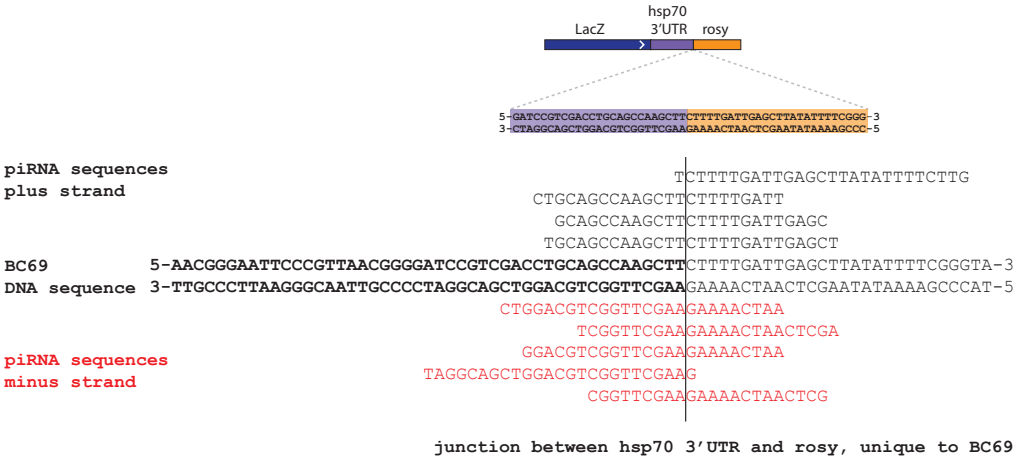

Figure 3.29: **BC69 unique region spanning piRNAs.** piRNAs spanning the unique region of BC69 in the first generation after maternal deposition of P1152 piRNAs. Sequences are mapping to both strands. Out of 9 sequences spanning this unique target region, 2 were detected in the total piRNA library and 7 were bound to Piwi

or in complex with Aub or AGO3. This strongly indicates that those piRNAs - even if it is only a few sequences- have to be products of primary piRNA biogenesis from the target.

Together, the Piwi bound primary piRNAs in MD progeny and the primary piRNAs from the BC69 target rose the question if transgenerationally inherited piRNAs generally induce primary processing at target loci. The P{lArB} insertion in the P1152 strain consists of the transgenic LacZ sequence, but also contains fragments of three endogenous genes in the *Drosophila* genome: hsp70 3'UTR, adh and rosy.

To establish the effect of inherited piRNAs on those endogenous targets, piRNAs from MD crosses and NMD crosses were mapped to the genome and target loci were visually inspected (figure 3.30).

As expected, the MD cross contains abundant piRNAs mapping to the regions, which are included in the transgenic piRNA cluster (indicated as a red bar in figure 3.30). The

same regions shows very few RNAs in the NMD cross.

Interestingly, the MD progeny produces piRNAs not only from the directly targeted endogenous regions, but also downstream of some target genes. At the rosy locus, piRNAs are exclusively mapping to the region inserted in the piRNA cluster. At the Adh locus piRNA production is spreading along the adh gene and the distribution of piRNAs on the plus strand of the MD cross suggests that the complete adh mRNA is processed into piRNAs, even downstream of the sequence contained in the piRNA cluster. An insertion of a transposable element at the 3' end of adh obscures the extend of the piRNA spreading. Since those piRNAs are only mapping to one strand and they were not contained in the pool of maternally deposited piRNAs, those small RNAs have to be produced *de novo* through primary piRNA biogenesis.

The same phenomenon of primary piRNA production downstream of target genes can be seen at the two Hsp70 loci shown in figure 3.30.

Primary piRNA biogenesis is generally thought to be a feature of piRNA clusters and this result suggests the intriguing idea that piRNA targets are turned into piRNA cluster themselves. To test this hypothesis, the target loci in this system were examined for the presence of known piRNA cluster "signatures": H3K9me3, Cutoff and Rhino.

Cutoff and Rhino are two nuclear components of the piRNA pathway, which were found to localize to piRNA cluster regions and they are believed to play a role in defining a genomic region as a piRNA cluster.

Cutoff was first isolated in a female sterile screen (Schupbach and Wieschaus, 1989, 1991) and mutations of Cutoff lead to TE upregulation (Chen et al., 2007) and a decrease in piRNA cluster transcripts (Pane et al., 2011).

Rhino is a germline specific homologue of the Heterochromatin Protein HP1. It localizes to double-stranded clusters (Klattenhoff et al., 2009; Mohn et al., 2014) together with Cutoff (Pane et al., 2011).

3.2.4.3 piRNA targets are enriched in H3K9me3, Rhino and Cutoff

In order to find out if exposure of a locus to inherited piRNAs attracts H3K9me3, Cutoff and Rhino to the target locus, I performed ChIP against those three factors in progeny of MD and NMD crosses. In the subsequent qPCR two primer pairs targeting the unique source junctions (S1 and S2 in figure 3.27) and two primer pairs targeting the unique BC69 junction (T1 and T2 in figure 3.27) were used. The region H1 is downstream of the endogenous Hsp70a locus, depicted in figure 6.1. IP over input was normalized to

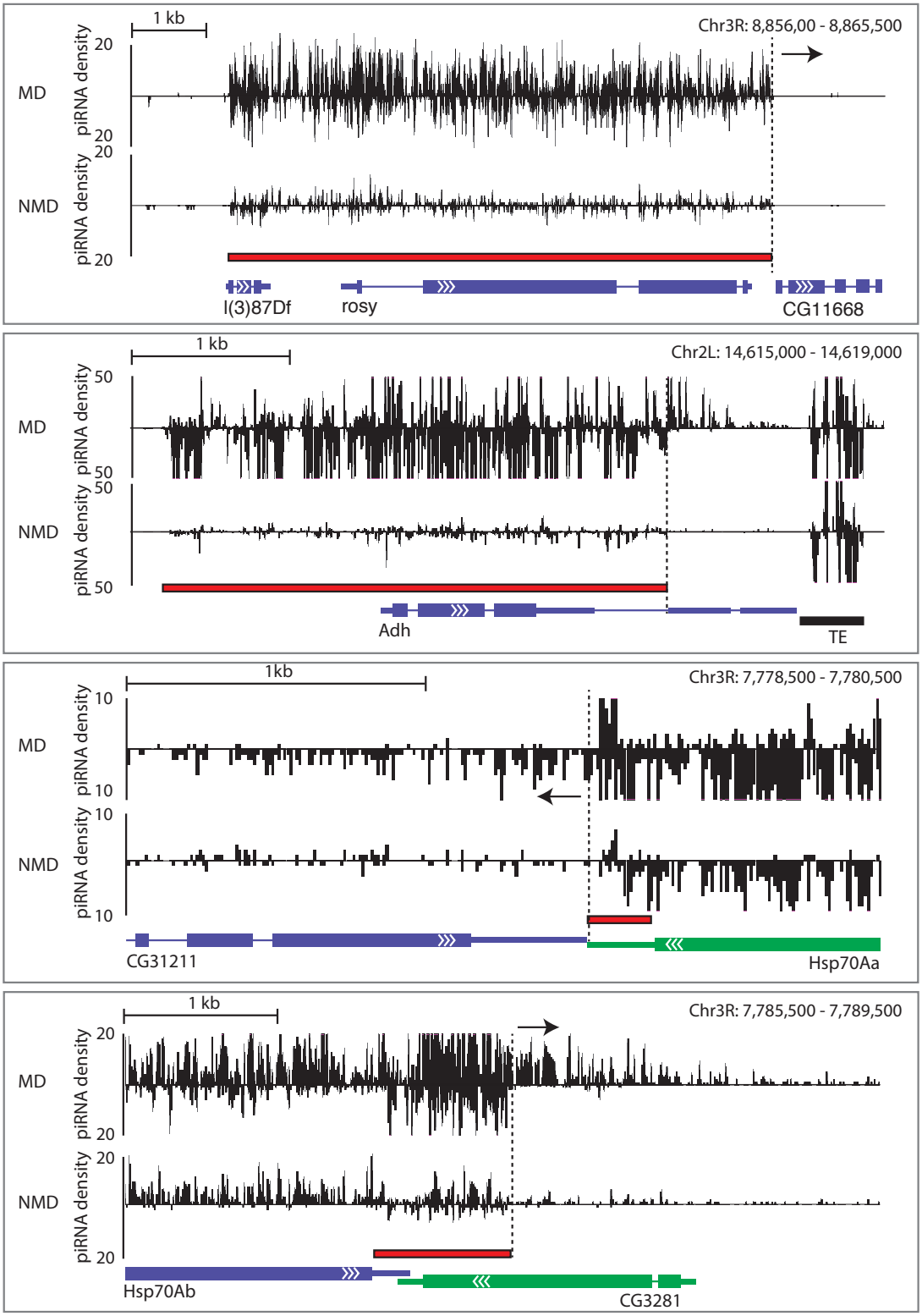

Figure 3.30: **Endogenous genes targeted by piRNAs start producing piRNAs.** The red bar indicates the sequence, contained in the telomeric piRNA cluster of P1152 and therefore the origin of those piRNAs can not be unabmiguously assigned to one source. The downstream regions of those endogenous genes are processed into piRNAs if piRNAs are maternally deposited (MD), but not if they are not (NMD). Shown are piRNA reads plotted with a sliding window (window size: 100 nt, step size: 200 nt)

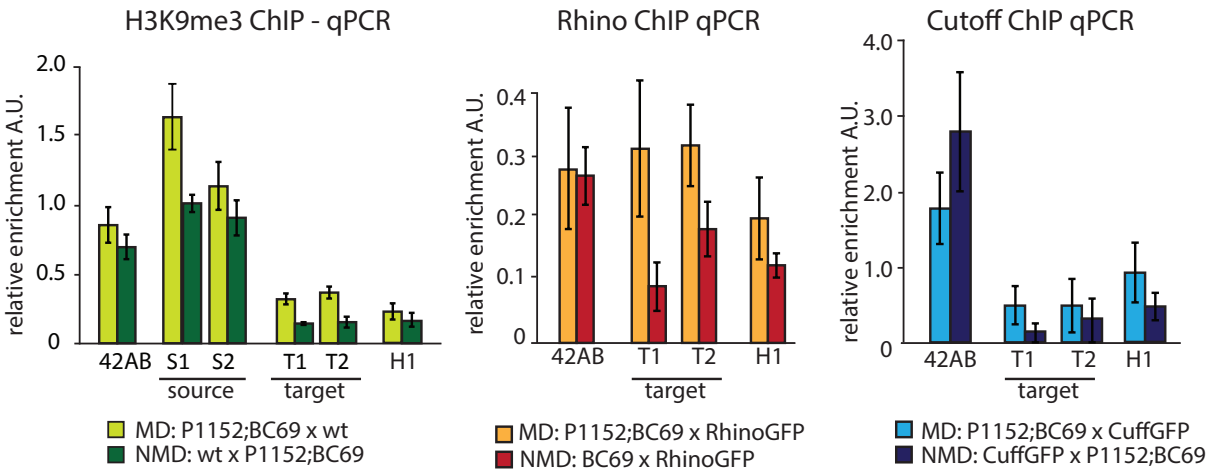

Figure 3.31: **H3K9me3, Rhino and Cutoff ChIPs upon maternal deposition.** ChIP qPCR for the three piRNA cluster associated components (A) H3K9me3, (B) Rhino and (C) Cutoff. The source and target primers are shown in 3.27. Genotypes of the flies are shown below the graphs in the order female x male. Shown is IP signal over input, normalized to a region within the piRNA cluster 42AB that is known to contain all three components.

a region within the piRNA cluster 42AB, with which all three components - H3K9me3, Rhi and Cuff - are known to associate.

A different region withing 42AB is shown as a positive control. As expected, there is no significant difference between the MD and the NMD cross within the piRNA cluster (figure 3.31).

The target locus, however, shows a higher relative enrichment in all three components when piRNAs were maternally transmitted. The nature of the crosses for those experiments allows the investigation of the source locus regarding H3K9me3 levels. Interestingly, even the source locus has elevated H3K9me3 levels when transmitted to the next generation through the maternal side and accordingly accompanied by its piRNAs.

Even though enrichment of H3K9me3, Rhino and Cutoff downstream of the hsp70a locus does not reach significance, it seems to be slightly higher in the MD crosses than in the NMD crosses. This suggests that targets of piRNAs start producing piRNAs and at the same time acquire features of piRNA clusters.

3.2.4.4 Read-through transcription on piRNA targets is increased

Loci targeted by piRNAs acquire features of piRNA clusters and piRNA production spreads downstream of the original target region. The uni-strandedness of the *de novo* piRNAs suggests that they are produced from read through transcripts. To investigate if maternal deposition of piRNAs indeed increases read through transcription at target

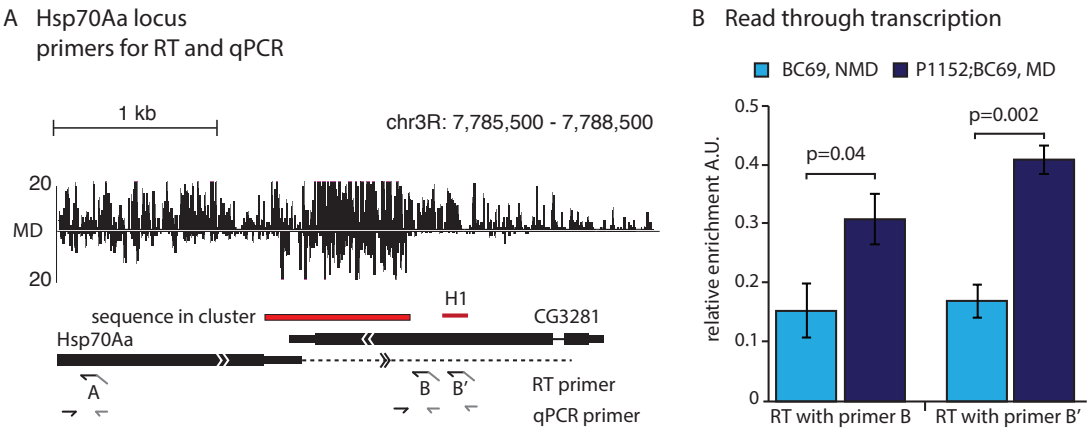

Figure 3.32: **Read through transcription upon maternal deposition.** (A) endogenous locus of Hsp70Aa that starts producing downstream piRNAs upon piRNA exposure. (B) Levels of read through transcripts measured by qPCR after strand specific reverse transcription (RT). Primers for RT and qPCR are shown in panel A. Shown are read-through transcript levels relative to the transcript levels of Hsp70Aa (see panel A). Shown are the averages of three biological replicates consisting of 3 technical replicates each. Error bars represent standard deviations between the three biological replicates. p-values were calculated with a two-sided t-test. Validation of the strand specific reverse transcription reaction is shown in figure 6.1

loci, I performed strand specific reverse transcription (RT) and qPCR to measure read through transcripts downstream of the Hsp70Ab gene.

The RT reaction contained a primer with a 5' end specifically annealing to a potential read-through transcript of Hsp70Aa and a universal adapter part at its 3' end (shown in figure 6.1). To ensure that only specifically reverse transcribed molecules were amplified, the qPCR was performed with a forward primer specific to the potential read-through transcript and a reverse primer, which can only anneal to the adapter sequence of the RT primer. The two different RT primers B and B' used in this experiments were suitable for specific reverse transcription (see also figure 6.1 in section 6.1.7); any RT primer further downstream of B' did not lead to any detectable product anymore. To control for the amount of RNA in the RT reaction and at the same time for potential differences in Hsp70 transcription levels, the RT reaction contained an additional primer annealing to the 5' end of Hsp70. Both experiments (RT primer B and B') show a strong increase in read through transcription upon maternal piRNA deposition.
So far the results suggest that maternally transmitted piRNAs ensure piRNA prodcution in the progeny through two mechanisms. First, they boost ping-pong processing at the source locus and therefore keep a piRNA cluster active. Second, they induce primary piRNA biogenesis downstream of target loci, which is accompanied by the attraction of H3K9me3, Cutoff and Rhino to the target locus. This suggests that H3K9me3, Cutoff

and Rhino induce transcription of loci beyond their normal range.

H3K9me3 is a cannonical histone modification, which can be recognized by chromodomain containing proteins such as Heterochromatin Proteins (HPs). Rhino is a homologue of the Heterochromatin Protein HP1 and contains a chromodomain, a hinge region and a chromoshadow domain.

Experiments in collaboration with the Patel lab at Memorial Sloan Kettering Cancer Center, New York, showed that Rhino's chromodomain indeed directly recognizes H3K9 methylated peptides. By isothermal titration calorimetry (ITC) binding affinities of the Rhino chromo-domain was measured to be 21.8 μM for H3K9me3 and 28.4 μM for H3K9me2. H3K9me1 and H3 on the other hand did not show any detectable binding affinity. This result suggests that Rhino localizes to H3K9me3 marked regions through its chromodomain. Next, co-immunoprecipitation experiments for Rhino and Cutoff performed in our laboratory confirmed the interaction between those two proteins. Together, the data provides compelling evidence for a model in which the chromatin of piRNA clusters and targets are modified with H3K9me3. This histone mark attracts Rhino, which in turn brings Cutoff to this specific genomic locus (Le Thomas et al., 2014).

To establish the functional role of Rhino and Cutoff, I turned to define their role in transcriptional regulation.

3.3 Transcriptional regulation of piRNA clusters

A genome wide study by Rangan and colleagues showed that piRNA clusters are enriched in H3K9me3 and that cluster transcription and piRNA production is severely decreased upon mutation of the histone methyl transferase SETDB1/Egg (eggless) (Rangan et al., 2011). Their data furthermore shows that not only piRNA production, but also siRNA production from piRNA clusters is affected, which suggests that H3K9 tri-methylation indeed is needed for transcription of piRNA clusters. Interestingly, pure siRNA source loci produce more siRNAs upon loss of SETDB1, which is in accordance with the traditional view of H3K9me3 as a represive mark. Contrary, in piRNA clusters the H3K9me3 mark acts as an activating mark. The data shown above suggests a model in which this interpretation of the activating mark could be facilitated through Cutoff and Rhino.

3.3.1 Objective

The next aim was to establish the role of Cutoff and Rhino in piRNA cluster transcription as well as their effects on piRNA targets. Therefore, I performed global nuclear run-on (GroSeq) in *Drosophila* ovaries depleted for those two factors. Through analysis of my

5' monophosphate RNA data and tethering experiments performed by another scientist in our lab, we were able to show that Cutoff ensures transcription of piRNA clusters through an anti-termination mechanism.

3.3.1.1 Rhino and Cutoff promote piRNA cluster transcription

Cutoff and Rhino were already included in the mini-screen for first processing factors (figure 3.16). Both factors show a strong effect on the levels of piRNA cluster transcripts when compared to the control knockdown line. In Rhino depleted ovaries, transcripts from double-stranded clusters are strongly decreased. Expression of 42AB and 80EF are diminished to roughly 20% of the expression levels in control flies, the two clusters on 38C drop to approximately 50%. Transcripts from the uni-stranded clusters flamenco and 20A are increased. Depletion of Cutoff has a similar effect on piRNA cluster transcripts. 42AB shows the greatest decrease in transcript level of 19% compared to the control. 80EF shows a 50% higher transcript level than the control and all other clusters produce less transcripts in the Cutoff knockdown. Total RNA sequencing only provides an insight into steady state levels of RNAs. Decreased levels of double-stranded cluster transcripts in Rhino and Cutoff knockdown therefore could also imply those two component in stabilization of piRNA precursor. To ultimately elucidate, if Rhino and Cutoff are involved in the actual transcription of piRNA clusters, I performed Gro-Seq on *Drosophila* ovaries.

An outline of the experimental procedure for Gro-Seq is shown in figure 3.33. Briefly, *Drosophila* ovaries of knockdown and control flies were dissected in PBS and nuclei were isolated. The run-on reaction was performed in the presence of Bromo-UTP (BrUTP), which incorporated into newly transcribed RNA. Subsequently, those newly transcribed RNAs were immunoprecipitated with an antibody againts BrUTP twice. Then, libraries were cloned with a commercially available RNA library cloning kit and subsequently sequenced.

A challenge in this experiment was the limited amount of material obtainable from *Drosophila* ovaries. Additionally, knockdown of Rhino and Cutoff lead to a change in ovary size and morphology. Thus, only small batches of material could be obtained in one go. In order to keep nuclei in a state in which they still actively transcribe RNA during the run-on reaction, the time from ovary dissection to the run-on reaction was kept below 60 minutes.

To first confirm that the method works comparably well with material from knockdown flies as from wild type flies, global mapping statistics of the three experiments were compared (table 3.4).

Since the material for library cloning was restricted, rRNA depletion was not performed prior to cloning of the RNA libraries. Accordingly, reads mapping to rRNA are

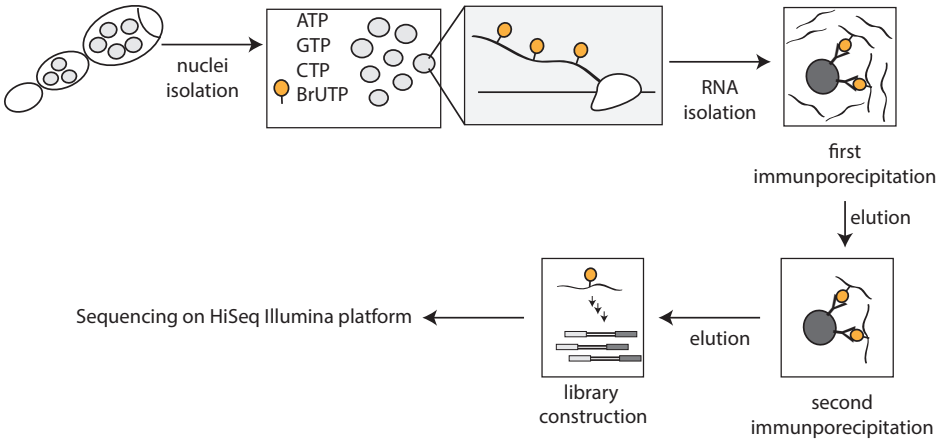

Figure 3.33: **Schematic overview over the nuclar run-on procedure.** The method is described in detail in the material section 6.2.2

Table 3.4: Gro-Seq sequencing statistics

	shCutoff	shRhino	control
total reads	**51,776,314**	**44,848,183**	**39,002,339**
mapping to rRNA [read numbers]	45,412,326	37,682,037	33,729,979
mapping to rRNA [in %]	88%	84%	86%
not mapping to rRNA [read numbers]	6,363,988	7,166,146	5,272,360
not mapping to rRNA [in %]	12%	16%	14%
not mapping to genome [read numbers]	2,449,183	2,722,278	1,992,161
not mapping to genome [in %]	38%	38%	38%
mapping to the genome [read numbers]	3,914,805	4,443,868	3,280,199
mapping to the genome [in %]	62%	62%	62%

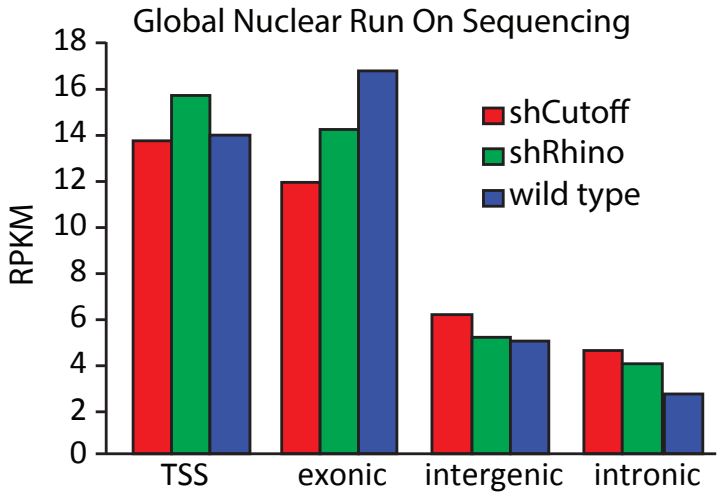

Figure 3.34: **GroSeq signal in shRhi, shCuff and control.** RPKMs calculated for the three genomic partitions transcription start site (TSS), exonic, intergenic and intronic for GroSeq data from shCutoff, shRhino and control

high, but with a percentage of 84% to 88% they are consistent within all three samples. Roughly 15% of all reads do not map to ribosomal RNA. Of those reads, 62% map to the genome, corresponding to approximately 4 million reads. Next, RPKMs for the following four genomic features (1) transcription start site (TSS), (2) exons, (3) intergenic space and (4) introns were calculated (see figure 3.34). Also here, the three libraries are comparable, indicating that knockdown of Cutoff and Rhino does not generally skew or alter overall transcription within any of those genomic features. piRNA clusters fall into intergenic regions and the fact that signal from intergenic space is not decreased in the Cutoff and Rhino knockdown ovaries implies that effects seen in piRNA clusters are specific to those and not a general defect in intergenic transcription.

Next, reads mapping to piRNA clusters were counted and RPKMs were calculated (see figure 3.35).

The strongest effect of Rhino and Cutoff on transcription can be observed in the two double-stranded clusters 42AB and 80EF. They show a dramatic decrease in transcription from the plus strand in both knockdown lines; the minus strand transcription seems only mildly affected. The first part of the 38C cluster does not show any change in transcription upon Rhino or Cutoff knockdown; the second 38C cluster part shows some decrease from the negative strand transcription and an increase in plus strand transcription upon Cutoff knockdown. The two uni-stranded clusters flamenco and 20A do not show a strong dependency on Cutoff and Rhino. This result was expected, since Cutoff and Rhino are exclusively expressed in germline cells and flamenco and 20A are predominatly expressed in somatic cells of the *Drosophila* ovary.

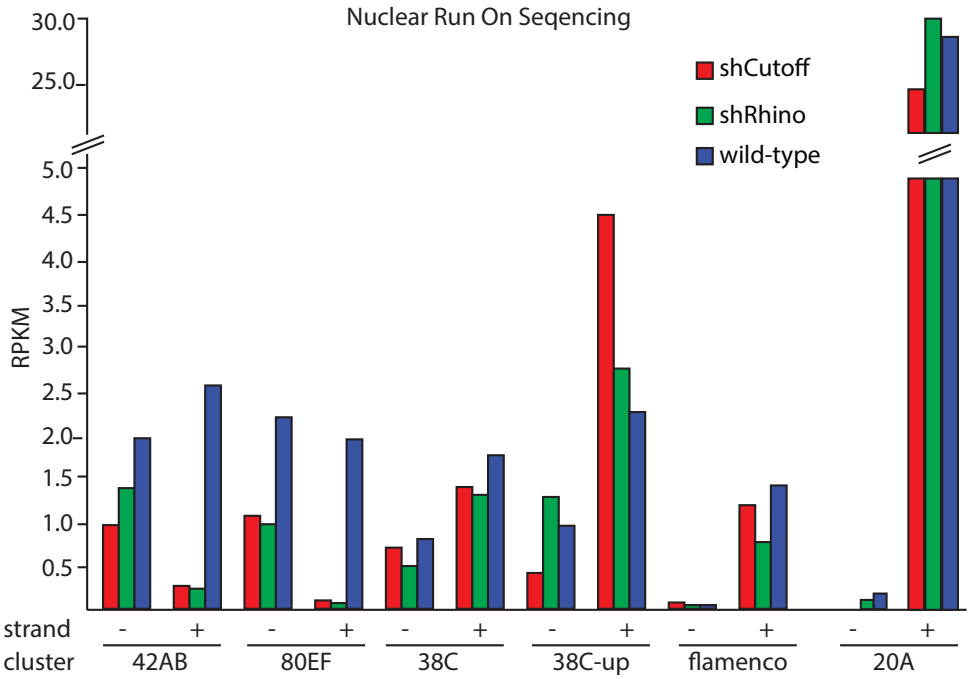

Figure 3.35: **GroSeq signal in piRNA clusters in sh lines.** GroSeq data from shCutoff, shRhino and control in piRNA clusters. Libraries were mapped strand specifically and shown are separate values for plus and minus strand clusters

Taken together, those data show that Cutoff and Rhino seem to be necessary for transcription of the germline specific double stranded piRNA clusters.

How exactly cluster transcription is altered in Rhino and Cutoff mutants can be seen in figures 3.36 and figure 3.37. To blot densities over the cluster regions, reads falling into the size of a specific sliding window were collapsed and the sliding window was moved over the genomic coordinates with a specific step size.

As already seen in figure 3.35, transcription over single stranded piRNA clusters does not change upon knockdown of Rhino and Cutoff. Cluster 80EF loses almost all plus strand transcription, while the minus strand seems to still initiate transcription at its beginning. In cluster 38C no dramatic decrease is observed upon knockdown of Rhino or Cutoff on any strand. Cluster 42AB shows an interesting residual transcription profile upon knockdown of Rhino. Judging from the total reads derived from this cluster (see figure 3.35) the overall transcription is decreased. The last, most downstream region of the minus strand, however, is transcribed stronger in the Rhino knockdown ovary. This result is surprising, since it first suggests that different regions of this cluster are regulated differentially. Furthermore, it shows that there indeed might be internal transciption initiation within this double stranded cluster. One posslibe explanation could be the activation of a transposable element inside this cluster when piRNA biogenesis is impaired in Cutoff and Rhino knockdown flies.

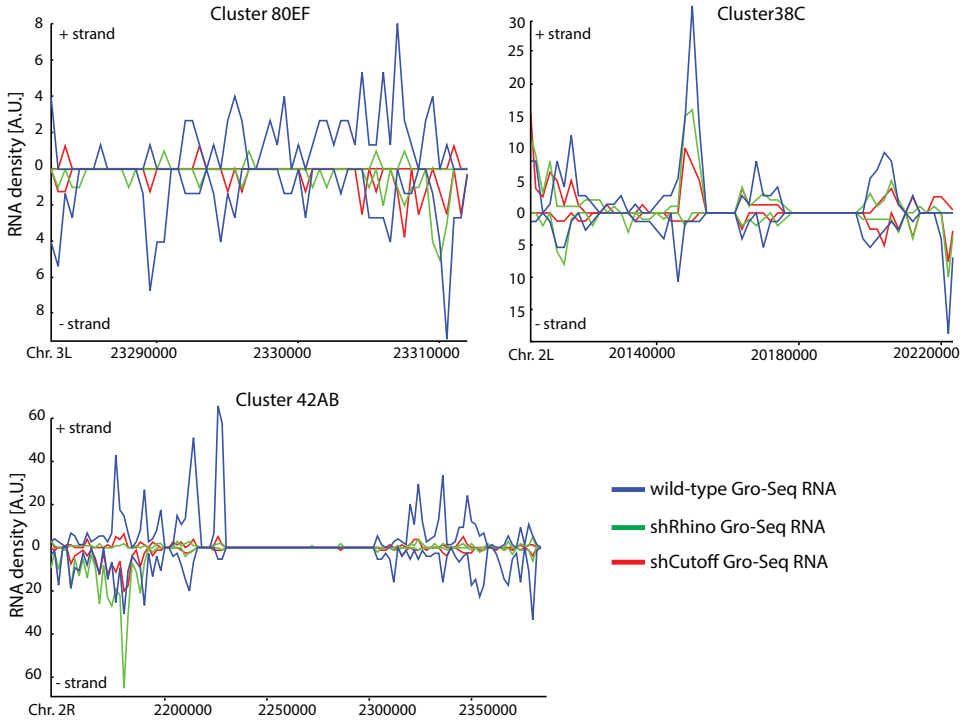

Figure 3.36: **GroSeq density in double-stranded clusters.** Mapping of GroSeq reads to the three major double stranded clusters 80EF, 38C and 42AB. RNA densities are normalized to the depth of the respective library. Reads are plotted with a sliding window (window size 500 nt, step size: 500 nt for cluster 80EF, window size: 2,000, step size: 2,000 for 38C and 42AB

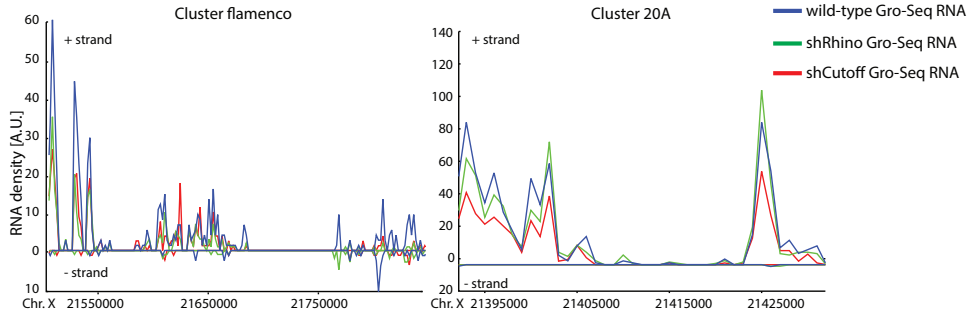

Figure 3.37: **GroSeq density in single-stranded clusters.** Mapping of GroSeq reads to single stranded clusters. RNA densities are normalized to the depth of the respective library. Reads are plotted with a sliding window (window size: 2,000 nt, step size: 2,000 nt for flamenco and window size: 1,000 nt, step size: 1,000 nt for 20A

Together, the run-on data show that the absence of Cutoff or Rhino results in a collapse of cluster transcription and Rhino and Cutoff are therefore acting very upstream of piRNA production, since they are securing transcription of piRNA source loci.

3.3.1.2 Cutoff is an anti-terminator and protects 5' monophosphate ends

The biochemical data from the Patel lab together with our co-immunoprecipitation of Cutoff with Rhino and recently published data (Mohn et al., 2014) suggest that Rhino is a mediator between H3K9me3 and Cutoff. Cutoff is a homologue to the yeast Rai1 and Dxo1 and the mammalian Dom3Z protein. All of these proteins are involved in transcription termination and in RNA quality control (Kim et al., 2004; West et al., 2004; Luo and Bentley, 2004; Xiang et al., 2009). The mammalian Dom3Z protein harbors RNA-binding, 5'-3' exonuclease, de-capping and pyrophosphohydrolase activities. Yeast Rai1 is a pyrophosphohydrolase and decapping enzyme and Dxo1 displays exonuclease activity. Homology modeling of Cutoff using the known structures of yeast Rai1 and the mammalian Dom3Z as a template suggest that Cutoff is an enzymatically inactive homologue. Rai1 and Dom3Z contain four conserved acidic residues to coordinate two magnesium ions, necessary for their enzymatic activity, which are not present in Cutoff (figure 3.38). Biochemical assays performed by Ariel Chen in our lab suggest that Cutoff does have RNA binding activity, but indeed does not posses any catalytical activity. Co-immunoprecipitation experiments moreover show that Cutoff interacts with the *Drosophila* homolog of Rat1/Xrn2.

Cutoff depletion results in a collapse of transcription of double-stranded piRNA clusters. One striking observation was made in our laboratory, when total RNA libraries from shCutoff ovaries were cloned after poly(A) selection: despite the general loss of transcription, few abundant transcripts appear in distinct positions along the piRNA clusters (figure 3.40). This indicates that while piRNA cluster transcripts are generally not polyadanylated in wild type flies, they become polyadenylated in Cutoff knockdown flies. Independent RT-qPCR confirmed the appearance of poly(A)+ RNA in the cuff mutant flies. Cloning and Sanger sequencing of the RT-PCR product confirmed that the identified poly(A)+ transcript, which can exclusively be detected in Cutoff depleted flies, spans several TE fragments and was transcribed from within the piRNA cluster. These results indicate that Cuff knockdown causes a switch from read-through transcription, which generates non-polyadenylated RNAs from the entire length of the piRNA cluster, to detection of polyA signals within the cluster and formation of polyadenylated RNAs.

Formation of polyadenylated RNA requires the recognition of the poly(A) signal on nascent RNA by the CPSF complex, which cleaves pre-mRNA. This is followed by addition of the poly(A) tail to the upstream RNA fragment by the nuclear poly(A)-

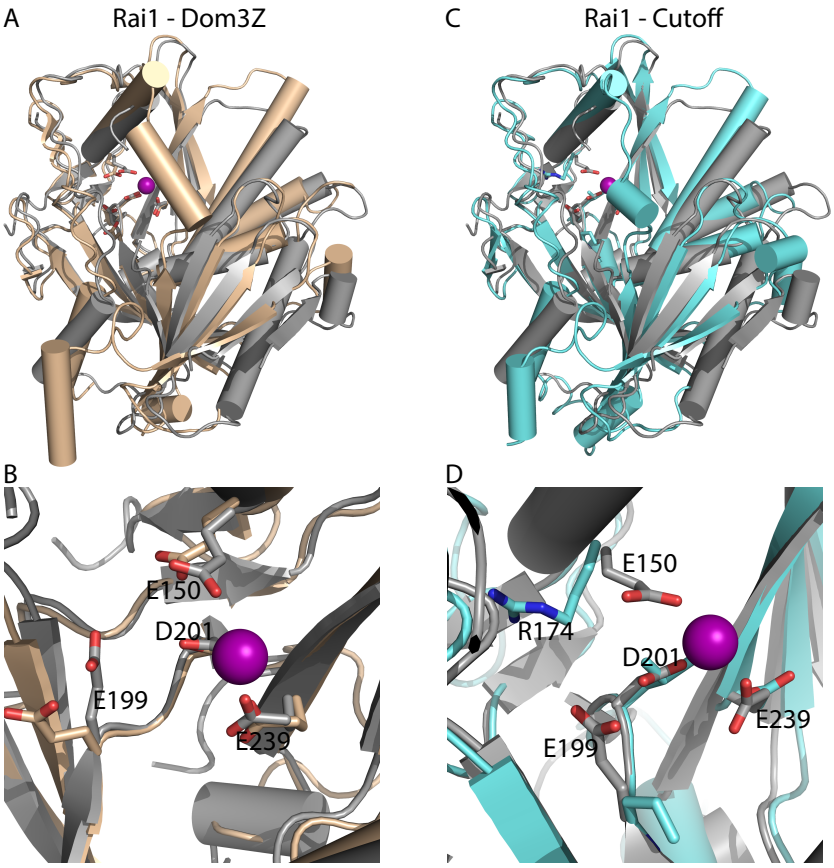

Figure 3.38: **Structural comparison between yeast Rai1, mammalian Dom3Z and *Drosphila* Cutoff.** (A) Superimposition of the two crystal structures of yeast Rai1 (3FQG) in gray and mammalian Dom3Z (3FQI) in wheat color. (B) The active site of Rai1 and Dom3Z coordinates a Magnesium ion, necessary for catalytic activity with four acidic residues. (C) Structural model of Cutoff (Cyan) onto Rai1. (D) Residues coordinating the Magnesium ion in Rai1 are not conserved in Cutoff.

polymerase (Proudfoot, 2011). Analysis of the distribution of the canonical poly(A) signal, AAUAAA, along piRNA clusters detected multiple motifs on both genomic strands. Recognition of a poly(A) site and subsequent cleavage of the nascent transcript leads to termination of transcription a few hundred nucleotides downstream (Proudfoot, 2011). Indeed, the formation of polyadenylated transcripts in shCuff ovaries correlates with the high density of polyadenylation sites. Together with the run-on data these results suggest that Cutoff controls transcription downstream of transcription initiation, as otherwise no poly(A)+ RNA would be formed. These data indicate that Cutoff facilitates read-through transcription by preventing termination when the RNA polymerase encounters a poly(A) signal.

To test if Cutoff is directly responsible for read-through transcription and suppressed termination Cutoff was tethered to the 3'UTR of an unrelated reporter mRNA by fusing Cuff to the λ-N RNA-binding domain, which recognizes four BoxB hairpins in the reporter (Keryer-Bibens et al., 2008) (figure 3.40B) . Tethering of Cuff to the reporter in fly ovaries led to a 5-fold increase in transcripts that span the poly(A) cleavage site (region B in figure 3.40). Furthermore, a similar increase was observed for read-through transcripts 805 bp downstream of the cleavage sites (region D in figure 3.40). Data shown in figure 3.40 was produced by Ariel Chen. They indicate that recruitment of the Cuff protein to nascent RNA suppresses pre-mRNA cleavage and termination of transcription.

When a poly(A) signal of a regular gene is recognized on a nascent RNA, CPSF mediated cleavage creates a 3' fragment with a monophosphorylated 5' terminus. This RNA end is a substrate for the nuclear exonuclease Rat1/XRN2, which degrades the 3' fragment and results in transcription termination as soon as it catches up with PolII. This mode of transcription termination was described as the "torpedo model" (West et al., 2004; Proudfoot, 2011)

Since Cutoff lacks any catalytic activity but binds RNA and associates with dRat1, we hypothesized that it might protect piRNA cluster transcripts against exonucleolytic degradation. Erroneous cleavage of nascent cluster transcripts at poly(A) sites would create a 5-monophosphorylated RNA terminus, which would render the transcript a substrate for dRat1/Xrn2 degradation.

The 5' monophosphate piRNA precurssor data supports this model (figure 3.39.) Depletion of Rat1 in *Drosophila* ovaries leads to an increase in piRNA precursor transcripts compared to the shWhite control. This indicates that Rat1 indeed is an exonuclease degrading 5' monophosphorylated piRNA precursors. In turn, when Cutoff was knocked down, 5'-monophosphorylated cluster transcripts were dramatically decreased. This effect was specific for double stranded clusters; single stranded clusters were not affected. Thus, Cutoff seems to protect 5'-monophosphorylated piRNA cluster transcripts against

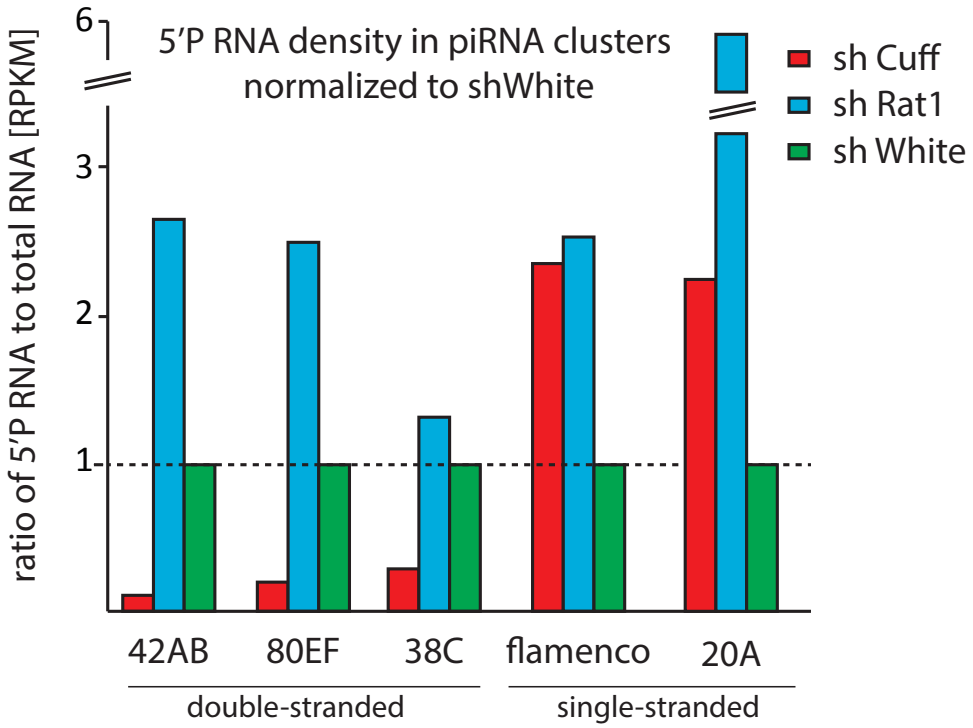

Figure 3.39: **piRNA precursor levels in shCuff, shRat1 and shWhite.** Ratio of 5' monophosphate to total RNA levels in Cutoff and Rat1 knockdowns. RPKMs were calculated by counting the number of mapped reads to the respective cluster and dividing by the total length of the cluster (in kb) and total number of mapped reads of the respective library (in million). Values of the Cutoff and Rat1 knockdown lines were normalized to the shWhite control through division.

exonucleolytic degradation. Since Cutoff specifically localizes to double stranded piRNA clusters (Pane et al., 2011; Mohn et al., 2014), it is likely that Cutoff shields free 5'-monophosphate termini on nascent transcripts and subsequently prevents termination of cluster transcription in the event of transcript cleavage.

In summary, the presented data suggest that Cutoff ensures transcription of piRNA source loci through prevention of transcription termination.

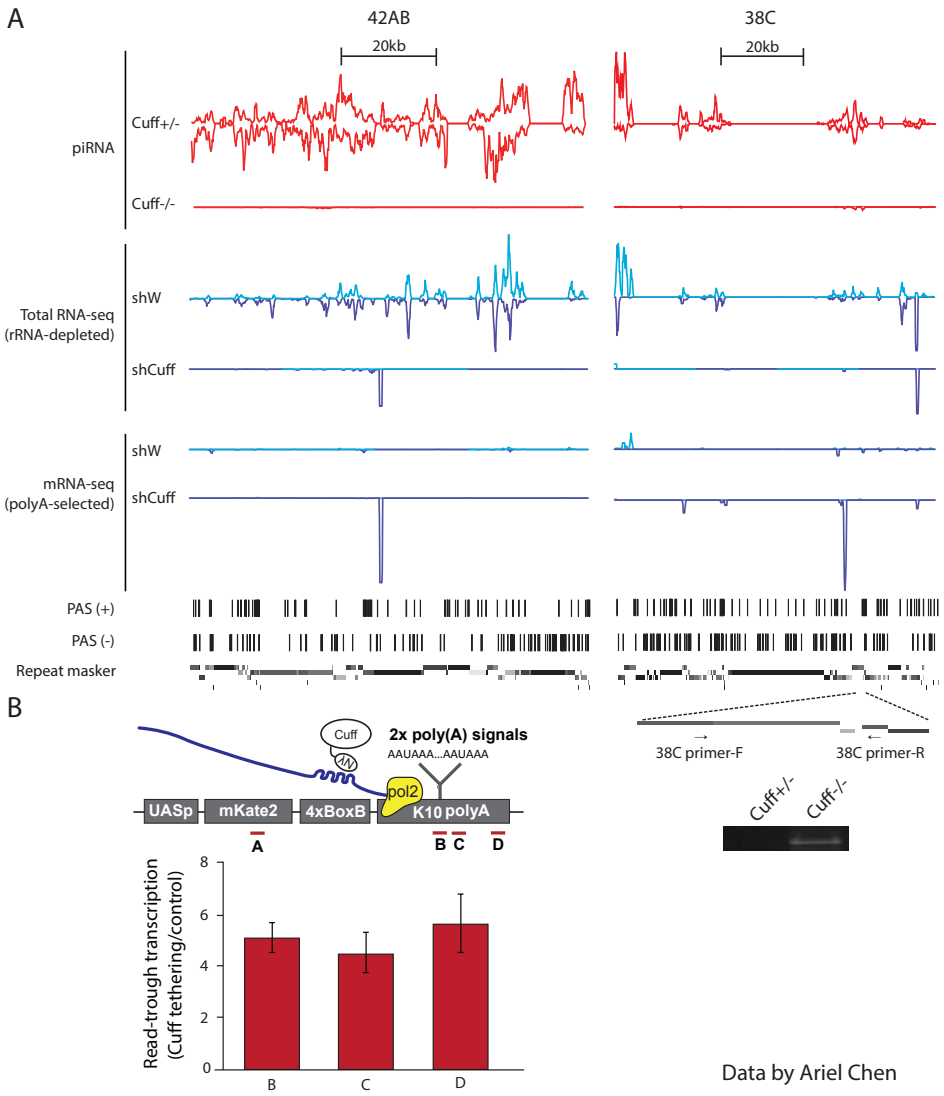

Figure 3.40: **Cuff acts as anti-cleavage factor.** (A)Profiles of strand specific piRNAs, strand-specific total RNA-seq after rRNA depletion and strand-specific polyA-selected RNA-seq from ovarie of Cutoff mutants and heterozygous control or shCutoff flies and shWhite control. Shown are two dual stranded clusters 42AB (chr2R:2,144,349-2,230,172) and 38C (chr2L:20,148,259-20,227,581). RNA-seq and ChIP-seq signals are normalized to total library depth (RPM, reads per million) and plotted with a sliding window (window size: 500 nt, step: 100 nt). The distribution of poly(A) signal (AAUAAA) on plus and minus genomic strands is shown in the PAS tracks. (B) Schematic diagram of the reporter used to study the effect of Cuff recruitment to RNA. The reporter has four BoxB sequences in the 3UTR of mKate2 that recruit the Cutoff protein fused to λN peptide. Both mKate2-4xBoxB reporter and λN-Cuff were expressed in ovaries of transgenic flies. Regions A to D show RT-PCR amplicons shown in the graph below. Cutoff tethering to a reporter causes increase in reporter read-through transcription. RTqPCR was used to measure the level of read-through transcripts (amplicons B to D) normalized to reporter mRNA (amplicon A). The result is shown as fold enrichment in Cuff tethering fly over control fly. Error bars are standard deviations from four biological replicates. All experiments to obtain those data were performed by Ariel Chen

CHAPTER *4*

DISCUSSION

4.1 Tracing piRNA precursors

The main objective of this work was to characterize piRNA cluster transcripts - from transcription to processing - with an RNA sequencing approach. Prior to the start of the investigation it was assumed that piRNA clusters are transcribed by Polymerase II. The reasoning behind this assumption stems from the fact that PolI mainly transcribes ribosomal RNAs and PolIII is responsible for transcription of some ribosomal RNAs in addition to short RNAs like spliceosomal, nucleolar or transfer RNAs. Furthermore, piRNA clusters are very large intergenic regions and of the three polymerases, PolII usually transcribes long transcripts. Furthermore, a mutation in the 5' region of the uni-stranded cluster flamenco abolishes piRNAs throughout the whole cluster (Brennecke et al., 2007), suggesting that piRNA clusters are transcribed as a single unit. Hence PolII was the candidate to perform the continuous piRNA cluster transcription. In fact, experimental data supported PolII dependent transcription of uni-stranded clusters in *Drosophila* and mouse (Li et al., 2013; Mohn et al., 2014; Zhang et al., 2014). But even if piRNA clusters in *Drosophila melanogaster* are transcribed by PolII, they are still very different from canonical gene coding regions. Until now, their transcriptional regulation is enigmatic: no conserved promoter regions have been found yet and several studies suggest that different clusters are transcribed in different ways: Transcription of flamenco seems to be initiated from one distinct promoter region (Brennecke et al., 2007; Mevel-Ninio et al., 2007; Goriaux et al., 2014), whereas 42AB might be transcribed through internal initiation (Pane et al., 2011). Generally, there are different possibilities how piRNA

cluster transcription can be initiated: either from (1) flanking genic promoters, (2) not-yet defined piRNA cluster promoters, (3) from internal promoters, like e.g. the remaining TE promoter sequences or (4) piRNAs are transcribed as individual transcriptional units like 21U-RNAs in nematodes (Gu et al., 2012), or other short capped RNAs (Affymetrix ENCODE Transcriptome and Cold Spring Harbor Laboratory ENCODE Transcriptome, 2009).

The presented work establishes that the two types of piRNA clusters - uni-stranded and double-stranded - greatly differ in the way they are transcribed.

4.1.1 Uni-stranded cluster transcripts are mRNA doppelganger

The biggest uni-stranded cluster "flamenco", was first described and named as a gene. Almost all TE remnants in flamenco are oriented antisense to the direction of transcription. Transcriptional initiation of this cluster was suspected to start at one promoter and result in one continuous transcript (Brennecke et al., 2007; Mevel-Ninio et al., 2007), since mutations in the region upstream of the piRNA signal leads to a complete loss of piRNAs throughout the whole cluster. Flamenco is expressed in somatic follicular cells (Lau et al., 2009; Malone et al., 2009) that only express one of the three Piwi Argonautes, Piwi. Therefore, flamenco encoded piRNAs do not engage in a ping-pong amplification loop. Their antisense nature, though, enables Piwi to directly pair with a nascent transposable element transcript and subsequently influence the chromatin state of the target site. The second big uni-stranded cluster is located upstream of flamenco and is called "20A" according to its cytological position. It is expressed in somatic as well as in germline cells. The main task of somatic piRNA clusters is the suppression of transposable elements, which potentially could infect the germline cells through transport vesicles. Indeed, the most abundant sequences in flamenco are derived from the retro-elements gypsy and ZAM that forms viral particles and invades the germline cells of *Drosophila* ovaries, if flamenco is impaired; even new insertions of gypsy genetic material were found in the F1 generation. (Pelisson et al., 1994; Chalvet et al., 1999; Brasset et al., 2006) The presented data in this thesis show that those two uni-stranded clusters are indeed similar to canonical genes. Cluster 20A shows a strong signal for a capped RNA at one end, which also coincides with a PolII ChIP signal. In case of flamenco the data set showes a weak signal for the 5' cap RNA library, which was confirmed to be the real TSS of flamenco by two recently published studies (Mohn et al., 2014; Goriaux et al., 2014).

Single-stranded piRNA clusters resemble mRNAs not only in *Drosophila*, but also in mouse. A recent study (Li et al., 2013) revealed that pachytene piRNA clusters, which are transcribed in a divergent, uni-stranded fashion (see figure 2.3) also possess a 5' cap. This study additionally observed the presence of polyA tails on those cluster

transcripts, which makes them resemble truly canonical mRNAs. Since somatic cells of *Drosophila* ovaries do not express the two canonical ping-pong Argonautes Aub and AGO3, only the linear, primary piRNA biogenesis pathway is available for processing of flamenco and somatic 20A transcripts. Since this biogenesis route does not depend on complementarity, it is still enigmatic how they are detected by the piRNA biogenesis machinery. If the transcript itself does not differ from canonical mRNAs, how is it channeled into the piRNA pathway? It is conceivable that the chromatin environment of uni-stranded clusters creates a predestining signal. The two most prominent chromatin-associated piRNA biogenesis factors Rhino and Cutoff, however, are not expressed in somatic cells. Possibly, other factors, play decisive roles in the fate of a somatic uni-stranded piRNA cluster transcript, but have yet to be identified.

4.1.2 Double stranded clusters do not have defined transcription start sites

Double stranded clusters are expressed in the germline cells of *Drosophila* ovaries and data of this thesis suggests that their transcription differs from the somatic uni-stranded clusters flamenco and 20A. The biggest double-stranded cluster 42AB is neighbored by the gene Pld, which transcribes into the direction of the cluster. A strong cap signal and a PolII peak clearly define the TSS of this neighboring gene. At the beginning of 42AB the signal for 5' capped RNA drops to approximately 30% intensity of Pld, but is still present throughout the whole cluster. One proposed mechanism of 42AB transcription is that Pld transcription towards the cluster does not properly terminate due to the presence of specific cluster components. In fact, experiments presented in the second part of this thesis support this hypothesis, since the presence of Cutoff in double stranded clusters might promote read-through transcription. This would, however, only explain initiation of transcription from one side of the double-stranded cluster 42AB. Furthermore, the second largest double-stranded cluster 80EF does not possess any converging flanking genes that could explain transcription of this cluster. One possibility is the existence of several mechanisms how different double-stranded clusters are transcribed. Generally, the signal for capped transcripts from double-stranded clusters appears to be at a low level, but can be seen throughout all clusters. A capped flamenco transcript was detected with 4 reads, which were confirmed to represent the actual TSS of this uni-stranded cluster. While it is very difficult to draw any conclusion in this low signal range, signal intensities between 4 and 17 reads throughout all three double-stranded clusters might still represent real 5' capped RNAs. piRNA clusters are strewn with remnants of transposable elements, which in some cases still contain promoter sequences. It

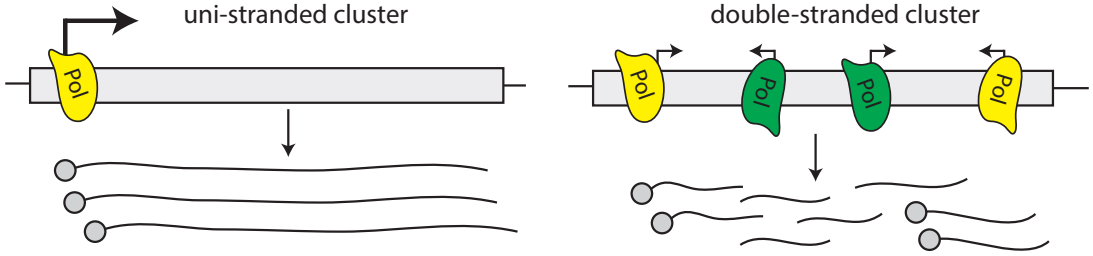

Figure 4.1: **Model of uni-stranded and double-stranded cluster transcription.** Transcription of uni-stranded clusters by PolymeraseII is initiated at a defined transcription start site giving rise to a cluster transcript with a 5' cap. Double-stranded clusters might be transcribed through internal initiation; it is not clear, which Polymerase transcribes those loci.

is therefore possible that the piRNA cluster environment is still permissive for transcription initiation at those promoters, which can be on either genomic strand. An additional anti-terminating environment would then allow transcription of longer regions throughout clusters. The double-stranded nature of clusters generally poses a challenge for transcribing polymerases. What exactly happens when two convergently transcribing polymerases collide, is a highly debated topic and experimental data is sparse. The prevailing model is that they can not bypass each other and while one Polymerase falls off its template, the second one can continue (Hobson et al., 2012). If this model is true and if double stranded clusters were transcribed from two outside promoter regions, transcript levels should be decreasing in both directions along the cluster from the point of transcription initiation. Such a biased distribution, thus, is not apparent. This supports the finding from the 5' cap data suggesting internal initiation of transcription. An alternative explanation for the sparse 5'cap signal throughout double-stranded piRNA clusters could be that those clusters are in fact not exclusively transcribed by PolII. It is known that some transposable elements are not PolII, but PolIII transcripts, like for example the mammalian Alu element (Kramerov and Vassetzky, 2005). Also in *Drosophila* one prominent transposable element, mdg-1, has been described to be PolIII dependent (Arkhipova, 1995). Therefore, internal TE promoters could potentially initiate PolIII transcription. Overall, transcription of double-stranded clusters seems to be non-canonical and future studies will be necessary to reveal regulatory mechanisms.

4.1.3 piRNA precursors are processed in intermediate steps

To answer the question if intermediate piRNA precursors exist, I cloned and sequenced long RNAs possessing 5' monophosphate termini. Those RNA molecules indeed fell within piRNA cluster regions and the quality controls supported that those RNA molecules can be regarded as real endonucleolytic products of piRNA cluster transcripts and not

random degradation products. Therefore, they represent bona fide piRNA precursors. This presence of long piRNA precursors suggests that piRNAs are in fact generated from long cluster transcripts through multiple endonucleolytic cleavage events. Those precursors did not show any apparent bias in the distribution of their 5' end throughout the clusters. The only apparent feature they possess is a nucleotide bias as seen for mature piRNAs. piRNA precursors from uni-stranded clusters show a 1U bias of approximately 60%, which is weaker than the 1U bias of Piwi-bound primary piRNAs of the same origin (roughly 90%), but higher than expected if precursor processing was random. This suggests that piRNA precursors might already be produced in a biased fashion through endonucleases exhibiting a preference for cleavage upstream of a Uridine nucleotide. In addition, 1U piRNA precursors could preferentially be stabilized by their Argonaute partners. Since the bias is still weaker than in a final piRNA population, piRNA biogenesis downstream of the detected precursors seems to further enrich for 1U molecules. The double-stranded cluster derived piRNA precursors also show a 1U bias and additionally already are enriched for 10A. The 10A bias is a feature of ping-pong processing and a recent publication suggests that it is introduced through preferential binding of Aub to 10A RNA molecules (Wang et al., 2014). The presence of the 10A bias in precursor molecules is a further support for the biogenesis model in which a long precursor is anchored in an Argonaute by its 5' end and trimmed from the 3' end.

4.1.4 Zucchini's role in piRNA precursor biogenesis

Detection of piRNA precursors enabled the investigation of piRNA factors that effect precursor biogenesis. One of the most interesting candidates was Zucchini. This protein is a member of the phospholipase-D/nuclease family, which is characterized by the presence of a characteristic HKD motif (Koonin, 1996). Members with two HKD domains fall into the phospholipase-D class, members with only one HKD domain, like Zucchini, are nucleases. Therefore, at the time when this project was started, Zucchini was a strong candidate for being the nuclease to cleave piRNA precursors.

Indeed, structural and biochemical studies published during the course of this project confirmed that Zucchini is an endonuclease cleaving ssRNA (Voigt et al., 2012; Nishimasu et al., 2012; Ipsaro et al., 2012). piRNA cluster transcripts were shown to accumulate in the absence of Zucchini (Haase et al., 2010). The involvement of Zucchini in cluster transcript processing was confirmed by my total RNA sequencing data: with the exception of transcripts from cluster 20A, RNA from all other clusters was elevated between 3 and 8 fold over levels of the shWhite control (see figure 3.16). Accumulated precursors, however, still seem to be endonucleolytically cleaved to an extend comparable to wild type flies (see figure 3.17), which was a very surprising result. If Zucchini was the only nuclease

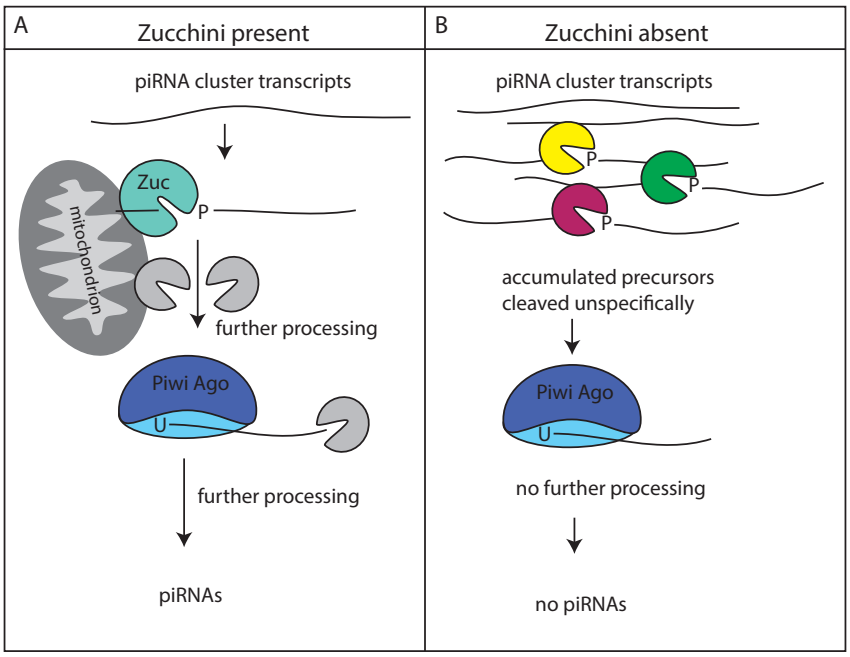

Figure 4.2: **Model for effect of Zucchini knockdown.** (A) In the presence of Zucchini first cleavage of piRNA cluster transcripts takes place at the mitochondrial surface. This specific localization allows for proper subsequent processing steps. (B) In the absence of Zucchini piRNA cluster transcripts accumulate and are cleaved through unspecific endonucleases. 1U molecules are still preferentially stabilized, but further processing is impaired, due to the improper localization.

cleaving cluster transcripts, 5' intermediates should disappear in Zucchini knockdowns. Most strikingly, the residual piRNA precursors still contain a 1U bias, comparable to control levels. One possible explanation for this result is that piRNA cluster transcripts accumulate if Zucchini is not present. Other endonucleases could then compensate for the lack of Zucchini and unspecifically cleave them, likewise leading to 5' monophosphate intermediates. The ones containing a 1U might also still be stabilized through association with an Argonaute, but they are not further processed into piRNAs. A reason for impaired downstream processing of unspecifically cleaved precursors might be the specific sub-cellular localization of Zucchini to the mitochondrial surface. Any other compensating processing could still happen, but further downstream factors needed for further processing (like for example trimmer), might not be attracted to the right subcellular compartment (see figure 4.2).

The precursor sequencing experiments reveal two other factors with an effect on 5' mono-phosphorylated piRNA precursors, Cutoff and Rat1. Those two nuclear factors will be discussed in more detail in section 4.2.6 and 4.2.7.

4.2 Trans-generational effect of piRNAs on source and target loci

4.2.1 maternally inherited piRNAs ensure piRNA biogenesis in progeny

The suppression of transposable elements through the piRNA pathway depends on a maternally inherited trans-generational cytoplasmic factor. When this work was started, the nature of this factor was not known. It could be a protein or small molecule. This PhD work shows that the maternal factors are in fact piRNAs themselves.

de Vanssay and colleagues described that activation of the inactive BX2 locus to produce piRNAs requires a cytoplasmic factor, contributed by the maternal side. Not only activation of a previously inactive locus, but even the maintenance of the active locus requires the presence of the maternal factor (de Vanssay et al., 2012). Those observations could still have been a peculiar special case in this transgenic system consisting of tandem-repeated insertions of heterologous sequences. However, the necessity for a maternally transmitted factor is not only a phenomenon in this one transgenic system, but rather a general principle in piRNA biology.

A dependency of piRNA biology on a maternal factor has been described in several systems of hybrid dysgenesis (Grentzinger et al., 2012; Khurana et al., 2011). Additionally, Brennecke and colleagues showed that piRNAs and Piwi proteins themselves are deposited into the early embryo from the maternal, but not from the paternal side (Brennecke et al., 2008). Also in the P1152/BC69 system, where a sigle-copy transgene is inserted into a piRNA cluster, a target locus was repressed only if the original source locus was present in the mother (see figure 3.27).

Whilst the nature of the maternally transmitted factor was not clear, elegant fly crosses by de Vanssay showed that it is of epigenetic, or non-chromosomal nature. To activate the BX2 locus in the next generation, the active T1 locus itself does not need to be present in the progeny. It is sufficient to transmit the cytoplasmic content of T1 mothers without the actual T1 locus. Hence, the inactive locus does not need any physical interaction with the active one in order to be activated. de Vanssay furthermore conducted experiments that showed that conversion of BX2 into BX2* relied on the presence of Aub, but not on Dicer2, which strongly points to piRNAs being necessary for the conversion (de Vanssay et al., 2012). Since the piRNA sources and targets generally are of the same DNA sequence, a sequence specific recognition of the target locus seems plausible. That piRNAs in fact are suitable to target homologous DNA regions was demonstrated by several studies showing the Piwi:piRNA guided installation of H3K9me3

(Wang and Elgin, 2011; Sienski et al., 2012; Le Thomas et al., 2013; Rozhkov et al., 2013). All those observation hinted towards piRNAs themselves being the necessary epigenetic factor for target silencing and conversion.

The data presented in this thesis suggests that piRNAs themselves indeed induce a plethora of changes on target loci after only one generation and therewith are the activators of piRNA biogenesis.

4.2.2 Maternally inherited piRNAs induce ping-pong amplification

Maternally inherited piRNAs fuel the ping pong amplification loop in the cytoplasm and therefore increase production of secondary piRNAs. This effect of maternally deposited piRNAs is apparent by the dramatic increase of ping-pong piRNAs in progeny of maternal deposition crosses (figure 3.28B). When the ping pong cycle for secondary piRNA biogenesis was originally proposed, a division between primary and secondary piRNA substrates was made: primary piRNAs are derived from piRNA cluster transcripts (since those contain antisense fragments of transposable elements) and secondary piRNAs were produced from mRNAs of active transposable elements (Aravin et al., 2007; Brennecke et al., 2007), since they are in sense-orientation.

Maternally inherited primary piRNAs can additionally induce secondary, ping-pong dependent piRNA biogenesis from any bi-directionally transcribed target locus. Since germline clusters are (with exception of cluster 20A) transcribed bidirectionally, this explains how maternally deposited piRNAs ensure the identification of cluster transcripts as substrates and therefore piRNA production. This mode of cytoplasmic piRNA production relies on the cytoplasmic ping-pong partners Aub and Ago3 and takes place in the perinuclear nuage. Aub has been shown to localize to the pole plasm of *Drosophila* eggs, from where the later germline cells start migrating. Therefore, the here presented data provide proof that those maternally deposited piRNAs indeed are not only targeting active transposable elements, but also utilize cluster transcripts to ensure piRNA production in the progeny.

4.2.3 piRNA clusters rely on inherited piRNAs to stay active

How crucial maternally inherited piRNAs are for the proper workings of piRNA clusters is illustrated when a piRNA producing locus is transmitted without its accompanying piRNAs. The T1 strain and BX2 strain contain the same 7 copy LacZ tandem array in the same cytoplogical position (Dorer and Henikoff, 1997). X-ray irradiation of BX2

converted the LacZ-array transgene into a piRNA producing locus (T1) (see figure 3.21). Characterization of the T1 and BX2 locus led to the conclusion that the X-ray irradiation did not lead to changes in the internal structure of the cluster (Ronsseray et al., 2001). BX2 can be converted into a piRNA producing locus upon exposure to cognate piRNAs from T1 and the conversion is stable in the subsequent generations (de Vanssay et al., 2012). Whatever distinguishes BX2 from T1 is overcome by only one exposure to piRNAs.

When T1 is transmitted to its progeny through the paternal germline, the progeny genome still contains the LacZ tandem array, but is not accompanied by congnate piRNAs. H3K9me3 levels on the T1 locus drop to approximately 50% (figure 3.22) and it does not silence a reporter anymore (figure 3.25). If there was some special genetic component of this tandem-array leading to the emergence of piRNA production by itself, the speculation was that it should slowly recover piRNA production over time. When transmitted through the female line in multiple generations, those piRNAs would be passed on into next generations and slowly lead to a stable re-activation of T1. Aging of female flies of dysgenic crosses showed that the amount of piRNAs transmitted into the next generation correlates with the age of the mother; dysgenic flies recover partial fertility when they age (Bucheton, 1979; Grentzinger et al., 2012). The progeny of aged females, in turn, has a higher fertility rate. Based on this observations, flies of the paternal T1 transmission lines were aged in one arm of the experiment (figure 3.24). However, none of the investigated generations starts accumulating H3K9me3 over the T1 transgene or is able to silence a reporter, both of which are measures for a recovery of piRNA production.

piRNA clusters in *Drosophila* are located in pericentromeric and sub-telomeric heterochromatic regions (Brennecke et al., 2007), which are generally decorated with H3K9me3. Position effect variegation experiments in *Drosophila* revealed that heterochromatin is able to spread (Muller, 1930) and thus, the piRNA clusters could acquire heterochromatic features from their neighbouring regions. It is concievable that the special location of piRNA clusters provided them with a required minimal initial amount of H3K9me3 that is in turn necessary for piRNA production. Since the LacZ insertions in T1 and BX2 are located in a euchromatic region, they do not have a nearby source of H3K9me3 and therefore those loci might not be able to acquire piRNA producing capability without exposure to cognate piRNAs. The idea that piRNA clusters need to be accompanied by their own piRNAs in order to stay active is furthermore supported by the results of H3K9me3 Chip on the P1152 strain (figure 3.31A). In the NMD cross not only the target locus loses H3K9me3 levels, but also the source, which is equally transmitted into the progeny through the paternal side.

Together, this once more illustrates that maternally transmitted piRNAs are the main factors for piRNA biogenesis themselves. The genetic content of the locus does not suffice

to start piRNA production. Possibly the right positioning within the genome close to heterochromatic regions, together with other unknown factors would facilitate the *de novo* establishment of a piRNA producing locus.

4.2.4 Maternally inherited piRNAs act as an epigenetic switch

A very interesting observation in the MD and NMD crosses of the P1152 system is the presence of Piwi associated primary piRNAs from targets only in the presence of maternally inherited piRNAs. This seemingly small detail suggests a whole new mode of action of transgenerationally inherited piRNAs. They do not only ensure piRNA production in the next generation through fueling of the ping-pong cycle, but additionally induce primary piRNA biogenesis from target loci. Exposure to maternally inherited piRNAs leads to an accumulation of H3K9me3 on the targeted regions in the T1/BX2 system as well as in the P1152 system (figure 3.31A, figure 3.22 and figure 3.29). This observation alone could also simply suggest transcriptional silencing of the target through heterochromatin formation. However, in the case of T1/BX2, transcription was not decreased and simultaneously, piRNA biogenesis was activated. H3K9me3 accumulation has two seemingly contradicting effects in the context of the piRNA pathway. It is present on piRNA clusters and there it is necessary for their activity. At the same time it is deposited onto piRNA targets and leads to their silencing (Sienski et al., 2012; Le Thomas et al., 2013; Rozhkov et al., 2013; Wang and Elgin, 2011). "Silencing" in this case, however, might not mean transcriptional silencing, but in fact induction of piRNA processing from the targeted region. The conversion of a target into a piRNA producing locus is apparent at endogenous loci targeted by transgenic piRNAs (figure 3.30) and has been confirmed in a study by Shpiz and colleagues, that investigated transposable element insertions (Shpiz et al., 2014). If targets of piRNA repression become a source of new primary piRNAs themselves, a distinction between piRNA clusters as sources and their targets is obsolete: every target is converted into a "mini piRNA cluster". As soon as a nascent transcript can be targeted by a complementary, Piwi bound piRNA, this transcript is channeled into the primary piRNA processing machinery. Therewith the output of activation of piRNA biogenesis on a locus and also its silencing are one and the same: cleavage of the target transcript ultimately prevents the expression of its gene product. The exact mechanism for piRNA-dependent deposition of the H3K9me3 mark on piRNA regions remains to be elucidated. However, recent studies suggest that it might occur through recognition of nascent transcripts by the nuclear Piwi/piRNA complex, which is known to be deposited by the mother into the developing egg (Megosh et al., 2006; Brennecke et al., 2008) and has been shown to install H3K9me3 on its genomic targets (Sienski et al., 2012; Le Thomas et al., 2013; Rozhkov et al., 2013; Wang

and Elgin, 2011).

4.2.5 H3K9me3 attracts piRNA biogenesis factors

Trimethylated histone tails of histone H3 (H3K9me3) are generally bound by HP1 (heterochromatin protein) family proteins, which - as their name already suggests - lead to heterochromatinization and subsequent transcriptional silencing. In the piRNA biogenesis pathway, however, H3K9me3 is correlated with an activation of cluster transcription and piRNA biogenesis from targets. A study by Rangan and colleagues showed in a genome wide H3K9me3 experiment, that piRNA clusters are enriched in H3K9me3 and that cluster transcription and piRNA production is severely decreased upon mutation of the histone methyl transferase SETDB1/Egg (eggless) (Rangan et al., 2011). Their data furthermore show that not only piRNA production, but also siRNA production from piRNA clusters is affected, which suggests that H3K9 tri-methylation indeed is needed for transcription of piRNA clusters. Interestingly, pure siRNA source loci produce more siRNAs upon loss of SETDB1, which agrees with the traditional view of H3K9me3 as a repressive mark. How can H3K9me3 compact chromatin in one case and keep it permissive in the other? One possible mechanism to define H3K9me3 as an activating mark could be its interpretation through specific factors. Indeed, not only H3K9me3 was increased on target loci in a maternal deposition dependent manner. Also the two piRNA biogenesis factors Rhino and Cutoff show an increased occupancy on targets when maternally deposited piRNAs are present (figure 3.31B and C). Rhino is a heterochromatin protein 1 (HP1) homolog, specific to the *Drosophila* germline. It binds to H3K9me3 through its chromodomain and was shown to localize to double-stranded clusters in complex with the linker protein Deadlock and Cutoff (RDC complex) (Mohn et al., 2014). In this complex Rhino seems to be the chromatin reader and Cutoff the actual effector protein. The effect of H3K9me3 as an activating mark has analogies to the siRNA system in the fission yeast *Saccharomyces pombe*. There, H3K9 methylation induces binding of Swi6/HP1, which then recruits the Jumonji protein Epe1 that promotes nucleosome turnover, resulting in increased transcription of heterochromatic repeats and generation of siRNAs (Zofall and Grewal, 2006; Grewal and Elgin, 2007). Results of this thesis suggest that Cutoff is the actual activator of piRNA biogenesis on piRNA clusters and targets.

4.2.6 Cutoff acts as an anti-terminator

The decreased transcription of double-stranded clusters upon knockdown of Rhino and Cutoff (figure 3.35) places those proteins directly in the process of cluster transcription and not in any downstream process, like for example stabilization of cluster transcripts.

Rhino depletion leads to mis-localization of Cutoff in *Drosophila* germ cells and is believed to be the recruiter for Cutoff to double-stranded clusters (Mohn et al., 2014). Therefore, the effects seen in Rhino-depleted ovaries are most likely caused by the lack of Cutoff at piRNA clusters. How does Cutoff promote transcription on piRNA clusters? Two possible mechanisms are conceivable. First, Cutoff could increase transcription at the initiation steps, or second, Cutoff could prevent transcription termination. In fact, the experimental evidence speaks in favor for the second mechanism of anti-termination. When Cutoff is recruited to endogenous genes that are targeted by piRNAs, they show increased read-through transcription (figure 3.30). This suggests transcription trespasses the termination signal of those genes, leading to failure of termination. Transcription termination of genes depends on detection and cleavage of a poly(A) signal. One model of transcription termination, called "Torpedo model" (Connelly and Manley, 1988; Proudfoot, 1989), suggests that the cleavage and polyadenylation specificity factor (CPSF) together with the cleavage stimulatory factor (CstF) recognize the poly(A) signal and then promote cleavage of the nascent transcript a few hundred nucleotides downstream (Proudfoot, 2011; Mandel et al., 2008). The upstream mRNA fragment subsequently obtains a poly(A) tail through a nuclear poly(A)-polymerase. The downstream 3' fragment is still connected to the transcribing PolII and contains an unprotected 5' monophosphate terminus. This fragment is degraded by a 5' - 3' exonlucleoleolytic activity, which induces dissociation of PolII from the DNA template (West et al., 2004). According to this model, recognition of poly(A) signals ultimately leads to transcription termination. Reciprocally, if Cutoff prevents detection of poly(A) signals, this would explain the read through transcription on piRNA target loci. Tethering experiments in our laboratory confirmed this hypothesis (figure 3.40). piRNA clusters are covered with transposable element fragments, and likewise with poly(A) signals. A prevention of Poly(A) site detection would therefore ensure cluster transcription. And indeed, in wild type *Drosophila* ovaries, no poly-adenylated cluster transcripts are detectable. Only after Cutoff depletion, a few prominent transcripts containing a poly(A) tail appear. Therefore, Cutoff indeed seems to suppress poly(A) site detection and/or cleavage and ultimately prevent transcription termination.

4.2.7 Cutoff protects free 5' monoposhate ends

Cutoff is a homolog of the yeast proteins Rai1 and Dxo1 and the mammalian protein Dom3Z, all of which are implied in transcription termination and RNA quality control (Kim et al., 2004; West et al., 2004; Luo and Bentley, 2004; Xiang et al., 2009). The yeast homolog Rai1 was found as an interacting partner of the nuclear 5' - 3' exonuclease Rat1/XRN2 (Rai1 stands for Rat1 interacting protein 1)(Xue et al., 2000). It

seems to stimulate the exonucleolytic activity of Rat1/XRN2. Rat1/XRN2 is also called "torpedo nuclease", since there is evidence for it to be the actual exonuclease leading to transcription termination after poly(A) site cleavage (Kim et al., 2004).

Interestingly, immunoprecipitation experiments in our lab show that Cutoff also seems to physically interact with Rat1/XRN2. One hypothesis is that Cutoff prevents Rat1/XRN2 from degrading free 5' monophosphate termini and therefore prevents displacement of PolII from the DNA template. Structure predictions of Cutoff suggest that while it lacks a catalytic site, it still has an RNA binding pocket. Efforts to detect an RNA binding activity specific to a 5' modification of RNA (like mono- or tripple-phosphate, caps or hydroxyl groups) have not yielded any definite result. The piRNA precursor sequencing data, however, suggest a protective effect of Cutoff against Rat1/XRN2 degradation. In the absence of Cutoff 5' monophosphorylated piRNA precursors are strongly reduced, which suggests a protective effect of Cutoff on 5' monophosphate precursors. In contrast, in the absence of Rat1/XRN2, precursor molecules are accumulated, which implies Rat1/XRN2 in the exonucleolytic degradation of piRNA precursors (figure 3.39). 5' monophosphorylated nascent piRNA cluster transcripts could occur if prevention of poly(A) site detection is incomplete, or if any other internal signal leads to endonucleolytic cleavage of the nascent transcript.

Together, the data lead to a model in which Cutoff ensures piRNA cluster transcription through two mechanisms 4.3. On the one hand, it prevents poly(A) site cleavage and subsequent transcription termination. On the other hand, it might prevent degradation of unprotected 5' monophosphate ends of cluster transcripts. Those could occur through incidental cleavage of nascent cluster transcripts or through other mechanisms related to the noncanonical transcriptional features of double stranded piRNA clusters.

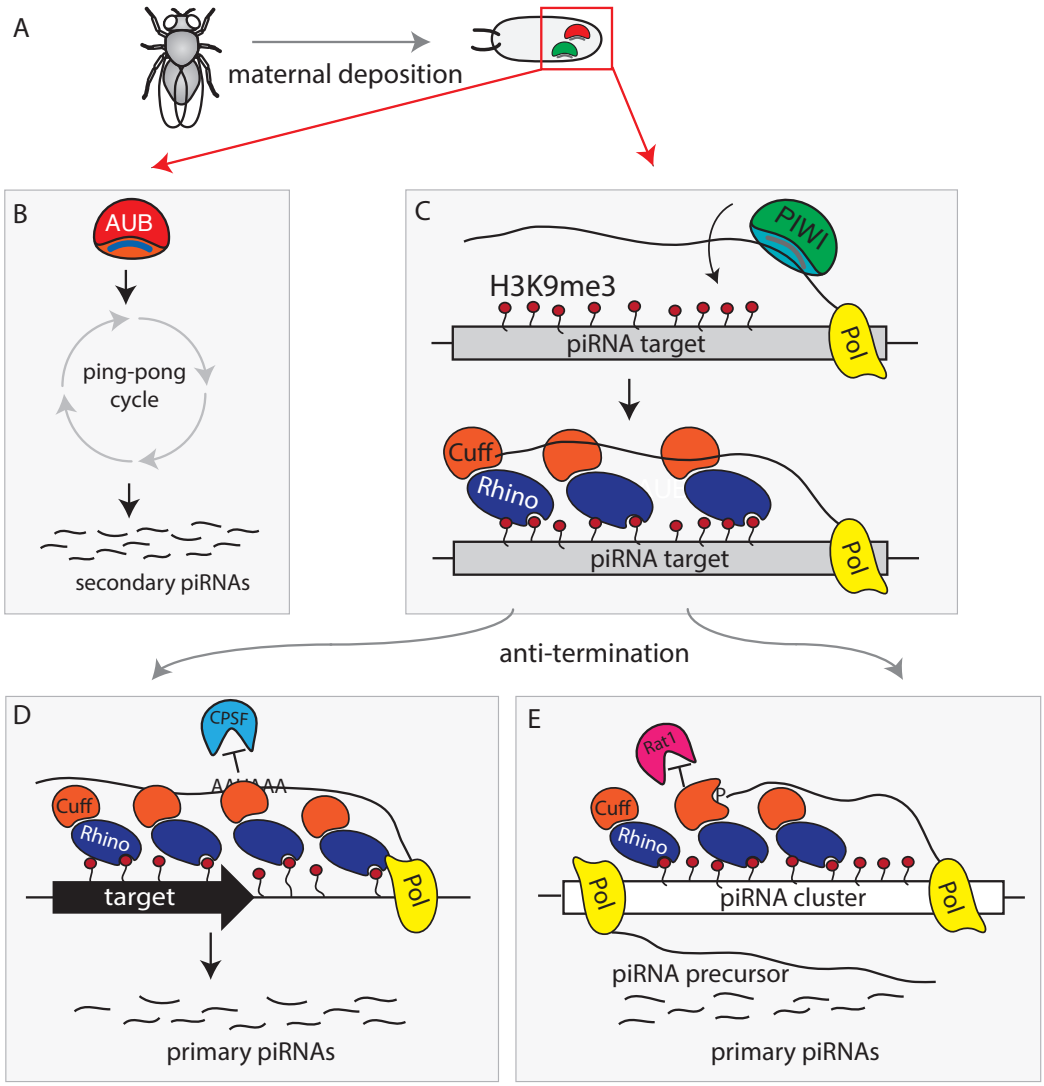

Figure 4.3: **Model for the transgenerational effect of piRNAs and subsequent changes on target loci.** (A) piRNAs derived from clusters are loaded into Aub and Piwi and maternally transmitted into the next generation. (B) Aub bound piRNAs feed into the ping-pong amplification loop, ensuring secondary piRNA biogenesis. (C) Piwi reinforces piRNA production from the source loci as well as from targets. Establishment of H3K9me3 on target regions attracts Rhino and Cutoff. Cutoff ensures transcription within clusters and on targets through suppression of polyA site detection and cleavage (exemplified on the target locus (D)). And through protection of free 5' monophosphorylated nascent RNA transcripts (exemplified on piRNA cluster (E)).

CHAPTER 5

CONCLUSIONS

This work provides insights into the early steps of piRNA biogenesis, as well as the transgenerational effect of piRNAs. It furthermore establishes the role of Cutoff as a piRNA biogenesis factor acting on a transcriptional level.

Results presented in this work establish that the two different classes of piRNA clusters, uni-stranded and double-stranded clusters, greatly differ in their transcription. Transcription of uni-stranded clusters is promoted at defined regions and produces long 5' m^7G-capped transcripts spanning the whole cluster. Double-stranded clusters might be transcribed through initiation within the cluster and not from one defined promoter. An alternative hypothesis is their transcription by RNA polymerase I or III, which do not give rise to 5' m^7G-capped transcripts.

piRNA cluster transcripts need to undergo processing to give rise to mature piRNAs. This work reveals the existence of intermediate piRNA precursor molecules, which are longer than 200 nucleotides and already possess characteristics of mature piRNAs: their 5' termini are monophosphorylated and exhibit a 1U bias, although weaker than the one observed for mature piRNAs. Therefore, downstream piRNA biogenesis steps possibly select for 1U precursors and thus enhance the 1U bias to reach the extend seen in mature piRNA populations.

Mature piRNAs guide target cleavage and are furthermore necessary to keep the piRNA biogenesis machinery active in subsequent generations.

Following a piRNA cluster that was paternally transmitted into the next generation and thus not accompanied by its homologous piRNAs, revealed the importance of ma-

ternally inherited piRNAs. The locus did not recover piRNA production, even after multiple generations. Therefore, maternally transmitted piRNAs are key components for the piRNA biogenesis machinery.

This work reveals the mechanism of piRNA production through maternally inherited piRNAs: both piRNA generating pathways are activated, primary as well as secondary piRNA biogenesis.

Secondary piRNA biogenesis is induced through activation of the cytoplasmic ping-pong cycle.

Concomitantly, primary piRNA biogenesis is induced at targeted loci. Inherited piRNAs lead to accumulation of H3K9me3, Rhino and Cutoff at homologous regions. This unexpected result challenges the traditional view of piRNA sources and targets. It suggests that there might in fact not be a mechanistic distinction. Any targeted locus converts into a piRNA cluster and therefore, piRNA production might be a mechanism of target silencing.

Overall, this work puts maternally inherited piRNAs at the core of the piRNA pathway and defines them as epigenetic components maintaining the piRNA biogenesis machinery.

The second part of this thesis establishes Cutoff as a key factor in piRNA biogenesis. It prevents polyA-site detection and cleavage and additionally protects free 5' termini from exonucleolytic degradation. Both mechanisms prevent transcription termination and subsequently lead to noncanonical transcription beyond stop-signals.

Cutoff can therefore be placed at the initial phase of piRNA biogenesis: without Cutoff, there are no cluster transcripts and hence no substrates for piRNA biogenesis.

CHAPTER 6

METHODS AND MATERIALS

6.1 Standard prcedures

6.1.1 RNA isolation

Depending on the experiment, RNA was isolated using two different methods.

6.1.1.1 Classic Phenol/Chloroform Extraction

The RNA sample (mostly enzymatic reaction samples) was brought to a volume of 500 µl, containing a final concentration of 300 mM NaCl. 500 µl of acidic Phenol/Chloroform was added to the sample and the mixture was vortexed. After a spin at 21,000 g at 4 °C for 15 min, the upper aequous phase was taken off and combined with an equal volume of Chloroform (450 µl). The mixture was vortexed, span at 21,000 g and 4 °C for 10 min. The upper aequous phase was carefully taken off and 1 µl co-precipitant Clyco-Blue was added to later visualize the RNA pellet. RNA was precipitated by adding 3 volumes of 100% Ethanol and incubation at -20 °C or -80 °C for one hour up to over night. The sample was span at 4 °C at 21,000 g for 1 hour. The RNA pellet was washed once in 1 ml of 70% Ethanol and air dried. RNA was then resuspended in the desired amount of nuclease-free DEPC treated water.

6.1.1.2 RNA extraction using Ribozol Reagent

If the sample consisted of a lysate, cell-pellet or tissue (like ovaries), I used Ribozol for RNA extraction. The volume of a liquid sample should be below 250 µl. Ribozol was

added to yield a total volume of 1 ml. If necessery, the sample was homogenized using a plastic pestle or a tissue grinder. After vortexing the sample, 200 µl Chloroform were added and the sample was vortexed again and let sit on ice for 5 min. The sample was then centrifuged at 21,000 g at 4 °C for 15 min. The upper aequous layer was carefully taken off and 1 µl of co-precipitant and 500 µl Isoopropanol was added. RNA was precipitated on ice or at -20 for 30 min up to over night. After spinning the sample at 21,000 g for 1 hour, the pellet was washed in 1 ml 70% Ethanol and subsequently air dried and resuspended in the desired amount of nuclease-free DEPC treated water.

6.1.2 Quantification of nucleic acids

The concentration of nucleic acids were measured with either a spectrophotometer (on a Nanodrop device) or fluorometrically (with a Qubit device).

6.1.2.1 Spectrophotometric DNA and RNA quantification

Quantification of nucleic acids on a spectrophotometer relies on the absorption of UV light of 260 nm by a DNA or RNA sample. However, any components of a sample, which also absorbs UV light in this wavelength range, can distort the measurement. The detection limit of a Nanodrop lies around 2 ng/µl. For standard nucleic acid quantification of samples with a sufficient amount of RNA or DNA, Nanodrop was used. For samples with low concentrations, or whenever a very precise determination of the concentration was necessary, nucleic acids were quantified on a Qubit.

6.1.2.2 Fluorometric DNA and RNA quantification

To measure the concentration of a sample on a Qubit device, the sample is mixed with a dye, which specifically interacts with RNA or DNA and exhibits fluorescence only upon binding of the target. Fluoresence of the sample is measured by the Qubit fluorometer and the sample concentration is calculated according to the measurements performed with reference samples. Fluorometric DNA and RNA quantification is more specific than specrtophotometric measurements and has a higher sensitivity. For quantification of RNA samples the Qubit RNA Assay with a concentration range of 250 pg/µl to 100 ng/µl was used. For quantification of dsDNA samples, like final libraries for Illumina sequencing, the Qubit dsDNA High Sensitivity Assay was used, which has a detection range between 10 pg/µl and 100 ng/µl.

6.1.3 PCR

For analytical PCR reactions a DNA polymerase from *Thermus aquaticus* (Taq-Polymerase) was used, preparative PCR reactions were performed with a DNA polymerase from *Thermococcus kodakaraensis* (KOD-Polymerase). This Polymerase contains a 3' - 5' proof-reading activity, which decreases the mutation rate in target amplification.

Typical reactions are shown in the tables below

component	final concentration
10 x KOD Buffer	1 x
dNTPs	0.2 mM
$MgCl_2$	1 mM
Template DNA	100 ng
forward primer	0.4 µM
reverse primer	0.4 µM
KOD DNA Polymerase	0.02 units
water	to 50 ul

component	final concentration
10 x Taq Buffer containing Mg^{2+}	1 x
dNTPs	0.2 mM
Template DNA	100 ng
forward primer	1 µM
reverse primer	1 µM
Taq DNA Polymerase	0.05 units
water	to 50 ul

6.1.4 DNase treatement

Prior to reverse transcription, all RNA samples were treated with DNase, Amp Grade from Invitrogen. The following components were mixed:

RNA	1 µg
10 x DNase I Reaction Buffer	1 µl
DNase I, Amp Grade, 1 U/µl	1 µl
DEPC-treated water	to 10 µl

The digest was incubated at RT for 15 minutes. To inactivate the DNase I, 1 µl of 25 mM EDTA was added to the reaction, which was then heated to 65 °C for 10 minutes.

6.1.5 Reverse transcription

Reverse transciption of RNA was performed with the SuperScriptIII Reverse Transcriptase from Invitrogen. For reverse transciption with random hexamers or oligodT primers, the reactions were performed according to the standard protocol from the supplier. For strand specific reverse transciption using gene specific primers containing a universal adapter sequence, reaction conditions were modified. For a typical reaction the following components were mixed in a tube:

DNase treated RNA	1 µg
random hexamers or oligo dT primer [50 µM]	1 µl
dNTP [10 mM]	1 µl
DEPC-treated water	to 13 µl

This mixture was heated to 65°C for 5 minutes to denature and linearize RNA molecules. Subsequently, the sample was chilled on ice for at least 1 minute. After a brief centrifugation in a microfuge, the following components were added:

5 x First-Strand Buffer	4 µl
DTT [100 mM]	1 µl
RNAseOUT	1 µl
SuperScript III RT	1 µl

The sample was mixed gently by pipetting up and down and was then incubeated at 25°C, if the reverse transcription was primed with random hexamers or oligo dT primers. Subsequently, the reaction was incubated at 50°C for 60 min. The reverse transcriptase was inactivated by heating the sample to 70°C for 15 min.

For experiments where reverse transcription was performed with gene specific primers (detection of read-through transcription upon maternal deposition of piRNAs), 2 pmol of each required specific primer was added instead of random hexamers or oligo dT primer. Before adding the SuperScript enzyme, the sample was heated to 60°C for 5 min. The enzyme was added and reverse transcription was performed at 60°C for 60 min.

6.1.6 Quantitative PCR (qPCR)

Quantitative PCR was performed using a 10,000 SYBR Green I Nucleic Acid Gel Stain and the MyTaq HS Mix on a Realplex Mastercycler, epgradientS machine from Eppendorff. The SybrGreen was used in a lower amount than recommended by the supplier. Prior to the experiment, the 10,000 x SybrGreen was diluted in DMSO 6,000 fold to yield a 1.7 x stock. Of this dilution 0.2 ul were used in a 20 ul reaction volume.

A typical reaction is listed in the table below.

2 x MyTaq Master Mix	10 µl
forward primer [10 µM]	1 µl
reverse primer [10 µM]	1 µl
SybrGreen	0.2 µl
template	1 µl
water	6.4 µl

Results were exported from the Mastercycler in excel format and analyzed in excel.

6.1.7 Strand specific RT and qPCR for detection of read through transcription

6.1.8 DEPC treatment of water

To ensure nuclease-free experimental conditions, all buffers and reagents for RNA work were made with DEPC treated water. Water was filtered through a Barnstead filter (D3750) with a pore size of 0.2 µm in a NanoPure apparatus from Thermo Scientific.

DEPC modifies -NH, -SH and -OH groups in RNases and other proteins and thus inactivates their enzymatic activities.

1 ml of DEPC was added to 1 liter of NanoPure filtered water, incubated at RT over night and then autoclaved.

6.1.9 T7 RNA transcirption

As template for T7 RNA transcirption a PCR product with a T7 promoter sequence at the 5' end of the forward primer was used. For a typical reaction 1-2 µg of PCR template was used. The reaction was set up as follows:

Template (PCR product)	1-2 µg
rNTPs [80 µM each]	2.5 µl
10x T7 RNA transcription buffer	10 µl
T7 RNA polymerase	2 µl
DEPC-treated water	up to 100 µl

The reaction was incubated at 37 °C for 2-4 hours. During the reaction pyrophosphate precipitate lead to a turbid appearance of the mixture, which did not affect the reaction. 15 µl EDTA, 200 µl water and 50 µl 5M Sodium-Acetate were added to the reaction and thoroughly mixed. RNA was extracted with Phenol/Chloroform as described in section

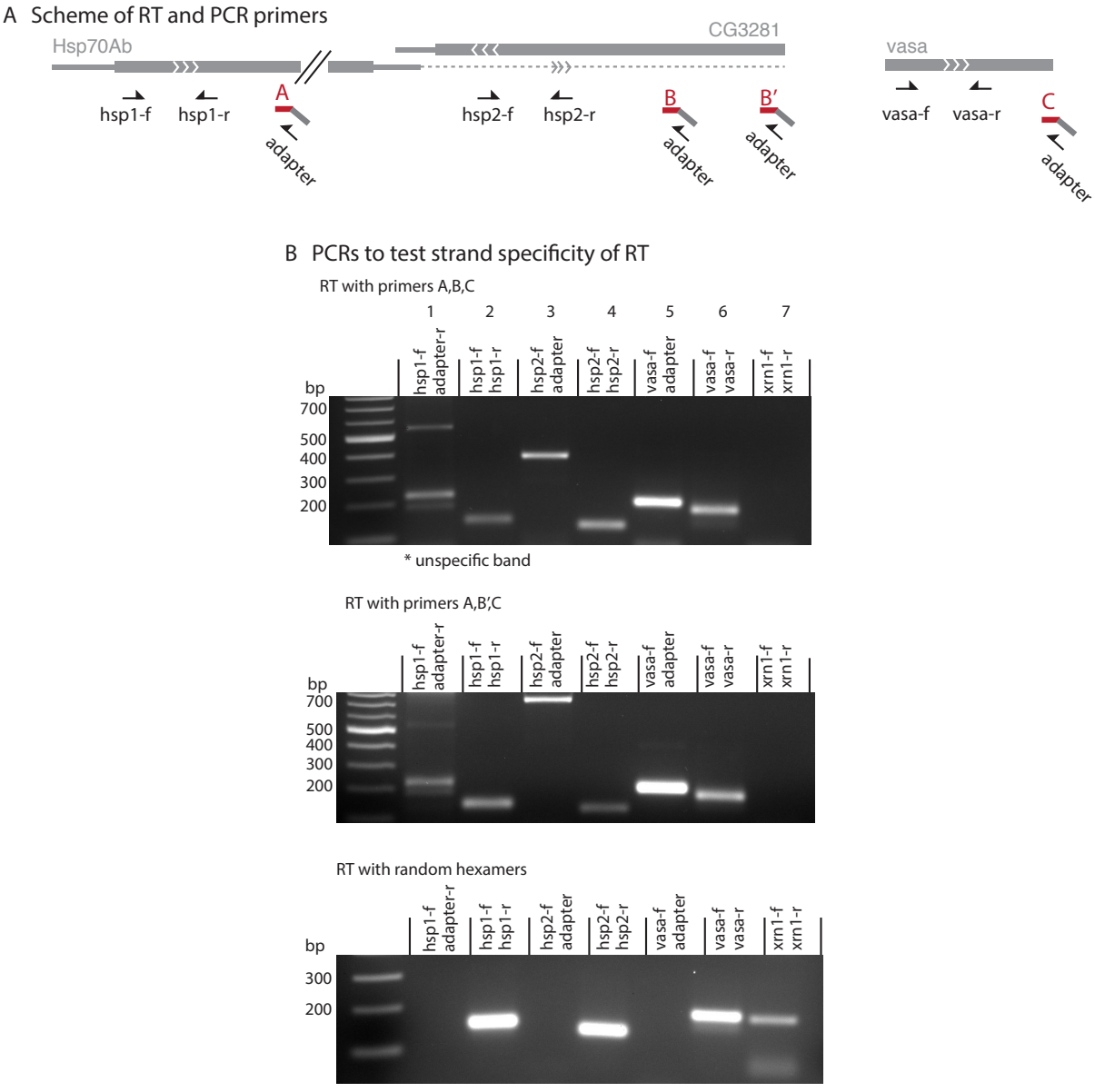

Figure 6.1: Primers for RT and test qPCR to ensure strand specific RT

6.1.1.1 with the modification that RNA was precipitated with isopropanol instead of ethanol. For some applications a DNase digest was performed to remove the template DNA.

6.2 Specific Experimental Methods

6.2.1 Chromatin Immunoprecipitation - ChIP

For each ChIP experiment, 100 flies were put on yeast for 2 to 3 days prior to ovary dissections. Ovaries were dissected in PBS and fixed using Paraformaldehyde (PFA) at a final concentration of 1% and incubated for 10 min at RT. Samples were quenched by adding Glycine to a final concentration of 25 mM. Quenching was performed for 5 min at RT, then ovaries were washed 3 times in PBS. Ovaries were afterwards washed twice in 1 ml Farnham Buffer prior to sonication. Ovaries were resuspended in 600 µl RIPA buffer, dounced with a tight pestle B in a Kontes Douncer with 25 strokes and were incubated on ice for 10 min. The lysate was divided into 300 µl aliquots into sonication tubes. Sonication was performed in a Bioruptor from Diagenode on Medium power for 20 cycles (30sec ON, 30sec OFF). Samples were centrifuged at 14,000 rpm for 15 minutes at 4 °C . The supernatant was collected and pre-cleared with 20 µl of magnetic Protein G coupled beads (Dynabeads, Invitrogen) for 2h at 4 °C . Meanwhile, antibodies were conjugated to Dynabeads Protein G for 2h at 4 °C . 5% of pre-cleared samples were saved for Input fraction and the rest was transferred to the antibody-conjugated beads and incubated for 2h at 4 °C . Beads were then washed with LiCl IP Buffer 5 times at 4 °C, then rinsed in TE and finally resuspended in 200 µl Proteinase K Buffer and digested with 100 µg proteinase K. Samples were incubated 3h at 55 °C then overnight at 65 °C. DNA was extracted following standard phenol-chlorophorm extraction and concentration was measured by Qubit (see section 6.1.2.2)

6.2.2 Nuclear Run-On / GRO-Seq

The nuclear run-on procedure was carried out as previously described (Shpiz et al., 2011) with slight modifications.

6.2.2.1 Preparations

To prepare a miracloth filter, the bottom of a 0.6 ml eppendorf tube was cut at 1/3 of the tube height and the cut side was covered with a piece of Miracloth. This sieve was put into a 1.0 ml eppendorf. To prepare the sucrose cushion, 800 µl of Buffer HB-B

was pipetted into a 1.5 ml tube and an upper phase of 300 µl Buffer HB-A was carefully layered on top.

6.2.2.2 Nuclei isolation

Flies were put on yeast 2-3 days prior to the experiment. For one experiment 150 ovaries were dissected in PBS and washed in PBS once. After washing, PBS was taken of and the ovaries were resuspenden in 300 µl of Buffer HB-A. They were transfered into an ice-cold Douncer and homogenize with 20-25 strokes of tight-fitting pestle B holding the homogenizer on ice. Subsequently, the lysate was filtered through the Miracloth filter in a microfuge. The filtrate was carefully pipetted onto a sucrose cushion and centrifuged for 10 min at 10,000 g at 4 °C . The supernatant was removed and the nuclei-pellet was resuspendet in 500 µl RB buffer and span at 6,000 g for 10 min to wash nuclei (to remove HB buffer and endogenous pool of triphosphates)

6.2.2.3 nuclear run on reaction

The washed nuclei were resuspended in 100 µl RB buffer containing 0.5 mM of each ATP, GTP, CTP and BrUTP. The run-on reaction was performed at 25 °C for 30 min. To stop the run-on reaction, 400 µl of Ribozol was added and the mixture was vortexed for 30 sec and let sit on ice for 3 minutes. Then, 100 µl Chloroform were added, vortexed again and let sit on ice for 5 min. The sample was span at maximal speed in a tabletop centrifuge at 4 °C. The top aequous layer was taken off and combined with 1 µl Glycogen co-precipitant (GlycoBlue) and 250 µl Isopropanol. The RNA was precipitated at -20 °C for 30 min and span for 30 min at 4 °C and 21,000 g in a tabletop centrifuge. The RNA was resuspendet in 50 µl nuclease-free water

6.2.2.4 Immunoprecipitation of run-on RNA

BrUTP labeled run-on-RNA was filtered through Illustra MicroSpin G25 columns twice to remove unincorporated BrUTP. After taking 10 µl of the sample as "Input", 260 µl of IP Buffer was added to the RNA. 2 µl anti-BrdU antibody and 1 µl RNAse inhibitor (RNAsin) were added and the sample was incubated on a rotator in the 4 °C room for 1 hour. At the same time, ProteinG Dynabeads were blocked at RT in 1 ml IP Buffer containing 0.1% PVP and 1 mg/ml BSA. After blocking, the beads were washed in IP buffer once and combined with the RNA-Antibody sample. Immunoprecipitation was performed for 1 hour at 4 °C followed by three washes in IP buffer for 5 min rotating at 4 °C. After the final wash, wash buffer was removed and the beads were resuspended in 500 µl Ribozol and RNA was extracted following the standard procedure. The RNA was

resuspended in 300 µl IP Buffer containing 1 µl RNAsin and the immunoprecipitation procedure was repeated once more to yield highly enriched BrUTP RNA. As a negative control the same procedure was performed on non-labeled, total *Drosophila* ovary RNA. As a quality control RT-qPCR was performed on 10% of the purified RNA of the negative control and experimental RNA with primers for Vasa, Rp49 and selected control genes.

Depending on the experiment, the run-on RNA was used to clone RNA-seq libraries with NEBNext Ultra Directional RNA Library Kit and sequenced on the Illumina HiSeq 2000 (50-bp reads) platform (for shCutoff and shRhino Gro-Seq experiments). Or the RNA was converted into cDNA and qPCR was performed (for T-1/BX2 experiments).

6.2.3 5' capped RNA cloning

6.2.3.1 Removal of RNAs shorter than 200 nt

The scheme for cloning of 5' capped RNAs is shown in figure 3.1 and it is adapted from the procedure described by Gu and colleagues (Gu et al., 2012). Pilot experiments showed that a significant proportion of 5' monophosphate libraries consists of small nucleolar RNAs (snoRNAs) and spliceosomal RNAs (U-RNAs), most of which are 200 nucleotides in length or shorter. In order to prevent an over-representation of those sequences, RNAs larger than 200 nucleotides were depleted from total RNA samples, using the commercially available MirVana kit (Life Technologies). The method is based on binding of RNAs smaller than 200 nucleotides to a glass-fiber filter, which results in a flow-through depleted for this size range. Visual inspection on polyacrylamide gels showed that depletion of very abundant RNA species smaller than 200 nucleotides requires two rounds of mirVana depletion. 40 µg of total RNA were used as input for two rounds of mirVana depletion and the depletion was performed following the supplier's manual.

The second large portion of contaminating RNAs, which were rRNAs, were removed from the sample using a commercially available kit (RiboZero Human/Mouse/Rat from epicentre). This kit contains biotinylated oligonucleotides complementary to rRNAs, which hybridize to the most abundant rRNA and remove them through binding to streptavidin beads. The composition of the oligonucleotides in this kit is optimized for mammalian samples and the supplier states that it is not ideal for rRNA depletion of *Drosphila* RNA. Therefore, I supplemented the pool of the kit's oligonucleotieds with custom designed biotinylated oligonucleotides. Those were designed to deplete the most abundant residual contaminants, which were left after conventional use of the kit's probes. They were mixed to obtain a mix of 11 oligonucleotides, each at a concentration of 5 µM. 5 µg of the size selected RNA was used as input for the rRNA depletion and the procedure was performed following the supplier's manual, with the modification of adding 2 µl of

the custom oligo mix to the reaction mix.

The concentration of the resulting RNA sample was measured on Qubit (see section 6.1.2.2).

6.2.3.2 CIP treatment

For 5' capped RNA libraries the RNA sample was treated with Calf Intestine Phosphatase. This step removes mono-, di- and tri-phosphates from the 5' ends of RNAs, RNA molecules containing a 5' m^7G Cap, however, are protected from this enzymatic conversion.

The reaction was assembled as follows and incubated at 37 °C for 60 min.

RNA sample	x µl
10 x Buffer No. 3 (NEB)	1 µl
Alkaline Phosphatase	1 µl
water	to 10 µl

25 µl of 3M Sodium Acetate and 215 µl water were added to the sample and the RNA was isolated with Phenol/Chloroform as described in the section 6.1.1. The RNA pellet was resuspendet in 8 ul water and treated with Tobacco Acid Pyrophosphatease (TAP) at 37 °C for 60 min.

6.2.3.3 TAP treatment

RNA sample	8 µl
10 x TAP Buffer	1 µl
Tobacco Acid Pyrophosphatase	1 µl

In this reaction, the TAP enzyme removes the 7mG 5' cap from all capped RNA molecules through hydrolysis of the phosphoric acid anhydride bond in the triphosphate bridge of the cap. The resulting RNA contains a 5'-monophosphate, which can be ligated to an RNA adapter in the subsequent reaction.

6.2.3.4 5' adapter ligation

The RNA adapter RA5 from the Illumina TruSeq Kit for small RNA cloning was used. The ligation was performed with a T4RNA ligase1, which ligates ssRNA 5'-monophosphorylated termini to 3'-hydroxylated ends. First, the RNA components were mixed and to linearize the RNA molecules the mixture was heated to 65 °C for 5 min, then put on ice. The

RNA sample	6 µl
RA5 [100 µM]	
10 X Ambion T4 RNA ligase buffer	1 µl

remaining components were added and mixed by gentle pipetting. PEG8000 enhances intramolecular ligation through an effective increase of the donor and acceptor concentration (Harrison and Zimmerman, 1984).

PEG8000	8 µl
ATP [10 mM]	2 µl
T4 ligase (5 U/ µl)	1 µl
RNAsin	1 µl

The ligation reaction was inubated at 37 °C for 2 hours, then 1 µl of fresh ATP and 1 µl of T4 RNA ligase 1 were added and the reaction was put at 4 °C over night.

6.2.3.5 Introduction of the 3' adapter

The 5' adapter ligation reaction was directly passed over into the next step of library preparation. The 3' adapter contained a random hexamer sequence at its 5' end and was introduced through reverse transcription. The reaction was set up as described in section 6.1.5, instead of 1 µl of random hexamers, 1 µl of the 10 µM 3' adapter was used.

6.2.3.6 Linear PCR

In order to enrich the sample for RNA molecules that were ligated to the 5' adapter, a linear PCR was performed. The short forward primer was complementary to the RA5 adapter:

RT Reaction	20 µl
forwad primer [2.5 µM]	0.5 µl
dNTPs	0.5 µl
10X Buffer	2.5 µl
Taq Polymerase	0.5 µl
H2O	1.0 µl

12 cycles of the PCR reaction were performed with the following conditions: 95 °C of initial denaturation, then 95 °C for 20 sec, 50 °C for 30 sec, 72 °C for 20 sec. In order to purify the PCR product from unused primer and from unligated adapter, the raction was ran on a 15% denaturing polyacrylamide/urea gel. As size marker, a DNA ladder for low molecular weight was loaded. The gel was ran until the Xylene Cyanol front had

reached the end of the gel (this corresponds to a nucleic acid size of approximately 20 nucleotides) and subsequently was stained with 1 x SybrGreen in TBE Buffer for 20 min. The size range between 130 and 170 nt was cut, which corresponds to an effective insert size of 66-106 nt of the linear PCR product. The gel pieces were placed in 200 µl of 300 mM NaCl in TE buffer and extracted on a shaking thermoblock over night at 16 °C and 1,000 rpm. Subsequently, the buffer was transferred into a fresh tube, 250 µl isopropanol and 1 µl Glycogen were added and the DNA was precipitated at -80 °C for 60 min. The sample was span at 21,000 g for 60 min, the pellet was washed in 70% ethanol and then air dried and resuspended in 25 µl water.

6.2.3.7 Final library amplification

To establish the optimal degree of amplification, three PCR reactions with different numbers of cycles were performed, containing 5 µl of the linear PCR product as template in each reaction. This PCR was performed with a 2 x hot start Taq Polymerase mix (Bioline).

Template	5 µl
2 X HS Taq MM	10 µl
RP1 primer [10 µM]	0.5 µl
indexed rev primer [10 µM]	0.5 µl
water	4 µl

The reverse primer contains a barcode sequence, which allows the simultaneous seuqencing of multiple libraries on one flowcell. Both primers in this PCR consist of a specific and a universal portion. The specific portion anneals to the respective adapter sequence, the universal portion is required for cluster formation during sequencing on the Illumina sequencer. Thus, in the first rounds of this final PCR only a short fragment with a low temperature anneals to the target, in later cycles the full primer can anneal to the target molecules and the melting temperature subsequently will be increased. Therefore, I performed final library amplification in a 2 step fashion, where the annealing temperature was raised in the second part of the PCR.

Final libraries were ran on a 1.8% agarose gel, cut in the size range of 250 bp - 450 bp and sequenced on the Illumina HiSeq 2000 platform with single end 50 base pairs read length.

Step	temperature	time
1	95 °C	3 min
2	95 °C	30 sec
3	50 °C	30 sec
4	72 °C	30 sec
5	go to step 2 6 / 7 / 8 times	
6	95 °C	30 sec
7	66 °C	45 sec
8	72 °C	30 sec
9	go to step 6 9 / 11 / 13 times	

6.2.4 5' mono-phosphate RNA cloning

To clone and sequence RNAs containing a mono-phosphorylated 5' terminus, the same general procedure like for cloning of 5' capped RNA was employed (see 6.2.3) without performing CIP and TAP treatments. 40 µg of total RNA were used as input for the size selection and subsequent rRNA removal. The resulting sample was directly used for the ligation of the 5' adapter and the following steps (see section 6.2.3.4 and onwards)

6.2.5 Nascent transcript isolation

Nascent transcripts were isolated following a protocol described by Wuarin and Schiebler (Wuarin and Schibler, 1994) with some modifications. For each experiment 150 wild type flies (Strain 333) were put on yeast 2-3 days prior to the experiment. Ovaries were dissected in PBS and subsequently nuclei were isolated.

6.2.5.1 Nuclei isolation

Nuclei were isolated following the same procedure described in section 6.2.2.2. Instead of a wash in Buffer RB, the nuclei were washed in PBS once.

6.2.5.2 Chromatin precipitation and nascent RNA isolation

The nuclei were resuspended in 300 µl Buffer N-A and transferred into a 15 ml conical tube. To lyse the cells and simultaneously precipitate the crhomatin fraction, 3 ml of Buffer N-B was added, the sample was briefly pipetted up and down, vortexed very briefly and put on ice. After a 20 min incubation on ice, the sample was split into 1.5 ml low adhesion tubes and span at 15,000 g and 4 °C for 30 minutes. The chromatin precipitate was then washed twice in a 1:10 mix of Buffer N-A : N-B. During the wash the precipitate was gently pipetted up and down with a P1000 pipette.

After the final wash 1 ml Ribozol was added to the pellet and the sample was incubated at 65 °C until the pellet had resolved (approximately 10 min). Subsequently, RNA was isolated following the standard procedure described in section 6.1.1.2

6.2.6 Oligo(dT) beads assay to determine polyA containing RNAs

To compare the amount of RNA molecules containing a polyA tail in a total RNA sample to the amount in the nascent transcript sample, both samples were passed over magnetic oligo(dT) beads and the unbound fractions were compared. Frist, the nascent RNA sample from section 6.2.5 was quantified fluorometrically (see 6.1.2.2) and a total RNA sample was brought to the same concentraton. Both RNA samples were DNase treated as described in 6.1.4. After heat inactivation of the DNase digest, the samples were immediately put on ice and brought to a total volume of 25 µl. 5 µl were kept as "input sample". 10 µl of oligo(dT) Dynabeads from the Dynabeads mRNA Purification Kit were washed in 50 µl supplied Binding Buffer twice and resuspended in 40 µl Binding Buffer. The sample was split into two and each RNA sample (total RNA and nascent transcript RNA) was added into one tube containing the 20 µl Bead Binding Buffer solution. The sample was incubated at RT for 5 min and occasionally mixed. Then, the mix was placed on a magnetic stand, the supernatant was taken of as "unbound" sample. After adding 20 µl of 3 M Sodium Acetate, pH 4.2, 1 µl Glycogen, 140 µl water and 600 µl 100% ethanol, RNA was precipitated at -80 °C for at least 60 min. After spinning at 21,000 g for 60 min and washing the pellet with 1 ml 70% ethanol, the RNA was resuspended in 5 µl.

The 5 µl "input" sample and the 5 µl "unbound" sample were reverse transcribed in a reaction described in section 6.1.5, primed with random hexamers.

6.2.7 Bioinformatic analysis of RNA sequencing data

6.2.7.1 RNA mapping

All data analysis was done either manually running subsequent programs, or in a pipeline set up by Sergei Manakov.

RNA libraries were sequenced on an Illumina 2000 or on an Illumina HiSeq2500 platform. In the first step 3' adapter seuqences were trimmed using Reaper tool, v. 12-205 (Davis et al., 2013). Reads were discarded if less than 15 nucleotides remained after trimming. Next, rRNA contamination was removed from libraries. To identify rRNA reads, libraries were aligned to the sequence of rRNA unit (Stage and Eickbush, 2007) and 5S rRNA sequences annotated in dm3 by RepeatMasker (http://www.reapeatmasker.org),

as available via the UCSC Genome Bioinformatics Site (Karolchik et al., 2014). Alignments were performed with Bowtie 0.12.7 (Langmead et al., 2009). Alignments were stored in BAM format (Li et al., 2009) and converted to UCSC Genome Browser (Kent et al., 2002) track files using BED tools (Quinlan and Hall, 2010), BigBed and BigWig tools (Kent et al., 2010)

Exons were defined as the region of the genome covered by any of the exons annotated by RefSeq (Pruitt et al., 2009) (RefSeq gene table was downloaded from the University of California at Santa Cruz [UCSC] genome browser (Meyer et al., 2013)); introns were defined as the regions between exonic partitions within the boundaries of a region covered by RefSeq transcripts belonging to any gene. Regions that were between exonic partitions but outside of the gene boundaries were defined as the intergenic space. The TSS of a given gene was defined as the 1-kb region around the 5' end of the most outstanding transcript annotated to that gene (500 bp).

6.3 Methods related to work with *Drosophila*

6.3.1 LacZ staining in *Drosophila* ovaries

Ovaries were dissected in PBS and fixed in Ovary Fixation buffer for 4 minutes. After two washes in PBS, ovaries were submerged in X-gal buffer and incubated at 37 °C until some staining was visible in the positive control and the negative control was still colorless. Depending on the experiment, this incubation can take from 30 minutes up to several hours or over night. The staining reaction was then sopped by washing ovaries in PBS. After the wash they were mounted onto a glass slide in 50% glycerol/PBS and inspected in the light microscope.

6.3.2 Fly crosses and stocks

- crosses for sh knockdowns
  female: driver fly expressing driver; male: fly expressing hairpin

The original strain P1152 carrying the insertion of the P{lArB} construct in telomeric sequences of the X chromosome (site 1A) is described in (Roche and Rio, 1998)

The strain BC69 with the insertion of the P{A92} construct at a euchromatic location on chromosome 2L (site 35B10-35C1) is described in (Lemaitre et al., 1993)

- crosses for H3K9me3 ChIP
  MD cross: females containing the P-1152 source and BC69 target locus and a

maternal-tubulin-Gal4 driver (P1152/P1152;BC69/Cy;MT:GAL4/TM6B) were crossed to wild-type males. NMD cross: flies were mated in the reciprocal direction.

- crosses for Rhino ChIP
  MD crosses: females containing the P-1152 source and BC69 target locus and a maternal-tubulin-Gal4 driver (P1152/P1152;BC69/Cy;MT:GAL4/TM6B) were crossed to males containing UASp-Rhino-GFP. NMD cross: females only contained the BC69 target locus and a maternal-tubulin-Gal4 driver (+/+;BC69/Cy;MT:GAL4/TM6B) and were crossed to males containing UASp-Rhino-GFP.

- crosses for Cutoff ChIP
  MD crosses: females containing the P-1152 source and BC69 target locus and a maternal-tubulin-Gal4 driver (P1152/P1152;BC69/Cy;MT:GAL4/TM6B) were crossed to males containing UASp-Cutoff-GFP. NMD cross: flies were mated in the reciprocal direction.

- crosses for sh knockdown
  For sh knockdown experiments female virgins expressing a GAL4 driver construct were crossed to males containing the hairpin. Hairpin flies from VDRC contain long hairpins, which requires additional expression of Dicer2 (in those cases the driver fly contained UAS-Dicer2 on the first chromosome)

Table 6.1: List of flies used for this work

Name	I Chr	II Chr	III Chr	received from	Aravin Stock Number	Lab Num-
Nos-Gal4	w[1118]		Pw[+mC]=GAL4::V nos.UTRCG6325[MV	Bloomington	4937	1210
UAS-Dcr2,Nos-GAL4	UAS-DCR2/hs-hid, y	nos-GAL4 (NGT40)		R. Lehmann Lab		1024
P1152-Nos	P1152/P1152	BC69/Cy	Nos:Gal4/TM6B			1444
P1152-MT	P1152/P1152	BC69/Cy	MT:GAL4/TM6B			1443
BC69-Nos		BC69/CyO	Nos:Gal4/TM6			1350
BC69/-MT		BC69/CyO	MT::Gal4/MT::Gal4			1764
333 Sequencing Strain	y[1]	Gr22b[1]Gr22d	LysC[1] Mst-Prox[1]GstD5[1]Rh6[	Bloomington	2057	333

Name	I Chr	II Chr	III Chr	received from	published	Aravin Lab Stock Number
sh-Yb988			Py[+t7.7] v[+t1.8]=TRiP.GL00204	Bloomington 35301		988
sh-Yb691			sh-fs(1)Yb/TM3	J. Brennecke		691
sh-Crm1	y[1] sc[*] v[1]		Py[+t7.7] v[+t1.8]=TRiP.HMS009	Bloomington 34021		692
sh-Piwi	y[1] sc[*] v[1]		Py[+t7.7] v[+t1.8]=TRiP.HMS006	Bloomington 33724		484
sh-Senataxin	y[1] sc[*] v[1]		Py[+t7.7] v[+t1.8]=TRiP.HMS011 Sb[1]	Bloomington 34683		930
sh-Rat1 (Xrn2)	w	CyO/Sp	shRat1-1a/ TM6	Rainbowgene 25710-9091		992
sh-Rai1 (CG9125)		CyO/Sp	shCG9125-2a/TM6	Rainbowgene 25710-9092		996
sh-SuVar205	y[1] sc[*] v[1]	Py[+t7.7] v[+t1.8]=TRiH		Bloomington 36792		905
sh-SuVar210	y[1] sc[*] v[1]		Py[+t7.7] v[+t1.8]=TRiP.HMS007	Bloomington 32956		85
sh-Rrp6	y[1] sc[*] v[1]		Py[+t7.7] v[+t1.8]=TRiP.HMS001 Sb[1]	Bloomington 34809		932
sh-Ago3	y[1] sc[*] v[1]; Py[+t7.7] v[+t1.8]=TRiP.G			Bloomington 35232		534
sh-Armi	y[1] sc[*] v[1]; Py[+t7.7] v[+t1.8]=TRiP.G			Bloomington 35343		535
sh-Aub	y1 sc* v1; PTRiP.GL00076a			Bloomington 35201		533
sh-Cuff549	y[1] sc[*] v[1]; Py[+t7.7] v[+t1.8]=TRiP.G			Bloomington 35318		549
sh-Cuff550	y1 sc* v1; PTRiP.GL00054a			Bloomington 35182		550
sh-Krimp	y1 sc* v1; PTRiP.GL00115a			Bloomington 35231		546

Name	I Chr	II Chr	III Chr	received from	published	Aravin Lab Stock Number
sh-Mael	y1 sc* v1; PTRiP.GL00077a			Bloomington 35202		548
sh-Papi	y1 sc* v1; PTRiP.GL00375a Sb1			Bloomington 35450		552
sh-SpnE	y1 sc* v1; PTRiP.GL00206a			Bloomington 35303		543
sh-Sq	y[1] sc[*] v[1]; Py[+t7.7] v[+t1.8]=TRiP.G Sb[1]			Bloomington 35574		540
sh-Tapas		If/CyO	sh-CG8920/TM3	J. Brennecke	Handler et al, 2011	700
sh-Tejas		If/CyO	sh-CG8589/TM3	J. Brennecke	Handler et al, 2011	703
sh-Tudor		If/CyO	sh-Tud/TM3	J. Brennecke	Handler et al, 2011	695
sh-Vasa	y[1] sc[*] v[1]; Py[+t7.7] v[+t1.8]=TRiP.H			Bloomington 32434		541
sh-Vret			sh-vret (CG4771)/TM3	J. Brennecke	Handler et al, 2011	701
sh-Yu		If/CyO	sh-yu/TM3	J. Brennecke	Handler et al, 2011	693
sh-Zuc538	y[1] sc[*] v[1]; Py[+t7.7] v[+t1.8]=TRiP.G			Bloomington 35227		538
sh-Zuc539	y[1] sc[*] v[1]; Py[+t7.7] v[+t1.8]=TRiP.G			Bloomington 35228		539
sh-CG11133			sh-CG11133/TM3	J. Brennecke	Handler et al, 2011	696
sh-CG14303		If/CyO	sh-CG14303/TM3	J. Brennecke	Handler et al, 2011	692
sh-CG13472		If/CyO	sh-CG13472/TM3	J. Brennecke	Handler et al, 2011	704
sh-CG31755			sh-CG31755/TM3	J. Brennecke	Handler et al, 2011	694
sh-CG9684			sh-CG9684/TM3	J. Brennecke	Handler et al, 2011	699

Name	I Chr	II Chr	III Chr	received from	published	Aravin Lab Stock Number
sh-CG9925			sh-CG9925/TM3	J. Brennecke	Handler et al, 2011	702
sh-UAP56	y[1] sc[*] v[1]		Py[+t7.7] v[+t1.8]=TRiP.HMS000	Bloomington 33666		703
sh-NEF-Sp			y[1] sc[*] v[1]; Py[+t7.7] v[+t1.8]=TRiP.HMS007	Bloomington 32931		1128
sh-smt3	y[1] sc[*] v[1]		Py[+t7.7] v[+t1.8]=TRiP.HMS015	Bloomington 36125		1131
sh-NXF2	y[1] sc[*] v[1]		Py[+t7.7] v[+t1.8]=TRiP.HMS009 Sb[1]	Bloomington 33985		935
sh-NXF1	y[1] sc[*] v[1]		Py[+t7.7] v[+t1.8]=TRiP.HMS002	Bloomington 34945		938
sh-NXT1	y[1] sc[*] v[1]		Py[+t7.7] v[+t1.8]=TRiP.GL00414	Bloomington 35485		939
shWhite	y[1] v[1]		Py[+t7.7] v[+t1.8]=TRiP.HMS000	Bloomington 33623		479
sh-Dis3 (CG6413)			w[1118]; PGD11917v35090	VDRC 108013		1362
sh-Dis3 (CG6413)			w[1118]; PGD11917v35091/TM3	VDRC 35091		1363
sh-Dis3 (CG6413)			PKK101473VIE-260B	VDRC 35090		1364
sh-CPSF73 (CG7698)		w[1118]; PGD3050v39557		VDRC 105388		1365
sh-CPSF73 (CG7698)		w[1118]; PGD3050v39558		VDRC 39558		1366
sh-CPSF73 (CG7698)		PKK100438VIE-260B		VDRC 39557		1367
sh-CPSF100 (CG1957)		w[1118] PGD14966v29		VDRC 29575		1368
sh-Pacman (Xrn1)			PKK108511VIE-260B	VDRC 105739		1369

6.4 Materials

Table 6.2: List of buffers

Buffers for ChIP		
Farnham Buffer	Hepes pH8.0	5mM
	KCl	85 mM
	NP-40 / Igepal	0.5%
	NaCl	
	Roche Protease Inhibitor Coctail	1 pill per 10 ml
	NaF	10mM
	Na3VO4	0.2 mM
RIPA Buffer	Tris-HCl pH 7.4	20 mM
	NaCl	150 mM
	NP-40/Igepal	1%
	Sodium Deoxycholate	0.5%
	SDS	0.1%
	Roche Protease Inhibitor Coctail	1 pill per 10 ml
	NaF	10mM
	Na_3VO_4	0.2 mM
LiCl IP Wash Buffer	Tris-HCl pH 7.5	10 mM
	LiCl	500 mM
	NP-40/Igepal	1%
	Sodium Deoxycholate	1%
Proteinase K Buffer	Tris-HCl pH7.4	200 mM
	EDTA	25 mM
	NaCl	300 mM
	SDS	2%
Buffers for Nuclear run-on		
Buffer HB-A	HEPES pH 7.5	15 mM
	KCl	10 mM
	$MgCl_2$	2.5 mM
	EDTA	0.1 mM
	EGTA	0.5 mM
	NP40	0.05%
	sucrose	0.35 M
	DTT	1 mM
	Roche Protease Inhibitor Coctail	1 pill per 10 ml
Buffer HB-B	HEPES pH 7.5	15 mM
	KCl	10 mM
	$MgCl_2$	2.5 mM
	EDTA	0.1 mM
	EGTA	0.5 mM
	NP40	0.05%
	sucrose	0.8 M
	DTT	1 mM
	Roche Protease Inhibitor Coctail	1 pill per 10 ml

Buffers for Nuclear run-on		
Buffer RB	Tris HCl pH 8.0	5 mM
	KCl	150 mM
	MgCl	5 mM
	DTT	1 mM
	Roche Protease Inhibitor Coctail	1 pill per 10 ml
IP Buffer	NaCl	150 mM
	Tris HCl pH 8.0	50 mM
	NP40	0.05%
	EDTA	1 mM
	Roche Protease Inhibitor Coctail	1 pill per 10 ml
Buffers for Nascent transcript isoalation		
Buffer N-A	Tris pH 7.9	20 mM
	NaCl	75 mM
	EDTA	0.5 mM
	DTT	0.85 mM
	PMSF	0.125 mM
	Glycerol	50%
Buffer N-B	HEPES pH 7.6	20 mM
	DTT	1 mM
	$MgCl_2$	7.5 mM
	EDTA	0.2 mM
	NaCl	0.3 M
	Urea	1 M
	NP40	1%
General Buffers		
2 x RNA loading dye	Formamide	47.5 %
	SDS	0.01 %
	Bromphenol Blue	0.01 %
	Xylene Cyanol	0.005 %
	EDTA	0.5 mM
TAE Buffer for agarose gels, pH 8.2	Tris-acetate	40 mM
	EDTA	1 mM
TBE Buffer for polyacrylamide gels	Tris Base	89 mM
	Boric Acid	89 mM
	EDTA	2 mM
TE Buffer	Tris-HCl, pH 7.5	10 mM
	EDTA	1 mM

General buffers		
T7 RNA transcription buffer (10X)	Tris pH7.9	400 mM
	MgCl2	60 mM
	NaCl	200 mM
	spermidine	20 mM
	Tris pH7.5	500 mM
	DTT	100 mM
Ovary Fixation Buffer	Glutaraldehyde	0.5%
	$MgCl_2$	1mM
	PBS (pH7.5)	1X
X-Gal Buffer	$K_4Fe(CN)_6$	3.5mM
	$K_3Fe(CN)_6$	3.5mM
	$MgCl_2$	1mM
	NaCl	150mM
	Na_2HPO_4	10mM
	NaH_2PO_4	10mM
	X-gal a stock solution of 2% X-gal was stored at -20°Cand added 1:10 to the X-Gal Buffer prior to use	0.2%
Phosphate Buffered Saline (PBS)	Na_2HPO_4	8.1 mM
	KH_2PO_4	1.5 mM
	NaCl	140 mM
	KCl	2.7 mM
	pH adjusted to 7.4 with HCl	
Saline Sodium Citrate (SSC) 20X		
	NaCl	3M
	Trisodium Citrate	30 mM
	pH adjusted to 7.0 with HCl	

Table 6.3: List of antibodies

Antibodies	**Clone**	**Supplier**
Histone H3 (tri methyl K9) antibody	polyclonal	abcam
RNA polymerase II CTD repeat YSPTSPS (phospho S5)	4H8	abcam
GFP (for Western Blot)	polyclonal	covance
GFP (for ChIP)	4C9	DSHB
BrdU	BU-33	Sigma

Table 6.4: List of enzymes

Enzyme	**Supplier**
Calf Intestinal Alkaline Phosphatase (CIP)	NEB
DNAse Amp Grade	Invitrogen
KOD DNA Polymerase R	Toyobo Novagen
MyTaq HS Mix	Bioline
MyTaq Red HS Mix	Bioline
SuperScriptIII TM Reverse Transcriptase	Invitrogen
T4 Polynucleotide Kinase (PNK)	NEB
T4 RNA Ligase 1 (ssRNA Ligase)	NEB
T4 RNA Ligase 2 (dsRNA Ligase)	NEB
T7 RNA polymerase	NEB
Tobacco Acid Pyrophosphatase (TAP)	Epicentre

Table 6.5: List of equipment

Equipment	**Supplier**
Balance	Mettler Toledo
Bioruptor	Diagenode
EasyPet Pipet Aid	Eppendorf
Espresso machine	Gaggia
Fly Incubator	VWR
Gel imaging, Alphaimager	Protein Simple
Mini Dry Bath	BioExpress
Minifuge	BioExpress
NanoDrop 2000c (Spectrophotometer)	Thermo Scientific
NanoPure water filter	Thermo Scientific
PH Meter: Accumet AB15 Plus Bio Kit	Fisher Scientific
Pro Flex Thermocycler, PCR machine	Life Technologies
Qubit 2.0 Fluormeter	Invitrogen
Realplex MasterCycler, qPCR machine	Eppendorf
Sorvall RC-6 Plus Superspeed Centrifuge	ThermoFisher Scientific
Tabletop centrifuge (5424)	Eppendorf
Tabletop centrifuge (5424D)	Eppendorf
Thermomixer, shaking thermoblock	Eppendorf
Vortex, standard, 120V	Fisher Scientific
Zeiss FL Cube EC P+C for microscope	Zeiss

Table 6.6: List of reagents

Reagents	**Supplier**
Acetic Acid, Glacial	Fisher
Agarose, AquaPor LM (Ultra Pure), low melt	National Diagnostics
Agarose, NuSieve GTG (low melt)	Lonza
Agencourt AMPure XP - PCR Purification	Beckman Coulter
Ammonium Acetate, Crystal	Mallinckrodt Chemicals
Ampicillin (sodium salt)	Sigma
ATP-EasyTide [y-32P]	PerkinElmer
Betaine, BioUltra	Sigma
Bromophenol Blue	Sigma
BrUTP (5-Bromouridine-5-triphosphate)	Sigma
BSA (Bovine Serum Albumin) Fraction 5	Rockland
Chloroform	Fisher
Complete, Mini EDTA-free Protease inhibitor cocktail	Roche
Coomassie Brilliant Blue, G-250	Research Organics/VWR
D-Glucose, Anhydrous, Granular	Avantor Performance Materials/VWR
D-Tube Dialyzer Maxi, MWCO 3.5kDa	
Diethylpyrocarbonate (DEPC)	Sigma
Dimethyl Sulfoxide	Mallinckrodt Chemicals
DNA ladder, 2-Log (100bp-10Kb)	NEB
DNA ladder, Low Molecular Weight (25bp-766bp)	NEB
dNTP Mix	NEB
DTT	Sigma
Dynabeads Protein G	Invitrogen
EDTA	Sigma
Ethidium Bromide	Sigma-Aldrich
Formaldehyde 37%	Fisher
Formaldehyde, Para certified	Fisher
Glycerol	Sigma-Aldrich
Glycine	Sigma
Glyco Blue	Bioline
GTP-Easy Tide [-32P]	PerkinElmer
HEPES	Millipore/VWR
IGEPAL (Substitute for NP40)	Sigma-Aldrich
Lithium Chloride	Sigma
MicroSpin G-25 Columns	Roche
Phenol/Chloroform	AMRESCO (Bioexpress)
Phenylmethylsulfonyl Fluoride (PMSF)	Fisher
Polyvinyl-pyrrolydone (PVP)	Fisher
Potassium Chloride, Granular	Mallinckrodt/Fisher
Protease Inhibitor Coctail Complete, Mini, EDTA-free	Roche
Protogel (30% acrylamide/bis (37.5:1 ratio))	National Diagnostics
Ribozol	Amresco

Table 6.7: List of reagents

Reagents	**Supplier**
RNA fragmentation reagent	Invitrogen/Ambion
SDS	Fisher
Sodium Azide	Sigma
ssRNA Ladder (500bp-9Kb)	NEB
ssRNA Ladder, Low Range (50bp-1Kb)	NEB
Sucrose (Rnase free, UltraPure)	MP Biomedicals/Fisher
SYBR Gold Nucleic Acid Stain	Invitrogen
SYBR Green I Nucleic Acid Gel Stain	Life technologies
Tris-HCl / Trizma	Sigma
Triton X-100	Sigma
Trizma Base	Sigma
TWEEN 20	Sigma
Xylene Cyanole FF	Bio Rad

Table 6.8: List of kits

Kit	**Supplier**
Agilent RNA 6000 Pico Kit for 275 samples	Agilent Technologies Inc.
DNA Clean ConcentratorTM -5	Zymo Reseach
Dynabeads mRNA Purification Kit	Life Technologies
IllustraTM Micro SpinTM G25	GE Healthcare Life Sciences
MinElute PCR Purification Kit (50)	Qiagen
MirVana, miRNA isolation Kit	Life technologies
NEBNext ChIP-Seq Library Prep Master Mix Set for Illumina	NEB
NEBNext mRNA Library Prep Master Mix Set for Illumina	NEB
NEBNext Ultra Directional RNA Library Prep Kit for Illumina	NEB
Novagen D-tube Dialyzer Midi	EMD Millipore
NucleoBond Xtra Midi	Macharey-Nagel
pGEM-T Vector System I	Promega (VWR)
Qubit DNA High Sensitivity Assay Kit	Invitrogen
Qubit RNA Assay Kit	Invitrogen
RiboZero rRNA Removal Kit	Epicentre
Zymoclean Gel DNA Recovery	Zymo Reseach
Zyppy Miniprep Kit	Zymo Reseach

CHAPTER 7

Acknowledgement

First and foremost, I would like to thank Alexei Aravin for giving me the opportunity to learn, work and laugh in his laboratory. I am grateful for the confidence he had in me, the freedom he gave me and the time he spent teaching me.

I also want to thank Gunter Meister for providing me with this chance to do my PhD, for his positive support at meetings and for his trust. I am grateful for all the contacts I had with him and his great lab.

I would also like to thank Gernot Längst for being part of my committee.

I am very thankful to Kata Fejes-Tóth for her support and advice throughout my PhD and I would also like to thank her for proof-reading this thesis. She encouraged me with her pep-talks and helped with flies. She was not only a support for me, but she is a great resource for the whole lab.

Furthermore, I would like to thank the Boehringer Ingelheim Fonds for providing me with a generous PhD stipend, which did not only support me with monetary resources, but also introduced me to an exceptional group of people. The seminars and workshops in Woods Hole and Cold Spring Harbor clearly were highlights of my PhD time.

RNA sequencing was performed at the Caltech Sequencing Facility and I would like to thank Igor Antoshechkin for help and advice. Some data analysis was performed on the platform of Ravi Sachidanandam and I want to acknowledge his support and patience.

A thank also goes to my summer students Pushpa and Elise, who helped especially in

the initial phase of my PhD time.

Thanks to the generous people from our neighboring labs, who always helped out with rare equipment, uncommon reagents or advice: Kostya and Chris from the Varshavsky lab and Dev and Michael from the Baltimore lab.

I had the pleasure to meet a lot of great people working in the Aravin Lab: Adrien, Alex, Alicia, Ariel, Can Li, Hamadi, Irina, Junho, Ken, Sergei, Sveta and Vanya. You were an extraordinary group, created a great work environment and made lab a good place to be.

A special thank goes to Sergei Manakov, who taught me everything I know about bioinformatics. He was a truly patient teacher and I am grateful for his help.

I would like to especially thank Alex Webster for being a great colleague and good friend and for being my companion through the highest highs and the lowest lows of a PhD.

One of the people to thank most is Dubravka Pezic, my baymate, colleague and friend. Her invaluable support reached from help at the bench over moral support in times of failed experiments to going for runs in beautiful San Marino. If there is one thing I am getting out of my time at Caltech (other than a doctoral degree), it is having her as a wonderful friend.

Furthermore, I would like to acknowledge the support and encouragement from the Caltech Masters Swim Team and especially my dear friends from lane No.1: Ricardo, Andrew, Nancy and Alice. Swimming with you under the Californian sun always made me happy, thanks for all the yards we spent together.

Also my friends from back home gave me invaluable support during the time of this PhD work: Lena, Andreas, Georg, Philipp. True friendship shows in times when we are more than 9,000 km and 9 time zones apart. It is invaluable to know that you are a part of my life and supported me from far away.

Furthermore, I would like to thank Neuhof: Josef, Beate, Henrike, Johannes and Evi. Thank you for your visits and for the emotional support.

A special thank goes to Silvia, my sister and best friend. She was an invaluable support and motivation and I would like to thank her not only for the surprise visit, but also for just being wonderful. I also want to thank Lars for his moral support and not least for help with the formatting of this thesis.

The one person that kept me going and whom I owe it all, is Tobias. Thank you for being everything in my life.

A last and very special thank goes to my parents. Thank you for your unconditional trust and support and for being my role-models. This work is for you.

CHAPTER 8

APPENDIX

8.1 Publications related to this thesis

Parts of this work are published in
Trans-generationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing
Adrien Le Thomas*, Evelyn Stuwe*, Sisi Li, Jiamu Du, Georgi Marinov, Nikolay Rozhkov, Ariel Yung-Chia Chen, Yicheng Luo, Ravi Sachidanandam, Katalin Fejes Tóth, Dinshaw Patel and Alexei A. Aravin
Genes Dev. 2014 Aug 1; Vol28, p1667-1680
* equal contribution

Parts of this work are included in a submitted manuscript
Cutoff acts as an anti-terminator in transcription of piRNA precursors
Yung-Chia Ariel Chen, Evelyn Stuwe, Katya Rozhavskaya, Sisi Li, Yicheng Luo, Dinshaw Patel, Katalin Fejes Tóth and Alexei A. Aravin

Parts of the introduction are published in a review article
Small but sturdy: small RNAs in cellular memory and epigenetics
Stuwe E, Tóth KF, Aravin AA; Genes and Dev. 2014, 28:423-431

and in a chapter of the book "Non-coding RNAs and the reproductive system" of the Springer AEMB book series.

8.2 Curriculum Vitae

Evelyn Christina Stuwe	
Maiden Name	Evelyn Christina Sawa
Date of birth	March 16, 1984
Place of birth	Schwetzingen

Education	
2010 - present	**PhD studies** California Institute of Technology in collaboration with Universität Regensburg
2009 - 2010	**Practical year (PJ)** EMBL (European Molecular Biology Laboratory), Heidelberg Bergheimer Apotheke, Heidelberg Degree: 3. Staatsexamen, Grade: 1.0
2003 - 2009	**Studies of Pharmaceutical sciences (Pharmazie)** Ruprecht-Karls Universität, Heidelberg Degree: 2. Staatsexamen, Grade: 1.2
2003	**Abitur** Carl-Friedrich Gauss Gymnasium, Hockenheim Grade 1.0

Positions and Employments	
2010 - present	PhD candidate California Institute of Technology, USA
04/2010 10/2010	Apothekerin Moerike Apotheke, Neuenstadt, Apothekerin
10/2009 04/2010	Intern (PJ) Bergheimer Apotheke, Heidelberg
04/2009 10/ 2009	Intern (PJ) European Molecular Biology Laboratory (EMBL), Heidelberg
10/2006 05/2007	Research Assistant University Clinic (Universitätsklinik) Heidelberg
09/2006 09/2007	Research Intern Rockefeller University, New York, USA

Fellowships and prizes

2009 - 2013	PhD fellowship from the Boehringer Ingelheim Fonds
2005	Reseach Fellowhip form the German National Academic Foundation (Studienstiftung des Deustchen Volkes)
2003 - 2008	Scholar of the German National Academic Foundation (Studienstiftung des Deustchen Volkes)

Publications

Le Thomas A ∗ , **Stuwe E** ∗ , Li S, Du J, Marinov G, Rozhkov N, Chen YC, Luo Y, Sachidanandam R, Tóth KF, Patel D, Aravin AA
Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing.
Genes Dev. 2014, 28:1667-1680
∗ equal contribution

Stuwe E, Tóth KF, Aravin AA
Small but sturdy: small RNAs in cellular memory and epigenetics
Genes and Dev. 2014, 28:423-431
Review

Lin DH, Zimmermann S, Stuwe T, **Stuwe E**, Hoelz A
Structural and functional analysis of the C-terminal domain of Nup358/RanBP2
J Mol Biol. 2013, 425:1318-1329

Weiss J, **Sawa E**, Riedel KD, Haefeli WE, Mikus G
In vitro metabolism of the opioid tilidine and interaction of tilidine and nortilidine with CYP3A4, CYP2C19, and CYP2D6
Naunyn Schmiedebergs Arch Pharmacol. 2008, 378:275-82

Westberg L, **Sawa E**, Wang AY, Gunaydin LA, Ribeiro AC, Pfaff DW
Colocalization of connexin 36 and corticotropin-releasing hormone in the mouse brain
BMC Neurosci 2009, 30:10-41

Ribeiro AC, **Sawa E**, Carren-LeSauter I, LeSauter J, Silver R, Pfaff DW
Two forces for arousal: Pitting hunger versus circadian influences and identifying neurons responsible for changes in behavioral arousal
PNAS 2007, 104:20078-20083

BIBLIOGRAPHY

Affymetrix ENCODE Transcriptome, P. and Cold Spring Harbor Laboratory ENCODE Transcriptome, P. (2009). Post-transcriptional processing generates a diversity of 5'-modified long and short rnas. *Nature* **457**, 1028–1032.

Anand, A. and Kai, T. (2012). The tudor domain protein kumo is required to assemble the nuage and to generate germline piRNAs in Drosophila. *EMBO J* **31**, 870–82.

Angell, S. M. and Baulcombe, D. C. (1997). Consistent gene silencing in transgenic plants expressing a replicating potato virus x rna. *EMBO J* **16**, 3675–3684.

Aravin, A., Gaidatzis, D., Pfeffer, S. A., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M. J., Kuramochi-Miyagawa, S., Nakano, T., Chien, M., Russo, J. J., Ju, J., Sheridan, R., Sander, C., Zavolan, M. and Tuschl, T. (2006). A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442**, 203–207.

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

Aravin, A., van der Heijden, G., Castañeda, J., Vagin, V., Hannon, G. and Bortvin, A. (2009). Cytoplasmic compartmentalization of the fetal piRNA pathway in mice. *PLoS genetics* **5**.

Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K. and Hannon, G. J. (2007). Developmentally Regulated piRNA Clusters Implicate MILI in Transposon Control. *Science* **316**, 744–747.

Argeson, A. C., Nelson, K. K. and Siracusa, L. D. (1996). Molecular basis of the pleiotropic phenotype of mice carrying the hypervariable yellow (Ahvy) mutation at the agouti locus. *Genetics* **142**, 557–67.

Arkhipova, I. R. (1995). Complex patterns of transcription of a drosophila retrotransposon in vivo and in vitro by rna polymerases ii and iii. *Nucleic Acids Res* **23**, 4480–4487.

Babiarz, J. E., Ruby, J. G., Wang, Y., Bartel, D. P. and Blelloch, R. (2008). Mouse es cells express endogenous shrnas, sirnas, and other microprocessor-independent, dicer-dependent small rnas. *Genes Dev* **22**, 2773–2785.

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

Baumberger, N. and Baulcombe, D. C. (2005). Arabidopsis argonaute1 is an rna slicer that selectively recruits micrornas and short interfering rnas. *Proc Natl Acad Sci U S A* **102**, 11928–11933.

Berezikov, E., Chung, W.-J., Willis, J., Cuppen, E. and Lai, E. C. (2007). Mammalian mirtron genes. *Mol Cell* **28**, 328–336.

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.

Bittencourt, D. and Auboeuf, D. (2012). Analysis of co-transcriptional rna processing by rna-chip assay. *Methods Mol Biol* **809**, 563–577.

Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M. and Benning, C. (1998). Ago1 defines a novel locus of arabidopsis controlling leaf development. *EMBO J* **17**, 170–180.

Bohnsack, M. T., Czaplinski, K. and Görlich, D. (2004). Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* **10**, 185–191. ISSN 1469-9001.

Boland, A., Tritschler, F., Heimstadt, S., Izaurralde, E. and Weichenrieder, O. (2010). Crystal structure and ligand binding of the mid domain of a eukaryotic argonaute protein. *EMBO Rep* **11**, 522–527.

Bologna, N. G. and Voinnet, O. (2014). The diversity, biogenesis, and activities of endogenous silencing small rnas in arabidopsis. *Annu Rev Plant Biol* **65**, 473–503.

Bourc'his, D. and Bestor, T. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431**, 96–99.

Brasset, E., Taddei, A. R., Arnaud, F., Faye, B., Fausto, A. M., Mazzini, M., Giorgi, F. and Vaury, C. (2006). Viral particles of the endogenous retrovirus zam from drosophila melanogaster use a pre-existing endosome/exosome pathway for transfer to the oocyte. *Retrovirology* **3**, 25.

Brennecke, J., Aravin, A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R. and Hannon, G. (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**, 1387–92.

Brzeski, J. and Brzeska, K. (2011). The maze of paramutation: a rough guide to the puzzling epigenetics of paramutation. *Wiley Interdiscip Rev RNA* **2**, 863–874.

Bucheton, A. (1979). Non-mendelian female sterility in drosophila melanogaster: influence of aging and thermic treatments. iii. cumulative effects induced by these factors. *Genetics* **93**, 131–142.

Bucheton, A. (1990). I transposable elements and i-r hybrid dysgenesis in drosophila. *Trends Genet* **6**, 16–21.

Buck, A. H. and Blaxter, M. (2013). Functional diversification of argonautes in nematodes: an expanding universe. *Biochem Soc Trans* **41**, 881–886.

Caporilli, S., Yu, Y., Jiang, J. and White-Cooper, H. (2013). The rna export factor, nxt1, is required for tissue specific transcriptional regulation. *PLoS Genet* **9**, e1003526.

Carmell, M., Girard, A., van de Kant, H., Bourc'his, D., Bestor, T., de Rooij, D. and Hannon, G. (2007). MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Developmental cell* **12**, 503–514.

Carmell, M. A., Xuan, Z., Zhang, M. Q. and Hannon, G. J. (2002). The argonaute family: tentacles that reach into rnai, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev* **16**, 2733–2742.

Carthew, R. W. and Sontheimer, E. J. (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136**, 642–55.

Castel, S. E. and Martienssen, R. A. (2013). Rna interference in the nucleus: roles for small rnas in transcription, epigenetics and beyond. *Nat Rev Genet* **14**, 100–112.

Chalvet, F., Teysset, L., Terzian, C., Prud'homme, N., Santamaria, P., Bucheton, A. and Pelisson, A. (1999). Proviral amplification of the gypsy endogenous retrovirus of drosophila melanogaster involves env-independent invasion of the female germline. *EMBO J* **18**, 2659–2669.

Chambeyron, S., Popkova, A., Payen-Groschêne, G., Brun, C., Laouini, D., Pelisson, A. and Bucheton, A. (2008). piRNA-mediated nuclear accumulation of retrotransposon transcripts in the Drosophila female germline. *Proceedings of the National Academy of Sciences* **105**, 14964–14969.

Cheloufi, S., Dos Santos, C. O., Chong, M. M. W. and Hannon, G. J. (2010). A dicer-independent mirna biogenesis pathway that requires ago catalysis. *Nature* **465**, 584–589.

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.

Chlebowski, A., Lubas, M., Jensen, T. H. and Dziembowski, A. (2013). Rna decay machines: the exosome. *Biochim Biophys Acta* **1829**, 552–560.

Churchman, L. S. and Weissman, J. S. (2012). Native elongating transcript sequencing (net-seq). *Curr Protoc Mol Biol* **Chapter 4**, Unit 4.14.1–17.

Cifuentes, D., Xue, H., Taylor, D. W., Patnode, H., Mishima, Y., Cheloufi, S., Ma, E., Mane, S., Hannon, G. J., Lawson, N. D., Wolfe, S. A. and Giraldez, A. J. (2010). A novel mirna processing pathway independent of dicer requires argonaute2 catalytic activity. *Science* **328**, 1694–1698.

Connelly, S. and Manley, J. L. (1988). A functional mrna polyadenylation signal is required for transcription termination by rna polymerase ii. *Genes Dev* **2**, 440–452.

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.

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

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

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

Davis, M. P. A., van Dongen, S., Abreu-Goodger, C., Bartonicek, N. and Enright, A. J. (2013). Kraken: a set of tools for quality control and analysis of high-throughput sequence data. *Methods* **63**, 41–49.

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**, 112–5.

Dengl, S. and Cramer, P. (2009). Torpedo nuclease rat1 is insufficient to terminate rna polymerase ii in vitro. *J Biol Chem* **284**, 21270–21279.

Deshpande, G., Calhoun, G. and Schedl, P. (2005). Drosophila argonaute-2 is required early in embryogenesis for the assembly of centric/centromeric heterochromatin, nuclear division, nuclear migration, and germ-cell formation. *Genes Dev* **19**, 1680–1685.

Ding, L. and Han, M. (2007). Gw182 family proteins are crucial for microrna-mediated gene silencing. *Trends Cell Biol* **17**, 411–416.

Ding, X. C. and Grosshans, H. (2009). Repression of c. elegans microrna targets at the initiation level of translation requires gw182 proteins. *EMBO J* **28**, 213–222.

Dong, Z., Han, M.-H. and Fedoroff, N. (2008). The rna-binding proteins hyl1 and se promote accurate in vitro processing of pri-mirna by dcl1. *Proc Natl Acad Sci U S A* **105**, 9970–9975.

Dorer, D. R. and Henikoff, S. (1997). Transgene repeat arrays interact with distant heterochromatin and cause silencing in cis and trans. *Genetics* **147**, 1181–1190.

Dougherty, W. G., Lindbo, J. A., Smith, H. A., Parks, T. D., Swaney, S. and Proebsting, W. M. (1994). Rna-mediated virus resistance in transgenic plants: exploitation of a cellular pathway possibly involved in rna degradation. *Mol Plant Microbe Interact* **7**, 544–552.

Doyle, M., Badertscher, L., Jaskiewicz, L., Guttinger, S., Jurado, S., Hugenschmidt, T., Kutay, U. and Filipowicz, W. (2013). The double-stranded rna binding domain of human dicer functions as a nuclear localization signal. *RNA* **19**, 1238–1252.

Elkayam, E., Kuhn, C.-D., Tocilj, A., Haase, A. D., Greene, E. M., Hannon, G. J. and Joshua-Tor, L. (2012). The structure of human argonaute-2 in complex with mir-20a. *Cell* **150**, 100–110.

Ender, C., Krek, A., Friedlander, M. R., Beitzinger, M., Weinmann, L., Chen, W., Pfeffer, S., Rajewsky, N. and Meister, G. (2008). A human snorna with microrna-like functions. *Mol Cell* **32**, 519–528.

Eulalio, A., Huntzinger, E. and Izaurralde, E. (2008). Gw182 interaction with argonaute is essential for mirna-mediated translational repression and mrna decay. *Nat Struct Mol Biol* **15**, 346–353.

Faehnle, C. R., Elkayam, E., Haase, A. D., Hannon, G. J. and Joshua-Tor, L. (2013). The making of a slicer: activation of human argonaute-1. *Cell Rep* **3**, 1901–1909.

Fagegaltier, D., Bouge, A.-L., Berry, B., Poisot, E., Sismeiro, O., Coppee, J.-Y., Theodore, L., Voinnet, O. and Antoniewski, C. (2009). The endogenous sirna pathway is involved in heterochromatin formation in drosophila. *Proc Natl Acad Sci U S A* **106**, 21258–21263.

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**, 806–11.

Frank, F., Hauver, J., Sonenberg, N. and Nagar, B. (2012). Arabidopsis argonaute mid domains use their nucleotide specificity loop to sort small rnas. *EMBO J* **31**, 3588–3595.

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.

Gagnon, K. T., Li, L., Chu, Y., Janowski, B. A. and Corey, D. R. (2014). Rnai factors are present and active in human cell nuclei. *Cell Rep* **6**, 211–221.

Ghildiyal, M., Seitz, H., Horwich, M. D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E. L. W., Zapp, M. L., Weng, Z. and Zamore, P. D. (2008). Endogenous

sirnas derived from transposons and mrnas in drosophila somatic cells. *Science* **320**, 1077–1081.

Ghildiyal, M., Xu, J., Seitz, H., Weng, Z. and Zamore, P. D. (2010). Sorting of drosophila small silencing rnas partitions microrna* strands into the rna interference pathway. *RNA* **16**, 43–56.

Ghildiyal, M. and Zamore, P. D. (2009). Small silencing RNAs: an expanding universe. *Nat Rev Genet* **10**, 94–108.

Golden, D. E., Gerbasi, V. R. and Sontheimer, E. J. (2008). An inside job for sirnas. *Mol Cell* **31**, 309–312.

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.

Gregory, R. I., Chendrimada, T. P. and Shiekhattar, R. (2006). Microrna biogenesis: isolation and characterization of the microprocessor complex. *Methods Mol Biol* **342**, 33–47.

Grentzinger, T., Armenise, C., Brun, C., Mugat, B., Serrano, V., Pelisson, A. and Chambeyron, S. (2012). pirna-mediated transgenerational inheritance of an acquired trait. *Genome Res* **22**, 1877–1888.

Grewal, S. I. and Elgin, S. C. (2007). Transcription and RNA interference in the formation of heterochromatin. *Nature* **447**, 399–406.

Grivna, S., Beyret, E., Wang, Z. and Lin, H. (2006). A novel class of small RNAs in mouse spermatogenic cells. *Genes & development* **20**, 1709–1714.

Gu, S. and Kay, M. A. (2010). How do mirnas mediate translational repression? *Silence* **1**, 11.

Gu, W., Lee, H.-C., Chaves, D., Youngman, E. M., Pazour, G. J., Conte, D. J. and Mello, C. C. (2012). Capseq and cip-tap identify pol ii start sites and reveal capped small rnas as c. elegans pirna precursors. *Cell* **151**, 1488–1500.

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

Hackett, J. A. and Surani, M. A. (2013). DNA methylation dynamics during the mammalian life cycle. *Philos Trans R Soc Lond B Biol Sci* **368**, 20110328.

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

Harrison, B. and Zimmerman, S. B. (1984). Polymer-stimulated ligation: enhanced ligation of oligo- and polynucleotides by t4 rna ligase in polymer solutions. *Nucleic Acids Res* **12**, 8235–8251.

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.

Hauptmann, J., Dueck, A., Harlander, S., Pfaff, J., Merkl, R. and Meister, G. (2013). Turning catalytically inactive human argonaute proteins into active slicer enzymes. *Nat Struct Mol Biol* **20**, 814–817.

Hauptmann, J., Kater, L., Loffler, P., Merkl, R. and Meister, G. (2014). Generation of catalytic human ago4 identifies structural elements important for rna cleavage. *RNA* **20**, 1532–1538.

Hobson, D. J., Wei, W., Steinmetz, L. M. and Svejstrup, J. Q. (2012). Rna polymerase ii collision interrupts convergent transcription. *Mol Cell* **48**, 365–374.

Houwing, S., Kamminga, L., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Filippov, D., Blaser, H., Raz, E., Moens, C., Plasterk, R., Hannon, G., Draper, B. and Ketting, R. (2007a). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* **129**, 69–82.

Houwing, S., Kamminga, L., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Filippov, D., Blaser, H., Raz, E., Moens, C., Plasterk, R., Hannon, G., Draper, B. and Ketting, R. (2007b). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* **129**, 69–82+.

Huang, C. R. L., Burns, K. H. and Boeke, J. D. (2012). Active transposition in genomes. *Annu Rev Genet* **46**, 651–675.

Huang, H.-Y., Houwing, S., Kaaij, L., Meppelink, A., Redl, S., Gauci, S., Vos, H., Draper, B., Moens, C., Burgering, B., Ladurner, P., Krijgsveld, J., Berezikov, E. and Ketting, R. (2011). Tdrd1 acts as a molecular scaffold for Piwi proteins and piRNA targets in zebrafish. *The EMBO journal* **30**, 3298–3308.

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

Ipsaro, J., Haase, A., Knott, S., Joshua-Tor, L. and Hannon, G. (2012). The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. *Nature* **491**, 279–283.

Izant, J. G. and Weintraub, H. (1984). Inhibition of thymidine kinase gene expression by anti-sense RNA: a molecular approach to genetic analysis. *Cell* **36**, 1007–15.

Jones, P. A. and Takai, D. (2001). The role of dna methylation in mammalian epigenetics. *Science* **293**, 1068–1070.

Josse, T., Teysset, L., Todeschini, A. L., Sidor, C. M., Anxolabehere, D. and Ronsseray, S. (2007). Telomeric trans-silencing: an epigenetic repression combining RNA silencing and heterochromatin formation. *PLoS Genet* **3**, 1633–43.

Kanno, T., Huettel, B., Mette, M. F., Aufsatz, W., Jaligot, E., Daxinger, L., Kreil, D. P., Matzke, M. and Matzke, A. J. M. (2005). Atypical rna polymerase subunits required for rna-directed dna methylation. *Nat Genet* **37**, 761–765.

Karolchik, D., Barber, G. P., Casper, J., Clawson, H., Cline, M. S., Diekhans, M., Dreszer, T. R., Fujita, P. A., Guruvadoo, L., Haeussler, M., Harte, R. A., Heitner, S., Hinrichs, A. S., Learned, K., Lee, B. T., Li, C. H., Raney, B. J., Rhead, B., Rosenbloom, K. R., Sloan, C. A., Speir, M. L., Zweig, A. S., Haussler, D., Kuhn, R. M. and Kent, W. J. (2014). The ucsc genome browser database: 2014 update. *Nucleic Acids Res* **42**, D764–70.

Kavi, H. H. and Birchler, J. A. (2009). Interaction of rna polymerase ii and the small rna machinery affects heterochromatic silencing in drosophila. *Epigenetics Chromatin* **2**, 15.

Kawamata, T., Yoda, M. and Tomari, Y. (2011). Multilayer checkpoints for microrna authenticity during risc assembly. *EMBO Rep* **12**, 944–949.

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

Kawaoka, S., Hara, K., Shoji, K., Kobayashi, M., Shimada, T., Sugano, S., Tomari, Y., Suzuki, Y. and Katsuma, S. (2013). The comprehensive epigenome map of piRNA clusters. *Nucleic acids research* **41**, 1581–1590.

Kawaoka, S., Izumi, N., Katsuma, S. and Tomari, Y. (2011). 3 end formation of PIWI-interacting RNAs in vitro. *Mol Cell* .

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M. and Haussler, D. (2002). The human genome browser at ucsc. *Genome Res* **12**, 996–1006.

Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S. and Karolchik, D. (2010). Bigwig and bigbed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204–2207.

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

Khodor, Y. L., Rodriguez, J., Abruzzi, K. C., Tang, C.-H. A., Marr, M. T. n. and Rosbash, M. (2011). Nascent-seq indicates widespread cotranscriptional pre-mrna splicing in drosophila. *Genes Dev* **25**, 2502–2512.

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**, 1551–63.

Khvorova, A., Reynolds, A. and Jayasena, S. D. (2003). Functional sirnas and mirnas exhibit strand bias. *Cell* **115**, 209–216.

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, M., Krogan, N. J., Vasiljeva, L., Rando, O. J., Nedea, E., Greenblatt, J. F. and Buratowski, S. (2004). The yeast rat1 exonuclease promotes transcription termination by rna polymerase ii. *Nature* **432**, 517–522.

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

Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Khurana, J. S., Zhang, F., Schultz, N., Koppetsch, B. S., Nowosielska, A., Seitz, H., Zamore, P. D., Weng, Z. and Theurkauf, W. E. (2009). The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell* **138**, 1137–49.

Klenov, M., Lavrov, S., Stolyarenko, A., Ryazansky, S., Aravin, A., Tuschl, T. and Gvozdev, V. (2007). Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the Drosophila melanogaster germline. *Nucleic acids research* **35**, 5430–5438.

Koonin, E. V. (1996). A duplicated catalytic motif in a new superfamily of phosphohydrolases and phospholipid synthases that includes poxvirus envelope proteins. *Trends Biochem Sci* **21**, 242–243.

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

Kramerov, D. A. and Vassetzky, N. S. (2005). Short retroposons in eukaryotic genomes. *Int Rev Cytol* **247**, 165–221.

Kuramochi-Miyagawa, S., Kimura, T., Ijiri, T., Isobe, T., Asada, N., Fujita, Y., Ikawa, M., Iwai, N., Okabe, M., Deng, W., Lin, H., Matsuda, Y. and Nakano, T. (2004). Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development (Cambridge, England)* **131**, 839–849+.

Kuramochi-Miyagawa, S., Kimura, T., Yomogida, K., Kuroiwa, A., Tadokoro, Y., Fujita, Y., Sato, M., Matsuda, Y. and Nakano, T. (2001). Two mouse piwi-related genes: miwi and mili. *Mechanisms of development* **108**, 121–133.

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T., Hata, K., Li, E., Matsuda, Y., Kimura, T., Okabe, M., Sakaki, Y., Sasaki, H. and Nakano, T. (2008). DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes & development* **22**, 908–917.

Kurihara, Y. and Watanabe, Y. (2004). Arabidopsis micro-rna biogenesis through dicer-like 1 protein functions. *Proc Natl Acad Sci U S A* **101**, 12753–12758.

Langmead, B., Trapnell, C., Pop, M. and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short dna sequences to the human genome. *Genome Biol* **10**, R25.

Lau, N., Seto, A., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D. and Kingston, R. (2006). Characterization of the piRNA complex from rat testes. *Science (New York, N.Y.)* **313**, 363–367.

Lau, N. C., Lim, L. P., Weinstein, E. G. and Bartel, D. P. (2001). An abundant class of tiny rnas with probable regulatory roles in caenorhabditis elegans. *Science* **294**, 858–862.

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**, 390–9.

Le Thomas, A., Stuwe, E., Li, S., Du, J., Marinov, G., Rozhkov, N., Chen, Y.-C. A., Luo, Y., Sachidanandam, R., Toth, K. F., Patel, D. and Aravin, A. A. (2014). Transgenerationally inherited pirnas trigger pirna biogenesis by changing the chromatin of pirna clusters and inducing precursor processing. *Genes Dev* **28**, 1667–1680.

Lee, H.-C., Li, L., Gu, W., Xue, Z., Crosthwaite, S. K., Pertsemlidis, A., Lewis, Z. A., Freitag, M., Selker, E. U., Mello, C. C. and Liu, Y. (2010). Diverse pathways generate microrna-like rnas and dicer-independent small interfering rnas in fungi. *Mol Cell* **38**, 803–814.

Lee, H. Y., Zhou, K., Smith, A. M., Noland, C. L. and Doudna, J. A. (2013). Differential roles of human dicer-binding proteins trbp and pact in small rna processing. *Nucleic Acids Res* **41**, 6568–6576.

Lee, R. C., Feinbaum, R. L. and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843–54.

Lemaitre, B., Ronsseray, S. and Coen, D. (1993). Maternal repression of the p element promoter in the germline of drosophila melanogaster: a model for the p cytotype. *Genetics* **135**, 149–160.

Li, C., Vagin, V. V., Lee, S., Xu, J., Ma, S., Xi, H., Seitz, H., Horwich, M. D., Syrzycka, M., Honda, B. M., Kittler, E. L., Zapp, M. L., Klattenhoff, C., Schulz, N., Theurkauf, W. E., Weng, Z. and Zamore, P. D. (2009). Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* **137**, 509–21.

Li, J., Yang, Z., Yu, B., Liu, J. and Chen, X. (2005). Methylation protects mirnas and sirnas from a 3'-end uridylation activity in arabidopsis. *Curr Biol* **15**, 1501–1507.

Li, X., Roy, C., Dong, X., Bolcun-Filas, E., Wang, J., Han, B., Xu, J., Moore, M., Schimenti, J., Weng, Z. and Zamore, P. (2013). An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. *Molecular cell* **50**, 67–81.

Lin, H. and Spradling, A. C. (1997). A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the drosophila ovary. *Development* **124**, 2463–2476.

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

Liu, J., Rivas, F. V., Wohlschlegel, J., Yates, J. R. r., Parker, R. and Hannon, G. J. (2005). A role for the p-body component gw182 in microrna function. *Nat Cell Biol* **7**, 1261–1266.

Liu, K., Chen, C., Guo, Y., Lam, R., Bian, C., Xu, C., Zhao, D., Jin, J., MacKenzie, F., Pawson, T. and Min, J. (2010). Structural basis for recognition of arginine methylated Piwi proteins by the extended Tudor domain. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 18398–18403.

Liu, Q., Rand, T. A., Kalidas, S., Du, F., Kim, H.-E., Smith, D. P. and Wang, X. (2003). R2d2, a bridge between the initiation and effector steps of the drosophila rnai pathway. *Science* **301**, 1921–1925.

Lund, E. and Dahlberg (2006). Substrate Selectivity of Exportin 5 and Dicer in the Biogenesis of MicroRNAs. *Cold Spring Harbor Symposia on Quantitative Biology* **71**, 59–66. ISSN 1943-4456.

Luo, W. and Bentley, D. (2004). A ribonucleolytic rat torpedoes rna polymerase ii. *Cell* **119**, 911–914.

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

Makarova, K. S., Wolf, Y. I., van der Oost, J. and Koonin, E. V. (2009). Prokaryotic homologs of argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biol Direct* **4**, 29.

Mallory, A. and Vaucheret, H. (2010). Form, function, and regulation of argonaute proteins. *Plant Cell* **22**, 3879–3889.

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**, 522–35.

Mandel, C. R., Bai, Y. and Tong, L. (2008). Protein factors in pre-mrna 3'-end processing. *Cell Mol Life Sci* **65**, 1099–1122.

Mathioudakis, N., Palencia, A., Kadlec, J., Round, A., Tripsianes, K., Sattler, M., Pillai, R. and Cusack, S. (2012). The multiple Tudor domain-containing protein TDRD1 is a molecular scaffold for mouse Piwi proteins and piRNA biogenesis factors. *RNA (New York, N.Y.)* **18**, 2056–2072.

Matzke, M., Kanno, T., Daxinger, L., Huettel, B. and Matzke, A. J. M. (2009). Rna-mediated chromatin-based silencing in plants. *Curr Opin Cell Biol* **21**, 367–376.

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.

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

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G. and Tuschl, T. (2004). Human argonaute2 mediates rna cleavage targeted by mirnas and sirnas. *Mol Cell* **15**, 185–197.

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

Meyer, L. R., Zweig, A. S., Hinrichs, A. S., Karolchik, D., Kuhn, R. M., Wong, M., Sloan, C. A., Rosenbloom, K. R., Roe, G., Rhead, B., Raney, B. J., Pohl, A., Malladi, V. S., Li, C. H., Lee, B. T., Learned, K., Kirkup, V., Hsu, F., Heitner, S., Harte, R. A., Haeussler, M., Guruvadoo, L., Goldman, M., Giardine, B. M., Fujita, P. A., Dreszer, T. R., Diekhans, M., Cline, M. S., Clawson, H., Barber, G. P., Haussler, D. and Kent, W. J. (2013). The ucsc genome browser database: extensions and updates 2013. *Nucleic Acids Res* **41**, D64–9.

Mirkovic-Hosle, M. and Forstemann, K. (2014). Transposon defense by endo-sirnas, pirnas and somatic pilrnas in drosophila: contributions of loqs-pd and r2d2. *PLoS One* **9**, e84994.

Moazed, D., Buhler, M., Buker, S. M., Colmenares, S. U., Gerace, E. L., Gerber, S. A., Hong, E.-J. E., Motamedi, M. R., Verdel, A., Villen, J. and Gygi, S. P. (2006). Studies on the mechanism of rnai-dependent heterochromatin assembly. *Cold Spring Harb Symp Quant Biol* **71**, 461–471.

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

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

Morgan, H. D., Sutherland, H. G., Martin, D. I. and Whitelaw, E. (1999). Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet* **23**, 314–8.

Motamedi, M. R., Verdel, A., Colmenares, S. U., Gerber, S. A., Gygi, S. P. and Moazed, D. (2004). Two rnai complexes, rits and rdrc, physically interact and localize to noncoding centromeric rnas. *Cell* **119**, 789–802.

Muerdter, F., Guzzardo, P., Gillis, J., Luo, Y., Yu, Y., Chen, C., Fekete, R. and Hannon, G. (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.

Muller (1930). Types of visible variations induced by x-rays in drosophila. *Journal of Genetics* **22**, 299–334.

Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N. and Yamanaka, S. (2008). Generation of induced pluripotent stem cells without myc from mouse and human fibroblasts. *Nat Biotechnol* **26**, 101–106.

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

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

Nishida, K., Okada, T., Kawamura, T., Mituyama, T., Kawamura, Y., Inagaki, S., Huang, H., Chen, D., Kodama, T., Siomi, H. and Siomi, M. (2009). Functional involvement of Tudor and dPRMT5 in the piRNA processing pathway in Drosophila germlines. *The EMBO journal* **28**, 3820–3831.

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

Noma, K.-i., Sugiyama, T., Cam, H., Verdel, A., Zofall, M., Jia, S., Moazed, D. and Grewal, S. I. S. (2004). Rits acts in cis to promote rna interference-mediated transcriptional and post-transcriptional silencing. *Nat Genet* **36**, 1174–1180.

Ohrt, T., Muetze, J., Svoboda, P. and Schwille, P. (2012). Intracellular localization and routing of mirna and rnai pathway components. *Curr Top Med Chem* **12**, 79–88.

Okamura, K., Hagen, J. W., Duan, H., Tyler, D. M. and Lai, E. C. (2007). The mirtron pathway generates microrna-class regulatory rnas in drosophila. *Cell* **130**, 89–100.

Okita, K., Ichisaka, T. and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313–317.

Olivieri, D., Sykora, M., Sachidanandam, R., Mechtler, K. and Brennecke, J. (2010). An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila. *The EMBO journal* **29**, 3301–3317.

Olovnikov, I., Chan, K., Sachidanandam, R., Newman, D. K. and Aravin, A. A. (2013). Bacterial argonaute samples the transcriptome to identify foreign dna. *Mol Cell* **51**, 594–605.

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.

Park, W., Li, J., Song, R., Messing, J. and Chen, X. (2002). Carpel factory, a dicer homolog, and hen1, a novel protein, act in microrna metabolism in arabidopsis thaliana. *Curr Biol* **12**, 1484–1495.

Parker, G. S., Eckert, D. M. and Bass, B. L. (2006). Rde-4 preferentially binds long dsrna and its dimerization is necessary for cleavage of dsrna to sirna. *RNA* **12**, 807–818.

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

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, J. C. and Karpen, G. H. (2007). H3k9 methylation and rna interference regulate nucleolar organization and repeated dna stability. *Nat Cell Biol* **9**, 25–35.

Pezic, D., Manakov, S. A., Sachidanandam, R. and Aravin, A. A. (2014). pirna pathway targets active line1 elements to establish the repressive h3k9me3 mark in germ cells. *Genes Dev* **28**, 1410–1428.

Pontier, D., Yahubyan, G., Vega, D., Bulski, A., Saez-Vasquez, J., Hakimi, M.-A., Lerbs-Mache, S., Colot, V. and Lagrange, T. (2005). Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct rna polymerases iv in arabidopsis. *Genes Dev* **19**, 2030–2040.

Proudfoot, N. J. (1989). How rna polymerase ii terminates transcription in higher eukaryotes. *Trends Biochem Sci* **14**, 105–110.

Proudfoot, N. J. (2011). Ending the message: poly(a) signals then and now. *Genes Dev* **25**, 1770–1782.

Pruitt, K. D., Tatusova, T., Klimke, W. and Maglott, D. R. (2009). Ncbi reference sequences: current status, policy and new initiatives. *Nucleic Acids Res* **37**, D32–6.

Qi, Y., Denli, A. M. and Hannon, G. J. (2005). Biochemical specialization within arabidopsis rna silencing pathways. *Mol Cell* **19**, 421–428.

Qi, Y., He, X., Wang, X.-J., Kohany, O., Jurka, J. and Hannon, G. J. (2006). Distinct catalytic and non-catalytic roles of argonaute4 in rna-directed dna methylation. *Nature* **443**, 1008–1012.

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

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**, 1373–9.

Ro, S., Park, C., Song, R., Nguyen, D., Jin, J., Sanders, K., McCarrey, J. and Yan, W. (2007). Cloning and expression profiling of testis-expressed piRNA-like RNAs. *RNA (New York, N.Y.)* **13**, 1693–1702.

Roche, S. E. and Rio, D. C. (1998). Trans-silencing by p elements inserted in subtelomeric heterochromatin involves the drosophila polycomb group gene, enhancer of zeste. *Genetics* **149**, 1839–1855.

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

Ronsseray, S., Boivin, A. and Anxolabehere, D. (2001). P-element repression in drosophila melanogaster by variegating clusters of p-lacz-white transgenes. *Genetics* **159**, 1631–1642.

Rozhkov, N. V., Hammell, M. and Hannon, G. J. (2013). Multiple roles for Piwi in silencing Drosophila transposons. *Genes Dev* **27**, 400–12.

Rubin, G. M., Kidwell, M. G. and Bingham, P. M. (1982). The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. *Cell* **29**, 987–94.

Ruby, J. G., Jan, C. H. and Bartel, D. P. (2007). Intronic microrna precursors that bypass drosha processing. *Nature* **448**, 83–86.

Ruiz, Voinnet and Baulcombe (1998). Initiation and maintenance of virus-induced gene silencing. *Plant Cell* **10**, 937–946.

Saito, Ishizu, Komai, Kotani, Kawamura, Nishida, H, S. and MC, S. (2010). Roles for the yb body components armitage and yb in primary pirna biogenesis in drosophila. *Genes Dev* **24**, 2493–2498.

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.

Schirle, N. T. and MacRae, I. J. (2012). The crystal structure of human argonaute2. *Science* **336**, 1037–1040.

Schraivogel, D. and Meister, G. (2014). Import routes and nuclear functions of argonaute and other small rna-silencing proteins. *Trends Biochem Sci* **39**, 420–431.

Schupbach, T. and Wieschaus, E. (1989). Female sterile mutations on the second chromosome of drosophila melanogaster. i. maternal effect mutations. *Genetics* **121**, 101–117.

Schupbach, T. and Wieschaus, E. (1991). Female sterile mutations on the second chromosome of drosophila melanogaster. ii. mutations blocking oogenesis or altering egg morphology. *Genetics* **129**, 1119–1136.

Schurmann, N., Trabuco, L. G., Bender, C., Russell, R. B. and Grimm, D. (2013). Molecular dissection of human argonaute proteins by dna shuffling. *Nat Struct Mol Biol* **20**, 818–826.

Schwarz, D. S., Hutvagner, G., Du, T., Xu, Z., Aronin, N. and Zamore, P. D. (2003). Asymmetry in the assembly of the rnai enzyme complex. *Cell* **115**, 199–208.

Shoji, M., Tanaka, T., Hosokawa, M., Reuter, M., Stark, A., Kato, Y., Kondoh, G., Okawa, K., Chujo, T., Suzuki, T., Hata, K., Martin, S. L., Noce, T., Kuramochi-Miyagawa, S., Nakano, T., Sasaki, H., Pillai, R. S., Nakatsuji, N. and Chuma, S. (2009). The TDRD9-MIWI2 Complex Is Essential for piRNA-Mediated Retrotransposon Silencing in the Mouse Male Germline. *Developmental Cell* **17**, 775–787.

Shpiz, S., Kwon, D., Rozovsky, Y. and Kalmykova, A. (2009). rasiRNA pathway controls antisense expression of Drosophila telomeric retrotransposons in the nucleus. *Nucleic acids research* **37**, 268–278.

Shpiz, S., Olovnikov, I., Sergeeva, A., Lavrov, S., Abramov, Y., Savitsky, M. and Kalmykova, A. (2011). Mechanism of the piRNA-mediated silencing of Drosophila telomeric retrotransposons. *Nucleic acids research* **39**, 8703–8711.

Shpiz, S., Ryazansky, S., Olovnikov, I., Abramov, Y. and Kalmykova, A. (2014). Euchromatic transposon insertions trigger production of novel pi- and endo-sirnas at the target sites in the drosophila germline. *PLoS Genet* **10**, e1004138.

Sienski, G., Dönertas, D. and Brennecke, J. (2012). Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* **151**, 964–980.

Sigova, A., Rhind, N. and Zamore, P. D. (2004). A single argonaute protein mediates both transcriptional and posttranscriptional silencing in schizosaccharomyces pombe. *Genes Dev* **18**, 2359–2367.

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.

Song, S. U., Kurkulos, M., Boeke, J. D. and Corces, V. G. (1997). Infection of the germ line by retroviral particles produced in the follicle cells: a possible mechanism for the mobilization of the gypsy retroelement of Drosophila. *Development* **124**, 2789–98.

Soper, S., van der Heijden, G., Hardiman, T., Goodheart, M., Martin, S., de Boer, P. and Bortvin, A. (2008). Mouse maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis. *Developmental cell* **15**, 285–297+.

Spradling, A. C., Bellen, H. J. and Hoskins, R. A. (2011). Drosophila p elements preferentially transpose to replication origins. *Proc Natl Acad Sci U S A* **108**, 15948–15953.

Stage, D. E. and Eickbush, T. H. (2007). Sequence variation within the rrna gene loci of 12 drosophila species. *Genome Res* **17**, 1888–1897.

Stuwe, E., Toth, K. F. and Aravin, A. A. (2014). Small but sturdy: small rnas in cellular memory and epigenetics. *Genes Dev* **28**, 423–431.

Swarts, D. C., Jore, M. M., Westra, E. R., Zhu, Y., Janssen, J. H., Snijders, A. P., Wang, Y., Patel, D. J., Berenguer, J., Brouns, S. J. J. and van der

Oost, J. (2014). Dna-guided dna interference by a prokaryotic argonaute. *Nature* **507**, 258–261.

Szakmary, A., Reedy, M., Qi, H. and Lin, H. (2009). The Yb protein defines a novel organelle and regulates male germline stem cell self-renewal in Drosophila melanogaster. *The Journal of Cell Biology* **185**. ISSN 0021-9525.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K. and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872.

Takahashi, K. and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676.

Thibault, S. T., Singer, M. A., Miyazaki, W. Y., Milash, B., Dompe, N. A., Singh, C. M., Buchholz, R., Demsky, M., Fawcett, R., Francis-Lang, H. L., Ryner, L., Cheung, L. M., Chong, A., Erickson, C., Fisher, W. W., Greer, K., Hartouni, S. R., Howie, E., Jakkula, L., Joo, D., Killpack, K., Laufer, A., Mazzotta, J., Smith, R. D., Stevens, L. M., Stuber, C., Tan, L. R., Ventura, R., Woo, A., Zakrajsek, I., Zhao, L., Chen, F., Swimmer, C., Kopczynski, C., Duyk, G., Winberg, M. L. and Margolis, J. (2004). A complementary transposon tool kit for drosophila melanogaster using p and piggybac. *Nat Genet* **36**, 283–287.

Tomari, Y., Du, T., Haley, B., Schwarz, D. S., Bennett, R., Cook, H. A., Koppetsch, B. S., Theurkauf, W. E. and Zamore, P. D. (2004). Risc assembly defects in the drosophila rnai mutant armitage. *Cell* **116**, 831–841.

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

van der Krol, A. R., Mur, L. A., de Lange, P., Mol, J. N. and Stuitje, A. R. (1990). Inhibition of flower pigmentation by antisense chs genes: promoter and minimal sequence requirements for the antisense effect. *Plant Mol Biol* **14**, 457–466.

Vogel, J. (2014). Biochemistry. a bacterial seek-and-destroy system for foreign dna. *Science* **344**, 972–973.

Voigt, F., Reuter, M., Kasaruho, A., Schulz, E., Pillai, R. and Barabas, O. (2012). Crystal structure of the primary piRNA biogenesis factor Zucchini reveals

similarity to the bacterial PLD endonuclease Nuc. *RNA (New York, N.Y.)* **18**, 2128–2134.

Volpe, T. A., Kidner, C., Hall, I. M., Teng, G., Grewal, S. I. S. and Martienssen, R. A. (2002). Regulation of heterochromatic silencing and histone h3 lysine-9 methylation by rnai. *Science* **297**, 1833–1837.

Wang, S. H. and Elgin, S. C. R. (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., Yoshikawa, M., Han, B. W., Izumi, N., Tomari, Y., Weng, Z. and Zamore, P. D. (2014). The initial uridine of primary pirnas does not create the tenth adenine that is the hallmark of secondary pirnas. *Mol Cell* **56**, 708–716.

Wang, Y., Juranek, S., Li, H., Sheng, G., Wardle, G. S., Tuschl, T. and Patel, D. J. (2009). Nucleation, propagation and cleavage of target rnas in ago silencing complexes. *Nature* **461**, 754–761.

Wang, Y., Sheng, G., Juranek, S., Tuschl, T. and Patel, D. J. (2008). Structure of the guide-strand-containing argonaute silencing complex. *Nature* **456**, 209–213.

Weinmann, L., Hock, J., Ivacevic, T., Ohrt, T., Mutze, J., Schwille, P., Kremmer, E., Benes, V., Urlaub, H. and Meister, G. (2009). Importin 8 is a gene silencing factor that targets argonaute proteins to distinct mrnas. *Cell* **136**, 496–507.

Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B. E. and Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent es-cell-like state. *Nature* **448**, 318–324.

West, S., Gromak, N. and Proudfoot, N. J. (2004). Human 5' - 3' exonuclease xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* **432**, 522–525.

White, E., Schlackow, M., Kamieniarz-Gdula, K., Proudfoot, N. J. and Gullerova, M. (2014). Human nuclear dicer restricts the deleterious accumulation of endogenous double-stranded rna. *Nat Struct Mol Biol* **21**, 552–559.

Wilson, C., Pearson, R. K., Bellen, H. J., O'Kane, C. J., Grossniklaus, U. and Gehring, W. J. (1989). P-element-mediated enhancer detection: an efficient method for isolating and characterizing developmentally regulated genes in drosophila. *Genes Dev* **3**, 1301–1313.

Wuarin, J. and Schibler, U. (1994). Physical isolation of nascent rna chains transcribed by rna polymerase ii: evidence for cotranscriptional splicing. *Mol Cell Biol* **14**, 7219–7225.

Xiang, S., Cooper-Morgan, A., Jiao, X., Kiledjian, M., Manley, J. L. and Tong, L. (2009). Structure and function of the 5'- 3' exoribonuclease rat1 and its activating partner rai1. *Nature* **458**, 784–788.

Xue, Y., Bai, X., Lee, I., Kallstrom, G., Ho, J., Brown, J., Stevens, A. and Johnson, A. W. (2000). Saccharomyces cerevisiae rai1 (ygl246c) is homologous to human dom3z and encodes a protein that binds the nuclear exoribonuclease rat1p. *Mol Cell Biol* **20**, 4006–4015.

Yamanaka, S., Siomi, M. C. and Siomi, H. (2014). pirna clusters and open chromatin structure. *Mob DNA* **5**, 22.

Yang, Z., Ebright, Y. W., Yu, B. and Chen, X. (2006). Hen1 recognizes 21-24 nt small rna duplexes and deposits a methyl group onto the 2' oh of the 3' terminal nucleotide. *Nucleic Acids Res* **34**, 667–675.

Yi, R., Qin, Y., Macara, I. G. and Cullen, B. R. (2003). Exportin-5 mediates the nuclear export of pre-micrornas and short hairpin rnas. *Genes Dev* **17**, 3011–3016.

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R. W., Steward, R. and Chen, X. (2005). Methylation as a crucial step in plant microrna biogenesis. *Science* **307**, 932–935.

Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R., Slukvin, I. I. and Thomson, J. A. (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920.

Zekri, L., Huntzinger, E., Heimstadt, S. and Izaurralde, E. (2009). The silencing domain of gw182 interacts with pabpc1 to promote translational repression and degradation of microrna targets and is required for target release. *Mol Cell Biol* **29**, 6220–6231.

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**, 572–84.

Zhu, J.-K. (2008). Reconstituting plant mirna biogenesis. *Proc Natl Acad Sci U S A* **105**, 9851–9852.

Zofall, M. and Grewal, S. I. (2006). Swi6/HP1 recruits a JmjC domain protein to facilitate transcription of heterochromatic repeats. *Mol Cell* **22**, 681–92.