

**On piRNAs and PIWI proteins in**  
*Schmidtea mediterranea*  
**and their role in mRNA surveillance of adult stem cells**



**UNIVERSITÄT  
BAYREUTH**

DISSERTATION

submitted to the Bayreuth Graduate School of Mathematical and Natural  
Sciences (BayNAT) of the University of Bayreuth for obtaining the academic  
grade of

*Doctor rerum naturalium (Dr. rer. nat.)*

**by**

**Iana V. Kim**

from Zhanatas, Kazakhstan

February 2020



This doctoral thesis was carried out in the laboratory "Gene regulation by non-coding RNA" at the University of Bayreuth from January 2015 until October 2019 and was supervised by Dr. Claus-D. Kuhn.

This is a full reprint of the thesis submitted to obtain the academic degree of Doctor of Natural Sciences (Dr. rer. nat.) and approved by the Bayreuth Graduate School of Mathematical and Natural Sciences (BayNAT) of the University of Bayreuth.

Date of submission: 28.10.2019

Date of defense: 31.01.2020

Acting director: Prof. Dr. Markus Lippitz

Doctoral committee:

Dr. Claus-Dieter Kuhn (reviewer)

Prof. Dr. Olaf Stemmann (reviewer)

Prof. Dr. Gerrit Begemann (chairman)

Prof. Dr. Andreas Möglich

additional reviewer: Prof. Dr. Gunter Meister



## Copyright information

Results from this dissertation were published in their entirety or in part in the following manuscripts:

1. Kim I.V., Duncan E.M., Ross E.J., Gorbovytska V., Nowotarski S.H., Elliott S.A., Alvarado A.S., Kuhn C.D. (2019). Planarians recruit piRNAs for mRNA turnover in adult stem cells. *Genes Dev.* 33:1575-1590. doi:10.1101/gad.322776.118.
2. Kim I.V., Ross E.J., Döring K., Dietrich S., Alvarado A.S., Kuhn C.D. (2019). Efficient depletion of ribosomal RNA for RNA sequencing in planarians. *BMC Genomics* 20, 909. doi.org/10.1186/s12864-019-6292-y.
3. Kim I.V., Riedelbauch S.H., Kuhn C.D. (2020). Connecting pieces: highlights of the piRNA pathway in planarian flatworms. *Biol Chem.* In revision (invited review).

All immunostaining experiments were conducted by Dr. Elizabeth M. Duncan, Dr. Stephanie H. Nowotarski and Dr. Sarah A. Elliott.

Dr. Elizabeth M. Duncan performed in situ hybridization experiments.

Eric J. Ross annotated planarian piRNA clusters and provided sequences of planarian ribosomal RNA.

Dr. Sean McKinney quantified the localization of foci on in situ hybridization experiment images.

Dr. Kristina Döring and Dr. Sascha Dietrich prepared and analyzed RNA-sequencing libraries from the gram-negative bacterium *Salmonella typhimurium*.



## Summary

PIWI proteins utilize small RNAs, called piRNAs, to silence transposable elements, thereby protecting the integrity of animal germline genome. In the planarian flatworm *Schmidtea mediterranea*, piRNAs and PIWI proteins are enriched in adult stem cells, called neoblasts. Neoblasts give rise to all planarian cell types ensuring tissue turnover and animal regeneration. The knockdown of two planarian PIWI proteins - SMEDWI-2 and SMEDWI-3 - impairs neoblasts development and deprives planarians of their ability to regenerate, resulting in a lethal phenotype.

This study is focused on the characterization of the planarian piRNA pathway and on understanding the essential functions of piRNAs in neoblast biology. To gain insights into the planarian piRNA machinery, a combination of immunostaining and small RNA sequencing of immunoprecipitated piRNAs was used. We discovered two functionally distinct classes of piRNAs in planarians: piRNAs that map to transposable elements (TE) and those that map to coding genes. TE-derived piRNAs are produced during the post-transcriptional degradation of TE transcripts in a heterotypic ping-pong cycle, which is predominantly constituted by SMEDWI-2 and SMEDWI-3. Genic piRNAs, on the other hand, originate from the homotypic ping-pong cleavage of coding transcripts by SMEDWI-3 only. The guide piRNAs that direct the degradation of both transposable elements and protein-coding mRNAs are encoded in uni-strand genomic clusters.

Further exploration of SMEDWI-3-targeted transcripts using combination of RNA-Seq, HITS-CLIP, CLASH and Degradome-Seq revealed the engagement of the piRNA pathway in mRNA surveillance in neoblasts. In particular, we find that one class of mRNAs is degraded in homotypic ping-pong cycle by SMEDWI-3, while another class of targeted transcripts evades SMEDWI-3-mediated endonucleolytic cleavage and processing into piRNAs. The two distinct modes of SMEDWI-3 activity are dictated by the degree of complementarity between a SMEDWI-3-bound guide piRNA and its target mRNA. In case guide piRNAs bind their targets with non-perfect complementarity, especially in the vicinity to the scissile phosphate bond between nucleotides 10 and 11, the degradation of the mRNA transcript by SMEDWI-3 is precluded. Otherwise, perfect or near-perfect base-pairing of guide piRNA to its target triggers the endonucleolytic cleavage of the targeted mRNA and the production of genic piRNAs. Further analysis of planarian mRNAs degraded in the piRNA pathway revealed the

presence of an active piRNA pathway in epidermal cells, in which these transcripts are likely used as piRNAs precursors.

Taken together, our data demonstrate that a piRNA pathway comprising three PIWI proteins (SMEDWI-1, -2, and -3) operates in neoblasts of *S. mediterranea*. However, only two PIWI proteins - SMEDWI-2 and -3 - are essential for regeneration and tissue homeostasis. Moreover, we find a somatic piRNA pathway that acts in epidermal cells with the use of SMEDWI-2 only. Another somatic pathway may also operate in neuronal cells, where we observed two SMEDWI proteins - SMEDWI-2 and -3. Functionally, the planarian piRNA pathway is responsible for silencing of transposable elements, which is the primary function of piRNA pathway in most animal. In addition, planarians have recruited piRNAs for mRNA surveillance - at least in neoblasts. Overall, this study sets the stage for studying the role of PIWI proteins and piRNAs in the planarian flatworm *Schmidtea mediterranea*. Moreover, it will serve as a catalyst for the use of numerous biochemical and sequencing techniques in planarians that were established during the course of this thesis.

## Zusammenfassung

PIWI-Proteine nutzen kleine RNAs, sogenannte piRNAs, um Transposons abzubauen und damit die Integrität des Genoms der tierischen Keimbahn zu schützen. In dem zu den Planarien zugehörigen Plattwurm *Schmidtea mediterranea* sind PIWI-Proteine für dessen faszinierende Regenerationsfähigkeit und das Überleben der Tiere unerlässlich. Die Regenerationsfähigkeit von *S. mediterranea* beruht dabei auf adulten Stammzellen, sogenannten Neoblasten. In Neoblasten werden drei SMEDWI-Proteine exprimiert, wobei nur zwei - SMEDWI-2 und -3 – essentiell für die Tiere und deren Regeneration sind.

Diese Studie konzentriert sich auf die Charakterisierung des piRNA-Weges in Planarien und auf das Verständnis der wesentlichen Funktionen von piRNAs in der Biologie der Neoblasten. Um einen Einblick in die piRNA-Maschinerie zu erhalten, wurde eine Kombination aus Immunfärbungen und der Sequenzierung immunpräzipitierter piRNAs verwendet. Die gewonnenen Daten zeigen, dass piRNAs im Genom von Planarien hauptsächlich in einzelsträngigen Clustern kodiert werden. Des Weiteren entdeckten wir zwei funktionell unterschiedliche Klassen von piRNAs in Planarien: piRNAs, die Sequenzen von Transposons (TE) aufweisen sowie solche, die aus Sequenzen protein-kodierender Gene bestehen. TE-basierte piRNAs werden während des post-transkriptionellen Abbaus von TE-Transkripten in einem heterotypischen Ping-Pong-Zyklus produziert, der überwiegend mithilfe von SMEDWI-2 und SMEDWI-3 betrieben wird. Von Genen abstammende piRNAs werden in einem homotypischen Ping-Pong-Zyklus von SMEDWI-3 hergestellt.

Weitere Untersuchungen an SMEDWI-3-interagierenden Transkripten unter Verwendung einer Kombination aus RNA-Seq, HITS-CLIP, CLASH-Seq und Degradome-Seq zeigten die Beteiligung der piRNA-Maschinerie am mRNA-Umsatz in Neoblasten. Unsere Ergebnisse zeigen insbesondere, dass eine Klasse von mRNAs im homotypischen Ping-Pong-Zyklus durch SMEDWI-3 abgebaut wird, während eine andere Klasse von Transkripten eine SMEDWI-3-vermittelte endonukleolytische Spaltung und Prozessierung in piRNAs vermeidet. Die zwei von uns entdeckten Modi der Aktivität von SMEDWI-3 werden durch das Ausmaß an Komplementarität zwischen Ziel-Transkript und PIWI-geladener piRNA reguliert. Im Falle einer nicht perfekten Komplementarität der piRNA zur Ziel-mRNA, insbesondere in der Region nahe der nukleolytischen Spaltstelle (zwischen Nukleotid 10 und 11), ist der Abbau der erkannten mRNA durch SMEDWI-3 nicht möglich. Im Falle nahezu perfekter Komplementarität von piRNA und mRNA wird die Ziel-mRNA gespalten. Die weitere

Analyse von mRNAs, die durch piRNAs abgebaut werden, zeigten das Vorhandensein eines aktiven piRNA-Pfades in epidermalen Zellen, in denen diese Transkripte wahrscheinlich als piRNAs-Vorläufer verwendet werden.

Insgesamt charakterisiert die vorliegende Arbeit den piRNA-Weg in Neoblasten von *S. mediterranea*. Darüber hinaus entdeckt diese Arbeit einen somatische piRNA-Weg in epidermalen Zellen, der nur unter Verwendung von SMEDWI-2 abläuft. Sie schlägt zudem einen piRNA-Weg in neuronalen Zellen vor, in denen wir zwei SMEDWI-Proteine - SMEDWI-2 und -3 - entdeckten. Der piRNA-Weg in Planarien ist funktionell auf die Stilllegung von Transposons ausgerichtet, beteiligt sich jedoch auch am Abbau bestimmter mRNAs in Neoblasten. Darüber hinaus wurden in dieser Arbeit zahlreiche biochemische und Hochdurchsatz-Techniken für deren Einsatz in Planarien angepasst. Insgesamt stellt diese Studie deshalb eine enorme Erweiterung unserer Kenntnisse über die Rolle von PIWI-Proteinen und piRNAs bei der Regeneration von Süßwasser-Planarien der Spezies *Schmidtea mediterranea* dar.

*The modern ideal has its type in art, and its means is science. It is through science that it will realize that august vision of the poets, the socially beautiful. Eden will be reconstructed  
by A+B.*

Victor Hugo, *Les Misérables*



## Acknowledgement

First of all, I would like to thank my supervisor Dr. Claus-D. Kuhn, who believed in me and gave me the opportunity to work on this project. His constant support and patience in guiding me through my PhD research were invaluable and highly appreciated. I am extremely grateful to him for his constructive criticism and his extensive discussions around my work, as well as for the proofreading of this thesis. Additionally, I thank my mentors, Prof. Dr. Birgitta Wörl and Prof. Dr. Olaf Stemmann, for their helpful comments and encouragement during our committee meetings.

This work would not have been possible without the contribution of collaborators I was so lucky to work with. I am extremely grateful to Prof. Dr. Alejandro Sánchez Alvarado for hosting me at the Stowers Institute for Medical Research in Kansas city, and for his insightful comments and suggestions. My deepest appreciation goes to Dr. Elizabeth M. Duncan, who provided ongoing scientific support, conducted immunostaining and in situ hybridization experiments, FACS sorted neoblasts and always warmly supported me. Her involvement in the project and assistance cannot be overemphasized. I also sincerely thank Eric J. Ross for his bioinformatics expertise and his perceptive suggestions. I would like to acknowledge Dr. Stephanie H. Nowotarski and Dr. Sarah A. Elliott for generating wonderful immunostaining images. In addition, I am very grateful to Stephanie for her enormous effort in sorting 8 million neoblasts cells for our Degradome-seq libraries. Many thanks to Vlada Gorbovytska for deciphering the structure of mRNAs and conducting SHAPE experiments for our paper. I would like to acknowledge Dr. Kristina Döring for making our sequencing runs smooth and flawless, and Dr. Sascha Dietrich for his input in testing our rRNA depletion workflow. I also thank Prof. Dr. Gunter Meister, Norbert Eichner and Gerhard Lehmann for their hospitality in Regensburg, for teaching me how to produce small RNA-Seq libraries and nurturing my first steps in bioinformatic analysis.

I would like to express my gratitude to Silke Spudeit. Her cheerful presence and technical support have helped me a great deal over these years.

Special thanks to Sebastian H. Riedelbauch for proofreading the introduction section of this thesis and for translating English summary into German.

I would like to acknowledge my colleagues, present and former, for sharing the struggle of daily laboratory life, for our scientific and philosophical discussions, and, of course, our many wonderful lunches.

Lastly, I would like to thank my mother and sister, Tatiana and Alena Kim, and Stavros Athanasopoulos. Their unconditional love and endless patience have helped me through the roller coaster of this research journey.

# Table of contents

<b>List of figures</b>	<b>xix</b>
<b>List of tables</b>	<b>xxiii</b>
<b>List of abbreviations</b>	<b>xxv</b>
<b>Preamble</b>	<b>xxvii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 piRNA biogenesis . . . . .	1
1.1.1 Discovery of small silencing RNAs . . . . .	1
1.1.2 Structural details and expression characteristics of PIWI proteins . .	2
1.1.3 Genomic origin of piRNAs . . . . .	3
1.1.4 Biogenesis of phased piRNAs . . . . .	7
1.1.5 The ping-pong cycle . . . . .	7
1.2 Biological function of piRNAs . . . . .	9
1.3 The piRNA pathway in <i>S. mediterranea</i> . . . . .	10
1.3.1 Planarians as a model organism . . . . .	10
1.3.2 Planarian PIWI proteins . . . . .	12
<b>2 Results part 1</b>	<b>15</b>
2.1 The role of piRNA pathway in <i>S. mediterranea</i> . . . . .	15
2.1.1 Immunoprecipitation of piRNAs associated with SMEDWI proteins	16
2.1.2 SMEDWI-2 and SMEDWI-3 cooperate to degrade planarian trans- posons . . . . .	20
2.1.3 Genic piRNAs bound to SMEDWI-3 are degradation products of planarian mRNAs . . . . .	24

2.1.4	Crosslinking immunoprecipitation confirms that SMEDWI-3 targets hundreds of planarian transcripts . . . . .	27
2.1.5	The dual role of SMEDWI-3 in neoblast mRNA surveillance . . . . .	32
2.1.6	The base-pairing pattern between SMEDWI-3-bound piRNAs and target mRNAs . . . . .	35
2.1.7	Transcripts degraded by SMEDWI-3 reveal a piRNA pathway in the planarian epidermis . . . . .	36
2.2	Discussion of Results part 1 . . . . .	40
2.2.1	Planarian piRNAs may function in cell cycle regulation and immune defense . . . . .	41
2.2.2	The potential role of SMEDWI-3 in determining neoblast mRNA fate . . . . .	42
<b>3</b>	<b>Results part 2</b>	<b>45</b>
3.1	rRNA depletion workflow . . . . .	45
3.1.1	Development of an efficient rRNA depletion protocol for planarians . . . . .	45
3.1.2	Detailed rRNA depletion workflow . . . . .	47
3.1.3	Ribosomal RNA depletion in planarian species related to <i>S. mediterranea</i> . . . . .	49
3.1.4	Comparison of RNA-seq libraries prepared using rRNA-depletion or poly(A) selection . . . . .	51
3.1.5	Non-specific depletion of coding transcripts in ribodepleted libraries . . . . .	54
3.1.6	Applicability of the described ribodepletion method to other organisms . . . . .	56
3.2	Discussion part 2 . . . . .	57
<b>4</b>	<b>Conclusion</b>	<b>59</b>
<b>5</b>	<b>Experimental procedures</b>	<b>63</b>
5.1	Experimental procedures of Results part 1 . . . . .	63
5.1.1	Chemicals reagents . . . . .	63
5.1.2	Primer's sequences . . . . .	63
5.1.3	Planarian culture . . . . .	63
5.1.4	Antibodies production and purification . . . . .	63
5.1.5	Expression of Strep-tagged SMEDWI . . . . .	64
5.1.6	Western blotting . . . . .	64
5.1.7	RNA interference . . . . .	64
5.1.8	Flow cytometry . . . . .	65

---

5.1.9	Slide-section immunofluorescence . . . . .	65
5.1.10	Whole-mount in situ hybridization . . . . .	66
5.1.11	Immunoprecipitation of SMEDWI proteins . . . . .	66
5.1.12	Small RNA library preparation . . . . .	67
5.1.13	Ribosomal RNA depletion . . . . .	67
5.1.14	RNA library preparation . . . . .	67
5.1.15	SMEDWI-3 HITS-CLIP . . . . .	68
5.1.16	SMEDWI-3 CLASH . . . . .	68
5.1.17	Degradome library preparation . . . . .	69
5.1.18	Processing of small RNA libraries . . . . .	69
5.1.19	Calculation of the ping-pong signature . . . . .	69
5.1.20	Sequence logos and Venn diagrams . . . . .	70
5.1.21	piRNA cluster prediction . . . . .	70
5.1.22	Processing of the ChIP-Seq libraries . . . . .	70
5.1.23	Processing of RNA-seq libraries . . . . .	70
5.1.24	Processing of HITS-CLIP libraries . . . . .	71
5.1.25	Genic piRNA density across SMEDWI-3 HITS-CLIP target sites . .	71
5.1.26	Gene-set enrichment analysis (GAGE) . . . . .	71
5.1.27	Processing of Degradome-seq libraries . . . . .	72
5.1.28	GO term enrichment analysis . . . . .	72
5.1.29	Identification of chimeric reads . . . . .	72
5.1.30	Data availability . . . . .	72
5.2	Experimental procedures of Results part 2 . . . . .	72
5.2.1	Ribosomal RNA depletion . . . . .	72
5.2.2	Planarian rRNA-depleted RNA-Seq dataset . . . . .	72
5.2.3	Publicly available RNA-Seq datasets . . . . .	73
5.2.4	Salmonella typhimurium SL1344 datasets . . . . .	73
5.2.5	Processing of RNA-Seq libraries . . . . .	73
5.2.6	Phylogenetic tree . . . . .	74
5.2.7	Analysis of DNA probe specificity . . . . .	74
	<b>References</b>	<b>75</b>
	<b>Appendix A Adapter and primer sequences</b>	<b>93</b>

---

<b>Appendix B Experimental procotols</b>	<b>99</b>
B.1 dsRNA synthesis for RNA interference . . . . .	99
B.2 Immunoprecipitation of SMEDWI . . . . .	104
B.3 TruSeq small RNA library preparation . . . . .	109
B.4 Planarian rRNA depletion . . . . .	117
B.5 SMEDWI-3 HITS-CLIP . . . . .	120
B.6 Degradome-seq library preparation . . . . .	136
<b>Appendix C Data processing pipelines</b>	<b>141</b>
C.1 small RNA-seq pipeline . . . . .	141
C.2 RNA-seq pipeline . . . . .	145
C.3 Processing of HITS-CLIP libraries . . . . .	146
<b>(Eidesstattliche) Versicherungen und Erklärungen</b>	<b>151</b>

# List of figures

1.1	Temporally controlled expression of PIWIs during spermatogenesis in mice	3
1.2	Three types of piRNA clusters	4
1.3	Production of phased piRNAs	6
1.4	Homotypic and heterotypic ping-pong cycles	8
1.5	Phylogenetic tree illustrating the similarity of amino acid sequences of PIWI proteins	12
1.6	Sequence alignment of PIWI domains of PIWI proteins	13
2.1	Specificity of anti-SMEDWI antibody	16
2.2	Co-immunostaining of SMEDWI proteins on cross-sections through planarian brain and pharynx	17
2.3	Immunoprecipitation of SMEDWI proteins	18
2.4	Characterization of piRNAs associated with SMEDWI proteins.	19
2.5	Local Z-score for the occurrence of homotypic and heterotypic ping-pong pairs.	20
2.6	Evaluation of SMEDWIs RNAi knockdown	21
2.7	Immunostaining of SMEDWI-1, -2 and -3 on cross-sections through the planarian pharynx in control ( <i>unc-22 RNAi</i> ) or knockdown ( <i>smedwi-1,-2,-3 RNAi</i> )	23
2.8	Differentially expressed transposable element families upon RNAi knockdown of <i>smedwi-2</i> or <i>smedwi-3</i>	24
2.9	SMEDWI-3 binds a diverse class of genic piRNAs	25
2.10	Genic piRNAs exhibit ping-pong signature	26
2.11	PAGE analysis of SMEDWI-3 HITS-CLIP and CLASH libraries	27
2.12	CIMS analysis and identification of chimeric reads	30
2.13	Identification of SMEDWI-3-targeted transcripts in HITS-CLIP and CLASH libraries	31

2.14	SMEDWI-3 degrades planarian mRNAs using the ping-pong cycle . . . . .	33
2.15	Three groups of SMEDWI-3 CLIP targets . . . . .	34
2.16	The base-pairing patterns between SMEDWI-3-bound piRNAs and target mRNAs . . . . .	36
2.17	Differential expression of the three distinct SMEDWI-3 target groups upon <i>smedwi-2</i> ) and <i>smedwi-3(RNAi)</i> . . . . .	37
2.18	GO term enrichment analysis of ping-pong degraded group 1 SMEDWI-3 targets. . . . .	37
2.19	SMEDWI-3 group 1 ping-pong targets form nuclear foci in the planarian epidermis . . . . .	39
2.20	Immunostaining of SMEDWI-1 and SMEDWI-2 in the epidermis on cross-sections through the pharynx of <i>S. mediterranea</i> upon differential RNAi knockdown . . . . .	40
3.1	Separation profile of total RNA and rRNA-depleted RNA of <i>S. mediterranea</i>	46
3.2	Schematic representation of rRNA removal protocol . . . . .	49
3.3	Phylogenetic tree showing the taxonomic position of the analyzed planarian species. . . . .	50
3.4	Total RNA separation profiles after the depletion of rRNA of other freshwater triclads . . . . .	51
3.5	Comparison of rRNA-depleted and poly(A)-enriched planarian RNA-seq libraries . . . . .	52
3.6	Difference in the expression value of histones and TEs in rRNA-depleted libraries . . . . .	53
3.7	Off-target analysis of DNA probes used for rRNA depletion . . . . .	55
3.8	Application of the developed rRNA workflow to <i>Salmonella typhimurium</i> using organism-specific probes . . . . .	57
4.1	Model of planarian piRNA pathway in neoblasts and epidermis . . . . .	61
B.1	Template amplification for in vitro transcription . . . . .	100
B.2	In vitro transcribed dsRNA. . . . .	102
B.3	Prepared small RNA-seq libraries . . . . .	111
B.4	Bioanalyzer electropherogram of prepared small RNA library . . . . .	116
B.5	Separation profile of planarian total RNA and rRNA depleted RNA . . . . .	119
B.6	SMEDWI-3 HITS-CLIP immunoprecipitated complexes . . . . .	125

---

B.7	PAGE analysis of SMEDWI-3 crosslinked RNA species from the excised upper or lower band . . . . .	131
B.8	PAGE analysis of prepared HITS-CLIP libraries . . . . .	133
B.9	Bioanalyzer electropherogram of prepared HITS-CLIP libraries . . . . .	135
B.10	Separation profile of degradome-seq libraries . . . . .	140



# List of tables

2.1	Sequenced HITS-CLIP and CLASH reads. . . . .	28
2.2	HITS-CLIP and CLASH reads mapped to the planarian genome. . . . .	29
A.1	Primer sequences used for cloning N-terminal 200 amino acids of SMEDWI-3 into pGEX vector . . . . .	93
A.2	Primer sequences used for cloning Strep-tagged SMEDWIs into pFL vectors	94
A.3	Primer sequences used for cloning gene fragments into pPR-T4P vector . .	94
A.4	Primer sequences used for in vitro transcription . . . . .	95
A.5	Primers and adapters for small RNA-seq library preparation . . . . .	96
A.6	Primers and adapters for HITS-CLIP . . . . .	96
A.7	Primers and adapters for Degradome-seq . . . . .	97



# List of abbreviations

*C. elegans* *Caenorhabditis elegans*

*D. melanogaster* *Drosophila melanogaster*

*M. musculus* *Mus musculus*

*S. mediterranea* *Schmidtea mediterranea*

dpf days post feeding

dsRNA double-stranded RNA

*E. coli* *Escherichia coli*

GST Glutathione S-transferase

kb kilobase

LTR long terminal repeat

miRNAs microRNAs

nt nucleotide

OSCs ovarian somatic cells

piRNAs PIWI-interacting RNAs

PIWI P-element Induced Wimpy testis

poly(A) polyadenylated

PTGS post-transcriptional gene silencing

RNAi RNA interference

rRNA ribosomal RNA

RT Room Temperature

siRNAs small interfering RNAs

TE transposable element

UTR untranslated region

# Preamble

This thesis contains five chapters that are either published (ch. 2 and parts of ch. 5) or submitted for publication in a peer-reviewed journal (ch. 3 and parts of ch. 5) or they are being prepared for publication (ch. 1).

The first chapter comprises a broad overview of our current knowledge of the piRNA pathway and its functions in different organisms. Although the architecture of the piRNA pathway is quite diverse throughout the animal kingdom, this thesis only describes the mechanistic aspects of the pathway in the fruit fly (*Drosophila melanogaster*) and in mouse (*Mus musculus*) as two examples of well-studied organisms. An overview of the functional aspects of the pathway attempts to cover the entire variety of piRNA functions. These functions reach from the conserved role of piRNAs in counteracting transposable elements to their role in mRNA surveillance, from specific genome editing in ciliates to antiviral defense mechanisms in mosquitoes. This thesis explores the piRNA pathway in the planarian flatworm *Schmidtea mediterranea*, a model organism with a large population of adult stem cells, called neoblasts. Neoblasts equip planarians with amazing regenerative abilities and a seemingly infinite life. Thus, the first chapter also illustrates why planarians are suitable for the study of piRNAs and summarizes current knowledge about the key molecular players involved in the planarian piRNAs pathway.

The second chapter of this thesis comprises the analysis of planarian piRNAs and explores the roles of PIWI proteins in neoblast biology. It describes the main principles of the piRNA pathway organization and localization in *S. mediterranea*. Additionally, this chapter demonstrates the involvement of planarian piRNA pathway in mRNA surveillance of adult stem cells.

In chapter three, a newly developed workflow for the efficient depletion of ribosomal RNA (rRNA) in planarian flatworms is presented. The protocol is based on the subtractive hybridization using organism-specific probes and can be applied to other freshwater triclad families. In addition, the presented rRNA-depletion procedure was successfully applied to the removal of ribosomal RNA from the gram-negative bacterium *Salmonella typhimurium*,

which indicates that the workflow can be adapted to any prokaryotic or eukaryotic species of choice.

The fourth chapter summarizes the main findings of this thesis.

The final chapter (chapter 5) describes the experimental and computational procedures used in this thesis.

# Chapter 1

## Introduction

### 1.1 piRNA biogenesis

#### 1.1.1 Discovery of small silencing RNAs

Small silencing RNAs are short (21 - 35 nt) non-coding RNAs associated with members of the Argonaute protein family. The small RNA thereby serves as a guide to target transcripts leading to the reduction in expression level of the latter. In 1993, two groups [97, 195] simultaneously reported that small RNA *lin-4* negatively regulates the level of LIN-14 protein in *Caenorhabditis elegans* (*C. elegans*) by partial complementary base pairing to multiple sites in the 3' untranslated region (UTR) of *lin-14* mRNA, thus inhibiting its translation [49]. This event marks the discovery of the first class of small silencing RNAs, micro RNAs (miRNAs) in eukaryotes. At present, thousands of miRNAs were characterized and found to be conserved from plants to humans [86].

Next in a line of revolutionary discoveries that shaped our current understanding of small RNA-mediated silencing was a report by A. Fire and C. Mello (1998) that described a phenomenon termed RNA interference (RNAi) [42]. RNAi denotes a type of post-transcriptional gene silencing (PTGS), in which double-stranded RNA (dsRNA) is processed into small interfering RNAs (siRNAs) that then direct target RNA recognition followed by target cleavage.

Finally, in 2001 a third class of small regulatory RNAs, PIWI-interacting RNAs (piRNAs), was discovered in *D. melanogaster* testis by A. Aravin and colleagues [7]. In the initial study, 25 - 27 nt long RNAs were detected and found responsible for silencing of the multicopy *Stellate* gene. Later, in 2006 these small RNAs, piRNAs (for PIWI-interacting RNAs), were found to be bound by PIWI-clade Argonaute proteins [6, 50, 55, 181, 93].

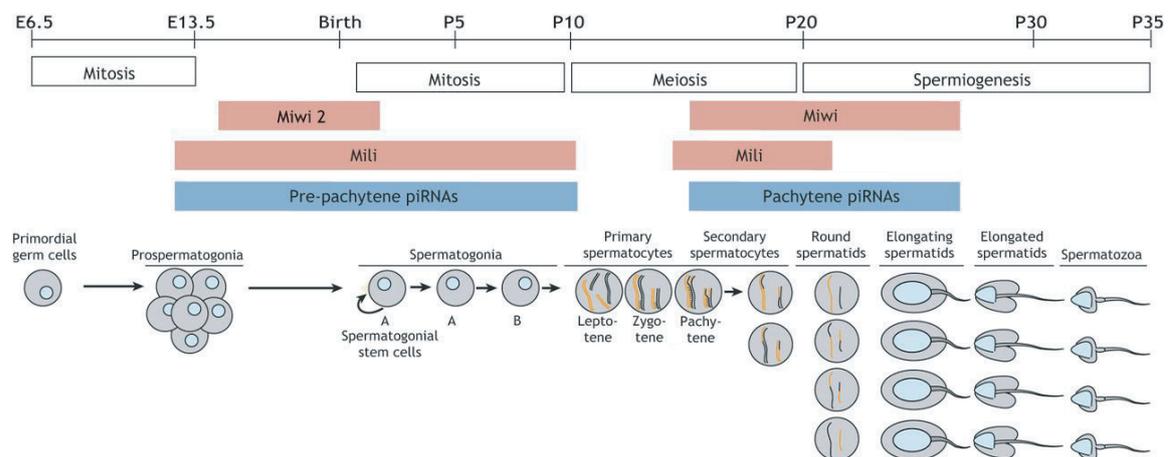
This is in contrast to other small silencing RNAs that associate with AGO-clade Argonautes. The knockdown of PIWI proteins in most organisms affects division and maintenance of germline and causes a sterility phenotype. The *piwi* (P-element induced wimpy testis) gene itself was identified during a P-element enhanced trap screen in *D. melanogaster*. It was named after a specific phenotype as mutant flies had a reduced number of egg chambers and sperm bundles [108, 25]. Since then, a constantly growing number of studies on piRNA biology have revealed a complex, diverse, and highly adaptive system employed by animal species, from sponges to humans. Its best-known functions are to safeguard the germline genome from transposable elements (TEs), to control gene expression, to ensure antiviral defense and to regulate sex determination in silkworms and genome rearrangement in ciliates [18, 69, 191, 162, 151, 148, 82, 40].

### 1.1.2 Structural details and expression characteristics of PIWI proteins

Argonaute proteins can be classified into two main clades, the Ago and the PIWI clade. The typical functional feature of both Ago- and PIWI-clade proteins is to bind small regulatory RNAs and to participate in small RNA-mediated gene silencing [119]. The structural organization of the proteins reflects this functional specification. Argonaute proteins consist of four domains: the N-terminal, the PAZ domain, the Mid domain and the PIWI domain [119, 36, 87, 117]. The crystal structure of a silkworm PIWI-clade Argonaute, Siwi, which was resolved by Matsumoto, Nishimasu and colleagues (2016), confirmed that the structure of PIWI proteins is very similar to its Ago-clade counterparts [117]. Siwi binds the 5'-monophosphate of a guide piRNA in a pocket between the Mid (from middle) and the PIWI domain in a  $Mg^{2+}$ -dependent manner. Moreover, its Mid domain is accommodated for the preferred binding of uridine as a first nucleotide. The 3'-end of the guide piRNA is anchored within the PAZ domain (named after Piwi-Ago-Zwille). The PIWI domain contains an RNase H fold with the catalytic tetrad DEDX (X is generally Asp or His), which ensures Siwi endonuclease activity [117].

Most animals express several PIWI proteins, which might reflect their participation in mechanistically distinct silencing modes, such as transcriptional and posttranscriptional gene silencing [9, 8, 172]. *D. melanogaster*, for example, expresses three PIWI proteins: Piwi, Aubergine (Aub) and Argonaute-3 (Ago3). Aub and Ago3 show a cytoplasmic localization and they are enriched in the perinuclear nuage, whereas Piwi is predominantly nuclear. All three PIWI proteins are expressed in the *Drosophila* germline. However, somatic ovarian

follicular cells (OSCs) possess a simplified piRNA pathway operated only by Piwi [62, 187]. In mouse (*M. musculus*) three PIWI are expressed as well, although in a temporally controlled manner (Fig. 1.1). Nuclear MIWI2 is expressed along with cytoplasmic MILI for a short time in the pre-pachytene stage during mouse spermatogenesis. Later, MILI is expressed together with MIWI, which is localized to the cytoplasm [148].

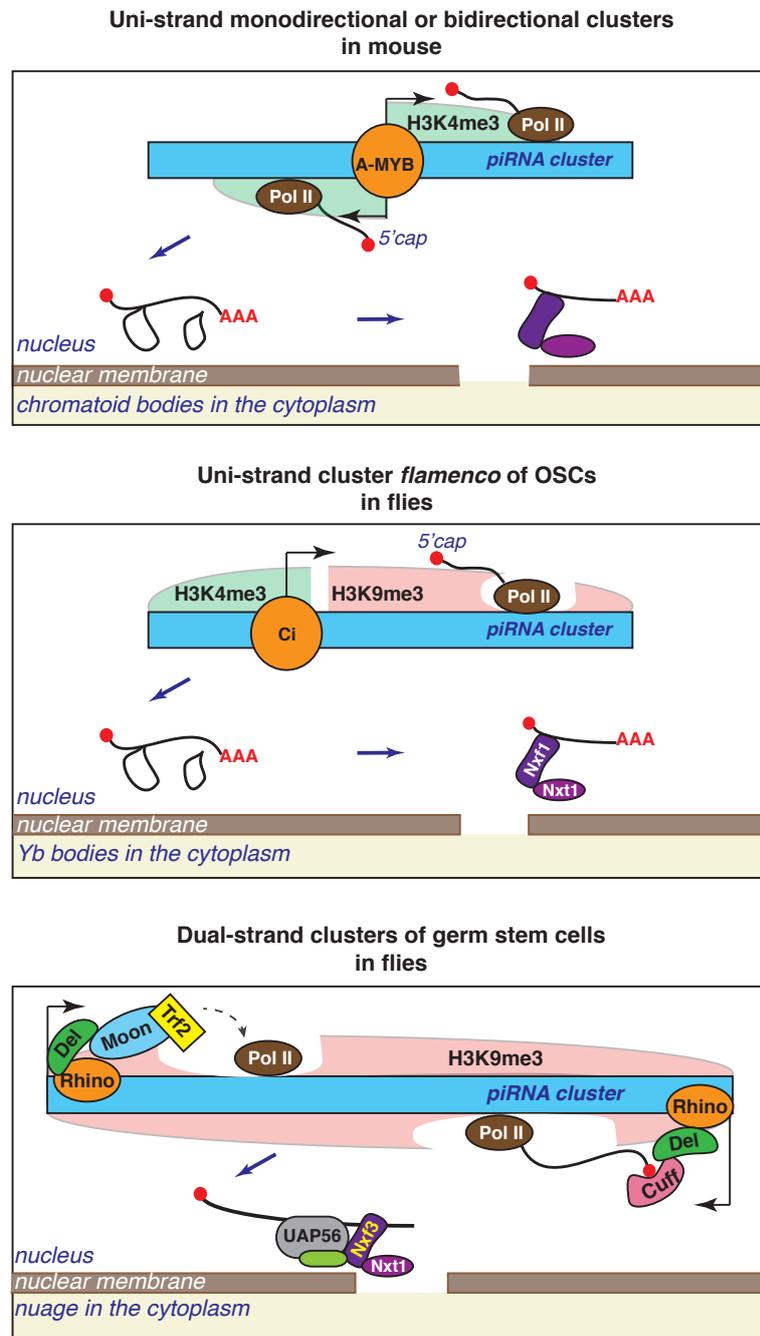


**Fig. 1.1** Temporally controlled expression of PIWI proteins during spermatogenesis in *Mus musculus*. The figure is adapted from Rojas-Ríos and Simonelig (2018) [148].

### 1.1.3 Genomic origin of piRNAs

piRNAs originate from specific genomic loci called piRNA clusters. piRNA clusters cover up to 200 kilobases (kb) and are transcribed by RNA polymerase II (Pol II) into long single-stranded precursors [95, 77]. Based on the mode of transcription and chromatin organization in the vicinity, clusters are divided into three known types: uni-strand clusters, bidirectional clusters and dual-strand clusters (Fig. 1.2) [199, 27, 132].

Uni-strand clusters are the most common type among animals and potentially the most ancient one [27, 132]. Their gene structure is similar to ordinary coding genes and includes exons, introns and regulatory sequences such as promoter and terminator. In mouse, uni-strand clusters are located in euchromatic regions of the genome. The promoter area and body of the clusters are enriched for trimethylation on lysine 4 of histone H3 (H3K4me3) [105], a histone mark that strongly correlates with active transcription [70, 71]. Mouse piRNA clusters that are expressed in the pre-pachytene stage of spermatogenesis harbor the remnants of TEs or are derived from 3'-UTRs of genes [9, 46]. piRNAs produced from these pre-pachytene clusters bound by MILI and MIWI2 (mouse PIWI proteins) ensure the silencing of TEs and *de novo* methylation of their loci in the genome [8].

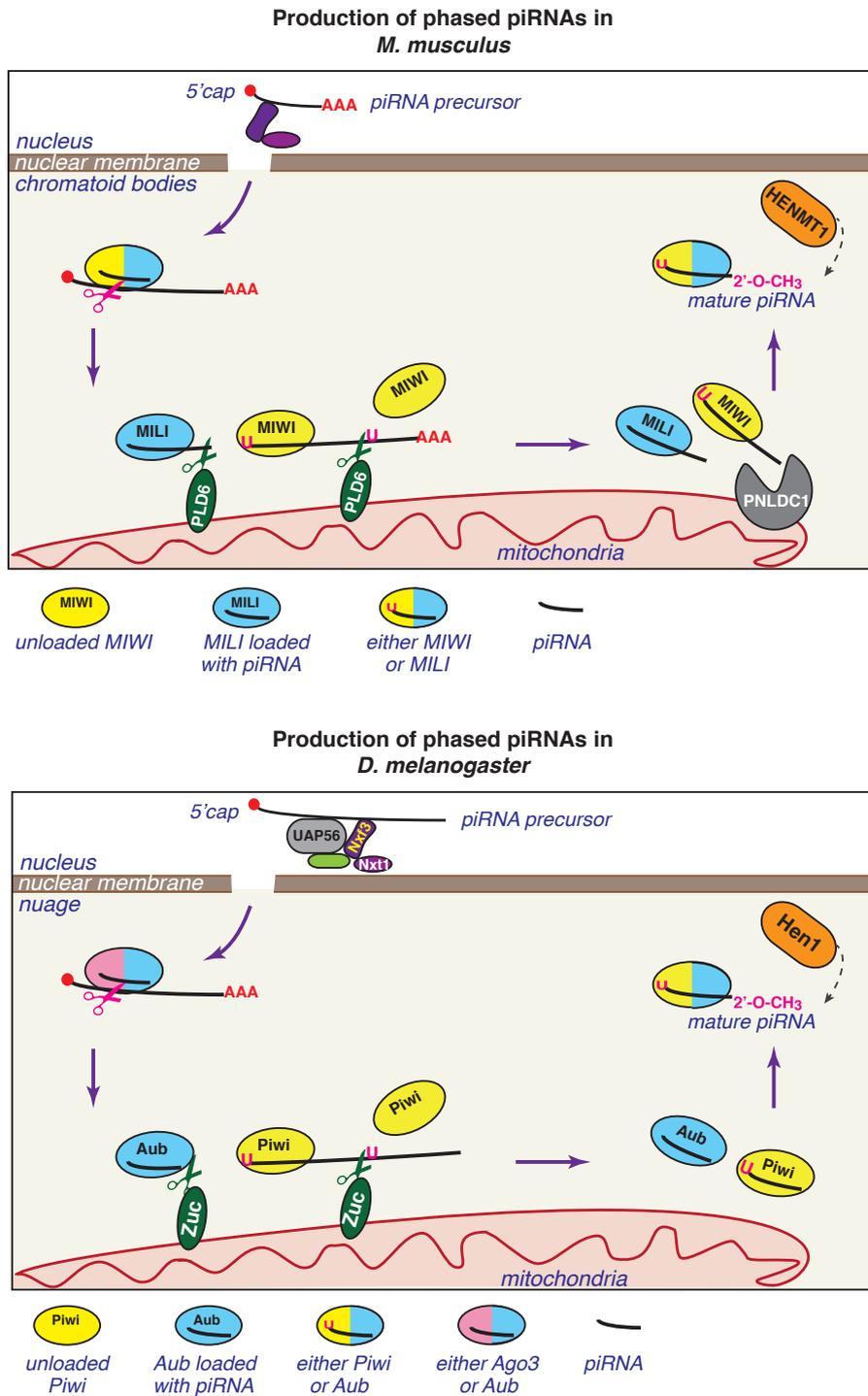


**Fig. 1.2** Simplified schematic representation of the three types of piRNA clusters. The transcription of mouse uni-strand clusters located in euchromatic region can be initiated from the promoter region in one direction or bidirectionally. Uni-strand cluster *flamenco* of *D. melanogaster* positioned at the border of eu- and heterochromatin. Dual-strand clusters of flies are transcribed from heterochromatin region. The Rhino-Del-Moon/Cuff complex marks transcription initiation sites of dual-strand cluster.

Pachytene piRNA precursors are produced from genic and intergenic regions at the pachytene stage of meiosis in mouse spermatocytes [53]. Their transcription is controlled by the transcription factor A-MYB. Moreover, the piRNA precursors transcribed from such clusters are capped, polyadenylated and spliced [105]. Pachytene clusters can be transcribed both uni- or bidirectionally, thus producing non-overlapping transcripts from the same promoter region (Fig. 1.2) [199]. Currently, it is not known how piRNA precursors are distinguished from other mRNAs, whose transcription is also regulated by A-MYB.

In flies, the uni-strand cluster *flamenco* is expressed in OSCs and responsible for the suppression of *gypsy-like* retrotransposons [18, 92]. The body of the *flamenco* gene is covered with histone 3 lysine 9 di- and trimethylation (H3K9me2/3) marks, which correlate with transcriptional repression and heterochromatin structure [71]. However, the promoter of *flamenco* is located in the euchromatic region of neighboring gene's 3'-UTR and is marked with H3K4me3. As in the case for mouse pachytene clusters, the transcription of *flamenco* is triggered by the transcription factor Cubitus interruptus (Ci) [52]. Produced piRNA precursors are capped, polyadenylated, spliced and exported from the nucleus by the nuclear export factor Nxf1 (Fig. 1.2) [52, 29, 180].

Dual-strand clusters were to date only found in arthropods. They are situated in heterochromatic regions and have no defined promoter [83, 142, 123, 5]. Therefore, transcription of dual-strand clusters is initiated by a non-canonical mechanism. Due to the absence of defined promoter region, transcription initiation occurs in the inexact place of the cluster, which produces precursors of diverse sizes and varied 5'-end starting points [27]. Transcription of dual-strand clusters relies on a protein complex that consists of Rhino, a variant of heterochromatin protein 1 (HP1), which recognizes and binds H3K9me3, the scaffolding protein Deadlock (Del) and Cutoff (Cuff), a protein with similarity to the Rai1 transcription termination factor [24, 123]. Rhino recruits Moonshiner (Moon), a paralog of the germline-specific transcription initiation factor IIA subunit 1. Moonshiner then forms a pre-initiation complex with TATA box-binding protein-related factor 2 (Trf2), which recruits other transcription factors and RNA polymerase II, thus allowing specific transcription in heterochromatin [5]. Premature transcription termination at canonical polyadenylation (poly(A)) sites, which are quite abundant in dual-strand clusters, is blocked by Cuff [207, 24]. This allows the read-through steady transcription and production of extended ( $\approx 250$  kb) piRNA precursors [123]. Splicing of piRNA precursors is also blocked by Del and Cuff [207, 24]. Transcribed non-spliced piRNA precursors from dual-strand clusters are exported from the nucleus using a specialized nuclear export factor, Nxf3. Nxf3 ensures the delivery of piRNA precursors to their cytoplasmic processing sites (Fig. 1.2) [38, 84].



**Fig. 1.3** Phased piRNAs in mice and flies are generated in a similar way. In contrast to phased piRNAs in flies, mouse pre-piRNAs are trimmed by exonuclease PNLDC1 before their 3'-end methylation.

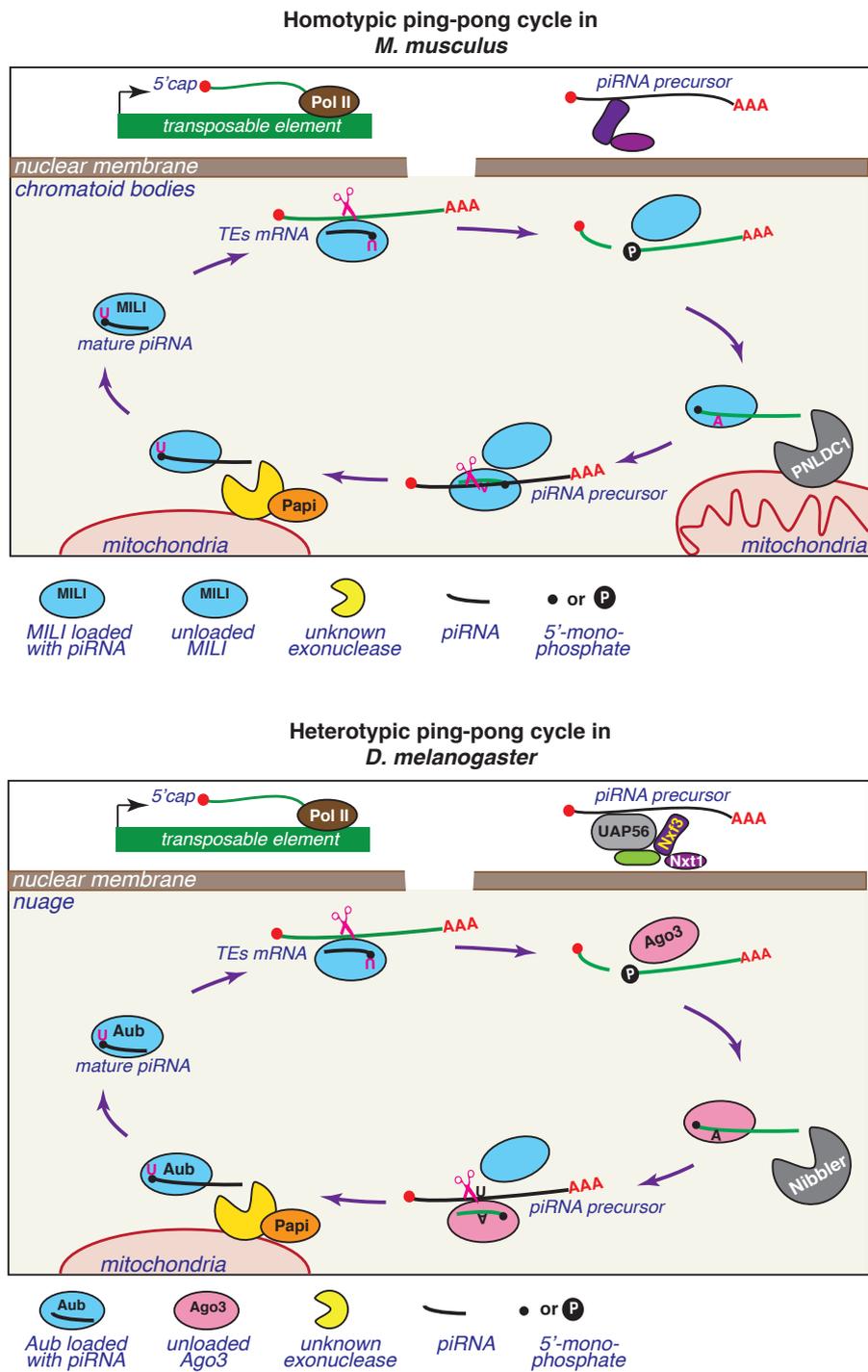
### 1.1.4 Biogenesis of phased piRNAs

The biogenesis of piRNAs is conventionally divided into a primary and a secondary pathway. Primary biogenesis entails for processing of precursors that were transcribed from piRNA clusters. This mechanism ensures the diversity and complexity of piRNA population, which is encoded in the genome. The secondary piRNA biogenesis pathway, also known as ping-pong cycle, is responsible for silencing of piRNA targeted transcripts. The targeted transcript thereby serve as templates for an additional population of piRNAs. The two pathways are interconnected and depend on each other.

Processing of piRNA precursors into mature piRNAs starts with the endonucleotic cleavage of the precursor initiated by a member of the PIWI protein family [122, 60, 68] (Fig. 1.3). PIWI is directed to its target by a co-bound piRNA. It cleaves the recognized piRNA precursor in a ping-pong manner opposite nucleotides 10 and 11 of the piRNA guide. This initial cleavage event removes the 7-methylguanosine cap that is typical for piRNA precursors in all animals, and, instead, generates a 5'-monophosphate at the newly defined precursor 5'-end [45, 132]. Now such transcript is bound by another PIWI family member and processed in a phased manner with the help of the endonuclease Zucchini (mitoPLD/PLD6 for mouse) [190, 78, 45]. In flies, a single cleavage event by Zucchini defines both the 5'-end and the 3'-end of phased piRNAs with a distance of 0 nucleotides [122, 60]. The mouse endonuclease PLD6 generates only 5'-ends of phased piRNAs. Their 3'-ends are trimmed by the exonuclease PNLDC1 [74]. The definition of piRNA 3'-ends is concluded by the 2'-O-methylation of the 3'-end by a S-adenosylmethionine (SAM)-dependent methyltransferase (Hen1 in flies, HENMT1 in mice) [132] (Fig. 1.3). In addition to the Zucchini or PLD6 and PNLDC1, multiple other proteins are involved in the processing and loading of phased piRNAs into the designated PIWI proteins [27, 132].

### 1.1.5 The ping-pong cycle

To initiate the first endonucleotic cleavage event in a piRNA precursor, a PIWI-bound guide piRNA needs to exhibit base-pair complementarity to the piRNA precursor for at least the 5'-terminal 10 - 11 nucleotides [122, 60, 169, 204]. However, the minimal base-pairing requirements between piRNA and its target that trigger the nucleolytic activity of the PIWI protein have not yet been determined [132]. The guide piRNA can be maternally deposited into the cell, derived from a piRNA precursor or obtained from transposon sequences during the ping-pong cycle, also known as the secondary piRNA biogenesis pathway [19, 45].



**Fig. 1.4** Simplified illustration of homotypic and heterotypic ping-pong cycle in mice and flies, respectively.

The ping-pong cycle in flies germ stem cells operates by the use of two PIWI proteins, Aub and Ago3, thus termed heterotypic cycle [18, 57]. The guide piRNA, or initiator piRNA is bound to Aub and identifies a targeted transcript, often the mRNA of a transposable element. Next, Aub cleaves the target in case the extend of base-pairing allows for that. The cleavage event occurs between the 10th and 11th nucleotides as counted from the 5'-end of the initiator piRNA and leaves a monophosphate at the 5'-end of the cleaved target. The cleaved target RNA is then loaded into another PIWI protein, Ago3. It is trimmed to its characteristic length by Nibbler or another unknown exonuclease that uses Papi as a co-factor, thereby generating a responder piRNA [64]. As the first 10 nt of the initiator piRNA are complementary to the first 10 nucleotide of the responder piRNA, this interdependence creates so-called ping-pong signature that can be used to identify ping-pong relationships between different PIWI proteins. Finally, the responder piRNA starts another "ping-pong" cycle by guiding Ago3 to transcripts with antisense orientation to the cleaved TEs [18, 57, 132]. In mouse spermatocytes only MILI constitutes a homotypic ping-pong cycle that operates at the pre-pachytene stage [28, 188] (Fig. 1.4). In addition to the PIWI proteins that are direct players of ping-pong cycle, many other proteins coordinate the target cleavage and loading of responder piRNAs onto their designated PIWI proteins [106, 173, 208, 138, 198].

The ping-pong cycle and piRNA phasing both take place in a specialized membrane-less cellular compartment [27]. In flies germinal stem cells, these compartment is called the nuage [106, 18]. In contrast, piRNA precursors in OSCs are processed in Yb-bodies [139]. The nuage of mouse germline is better known as chromatoid bodies [177]. Apparently, the existence of these spatially confined piRNA processing sites not only increases the local concentration of proteins participating in the pathway, but also prevents the unwanted degradation of other cellular transcripts [132].

## 1.2 Biological function of piRNAs

Silencing of TEs in the animal germline is considered the ancestral function of piRNAs [132]. The suppression of TEs is thereby ensured both at the level of transcription by directing heterochromatin formation [94, 172, 39, 126, 14], and by the post-transcriptional degradation of TEs in the piRNA-mediated ping-pong cycle [18, 57].

In addition to their predominant function in transposon control, piRNAs also regulate the expression of coding genes. In fact, the piRNA-mediated silencing of the *Stellate* gene in *D. melanogaster* testes by *Su(Ste)* led to the discovery of piRNAs [7]. piRNAs also direct the degradation of selected maternal mRNAs in early embryos of *D. melanogaster* [151, 13,

148]. Moreover, pachytene piRNAs were found to eliminate mRNAs in murine elongating spermatids by their direct cleavage or by guiding the recruitment of the deadenylase complex [53, 26]. Interestingly, in female silkworms, *Bombyx mori*, piRNAs are responsible for the degradation of a single mRNA that defines the male-specific phenotype using the ping-pong cycle [82].

The degradation of TEs and of protein-coding mRNAs at distinct developmental stages is considered as a reaction of piRNAs to "non-self" sequences. In a similar fashion, the ciliate *Tetrahymena* exploits piRNAs to mark entire regions of its genome that will then be eliminated during a genome rearrangement [121]. Ciliates have a germline micronucleus for reproduction and a large somatic macronucleus that harbors only genetic information required for protein coding and cell regulation. Thus, formation of the macronucleus requires elimination of repetitive content and intergenic sequences found in the micronucleus [73]. In contrast to *Tetrahymena*, piRNAs of another ciliate species, *Oxytricha trifallax*, map to somatic parts in the macronucleus, thus defining "self" sequences that are kept during genome rearrangement [40].

The nematode *Caenorhabditis elegans* combines the best of the two worlds of ciliates. Non-self transcripts are recognized by piRNAs associated with the PIWI protein PRG-1. This leads to the production of secondary siRNAs bound by WAGOs, a third clade of the Argonaute superfamily of proteins [119]. WAGOs subsequently silence non-self targets both by their direct cleavage and by the establishment of heterochromatin. In contrast, another PIWI protein, CSR-1, with its co-bound piRNAs recognizes and protects mRNAs that need to be retained [171, 96, 11, 166, 193].

Last, the expression of piRNAs is not restricted to the germline. piRNAs were found in somatic cells, such as nerve cells, where they, for example, play a role in long-term memory formation in the sea slug *Aplysia* [141], or during inhibition of axonal regeneration in *C. elegans* [81]. In addition, piRNAs and PIWI proteins were found to be essential for the regeneration capacity of several species, including hydra, colonial ascidians, and planarian flatworms [145, 133, 150, 183, 107].

## 1.3 The piRNA pathway in *S. mediterranea*

### 1.3.1 Planarians as a model organism

Freshwater planarians of the species *Schmidtea mediterranea* are well known for their extraordinary ability to regenerate. This ability is enabled by a large population of adult

pluripotent stem cells, called neoblasts [37]. Neoblasts are capable of giving rise to all planarian cell types [201]. Moreover, they preserve their potency over the entire lifespan of the animal [160]. Therefore, planarians embody an excellent model to study regeneration, aging and stem cell-based diseases.

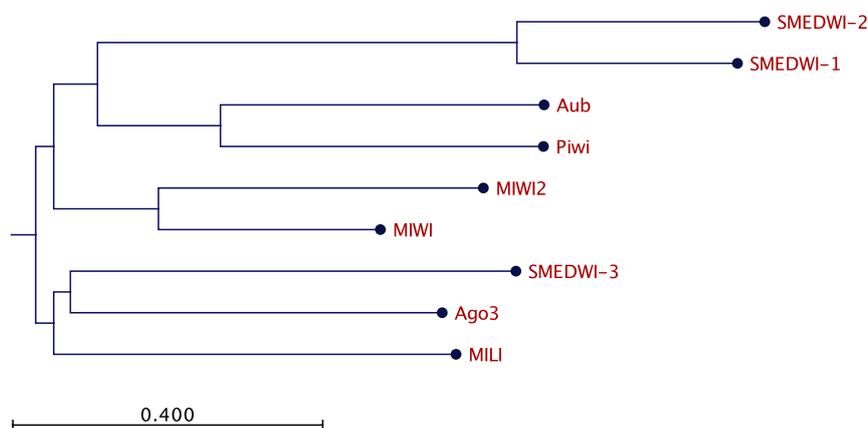
To address a specific research question, a model organism not only needs to possess the studied feature but it also needs to be accessible for genetic manipulation. In planarians, RNA interference followed by transcriptome-wide gene expression analysis by RNA-seq is a popular and robust technique for the investigation of gene function. Data from these systems biology techniques are often combined with in situ hybridization experiments. As an increasing number of research groups focus their interest on planarians, more organism-specific antibodies are being generated by the planarian community. This makes it possible to apply fluorescent immunostaining and different immunoprecipitation experiments including CHIP-Seq and CLIP-Seq [34, 80]. Moreover, to focus on a particular tissue or cell population, simple fluorescence-activated cell sorting (FACS) protocols can be applied to sort nerve cells, to separate neoblasts from differentiated cells or even to isolate sub-populations of totipotent neoblasts [145, 66, 201, 65]. Single-cell sequencing is also widely used in planarians to dissect the molecular basis of tissue regeneration and development [41]. Although genetic engineering is currently not possible in planarian flatworms, a number of research groups are actively working to establish robust planarian transgenesis protocols [72, 98].

In addition to the aforementioned advantages, the planarian flatworm *Schmidtea mediterranea* is a stable diploid ( $2n = 8$ ), unlike many planarian species that exist as mixoploids or polyploids [127]. *S. mediterranea* possesses a haploid genome with a size of  $\sim 1 - 2$  gigabase pairs (Gb) [56] and a very high AT-content ( $\sim 70\%$ ). Another interesting feature of the planarian genome is its high degree of repetitiveness. 61.7% of the planarian genome is repetitive, while this value does not exceed 38% or 46% in mouse and human, respectively. Interestingly, during genome re-sequencing and reassembly, giant retrotransposons named Burro elements were discovered. Burro elements span more than 30 kb genome regions compared to 5 - 10 kb for long terminal repeats in vertebrates [56]. This poses the fascinating question of how and when these elements have evolved in planarians. Thus, hand in hand with the development of new tools for gene manipulation, a new genome assembly and improved genome annotation [156, 157], planarians are becoming an attractive model to study multiple aspects of regeneration and stem cell biology.

### 1.3.2 Planarian PIWI proteins

In *S. mediterranea* three PIWI proteins were identified, termed SMEDWI-1, -2 and -3 [145, 133]. The *smedwi-1* gene, encoding the SMEDWI-1 protein, is expressed exclusively in neoblasts and is used as a marker of stem cells in numerous planarian studies [145]. In contrast, *smedwi-2* and *smedwi-3* expression is not limited to neoblasts. The *smedwi-2* transcript, for example, was also detected in the central nervous system and in the epidermis [34, 175], while *smedwi-3* appears to be present in the pharynx, nervous system and anterior of the photoreceptors, albeit at low levels [133]. Moreover, six additional PIWI-clade genes were identified in the planarian genome. These PIWI variants are highly similar to *smedwi-2* at the nucleotide level, making it difficult to determine their actual expression values [133].

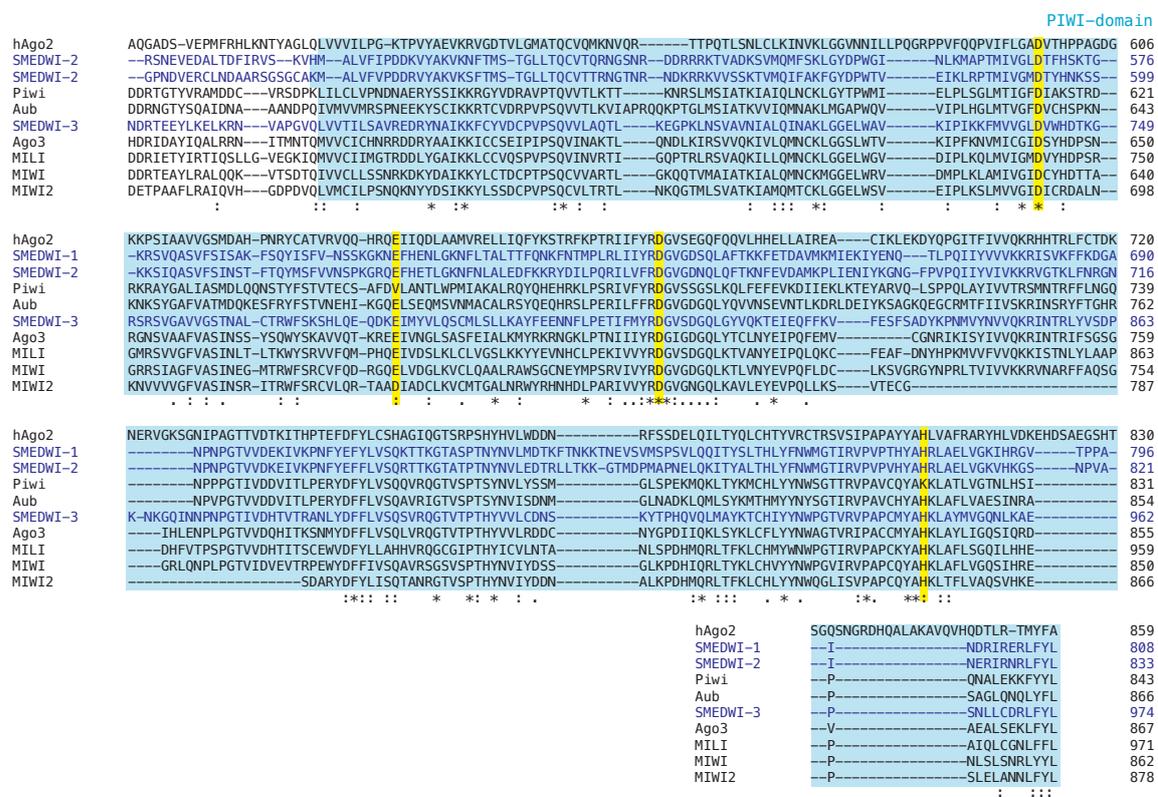
The localization of SMEDWI proteins mainly follows their gene expression patterns. SMEDWI-1 is enriched in the cytoplasm of neoblasts [59]. SMEDWI-2 has a nuclear localization in both neoblasts and differentiated cells (shown in this thesis) [202, 170, 80]. Moreover, a PIWI protein in a related planarian species *Dugesia japonica*, DjPiwiB, is present in the nuclei of almost all differentiated cells. Moreover, at least part of this DjPiwiB pool was inherited by somatic cells during their differentiation [170]. The cytoplasmic localization of SMEDWI-3 in neoblasts and head cephalic ganglions is demonstrated in this thesis (Fig. 2.2).



**Fig. 1.5** Phylogenetic tree illustrating the similarity of amino acid sequences of PIWI proteins for *D. melanogaster* (Piwi, Aub, Ago3), *M. Musculus* (MILI, MIWI, MIWI2) and *S. mediterranea* (SMEDWI-1, -2, -3). The phylogenetic distance was calculated using CLC Main Workbench 7 (<https://www.qiagenbioinformatics.com>).

At amino acid sequence level, SMEDWI-1 and -2 have higher similarity to each other (53% identity, 71% similarity) rather than to PIWI proteins of other species. However,

SMEDWI-3 is similar to Ago3 of fruit fly (34% identity, 52% similarity) and to mouse MILI (33% identity, 50% similarity) (Fig. 1.5) [133, 112]. The domain architecture of SMEDWI proteins is analogous to all Argonaute family members. They contain N-terminal, PAZ-, MID- and PIWI-domain. In addition, all three planarian PIWIs are active endonucleases that harbor catalytic DEDH tetrad in their PIWI-domain (Fig. 1.6). Furthermore, all SMEDWIs possess RG-motifs that are recognized by the arginine methyltransferase PRMT5. PRMT5 places the symmetrical dimethylarginine modification of these RG motifs, which enables the interaction of SMEDWIs and Tudor domain-containing proteins. SMEDWI-3 has the highest density of such RG-motifs among planarian PIWIs [153].



**Fig. 1.6** Sequence alignment of PIWI domains of PIWI proteins from *D. melanogaster* (Piwi, Aub, Ago3), *M. musculus* (MILI, MIWI, MIWI2) and *S. mediterranea* (SMEDWI-1, -2, -3 are in blue), and human Argonaute Ago2 (hAgo2). DEDH catalytic tetrad marked in yellow, PIWI-domain is in cyan. The alignment was performed using multiple sequence alignment program Clustal Omega [112].

SMEDWI-2 and SMEDWI-3 are essential for planarian regeneration and tissue homeostasis [145, 133]. Their knockdown abolishes the regenerative capacity of planarians and results in a lethal phenotype. The *smedwi-2(RNAi)* phenotype resembles that of irradiated animals, which are deprived of all neoblasts. They display head regression at day 9 post RNAi feeding (dpf) and body curling around their ventral surface at 11 dpf [145]. By comparison,

lethal irradiation at 6000 rads eliminates all neoblasts 24 hours post exposure, causing tissue regression at day 9 and curling at day 15 [144]. In contrast to irradiated animals, neoblasts are present and even dividing upon *smedwi-2(RNAi)*. However, they are incapable of giving rise to differentiated cells [145]. *Smedwi-3(RNAi)* worms also experience regeneration defects. However, head regression and curling, typical signs of *smedwi-2(RNAi)* worms, appear later and lead to lethality by days 22-24 post feeding. In addition, all SMEDWI-3 knockdown animals experience skin lesions starting at 12 dpf [133]. Thus, the appearance of lesions may indicate the involvement of SMEDWI-3 in the development of cells in the epidermal lineage or the importance of SMEDWI-3 in cell differentiation. A knockdown of SMEDWI-1 did not cause any apparent phenotypic defects.

Thus, planarian regeneration and tissue homeostasis depend on a functional piRNA pathway. The knockdown of planarian SMEDWI-2 and -3 leads to impair neoblasts development, in line with findings in fruit flies and mice, where mutations in piRNA pathway are detrimental for germline development. The present study therefore aims to investigate and expand our knowledge about the architecture of planarian piRNA pathway, its function and the role of its key components in neoblasts biology.

# Chapter 2

## Results part 1

### 2.1 Planarians recruit piRNAs for mRNA turnover in adult stem cells

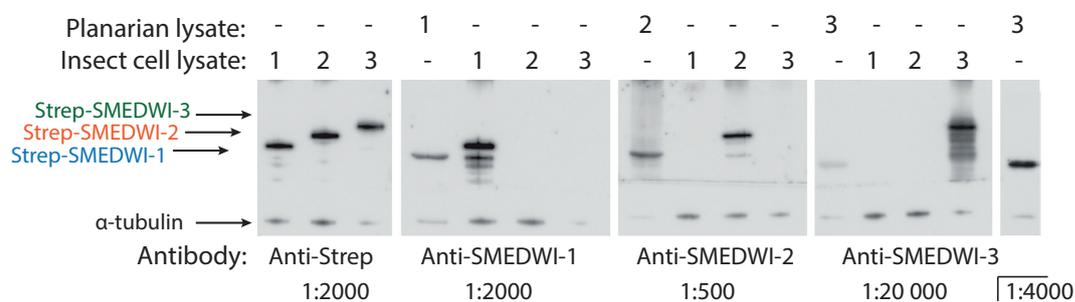
#### Aim and Objective

The aim of this study is to characterize the piRNA pathway in *S. mediterranea* and decipher the role of PIWI proteins in stem cell differentiation and development.

To unravel the principles of piRNA biogenesis in planarians, immunoprecipitation protocols were established for all planarian PIWI proteins. Deep sequencing of co-bound piRNAs allowed us to describe the major targets of the planarian piRNA pathway and the interaction of SMEDWI proteins in the ping-pong cycle. The exact cellular localization of SMEDWI proteins was determined using slide-section immunofluorescence. The combination of small RNA-Seq and immunofluorescence enabled us to reconstruct the main organizational principles of the planarian piRNA pathway. To further investigate the role of the piRNA pathway, we applied RNA sequencing techniques and successfully adapted HITS-CLIP, CLASH and Degradome-seq to planarian flatworms as well. The methods mentioned before were implemented in planarian flatworms for the first time. In combination with RNA-seq and RNAi knockdown experiments, newly adapted techniques uncovered the involvement of SMEDWI-3 in mRNA surveillance in planarians.

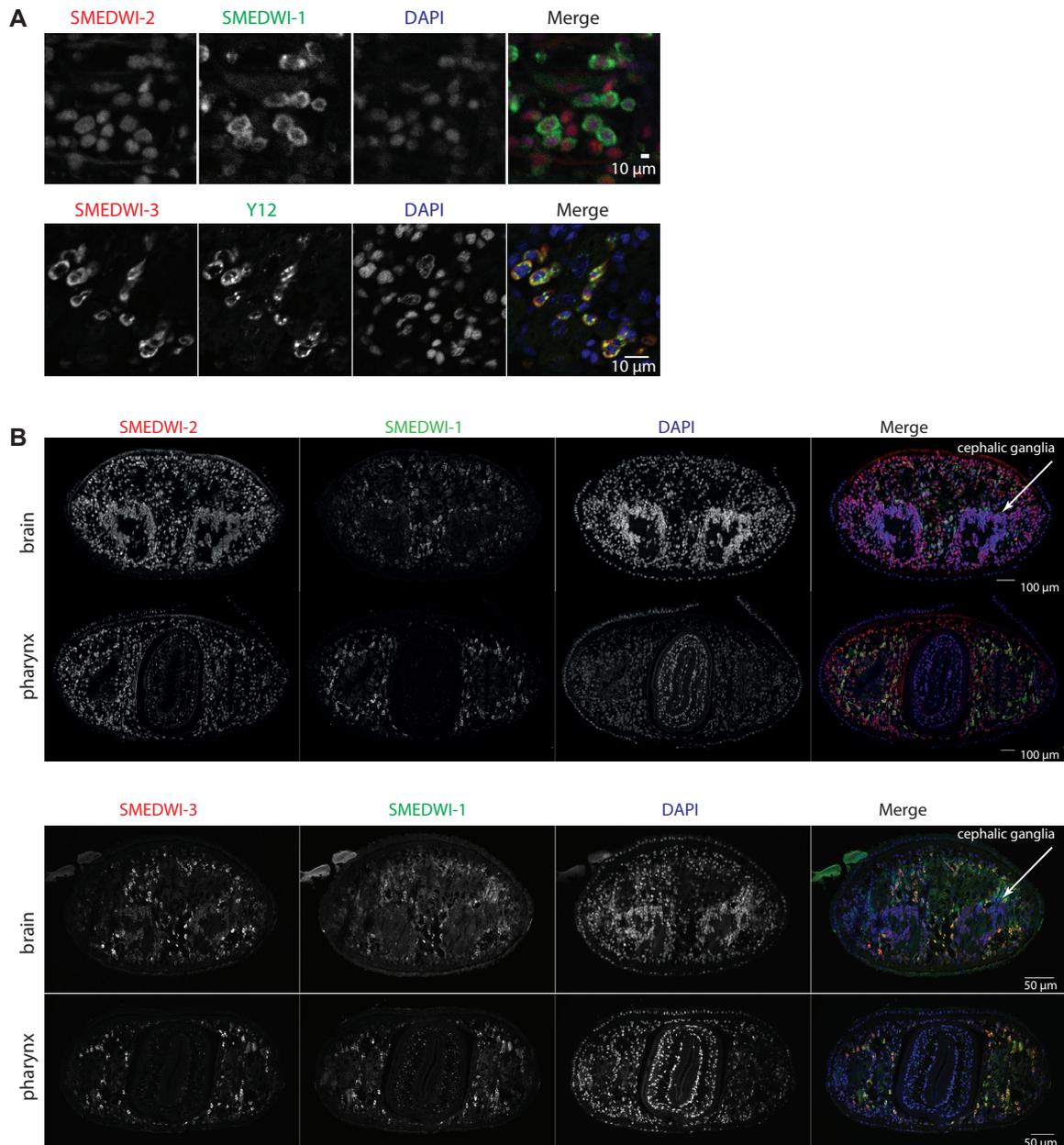
### 2.1.1 Immunoprecipitation of piRNAs associated with SMEDWI proteins

In planarian flatworms, three PIWI proteins are being expressed: SMEDWI-1, SMEDWI-2 and SMEDWI-3. As a prerequisite for their characterization and for the investigation of co-bound piRNAs, we raised polyclonal antibodies against SMEDWI-2 and SMEDWI-3. The anti-SMEDWI-1 antibody used here was previously described [59]. The specificity of the antibodies for each protein was confirmed using full-length Strep-SUMO-tagged SMEDWI-1, SMEDWI-2 and SMEDWI-3 expressed in insect cell (Fig. 2.1).



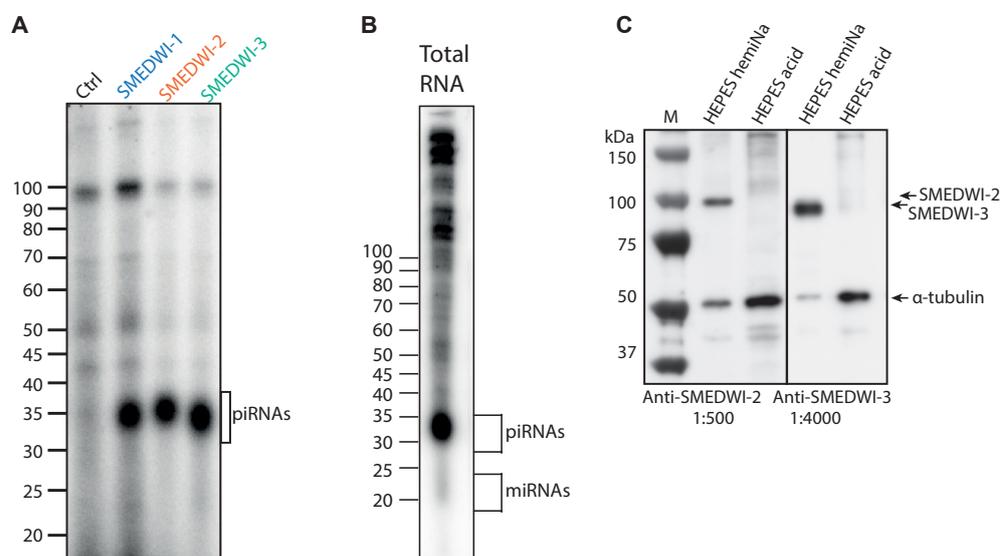
**Fig. 2.1** Western blot analysis of SMEDWI-1, -2 and -3 antibodies confirms their specificity. Test samples were prepared from insect cells lysates expressing Strep-SUMO-tagged SMEDWI proteins or directly from planarian lysate.  $\alpha$ -Tubulin was used as a loading control.

To determine the cellular localization of SMEDWI proteins, we performed slide-section immunofluorescence. Whole-mount immunofluorescence confirmed that SMEDWI-2 is enriched in the nuclei of neoblasts and differentiated cells [202] (Fig. 2.2A). SMEDWI-2 is present in neoblasts, as its signal overlaps with SMEDWI-1, a marker of neoblasts (Fig. 2.2A). In addition, SMEDWI-2 is detectable in the nuclei of differentiated cells, and especially in cephalic ganglia and pharynx (Fig. 2.2B). SMEDWI-1 and SMEDWI-3 are predominantly cytoplasmic (Fig. 2.2A). SMEDWI-1 is enriched in neoblasts and often used as a neoblasts marker [59]. The localization of SMEDWI-3 mainly matches that of SMEDWI-1. However, SMEDWI-3 is also expressed at low levels in neuronal cells of cephalic ganglia along with SMEDWI-2 (Fig. 2.2B), suggesting the presence of a neuronal piRNA pathway. In addition, SMEDWI-3 signal overlapped with Y12 antibody staining (Fig. 2.2A), a marker of planarian chromatoid bodies, as shown previously [153]. The Y12 antibody recognizes symmetric dimethylarginine (sDMA), a characteristic modification of the protein components of the chromatoid bodies.



**Fig. 2.2** (A) Subcellular localization analysis of SMEDWI-1, -2, and -3 using immunofluorescence. SMEDWI-1 (in green) is cytoplasmic, while SMEDWI-2 (in red) shows a nuclear localization. SMEDWI-3 signal (in red) is enriched in the cytoplasm and overlaps with chromatoid bodies stained by anti-Y12 antibodies (in green). Nuclei stained with DAPI are in blue. (B) Co-immunostaining of SMEDWI-1 and SMEDWI-2 and of SMEDWI-1 and SMEDWI-3 on cross-sections through the planarian brain and the pharynx. SMEDWI-1 is in green, SMEDWI-2 and SMEDWI-3 are in red. Nuclear DAPI staining is shown in blue.

To further investigate the small RNA population bound to SMEDWI proteins, immunoprecipitation protocols were developed for all three proteins (Fig. 2.3A). In comparison to miRNAs piRNAs are abundant in planarians (Fig. 2.3B), especially due to the fact that SMEDWI proteins are highly expressed in neoblasts, which constitute around  $\approx 20\%$  of all cells in adult worms [75]. During the development of specific immunoprecipitation protocols, the low osmolality of planarian tissues rendered it difficult to adjust the ionic strength of the lysis buffer, especially for highly concentrated whole-worm lysate solution ( $>3$  mg/ml). Planarian osmolality is estimated at 125 - 128 mosmol/l [165] in contrast to  $\approx 285$  mosmol/l for mammalian body fluids [192].

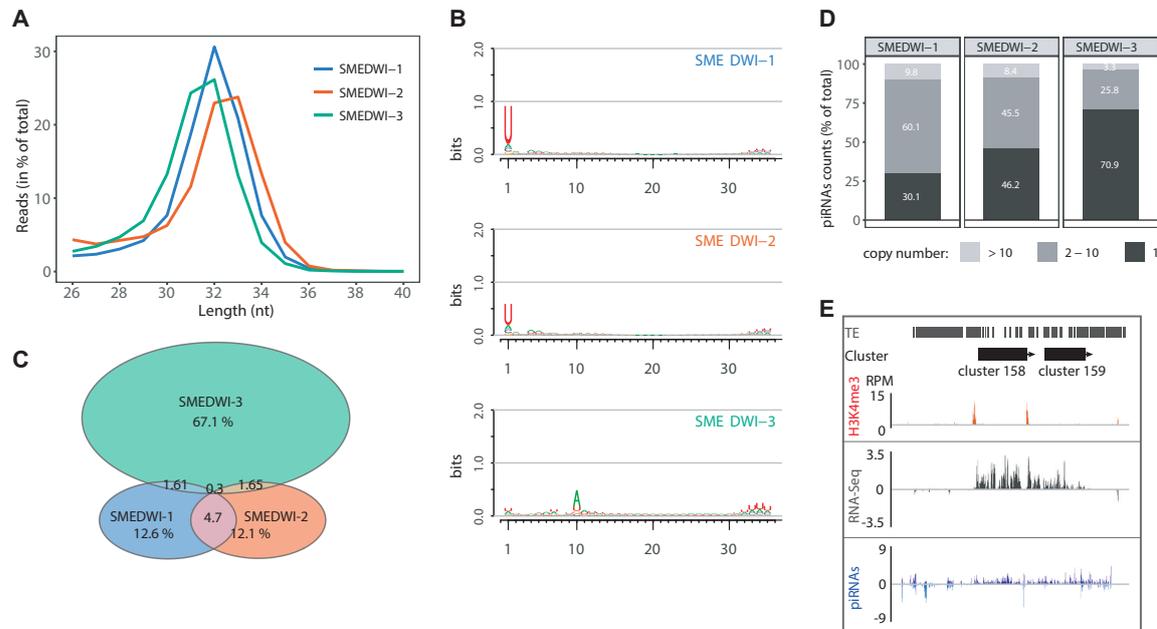


**Fig. 2.3** (A) piRNAs were immunoprecipitated from total worm lysate using anti-SMEDWI-1, anti-SMEDWI-2 and anti-SMEDWI-3 antibodies, labeled with  $[\gamma\text{-}^{32}\text{P}]\text{-ATP}$  and separated on a 10% Urea PAGE. Pre-immune serum served as negative control. (B) Radioactive labeling of total RNA isolated from the whole animal. piRNAs are predominantly 30-35 nt long, while miRNAs are less abundant and have a length of  $\sim 22$  nt. (C) Western blot analysis of SMEDWI-2 and SMEDWI-3 in the presence of HEPES hemisodium salt or NaOH-buffered HEPES acid confirms the sensitivity of SMEDWI proteins to ionic strength.  $\alpha$ -Tubulin was used as a loading control.

As a result, only HEPES in its hemisodium form was able to sufficiently stabilize planarian PIWI proteins in the worm lysate (Fig. 2.3C). Thus, using optimized conditions of the lysis buffer individual PIWI proteins with co-bound piRNAs were successfully immunoprecipitated, allowing for the construction and deep sequencing of small RNA libraries.

Subsequent analysis of piRNA sequences revealed that SMEDWI-1 binds piRNAs with a distribution maximum of 32 nucleotides (nt) (Fig. 2.4A). The length distribution of SMEDWI-2 and SMEDWI-3 associated piRNAs lies within a range of 32 to 33 and 31 to 32

nt, respectively. Moreover, sequence conservation analyses revealed that SMEDWI-1 and SMEDWI-2 bind piRNAs with a strong preference for a 5'-terminal uridine (78% and 67%, respectively) (Fig. 2.4B). In contrast, SMEDWI-3 binds piRNAs that show a +10 A bias (63%) and a slight conservation of a 5'-terminal uridine (41%).



**Fig. 2.4** (A) Size distribution of piRNAs bound to different planarian PIWI proteins. (B) Nucleotide composition of immunoprecipitated piRNAs represented as a position weight matrix with seqLogo. (C) Percentage of unique sequences bound by different SMEDWI proteins. (D) Percentage of unique sequences bound by different SMEDWI proteins. Only genome-mapped piRNAs that appeared in at least two replicates were counted. (E) Coverage profile of H3K4me3 CHIP-seq, RNA-Seq and total piRNAs on unidirectional piRNA clusters 158 and 159.

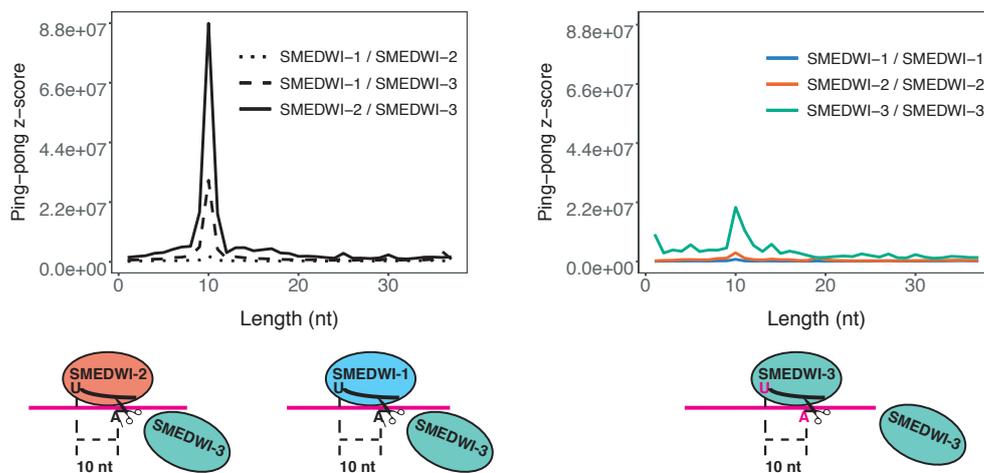
To assess the complexity of planarian piRNAs and to visualize potential sequence overlap between different PIWI proteins, we analyzed unique piRNAs bound by all three PIWI proteins that appeared in at least two replicates (Fig. 2.4C). Intriguingly, while SMEDWI-1 and SMEDWI-2 showed significant overlap in uniquely mapped piRNAs, SMEDWI-3 binds a distinct and very diverse pool of piRNAs. Moreover, 71% of SMEDWI-3-bound piRNAs were detected only once, indicating that this piRNA pool is even more complex than we could resolve with applied sequencing depth (Fig. 2.4D). Altogether, obtained results suggest that SMEDWI-3 has a distinct role in planarians that sets it apart from the two other PIWI proteins.

To examine how planarians are able to produce a highly complex and highly abundant piRNA repertoire, we determined the genomic location of planarian piRNA clusters using

proTRAC [149]. We were able to assign 270 piRNA clusters and found 268 of them to be unidirectional (Supplemental Table2 [80]). Moreover, 14% of planarian piRNA clusters carry histone H3K4me3 marks, suggesting their euchromatic localization (Fig. 2.4E). This confirms an earlier report [44] and resembles murine pachytene clusters and piRNA clusters in ovarian follicle cells in *D. melanogaster*. In both cases piRNA clusters show a strong strand bias [6, 26].

### 2.1.2 SMEDWI-2 and SMEDWI-3 cooperate to degrade planarian transposons

The recognition of an active transposon by antisense piRNAs can initiate the amplification of transposon silencing in the ping-pong cycle [18, 57]. When analyzing planarian piRNAs for the presence of such signatures, SMEDWI-2 and SMEDWI-3 were found to be the main ping-pong players in planarians, accounting for the majority of ping-pong pairs detected in our piRNA libraries (Fig. 2.5).

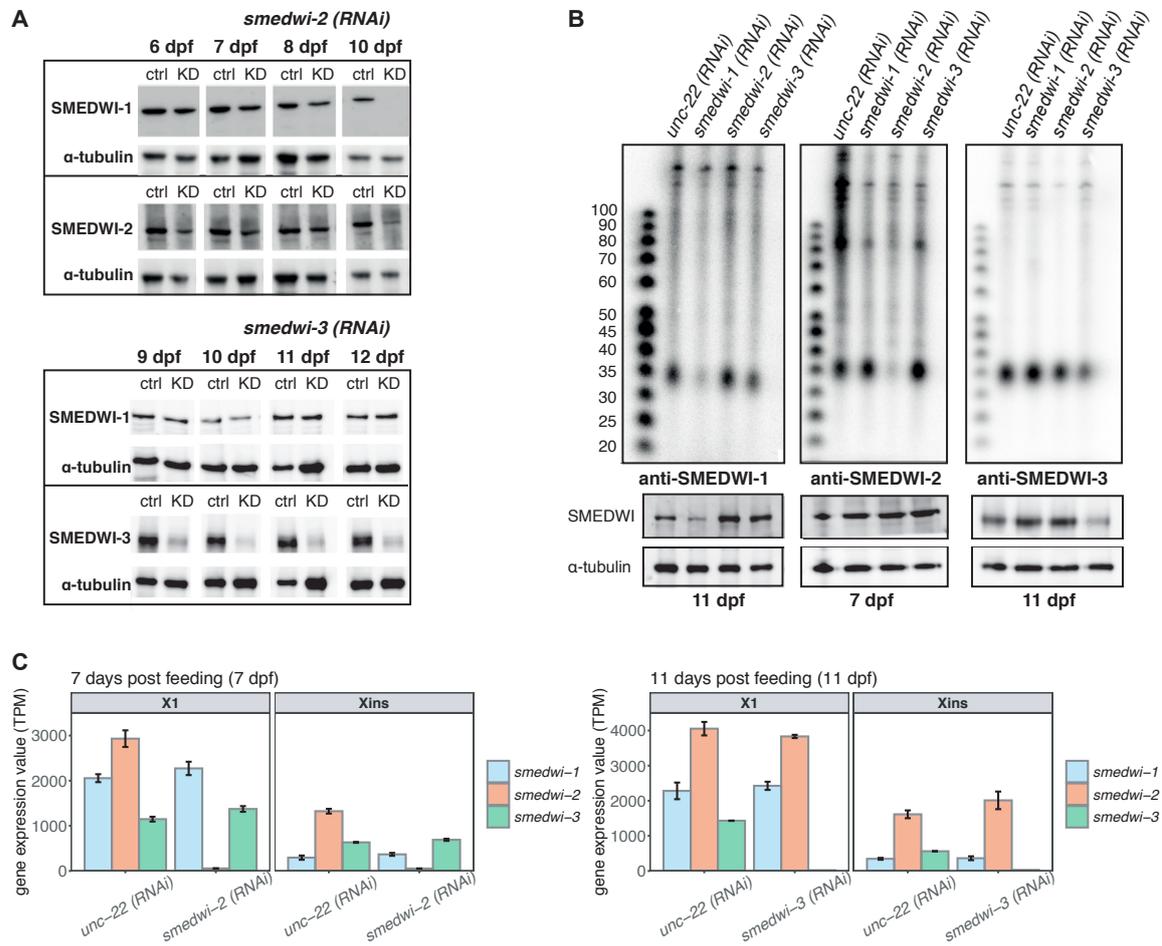


**Fig. 2.5** Local Z-score for the occurrence of homotypic and heterotypic ping-pong pairs.

We also observed a less pronounced ping-pong signature for SMEDWI-1 and SMEDWI-3, indicating that those two PIWI proteins also work together in degrading transposable elements. Additionally, SMEDWI-3-bound piRNAs show a homotypic ping-pong signature, whereas piRNAs bound to the other PIWI proteins do not (Fig. 2.5).

Since the degradation of transposable elements is the primary function of PIWI proteins in other organisms, we set out to study transposon levels in planarians upon loss of SMEDWI-2 and SMEDWI-3. As both PIWI proteins are essential for planarian homeostasis, we devised

an RNA interference (RNAi) knockdown strategy for SMEDWI-2 and SMEDWI-3 without depleting neoblasts.



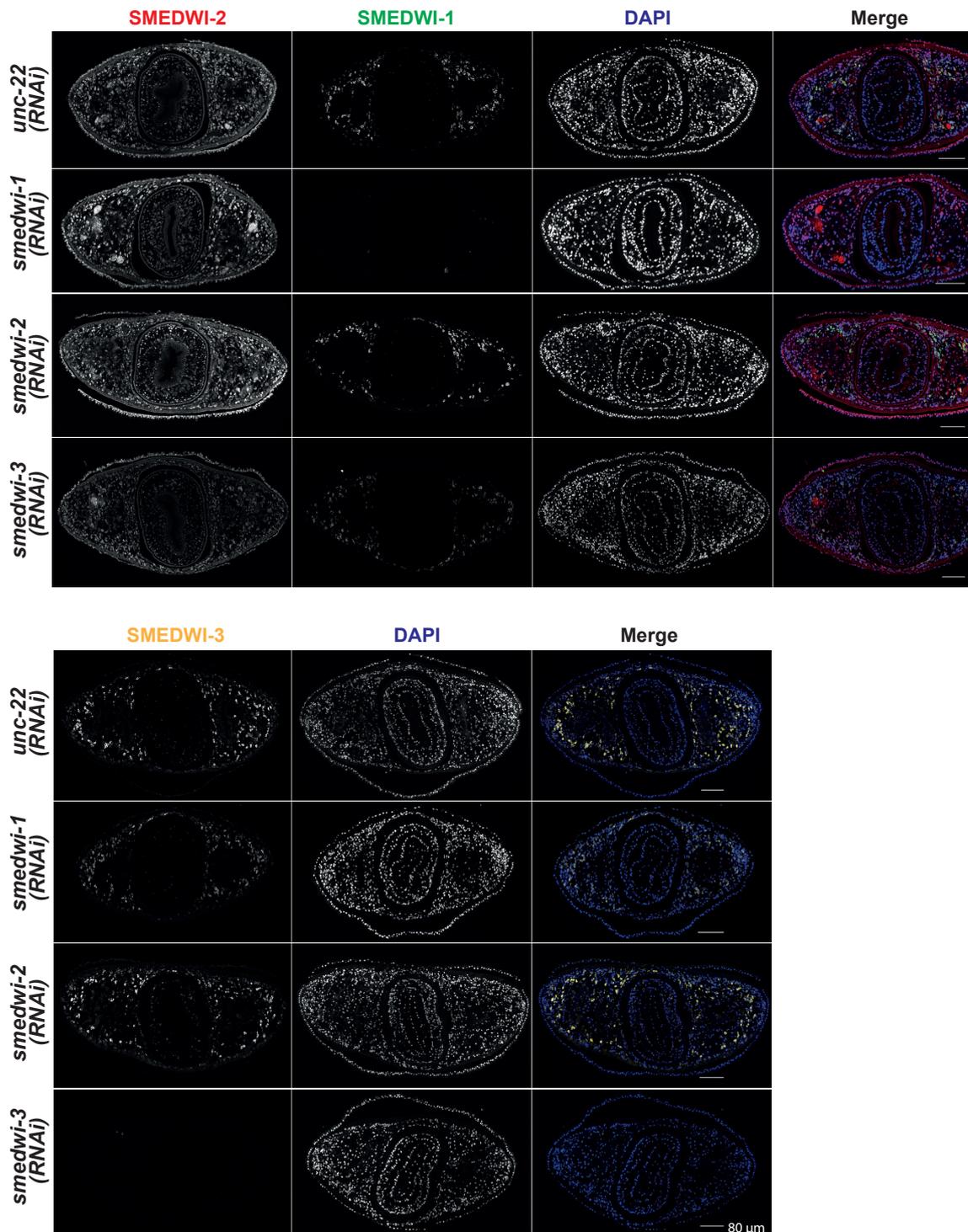
**Fig. 2.6** (A) Evaluation of SMEDWI-1, -2 and -3 protein levels by Western blot upon *unc-22(RNAi)*, *smedwi-2(RNAi)* or *smedwi-3(RNAi)* at different days post feeding (dpf) with dsRNA.  $\alpha$ -Tubulin was used as a loading control. (B) piRNAs co-immunoprecipitated from whole-worm lysates of control (*unc-22 RNAi*) or knockdown (*smedwi-1,-2,-3 RNAi*) animals using antibodies against SMEDWI-1, -2 or -3. Below: western blot showing SMEDWI-1, -2 and -3 protein levels in different RNAi knockdown backgrounds. (C) Gene expression levels in TPMs (transcript per million) of *smedwi-1*, -2 and -3 transcripts upon *unc-22(RNAi)* and *smedwi-2(RNAi)* or *smedwi-3(RNAi)* in neoblasts (X1) and differentiated (Xins) cells.

Optimal time points for RNAi knockdown experiments were defined by achieving the maximum possible depletion of SMEDWI-2 or SMEDWI-3 protein levels with only minor accompanying effects on SMEDWI-1, a marker of neoblast maintenance (Fig. 2.6A). Whereas we observed a significant reduction in SMEDWI-3 protein levels upon *smedwi-3(RNAi)* on day 11 post feeding without affecting SMEDWI-1 expression (Fig. 2.6A-C,

Fig. 2.7), SMEDWI-2 largely persisted in both neoblasts and differentiated cells on day 7 post feeding. However, the amount of co-immunoprecipitated piRNAs after SMEDWI-2 immunoprecipitation was drastically reduced upon *smedwi-2(RNAi)* (Fig. 2.6B). Moreover, we observed the accumulation of SMEDWI-2 in the cytoplasm of planarian cells upon *smedwi-2(RNAi)* (Fig. 2.7). As piRNA loading is required for the nuclear entry of PIWI protein [161, 130], we speculate that at least part of the persisting pool of SMEDWI-2 is not loaded with piRNAs. Overall, we found the levels of piRNAs co-immunoprecipitated with SMEDWI-3 not altered under *smedwi-2(RNAi)* conditions and vice versa.

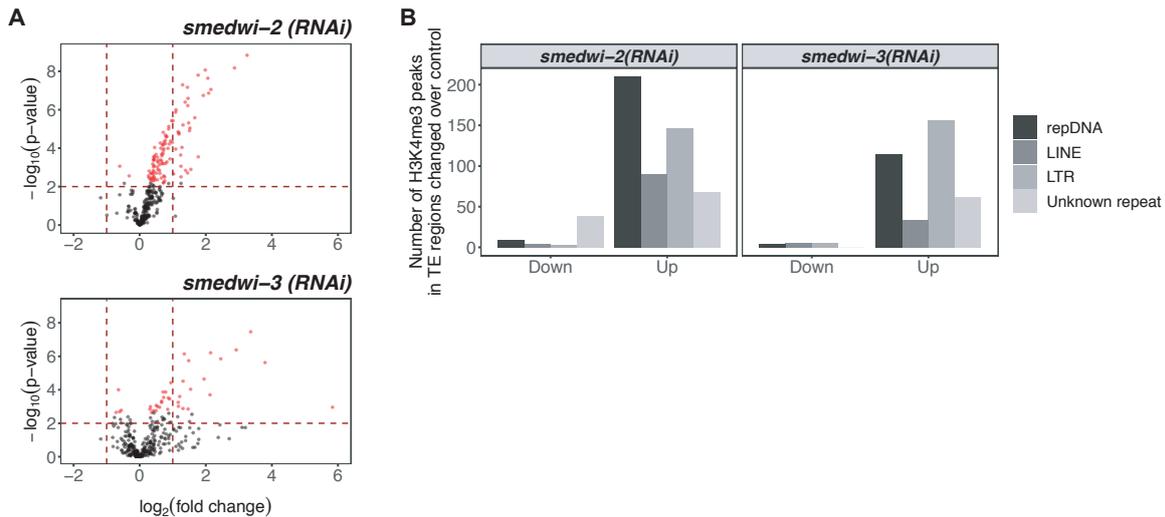
Following RNAi, worms were dissociated and separated by FACS into neoblasts (X1) and differentiated cells (Xins). In the asexual *S. mediterranea*, neoblasts are the only dividing cells and hence can be separated from differentiated cells by sorting for DNA content [66, 145]. Total RNA from 100,000 FACS-isolated neoblasts was prepared and subjected to a custom ribosomal RNA (rRNA) depletion procedure. We developed this procedure to retain non-coding transcripts and those lacking long polyA tails (see Results part 2). Our protocol is robust and removes nearly 99% of the rRNA. Following ribodepletion, we prepared strand-specific libraries for knockdown and control conditions as described [206]. The expression levels of transposable elements after both SMEDWI-2 and SMEDWI-3 knockdown were analyzed by measuring the abundance of 316 consensus transposon families [12]. Upon SMEDWI-2 knockdown, 109 transposon families were upregulated, whereas the SMEDWI-3 knockdown increased the expression of only 38 families (Fig. 2.8A).

As epigenetic silencing of transposable elements by PIWI proteins has been reported in other organisms [200, 172], we asked whether this is also the case in planarians. We performed CHIP-seq experiments for trimethylated histone H3 Lysine 4 (H3K4me3), a chromatin mark associated with active transcription [70]. We found a considerable increase of H3K4me3 peaks at transposable elements upon *smedwi-2(RNAi)* and *smedwi-3(RNAi)* (Fig. 2.8B). Out of 515 upregulated peaks overlapping with transposable elements upon *smedwi-2(RNAi)*, 40% were located at DNA transposons and 28% at long terminal repeat elements (LTRs) (Fig. 2.8B). *Smedwi-3(RNAi)* also led to an increase in the active transcription mark at DNA transposons (115 out of 367 upregulated peaks mapped to transposable elements) and LTRs (156 peaks) (Fig. 2.8B).



**Fig. 2.7** Immunostaining of SMEDWI-1, -2 and -3 on cross-sections through the planarian pharynx in control (*unc-22 RNAi*) or knockdown (*smedwi-1,-2,-3 RNAi*) animals. SMEDWI-1 is in green, SMEDWI-2 is in red, SMEDWI-3 is in yellow. Nuclear staining with DAPI is shown in blue.

Altogether, the enrichment of SMEDWI-2 in the nucleus (Fig. 2.3C) and the observed increase in the expression of transposable elements upon SMEDWI-2 and SMEDWI-3 RNAi knockdown (Fig. 2.8A, 2.8B), indicate that SMEDWI-2 is likely directly involved in epigenetic silencing of transposable elements. Why a knockdown of SMEDWI-3 also leads to an increase in H3K4me3 peaks remains to be investigated.



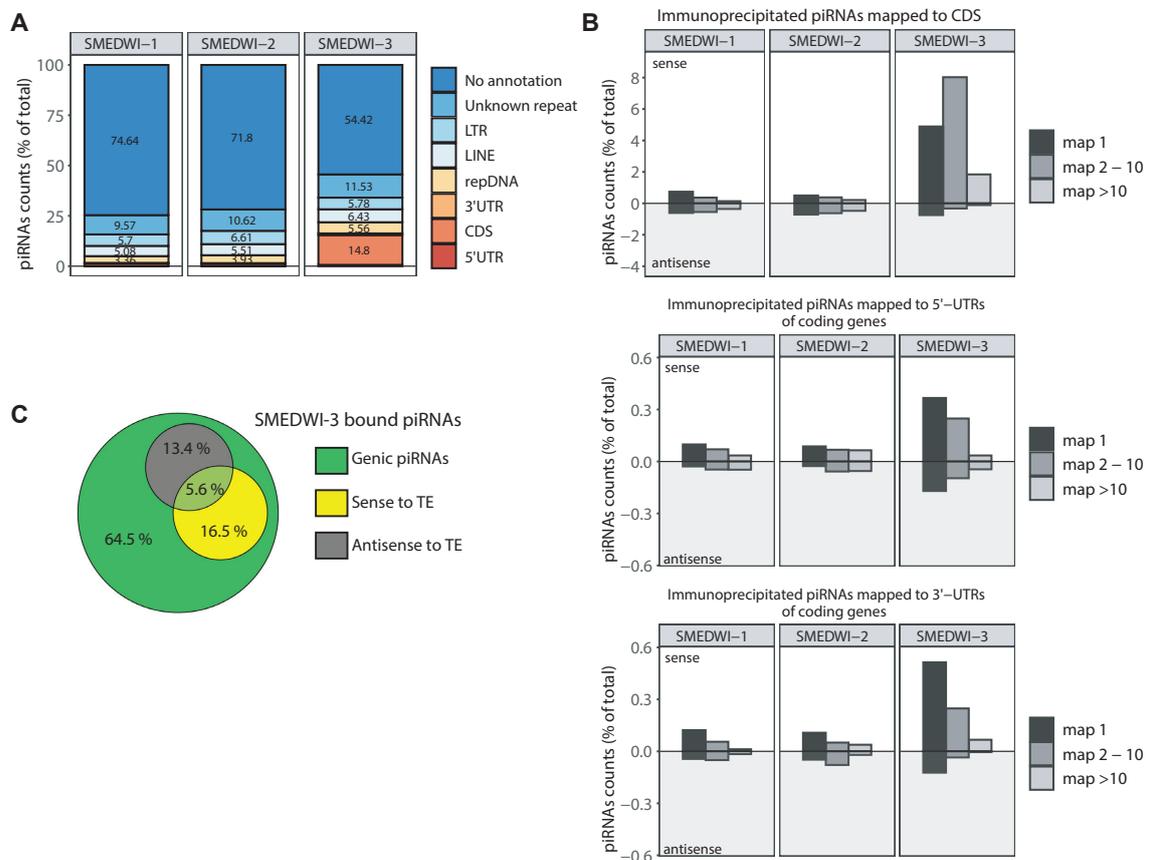
**Fig. 2.8** (A) Differentially expressed transposable element families upon RNAi knockdown of *smedwi-2* or *smedwi-3* (in red). The dashed horizontal red line represents the p-value threshold (p-value < 0.01). The two dashed vertical lines illustrate the threshold of  $\log_2$  fold changes > 1 or < -1. (B) Numbers of up- and downregulated H3K4me3 peaks mapping to transposable elements upon *smedwi-2*(RNAi) and *smedwi-3*(RNAi).

### 2.1.3 Genic piRNAs bound to SMEDWI-3 are degradation products of planarian mRNAs

To decipher which genomic targets apart from transposable elements are subjected to piRNA-mediated regulation, all immunoprecipitated piRNAs were mapped to the planarian genome (see chapter 4). By allowing no mismatches and retaining multiple alignments we achieved mapping rates of 52%, 56% and 43% for SMEDWI-1, -2 and -3-bound piRNAs, respectively. The mapping rates we obtained are rather low, likely due to the high degree of genome heterozygosity in planarians [128, 58]. Nonetheless, the alignment of SMEDWI-1, -2, and -3 co-bound piRNAs resulted in 2.8, 3.0 and 9.9 million uniquely mapped sequences, respectively. We found that the majority of planarian piRNAs mapped to unannotated regions. Only 25 - 30% of piRNAs were assigned to repetitive loci associated with transposable elements (Fig. 2.9A). This fact is in agreement with previous reports [44] and resembles a

genomic distribution of the pachytene piRNAs (74% pachytene piRNAs map to unannotated regions and 17% to TEs) [9]. In *D. melanogaster* over 70% of immunoprecipitated piRNAs were annotated to TEs [18].

However, a significant fraction of SMEDWI-3-bound piRNAs mapped to genic features, especially to coding regions, also displaying a strong sense bias and comparatively little multimapping (Fig. 2.9A-B). What is more, SMEDWI-3-bound genic piRNAs do not map to annotated transposable elements, further strengthening our conclusion that these piRNAs are directly derived from mRNA transcripts (Fig. 2.9C).

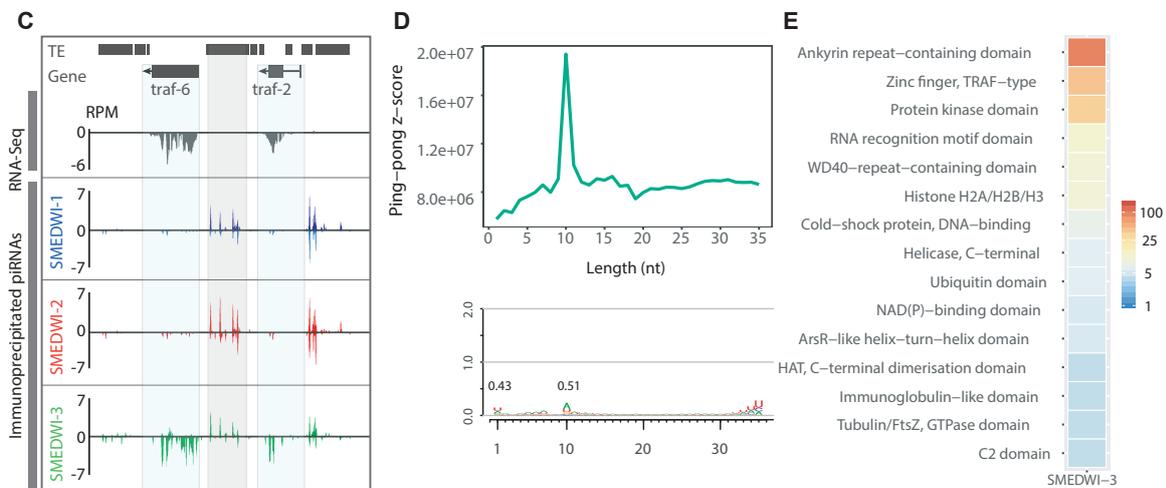


**Fig. 2.9** (A) Genomic locations of piRNAs associated with different SMEDWI proteins. piRNA counts weighted by the number of places they map to were used for the genomic annotation of piRNAs. (B) piRNAs co-immunoprecipitated with SMEDWI protein mapped to 5'-UTRs, coding regions or 3'-UTRs of mRNAs. (C) Percentages of SMEDWI-3-bound genic piRNAs mapping in sense and antisense orientation to transposable elements.

We then focused our attention on mRNAs to which SMEDWI-3-bound piRNAs mapped with an average of 10 reads per million (RPM) in all three immunoprecipitation replicates.

Using this cutoff, we identified 925 transcripts that are processed into SMEDWI-3 bound piRNAs (Supplemental Table 1 [80]).

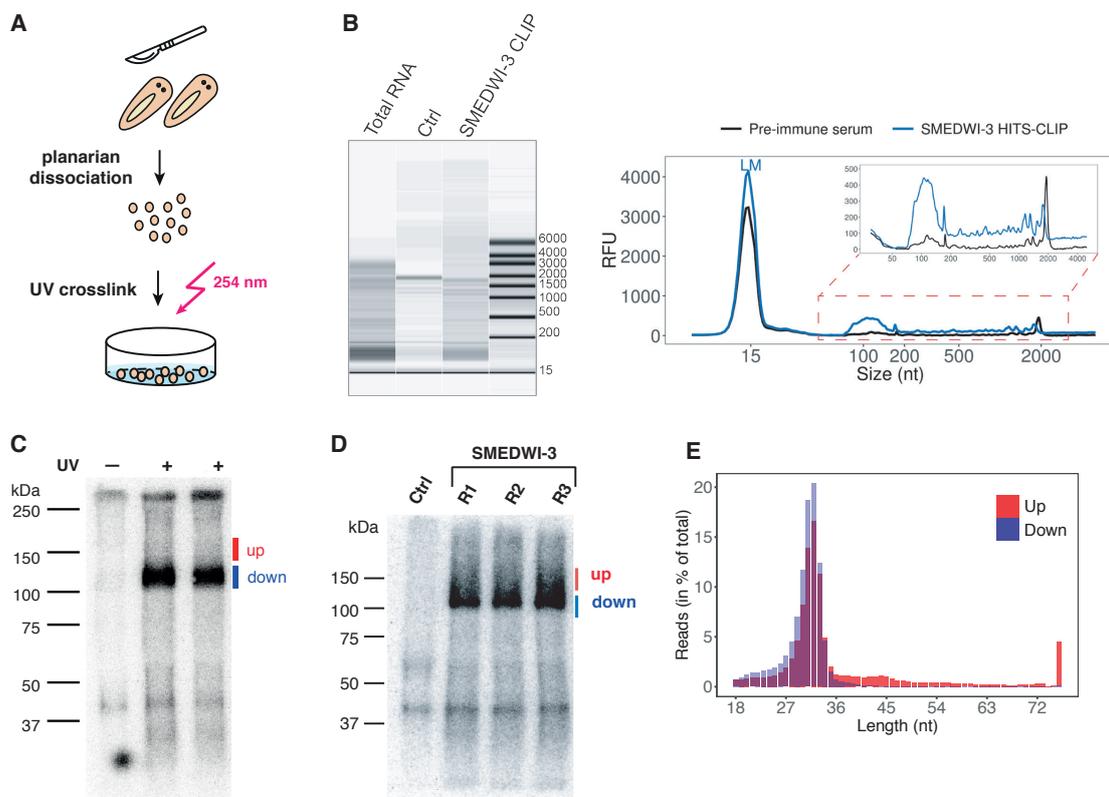
Figure 2.10A shows a representative gene, *traf-6* (SMESG000000371.1), which is degraded into great numbers of SMEDWI-3 bound piRNAs. The resulting genic piRNAs map in a sense orientation exclusively to coding regions of the targeted genes. In contrast, we only detected negligible amounts of *traf-6* piRNAs bound to SMEDWI-1 and SMEDWI-2. Neither of the two proteins showed the strong sense bias of mapped piRNAs that we find to be characteristic for SMEDWI-3. Furthermore, transposon-related elements near genes are not degraded into strand-specific SMEDWI-3-bound piRNAs, yet can be detected bound by all three PIWI proteins (Fig. 2.10A). We also detected significant ping-pong signatures and the characteristic 5' 1U and 10A nucleotide biases that accompany ping-pong amplification for our list of genic SMEDWI-3 bound piRNAs (Fig. 2.10B). These data strongly suggest that the piRNA-mediated mRNA turnover results in genic piRNAs specifically bound to SMEDWI-3. Notably, the large number of targeted mRNAs contains ankyrin-repeat-containing domain, zinc-finger TRAF-type domain, histone fold etc. (Fig. 2.10C), suggesting the possibility that piRNAs might regulate the expression levels of entire protein domain families.



**Fig. 2.10** (A) Coverage profile of RNA-seq and immunoprecipitated piRNAs for *traf-6* (SMESG000000371.1). The gene loci are highlighted in blue. A neighboring transposable element is highlighted in gray. (B) Ping-pong analysis of transcripts producing sense piRNAs. Nucleotide composition analysis of total piRNAs mapped to processed transcripts confirms the involvement of the ping-pong cycle. (C) Heatmap showing the enrichment of predicted InterPro protein domains in transcripts with mapped genic piRNAs.

### 2.1.4 Crosslinking immunoprecipitation confirms that SMEDWI-3 targets hundreds of planarian transcripts

To gain direct evidence of the involvement of SMEDWI-3 in planarian mRNA decay, we established a crosslinking immunoprecipitation (CLIP) protocol for SMEDWI-3 in planarians [185]. Since UV irradiation does not penetrate tissues efficiently, we first rapidly dissociated planarians into a single cell suspension and then crosslinked the dissociated cells on ice at  $\lambda = 254 \text{ nm}$  (Fig. 2.11A).



**Fig. 2.11** (A) Schematic representation of HITS-CLIP in planarians. (B) Fragment Analyzer digital gel image and electropherogram of RNA fragments crosslinked and co-immunoprecipitated along with SMEDWI-3. Pre-immune serum was used as a negative control. The gel electrophoresis profile for total planarian RNA demonstrates the high level of RNase activity in planarian lysates. (C) Immunoprecipitated HITS-CLIP SMEDWI-3-RNA complexes were radioactively labeled, separated on an 8% Bis-Tris gel and blotted onto a nitrocellulose membrane. CLIP libraries were prepared from RNA extracted separately from the upper and lower bands. Pre-immune serum served as negative control. (D) The same as (C) for SMEDWI-3 CLASH libraries. (E) Length distribution of CLIP reads from the upper and lower band.

Crosslinked protein-RNA complexes were isolated using the anti-SMEDWI-3 antibody and resolved by PAGE. The average size of RNAs crosslinked to SMEDWI-3 was below

200 nts, hinting at strong nucleolytic activity in planarian lysates (Fig. 2.11B). We also used a modified CLASH (cross-linking ligation and sequencing hybrids) protocol to capture chimeric piRNA-mRNAs reads to gain direct experimental evidence of piRNA-target interactions [169, 67]. Both HITS-CLIP (Fig. 2.11C) and CLASH (Fig. 2.11D) libraries were prepared from two crosslinked RNA species that were separately isolated: a shorter, very prominent species crosslinked to SMEDWI-3 (labeled "down"), and a smear of longer RNA fragments (labeled "up") (Fig. 2.11E).

Sample	Raw reads	Collapsed	rRNA	Others
CLASH rep1 Up	24 491 573	4 852 114	23 402 (0.48%)	4 828 712 (99.52%)
CLASH rep1 Down	26 355 748	7 165 640	10 563 (0.15%)	7 155 077 (99.85%)
CLASH rep2 Up	27 365 590	6 868 529	26 767 (0.39%)	6 841 762 (99.61%)
CLASH rep2 Down	25 765 796	7 680 635	10 827 (0.14%)	7 669 808 (99.86%)
CLASH rep3 Up	21 465 736	4 734 867	26 630 (0.56%)	4 708 237 (99.44%)
CLASH rep3 Down	24 209 886	7 526 984	11 138 (0.15%)	7 515 846 (99.85%)
CLASH rep4 Up	20 977 519	5 260 232	20 596 (0.39%)	5 239 636 (99.61%)
CLASH rep4 Down	25 279 223	9 016 828	12 621 (0.14%)	9 004 207 (99.86%)
CLASH rep5 Up	19 402 092	1 125 858	154 604 (13.73%)	971 254 (86.27%)
CLASH rep5 Down	28 465 247	741 847	6 663 (0.90%)	735 184 (99.10%)
CLASH rep6 Up	25 713 339	1 790 701	133 178 (7.44%)	1 657 523 (92.56%)
CLASH rep6 Down	28 902 373	940 761	12 647 (1.34%)	928 114 (98.66%)
HITS-CLIP rep1 Up	22 658 032	5 740 698	33 265 (0.58%)	5 707 433 (99.42%)
HITS-CLIP rep1 Down	25 275 256	8 037 363	12 600 (0.16%)	8 024 763 (99.84%)
HITS-CLIP rep2 Up	27 914 812	1 992 586	197 431 (9.91%)	1 795 155 (90.09%)
HITS-CLIP rep2 Down	29 016 966	1 554 661	14 472 (0.93%)	1 540 189 (99.07%)
HITS-CLIP rep3 Up	25 854 504	1 203 988	247 219 (20.53%)	956 769 (79.47%)
HITS-CLIP rep3 Down	26 441 640	1 426 598	5 201 (0.36%)	1 421 397 (99.64%)

**Table 2.1** Sequenced HITS-CLIP and CLASH reads.

In total, we analyzed nine CLIP replicates. After removal of PCR duplicates and rRNA filtering (Table 2.1), fragments from "up" and "down" libraries were pooled and jointly analyzed. CLIP reads were mapped to the planarian genome, resulting in a mapping rate

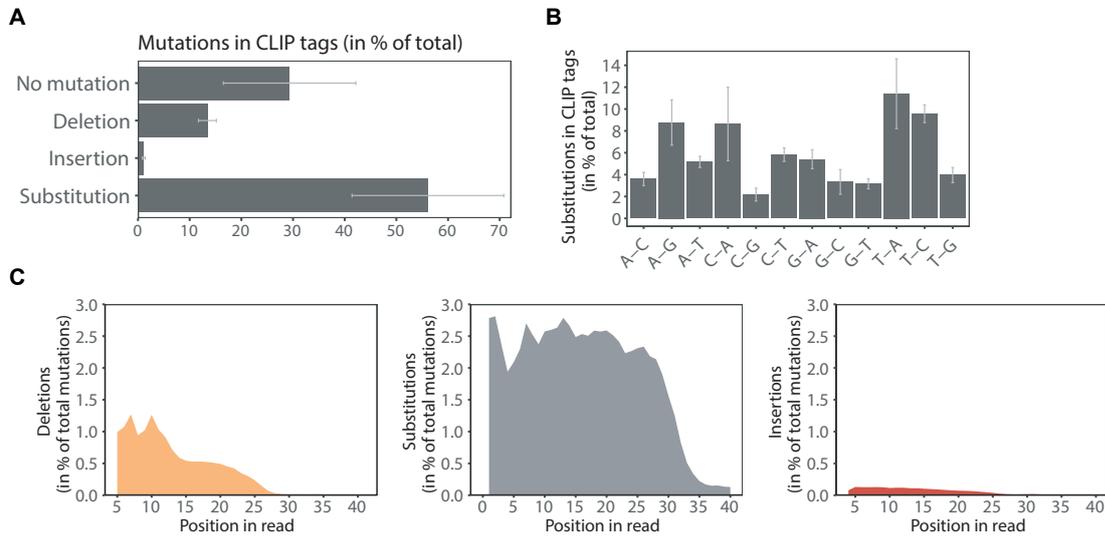
of 55 – 65%. This corresponds to between 0.4 and 6.6 million uniquely assigned reads per replicate (Table 2.2).

Sample	Total reads (Up + Down)	Mapped to the genome	Unique mappers	Multi- mappers	Chimeric reads
CLASH rep1	11 983 789	7 467 622 (62.3%)	5 123 574 (42.75%)	2 344 048 (19.56%)	3 836 (0.03%)
CLASH rep2	14 511 570	9 685 201 (66.7%)	6 681 087 (46.04%)	3 004 114 (20.7%)	3 851 (0.026%)
CLASH rep3	12 224 083	7 785 957 (63.7%)	5 360 248 (43.85%)	2 425 709 (19.84%)	5 553 (0.045%)
CLASH rep4	14 243 843	9 336 396 (55.5%)	6 412 284 (45.02%)	2 924 112 (20.53%)	3 885 (0.027%)
CLASH rep5	1 706 438	632 053 (37%)	424 656 (24.89%)	207 397 (12.15%)	33 418 (1.95%)
CLASH rep6	2 585 637	1 110 606 (65.2%)	750 874 (29.04%)	359 732 (13.91%)	26 125 (1.01%)
HITS-CLIP rep1	13 732 196	9 058 729 (65.9%)	6 226 614 (45.34%)	2 832 115 (20.62%)	290 (0.002%)
HITS-CLIP rep2	3 335 344	2 202 502 (66%)	1 539 741 (46.16%)	662 761 (19.87%)	9 979 (0.3%)
HITS-CLIP rep3	2 378 166	1 324 806 (55.7%)	907 520 (38.16%)	417 286 (17.55%)	13 646 (0.57%)

**Table 2.2** HITS-CLIP and CLASH reads mapped to the planarian genome.

Next, we applied crosslinking induced mutation site analysis (CIMS) to define CLIP sites in planarian mRNAs [203, 168]. CIMS analysis identifies crosslink-induced deletions, substitutions and insertions, which are introduced at the protein-RNA crosslink sites by reverse transcriptase [203]. As reported for human Argonaute-2 and others, mutation-induced base substitutions were most frequent in our data (present in 56% of CLIP tags), while deletions accounted only for 13% of all mutations (Fig. 2.12A-B) [124]. We focused our analysis on CLIP tags that carry both deletions ( $FDR \leq 0.001$ ) and substitutions ( $FDR \leq 0.001$ ) and that appear in at least two replicates, since both types of mutations were evenly

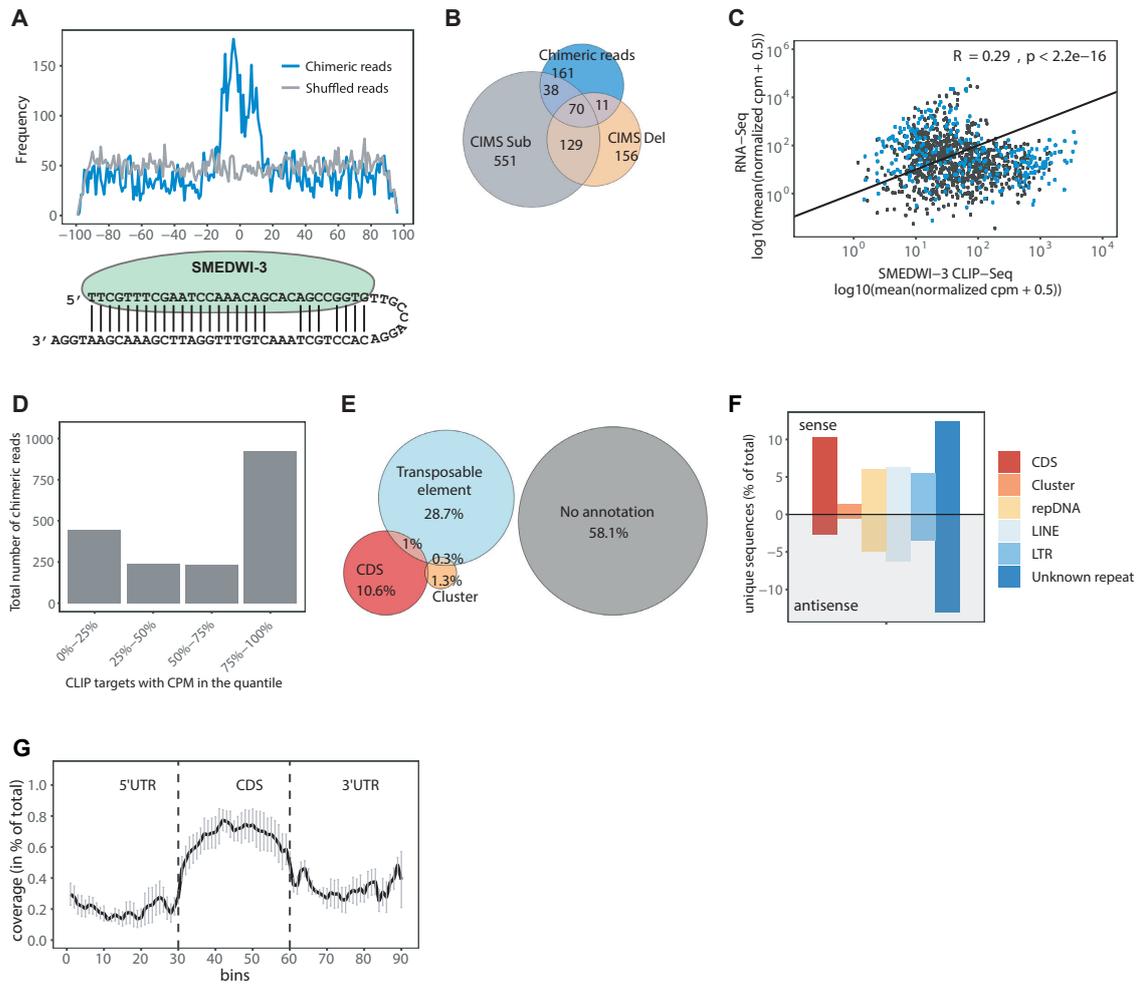
distributed throughout the length of the CLIP read (Fig. 2.12C). We detected 745 deletions in 366 mRNAs and 1629 substitutions in 788 mRNAs. All final CLIP tags comprise > 30 unique reads in at least two replicates per site of mutation.



**Fig. 2.12** (A) Percentage of mutations in CLIP tags. (B) Nucleotide substitution frequency observed in CLIP reads. (C) Distribution of deletions (in orange), substitutions (in grey) and insertions (in red) along CLIP reads.

Using the CLIP-chimaeric pipeline, our data also revealed thousands of chimeric piRNA-mRNA reads (Table 2.2) [3, 184]. The piRNA parts of these chimeric reads were mapped on the extended mRNA fragments, and chimeric reads with alignments of piRNA parts within  $\pm 20$  bp around the midpoint of the mRNA fragments were kept for the subsequent analysis (Fig. 2.13A). In total, we discovered chimeric reads for 280 planarian mRNAs (Supplemental Table 3 [80]). The mRNAs carrying chimeric reads were also detected by our CIMS analysis (Fig. 2.13B).

After combining chimeric reads and CIMS-analyzed CLIP tags we were able to define 1,116 transcripts as SMEDWI-3 targets (Fig. 2.13B, Supplemental Table 4 [80]). To compare the distribution of the reads mapping to genes in CLIP-Seq and RNA-Seq experiments, the Spearman correlation coefficient was calculated between both types of experiments. It revealed that the correlation between the gene expression and CLIP tag abundance is low ( $R=0.29$ ,  $p < 2.2e-16$ ) strengthening the specificity of our CLIP experiment (Fig. 2.13C). Moreover, we note that about 50% of all chimeric reads are derived from the top 25% of highly abundant transcripts in the CLIP libraries (Fig. 2.13D).



**Fig. 2.13** (A) Alignment of the piRNAs parts of chimeric reads onto the  $\pm 100$  nt extended midpoint of mRNA fragments. An alignment of piRNAs to random mRNA fragments was used as a negative control (in grey). (B) Venn diagram representing the identified SMEDWI-3 CLIP targets carrying deletions, substitutions and chimeric reads. (C) Scatter plot comparing  $\log_2$  normalized CPM (counts per million) of mapped reads for CLIP-Seq and RNA-Seq data. Transcripts carrying chimeric reads are highlighted in blue. The Spearman's correlation coefficient is indicated. (D) Cumulative sum of chimeric reads in nine replicates of CLIP targets. All CLIP targets were divided into bins based on their expression value (in CPMs) in the four quantiles. (E) The annotation of CLIP fragments to the planarian genome confirms the involvement of SMEDWI-3 in the post-transcriptional regulation of transposable elements and coding transcripts. (F) Percentage of chimeric reads mapped to different genomic features in sense and antisense orientation. (G) Density of the CLIP fragments mapping to 5'-UTRs, coding regions and 3'-UTRs of coding genes. Each feature was divided into 30 bins. The mean density of CLIP fragments was calculated over the corresponding features. Error bars represent SE.

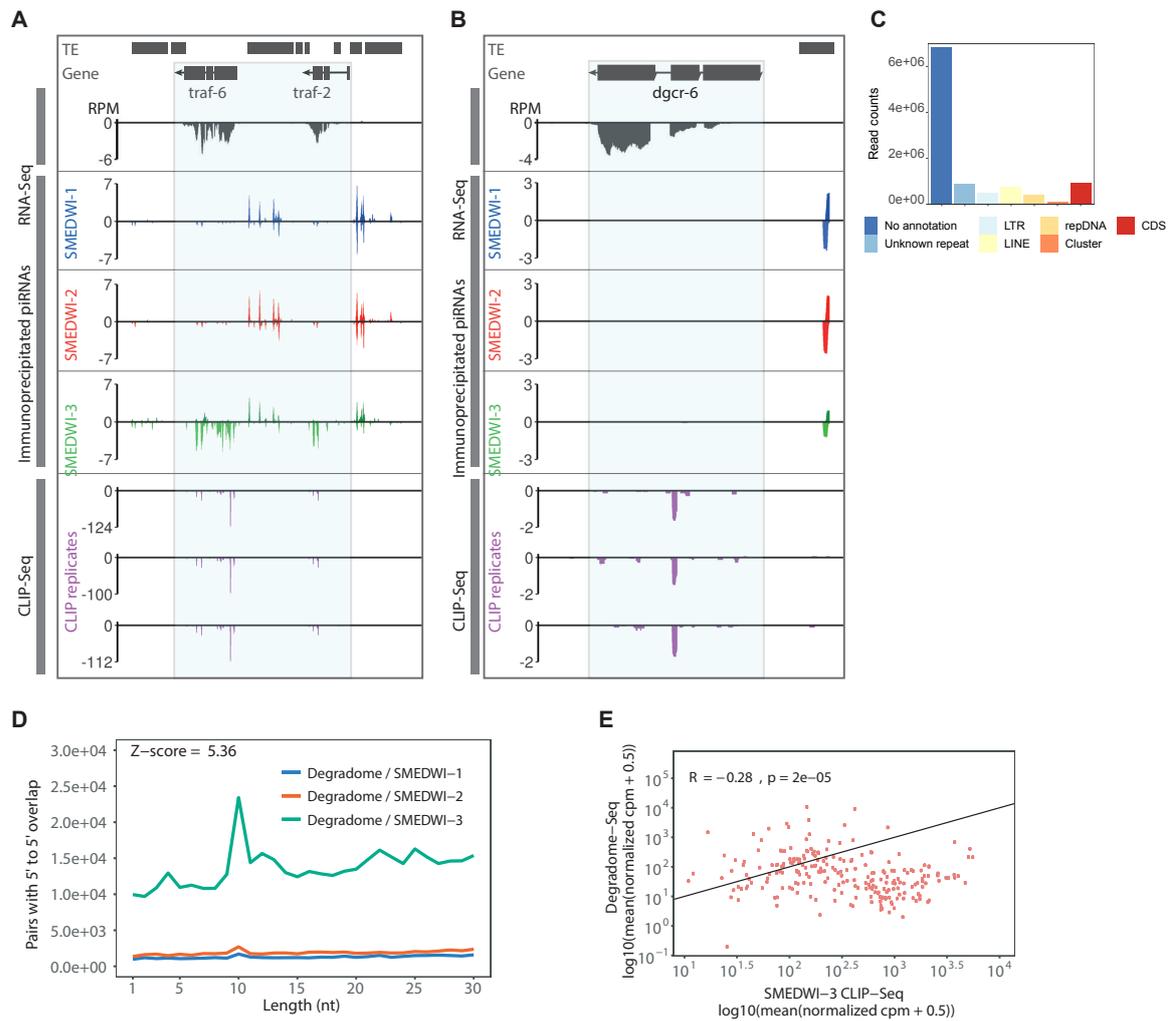
As expected, a significant fraction of CLIP reads (10%) could be traced back to coding regions of genes and not UTRs (Figs. 2.13E-G). They preferentially map in sense orientation,

confirming that SMEDWI-3 indeed targets coding transcripts (Fig. 2.13F). In addition, 28% of our CLIP reads are derived from transposable elements, a fact that supports SMEDWI-3's central role in the ping-pong cycle (Figs. 2.13E).

### 2.1.5 The dual role of SMEDWI-3 in neoblast mRNA surveillance

To dissect how planarian mRNAs are converted into piRNAs, we analyzed the abundance of SMEDWI-3-bound genic piRNAs in the identified CLIP regions. piRNAs were mapped to CLIP regions allowing one mismatch to account for planarian heterozygosity [58, 128]. Next, we computed the piRNA density per bp in all CLIP regions by normalizing weighted piRNA counts to the length of the corresponding CLIP regions. Unexpectedly, not all CLIP regions were densely covered with genic piRNAs. In contrast, our analysis revealed two distinct modes of SMEDWI-3 transcript targeting. A first class of transcripts displayed significant CLIP peaks across their coding regions and showed SMEDWI-3-bound piRNA coverage in the sense direction (Fig. 2.14A). A second class of transcripts showed CLIP signals, yet hardly any piRNA coverage (less than 0.5 piRNA counts per bp of CLIP region) (Fig. 2.14B).

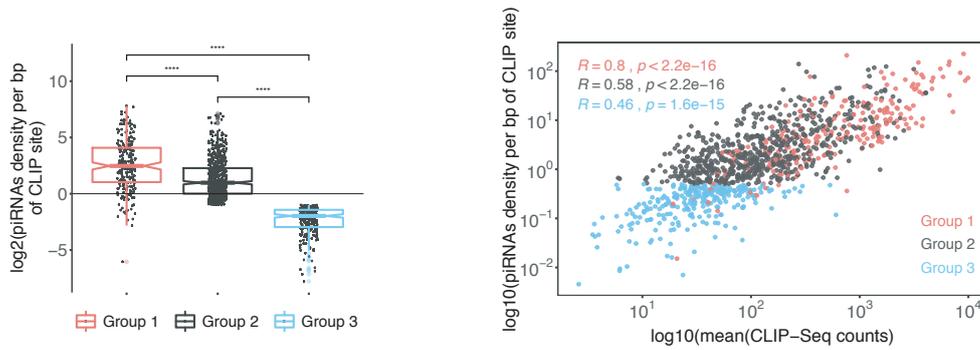
The presence of genic piRNAs in CLIP regions suggested that these targets might be cleaved and processed into piRNAs. To confirm that the degradation of SMEDWI-3 CLIP targets is indeed piRNA-dependent, we performed 5'-monophosphate-dependent cloning of mRNA fragments (Degradome-seq) in sorted planarian neoblasts (X1 cells). We focused on X1 cells since SMEDWI-3 is predominantly present in these cells (Fig. 2.2). Degradome-seq allows the capture of RNA cleavage products that result from piRNA targeting [188, 146]. In total, we were able to map 10,230,560, 10,346,869 and 10,372,010 unique degradome reads to the planarian genome for three biological replicates, respectively. Approximately 9% of all degradome reads mapped to gene coding regions, with 1% possessing a 5'-5' 10-nt overlap with immunoprecipitated piRNAs (Fig. 2.14C). The vast majority of ping-pong pairs between degradome reads and PIWI-bound piRNAs were formed with piRNAs bound by SMEDWI-3 (Z-score = 5.36). This suggests that SMEDWI-3 plays a leading role in degrading planarian transcripts that are not transposable elements.



**Fig. 2.14** (A) RNA-Seq, piRNA-seq and HITS-CLIP coverage profiles for *traf-6* (SMESG00000371.1). The presence of both genic SMEDWI-3-bound piRNAs and CLIP fragments suggests their piRNA-mediated degradation by SMEDWI-3. (B) The presence of CLIP fragments along with the simultaneous absence of sense piRNAs for *dgcr-6* (SMESG000021088.1) indicates that those transcripts are regulated by SMEDWI-3 in a cleavage-independent manner. (C) Genomic annotation of Degradome-Seq reads showing 5' to 5' overlap with piRNAs immunoprecipitated with anti-SMEDWI-1, -2, and -3 antibodies. (D) Number of unique pairs with 5' to 5' 10 nt overlaps between degradome sequences and piRNAs immunoprecipitated with anti-SMEDWI antibodies. (E) Correlation of normalized CPMs of CLIP-Seq and Degradome-Seq data. Each dot represents a transcript with ping-pong degradation signature. Spearman's correlation coefficient is indicated.

Next, we intersected the list of genes showing piRNA-dependent degradation with our list of 1,116 CLIP targets. Using at least five unique ping-pong pairs per transcript as cutoff, we identified 232 SMEDWI-3 CLIP targets to be processed into genic piRNAs (Fig. 2.14D, Supplemental Table 5 [80]). CLIP targets with stronger degradation signal were slightly

less abundant in our CLIP-seq libraries ( $R = -0.28$ ,  $p = 2e-05$ ), a fact likely attributed to their efficient degradation into genic piRNAs. In agreement with the latter, the majority of degraded transcripts showed a high density of genic piRNAs across their CLIP sites (Fig. 2.15).



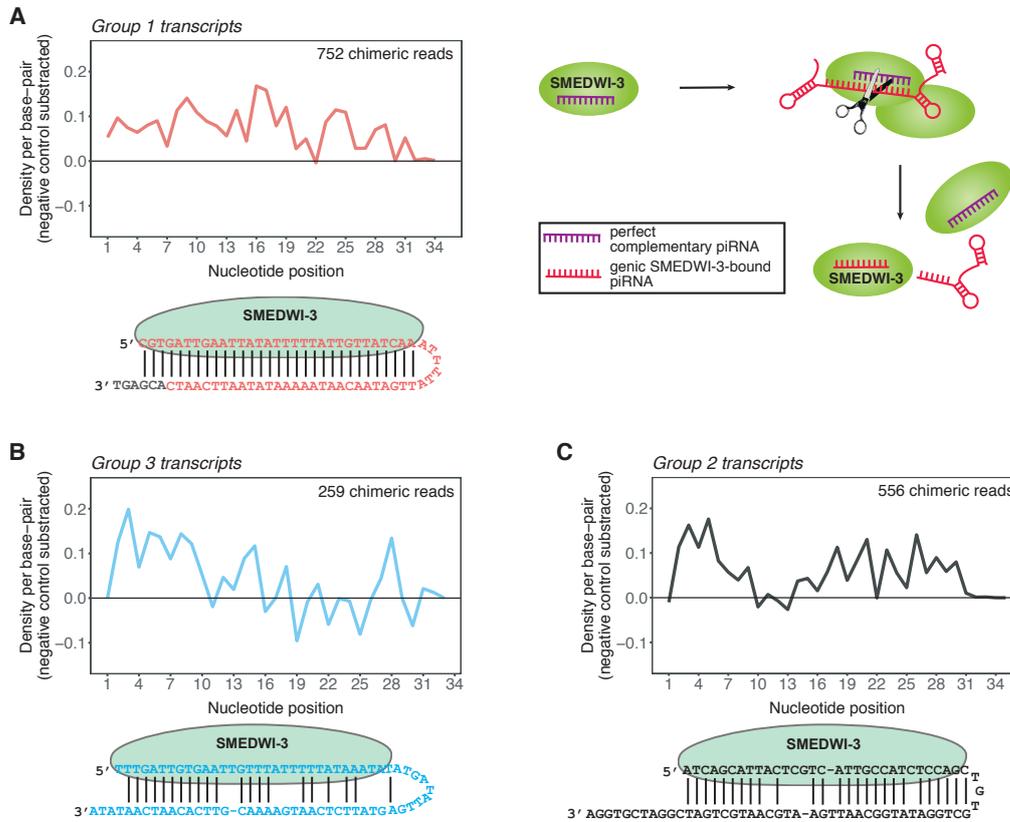
**Fig. 2.15** Left: Boxplots of three groups of CLIP transcripts that are defined based on SMEDWI-3-bound genic piRNA density per bp of CLIP site and the presence of a piRNA-dependent degradation signature. The unpaired t-test was applied to estimate significant differences between groups. P-value  $\leq 0.0001$  is marked with \*\*\*\*. Right: Scatter plot showing the cumulative density of genic piRNAs across the CLIP sites of SMEDWI-3 targets and the corresponding abundance of CLIP counts assigned to the transcript. Group 1 includes transcripts that display a ping-pong degradation pattern along with high genic piRNAs density. Group 2 comprises transcripts with piRNAs density  $\geq 0.5$ /bp CLIP and without a piRNA-dependent degradation signature. Transcripts with a piRNA density  $< 0.5$ /bp CLIP are sorted in group 3. Spearman's correlation coefficients for all three groups are indicated in different colors.

We therefore combined these SMEDWI-3 targets into a first group of targets (group 1). Subsequently, we assigned 621 transcripts to a second group of SMEDWI-3 targets (group 2). Group 2 targets exhibit substantial density of genic piRNAs ( $> 0.5$  piRNAs/bp of CLIP region), however, no ping-pong degradation pairs were detected for these transcripts in the degradome data (Fig. 2.15). Most likely, group 2 transcripts are degraded in the ping-pong cycle as well, however, their degradation might be triggered by piRNAs that exhibit non-perfect base-pairing for target recognition. As we only considered piRNAs mapping to the genome with a single mismatch, we have no means to detect these cases. Combining all SMEDWI-3 CLIP targets with significant genic piRNA density ( $> 0.5$  piRNAs/CLIP bp) left 263 transcripts. For these group 3 transcripts we detected significant SMEDWI-3 binding, yet very little piRNA mapping ( $< 0.5$  piRNAs/CLIP bp) (Fig. 2.15, Supplemental Table 6 [80]). The correlation coefficients for the three groups of transcripts support our classification, as the positive correlation between the number of CLIP reads and the density of genic piRNAs decreases from group 1 through group 3. We therefore conclude that SMEDWI-3 binds

group 3 transcripts, but does not trigger their degradation into significant numbers of genic piRNAs.

### **2.1.6 The base-pairing pattern between SMEDWI-3-bound piRNAs and target mRNAs**

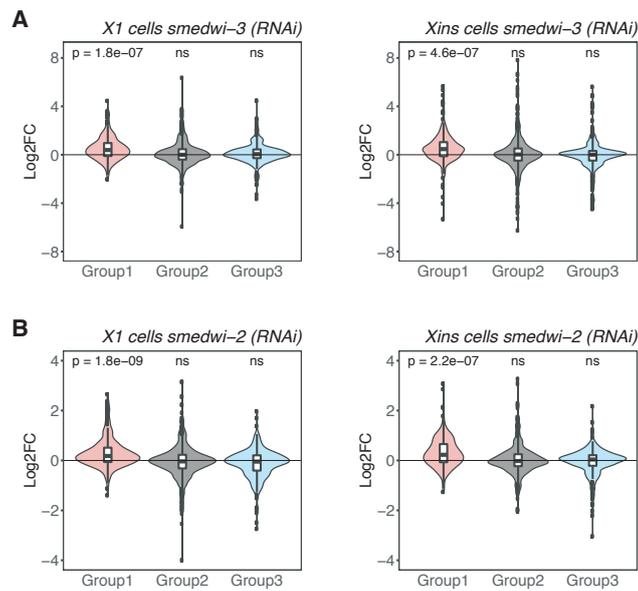
To investigate whether piRNAs are responsible for guiding SMEDWI-3 to group 3 transcripts, we examined the presence of chimeric reads mapping to this group of transcripts. In total, we identified 259 chimeric reads mapping to 75 group 3 transcripts. In addition, our data revealed 752 chimeric reads for 88 group 1 transcripts and 556 for 117 group 2 transcripts. To determine the base-pairing patterns underlying the interaction of SMEDWI-3-bound piRNAs with all three groups of transcripts, the unweighted local read aligner as implemented in the CLIP-chimeric pipeline was used [3, 184]. After subtraction of a randomized control, we observed continuous base pairing over the whole length of antisense piRNAs for group 1 targets (Fig. 2.16A). This observation confirms that antisense piRNAs need to recognize their targets with perfect or near-perfect complementarity to initiate target cleavage. It also confirms prior knowledge on cleavage-competent Argonaute proteins, which require continuous base-pairing across the scissile phosphate between target nucleotides 10 and 11 [163]. Meanwhile, piRNAs that target group 3 transcripts show a significant drop in base-pairing at nucleotides 10 and 11 and a general lack of base-pairing between nucleotide positions 16 and 26 (Fig. 2.16B). These data argue that group 3 transcripts are recognized by SMEDWI-3 with piRNAs as guides. However, target cleavage by SMEDWI-3 is likely precluded due to the multitude of mismatches between piRNA and mRNA target. In support of this notion group 3 transcripts show a low density of genic piRNAs across their CLIP sites. Last, when analyzing the base-pairing patterns of group 2 chimeras, we noticed almost continuous piRNA-mRNA base-pairing apart from a significant drop thereof between nucleotide positions 10 and 13 (Fig. 2.16C). Because we detected significant genic piRNAs for this group of transcripts yet no degradome signal (Fig. 2.15), we speculate that this drop may illustrate that efficient target cleavage is impaired for these transcripts due to a lack of base-pairing over the scissile phosphate group. Alternatively, due to limitations of our classification of group 2 transcripts, this group might include transcripts from both group 1 and group 3. Taken together, our results are in agreement with the previously described piRNA targeting rules in *C. elegans* [169, 204]. Moreover, they are also in line with the rules established for Argonaute targeting and underline the importance of continuous base-pairing over the scissile phosphate bond to initiate target cleavage [163] (Fig. 2.16A).



**Fig. 2.16** (A) Average density of base-pairing events per nucleotide position of the piRNA parts of chimeric reads. Chimeric piRNA parts were mapped onto targeted mRNA fragments in the vicinity of *pm20* nt from the mRNA fragment midpoint. Only unique chimeric reads mapping to transcripts of Group 1 were included from all nine replicates. Random piRNA mapping densities were subtracted as negative control. Below, an illustration of one exemplary chimeric read from group 1 transcripts is shown. The extended mRNA part of the chimeric read is shown in gray. Left: schematic representation of mRNAs degradation in a homotypical ping-pong cycle by SMEDWI-3. (B) Same as in (A) for group 3 SMEDWI-3 targets. (C) Same as in (A) for group 2 SMEDWI-3 targets.

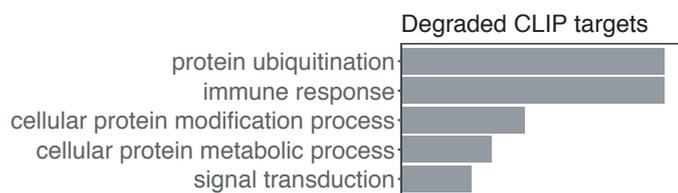
### 2.1.7 Transcripts degraded by SMEDWI-3 reveal a piRNA pathway in the planarian epidermis

To decipher the impact of SMEDWI-3 on its target mRNAs, we examined the expression levels of transcripts in all three previously defined groups after *smedwi-3(RNAi)* knockdown. A gene set enrichment analysis [111] revealed that the expression of ping-pong degraded group 1 targets increased significantly in both *smedwi-3(RNAi)* neoblasts and differentiated cells (Fig. 2.17A). However, we also observed a significant upregulation of group 1 targets upon *smedwi-2(RNAi)* (Fig. 2.17B).



**Fig. 2.17** (A) Violin plots showing the log<sub>2</sub> fold differential expression changes of the three distinct SMEDWI-3 target groups upon *smedwi-3* knockdown in neoblasts (X1) and differentiated cells (Xins). Statistical significance of differential expression of the gene sets was assessed using Generally Applicable Gene-set Enrichment (GAGE) analysis with a two-sample t-test ("ns" = not significant). (B) The same as (A) for *smedwi-2*(RNAi).

We therefore speculate that piRNA pathway-dependent factors contribute to the degradation of these mRNAs or that SMEDWI-2 is involved in piRNAs precursor transcription and processing. Alternatively, a knockdown of either PIWI protein might disturb neoblasts fate and lead to the upregulation of SMEDWI-3 ping-pong targets (group 1). We note here that group 1 targets are enriched in transcripts involved in protein ubiquitination and immune response (Fig. 2.18).



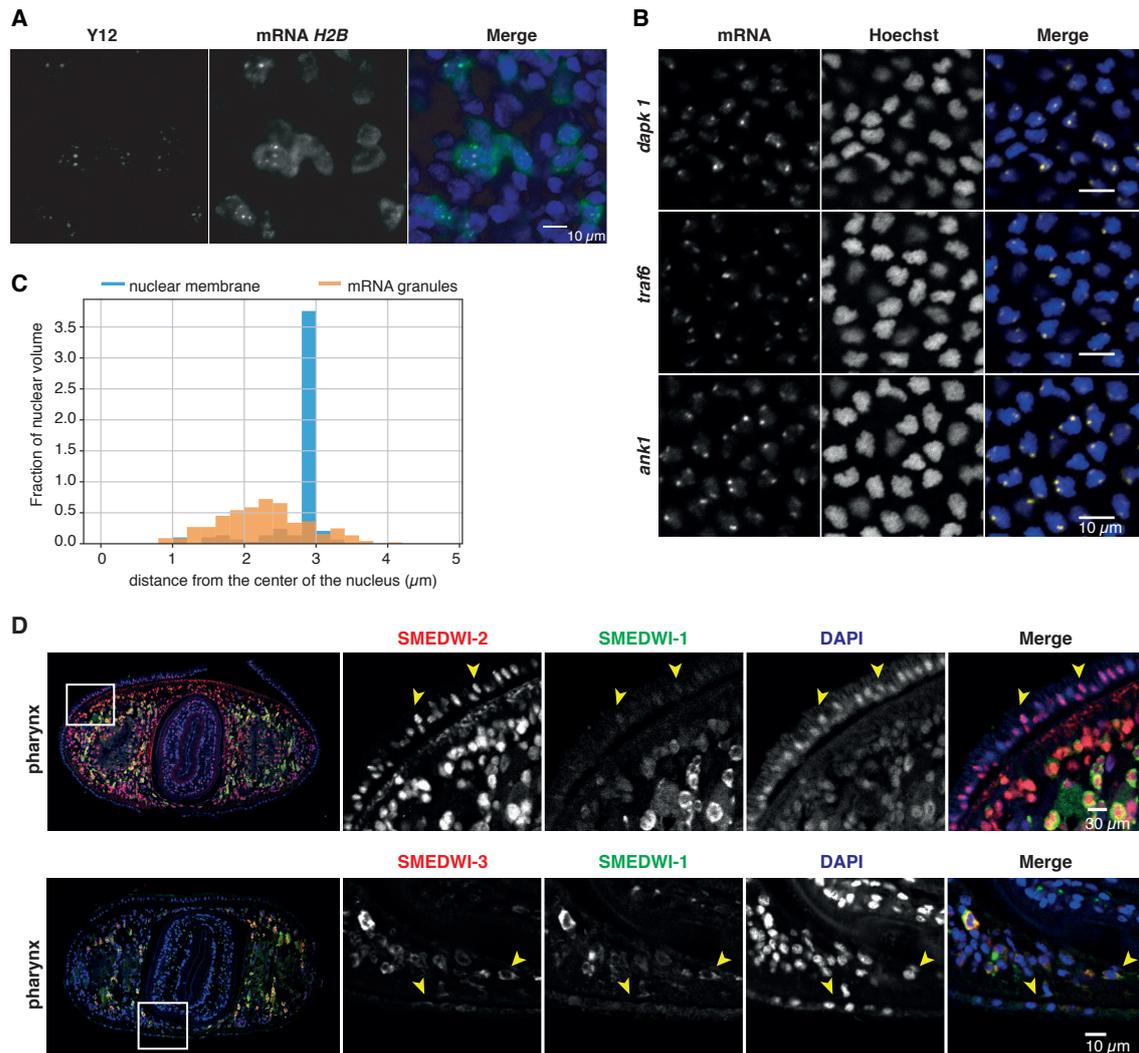
**Fig. 2.18** GO term enrichment analysis of ping-pong degraded group 1 SMEDWI-3 targets.

Overall, group 2 and group 3 targets do not show significant expression changes upon SMEDWI-2 and SMEDWI-3 knockdown. This suggests that SMEDWI-3 binding to these mRNAs does not have an immediate impact on their stability. However, their interaction

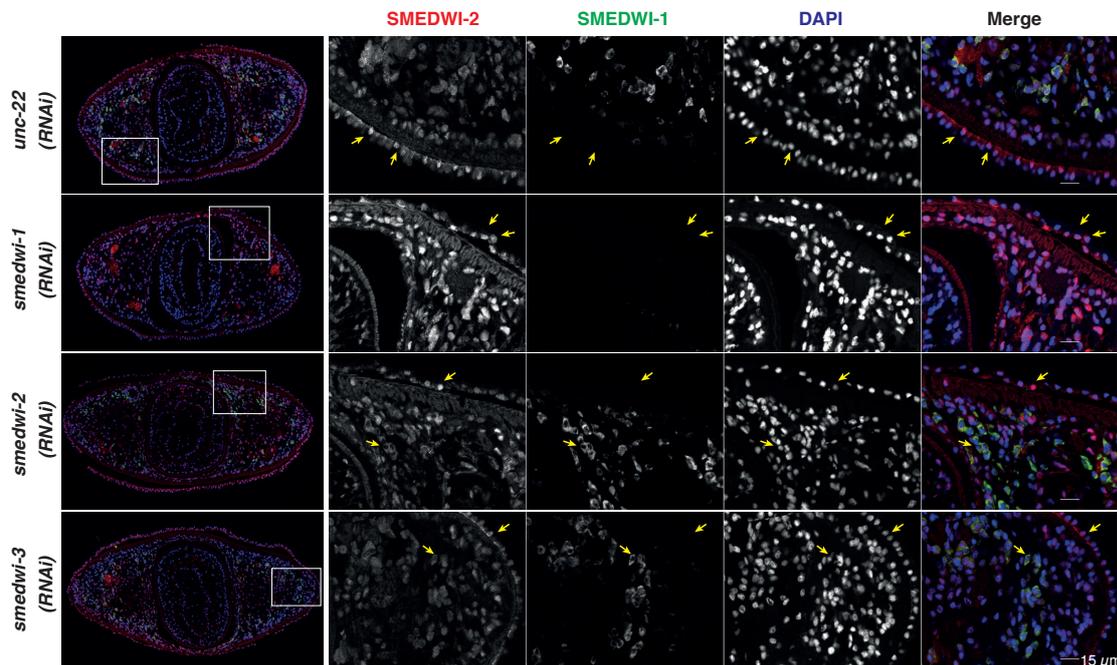
with SMEDWI-3 might still be important by either licensing these transcripts in a type of surveillance pathway or by regulating their translation [167, 186].

Given the significant upregulation of SMEDWI-3 group 1 targets in neoblasts and differentiated cells, we asked where these targets are expressed in planarians. We carried out whole-mount fluorescent in situ hybridization (WISH) of four group 1 transcripts that we found targeted by the ping-pong cycle (*histone H2B* (SMESG000052758.1), *ank1* (SMESG000076223.1), *dapk1* (SMESG000043474.1), *traf-6* (SMESG000000294.1)). Staining for *histone H2B* confirmed an earlier report that found histone mRNAs to be enriched in chromatoid bodies of neoblasts [155] (Fig. 2.19A). However, we were unable to detect the other three transcripts in neoblasts, a fact likely owed to their low expression levels (average expression level 23 TPM) when compared with histone H2B mRNA (22,677 TPM). Instead, we detected strong punctuated staining for all three transcripts in the epidermal cell layer in planarians (Fig. 2.19B). This finding was unexpected, since the expression levels of these transcripts in differentiated cells (Xins) is low (*ank1* = 2.3 TPM, *traf6* = 37 TPM, *dapk1* = 97.7 TPM). To understand the nature of the foci, we quantified their exact localization and found them to be predominantly nuclear. More precisely, they were enriched in the nuclear periphery (Fig. 2.19C). Moreover, using immunostaining on sectioned animals, we could demonstrate that SMEDWI-2 is present in the epidermis, whereas the amount of SMEDWI-3 is negligible (Fig. 2.19D, Fig. 2.20).

We found the observed nuclear foci to be reminiscent of the accumulation of piRNA precursors in Dot COM structures in somatic cells in *Drosophila* [30]. Thus, we suspect that the three tested transcripts might actually be processed just like piRNA precursors in the epidermis. Taken together, our in situ data on group 1 SMEDWI-3 targets demonstrates that planarians possess an epidermal somatic piRNA pathway involving likely only a single PIWI protein, SMEDWI-2. That is in addition to the piRNA pathway, which operates in neoblasts and requires the presence of at least SMEDWI-2 and SMEDWI-3 [145, 133].



**Fig. 2.19** (A) *Histone H2B* (SMESG000052758.1) mRNA (in green) localizes to chromatoid bodies stained with anti-Y12 (in magenta). Nuclei stained with Hoechst are in blue. (B) Expression and localization pattern of three exemplary ping-pong targets of SMEDWI-3 (*dapk1* (SMESG000043474.1), *traf6* (SMESG000000294.1), *ank1* (SMESG000076223.1)) in the planarian epidermis analyzed with whole-mount in situ hybridization (WISH). Nuclei stained with Hoechst are in blue, mRNAs are in yellow. (C) Nuclear localization of foci was calculated as a distance between each in situ signal (in orange) and the nearest nuclear center (x-axis = 0). The nuclear membrane data is plotted in blue. The analysis was performed on WISH signal for *traf6*. (D) Co-immunostaining of SMEDWI-1, -2, and -3 in the epidermis on a cross-section through the planarian pharynx. SMEDWI-1 is in green, SMEDWI-2, and -3 are in red. Nuclei stained with DAPI are in blue.



**Fig. 2.20** Immunostaining of SMEDWI-1 (in green) and SMEDWI-2 (in red) in the epidermis on cross-sections through the pharynx of *S. mediterranea* upon differential RNAi knockdown. Nuclei are stained with DAPI and are shown in blue.

## 2.2 Discussion of Results part 1

The piRNA pathway was discovered in the germline, where it orchestrates germline development and ensures genome integrity by silencing of transposable elements [6, 25, 50, 55]. However, piRNAs were recently found to also have important roles in somatic cells such as neurons, cancer cells, embryonic stem cells and adult somatic stem cells [148, 183]. Somatic piRNAs are widely expressed across arthropods, where they target both transposable elements and protein-coding transcripts. Yet, the best-studied arthropod, *D. melanogaster*, only expresses piRNAs in the germline, thus representing an evolutionary exception in its phylum [100]. Therefore, much remains to be understood about the role of piRNAs in somatic tissues of animals. In this study, we developed an arsenal of biochemical tools that will now make it possible to further dissect piRNA function in the planarian *S. mediterranea* - an organism that neither senesces nor develops stem cell-based diseases like cancer [134, 160].

### 2.2.1 Planarian piRNAs may function in cell cycle regulation and immune defense

Embryonic stem cells are transcriptionally hyperactive. This allows for the expression of repetitive sequences and mobile elements as well as the expression of lineage- and tissue-specific genes at low levels [35]. In addition to their role as a counteracting force to detrimental transposon activity (Figs. 2.8), we show that the planarian PIWI protein SMEDWI-3 utilizes a diverse set of piRNAs to degrade, among others, a multitude of *traf* mRNAs, *histone* mRNAs, mRNAs coding for E3 ubiquitin ligases and transcripts containing ankyrin repeats (Supplemental Table 5 [80]). These results suggest that piRNA-mediated degradation of SMEDWI-3 group 1 targets might be important for planarian regeneration and neoblasts cell cycle control. In addition, a modulation of the histone mRNA metabolism has the potential to directly control cell cycle progression and cell division [114]. Moreover, E2- and E3-ubiquitin ligases are known to regulate the proteolysis of key cell cycle-regulatory proteins, chromosome separation, cytokinesis and cell differentiation [179]. The transcripts of group 2, which show significant SMEDWI-3 binding and high genic piRNA density, although no SMEDWI-3-mediated cleavage, might be degraded into piRNAs using a mechanism distinct from direct cleavage. For example, in the early embryos of *Drosophila*'s during maternal-to-zygotic transition numerous mRNAs undergo Aub-dependent mRNA deadenylation and decay in the soma through recruitment of the CCR4-NOT deadenylation complex [13, 148, 151].

As in other organisms secondary piRNA production operates in chromatoid bodies or the nuage [106], we envision that SMEDWI-3 needs to be localized to chromatoid bodies for the degradation of transposons and mRNAs in the ping-pong cycle (Fig. 2.2A, 2.19A) [76, 106, 155]. In support of this, RNAi knockdown of SMEDWI-1 and SMEDWI-3 leads to the increased expression level of *histone* mRNAs and their delocalization from chromatoid bodies [155]. Apart from SMEDWI-3, planarian chromatoid bodies contain homologs of Tudor proteins [176], helicase VASA [152] and methylated substrates of the protein arginine methyltransferase 5 (PRMT5) [153]. These results strongly suggest that chromatoid bodies are the host of the piRNA pathway.

In addition to the piRNA pathway in neoblasts, we also found an active piRNA pathway in epidermal cells operating with the use of SMEDWI-2 only (Fig. 2.19). As epidermal cells do not divide and rapidly turn over, we hypothesize that the function of the epidermal piRNA pathway goes beyond the maintenance of genome integrity. We propose that piRNAs in the planarian epidermis might play a role in innate immunity, along with the role of the epidermis

in phagocytic cell responses, the secretion of anti-microbial mucus [135] and the response to bacterial infection [10]. The piRNA pathway also plays an important role in epithelium cell metabolism and development. In support, planarian *myb-1*, whose mouse homolog initiates pachytene piRNA production in testes [105], was recently shown to be responsible for the progression of the epidermal lineage, in particular for the specification of the early progeny cell state (*prog-1*) [209]. Moreover, we note that *smedwi-3(RNAi)* treatment leads to skin lesions in planarians on days 13 - 14 post RNAi, well before the animal lyses (21 days post RNAi) [133]. This might be due to reduced SMEDWI-2 protein levels in the epidermis of *smedwi-3(RNAi)* animals (Fig. 2.7, 2.20).

Furthermore, two SMEDWI proteins - SMEDWI-2 and -3 - were found to be expressed in neuronal cells of planarian brain (Fig. 2.2) [133, 170], suggesting the presence of an additional somatic piRNA pathway in planarians operated by two SMEDWIs. Planarians possess a relatively primitive brain composed of glia and multiple neuronal subtypes [136]. However, planarian brain is a highly dynamic structure that undergoes constant neuronal turnover at the level of 25% per week. Upon injury or amputation, the planarian brain can regenerate and functionally reintegrate new tissue in 7 days without scarring [20]. This raises a fascinating question about the role of piRNA pathway in planarian brain metabolism. Especially, since in other organisms piRNAs play a role in the long-term memory formation and axon regeneration [81, 141].

### 2.2.2 The potential role of SMEDWI-3 in determining neoblast mRNA fate

SMEDWI-3, apart from degrading planarian mRNAs by targeting their coding sequences in the ping-pong cycle, specifically binds numerous planarian mRNAs, including its own. The key to these seemingly opposed SMEDWI-3 activities might lie in the presence or absence of antisense piRNAs that exhibit distinct base-pairing patterns when recognizing their mRNA targets (Fig. 2.16). Moreover, piRNAs render it unnecessary to evolve conserved nucleotide binding motifs for each mRNA to be targeted.

We hypothesize that the degradation-independent interaction of SMEDWI-3 with group 3 transcripts might be a sign of transcript binding by SMEDWI-3, analogous to mRNA surveillance in the *C. elegans* germline. There, PRG-1 and CSR-1 act out a functional rivalry to achieve the degradation of transcripts that are rated foreign and to establish an epigenetic memory of this distinction [11, 96, 166, 171]. However, at this point it is not certain whether SMEDWI-3-mediated piRNA-mRNA interactions that do not trigger mRNA degradation

persist long enough to have a biological function in planarians. Whether SMEDWI-3 is involved in recognizing and licensing neoblast mRNAs in the planarian *S. mediterranea* will thus require a thorough study of the piRNA pathway response to the invasion of foreign nucleic acids.

Furthermore, to dissect the molecular mechanism of piRNA-mediated mRNA turnover biochemical studies addressing the function and composition of planarian chromatoid bodies as well as a thorough investigation of the role of the planarian epidermis in immune defense are necessary.



# Chapter 3

## Results part 2

### 3.1 Efficient depletion of ribosomal RNA for RNA sequencing in planarians

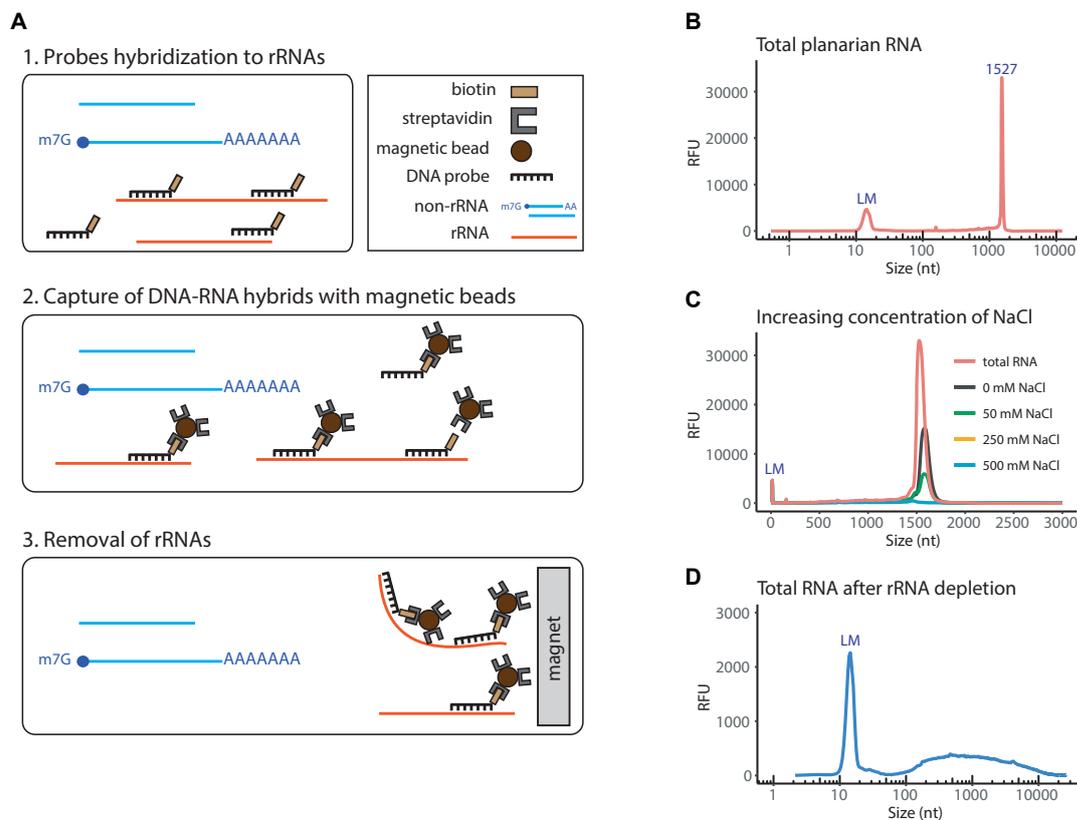
#### Aim and Objective

The aim of this study is to develop a workflow for the efficient depletion of rRNA in the planarian flatworm *S. mediterranea*.

To establish rRNA depletion protocol, we designed 40-mer biotinylated DNA oligonucleotide probes antisense to rRNA. We based our protocol on the hybridization of synthesized probes to rRNA followed by the subtraction of formed DNA-RNA hybrids. Specificity of the developed protocol was tested on total RNA isolated from stem cells of *S. mediterranea*. The expression level of genes in obtained rRNA-depleted libraries was compared with publicly available poly(A)-enriched ones.

#### 3.1.1 Development of an efficient rRNA depletion protocol for planarians

To deplete ribosomal RNA from planarian total RNA, we chose to develop a protocol based on the hybridization of rRNA-specific biotinylated DNA probes to ribosomal RNA and the capture of the resulting biotinylated rRNA-DNA hybrids using streptavidin-coated magnetic beads (Fig. 3.1A).



**Fig. 3.1** (A) Schematic representation of rRNA depletion workflow. Biotinylated DNA probes are hybridized to rRNA, followed by subtraction of DNA-rRNA hybrids using streptavidin-coated magnetic beads. (B) Separation profile of planarian total RNA. The large peak at 1527 nts corresponds to the co-migrating 18S rRNAs and the two fragments of processed 28S rRNA. LM denotes the lower size marker with a length of 15 nts. (C) Increasing concentration of NaCl improves the efficiency of rRNA removal. (D) Total planarian RNA after rRNA depletion.

To that end, we synthesized a pool of 88 3'-biotinylated 40-nt long DNA oligonucleotide probes (siTOOLS Biotech, Martinsried, Germany). We chose probes with a length of 40 nucleotides since their melting temperature in DNA-RNA hybrids was shown to be  $80 \pm 6.4^\circ\text{C}$  in the presence of 500 mM sodium ions [194]. This would allow probe annealing at  $68^\circ\text{C}$  in agreement with generally used hybridization temperatures [54]. The probes were devised in antisense orientation to the following planarian rRNA species: 28S, 18S type I and type II, 16S, 12S, 5S, 5.8S, internal transcribed spacer (ITS) 1 and ITS 2 (Additional file 1).

To assess RNA quality and the efficiency of rRNA removal, we used capillary electrophoresis (Fragment Analyzer, Agilent). The separation profile of total planarian RNA only shows a single rRNA peak at about 1500 nt (Fig. 3.1B). This single rRNA peak is the result of the 28S rRNA being processed into two fragments that co-migrate with the peak of 18S rRNA [178]. Planarian 28S rRNA processing usually entails the removal of a short

sequence located in the D7a expansion segment of 28S rRNA. The length of the removed fragment thereby varies between 4 nt and 350 nt amongst species (e.g. in *Dugesia japonica* 42 nt are removed) [178]. Intriguingly, a similar rRNA maturation process was observed in particular protostomes, in insects such as *D. melanogaster* and in other Platyhelminthes [178, 182, 189]. In addition to the 28S rRNA maturation phenomenon, *S. mediterranea* possesses two 18S rDNA copies that differ in about 8% of their sequence. However, only 18S rRNA type I was reported to be functional and predominantly transcribed [21, 22].

As a first step during rRNA removal all 88 DNA probes were annealed to total planarian RNA. Since RNA molecules are negatively charged, the presence of cations facilitates the annealing of probes to RNA by reducing the repulsion of phosphate groups [33, 90]. Although  $Mg^{2+}$  ions are most effective in stabilizing the tertiary structure of RNA and in promoting the formation of DNA-RNA hybrids, they are also cofactors for multiple RNases [47] and hence should not be included during ribodepletion. Therefore, we tested several hybridization buffers with varying concentrations of sodium ions (Fig. 3.1C). In the absence of sodium ions we could only accomplish an incomplete removal of rRNA. However, hybridization buffers with a sodium concentration  $>250$  mM led to the complete depletion of rRNA from planarian total RNA (Fig. 3.1C, 3.1D). Thus, optimal rRNA removal requires the presence of  $>250$  mM NaCl in the hybridization buffer. As we obtained the most consistent results in the presence of 500 mM NaCl, we decided to utilize this salt concentration in our procedure (Fig. 3.1D).

### 3.1.2 Detailed rRNA depletion workflow

#### Required buffers:

Hybridization buffer (20 mM Tris-HCl (pH 8.0), 1 M NaCl, 2 mM EDTA)

Solution A (100 mM NaOH, 50 mM NaCl, DEPC-treated)

Solution B (100 mM NaCl, DEPC-treated)

2xB&W (Binding&Washing) buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 2 M NaCl)

Dilution buffer (10 mM Tris-HCl (pH 7.5), 200 mM NaCl, 1 mM EDTA)

#### Protocol:

##### 1. RNA input

The following protocol efficiently depletes ribosomal RNA from 100 ng up to 1.5  $\mu$ g of total RNA (Fig. 3.2). The procedure can be scaled up for higher RNA input.

2. Hybridization of biotinylated DNA oligonucleotides (40-mers) to ribosomal RNA (step 1 Fig. 3.1A).

- For oligonucleotide annealing the following reaction is set up:
  - 10  $\mu$ l hybridization buffer
  - 10  $\mu$ l RNA input (1  $\mu$ g)
  - 1  $\mu$ l 100  $\mu$ M biotinylated DNA probes
- Gently mix the solution by pipetting and incubate at 68°C for 10 min.
- Immediately transfer the tubes to 37°C for 30 min.

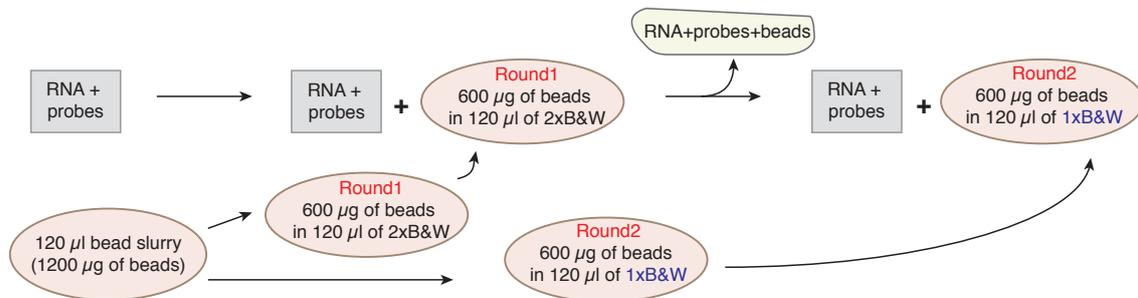
3. Prepare Dynabeads MyOne Streptavidin C1 (Invitrogen) according to the manufacturer's instruction as follows:

- For each sample use 120  $\mu$ l (10  $\mu$ g/ $\mu$ l) of bead slurry.
- Wash the beads twice with an equal volume (or at least 1 ml) of Solution A. Add Solution A and incubate the mixture for 2 min. Then, place the tube on a magnet for 1 min and discard the supernatant.
- Wash the beads once in Solution B. Split the washed beads into two separate tubes for two rounds of subtractive rRNA depletion (Round1 and Round2). Place the beads on a magnet for 1 min and discard Solution B.
- Resuspend the beads for Round1 in 2xB&W buffer to a final concentration of 5  $\mu$ g/ $\mu$ l (twice the original volume). The beads for Round1 will be used during the first round of rRNA depletion. For the second round of depletion, resuspend the beads for Round2 to a final concentration of 5  $\mu$ g/ $\mu$ l in 1xB&W buffer. The beads for Round2 will be used during a second depletion step. Keep the beads at 37°C until use.

4. Capture of DNA-RNA hybrids using magnetic beads (step 2 Fig. 3.1A).

- Briefly spin the tubes containing total RNA and probes. Then, add the following:
  - 100  $\mu$ l dilution buffer
  - 120  $\mu$ l washed magnetic beads (5  $\mu$ g/ $\mu$ l) in 2xB&W (Round1)
- Resuspend by pipetting up and down ten times. The final concentration of NaCl during this step is 1 M. Incubate the solution at 37°C for 15 min. Gently mix the sample by occasional tapping.

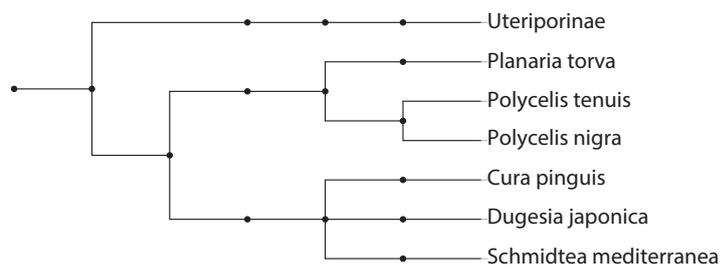
- Place on magnet for 2 min. Carefully remove the supernatant and add it to the additional 120  $\mu\text{l}$  of washed magnetic beads in 1xB&W (Round2). Incubate the mixture at 37°C for 15 min with occasional gentle tapping.
  - Place on magnet for 2 min. Carefully transfer the supernatant into a new tube and place on magnet for another 1 min to remove all traces of magnetic beads from the sample (step 3 Fig. 3.1A).
  - Transfer the supernatant into a fresh tube.
5. Use the RNA Clean & Concentrator-5 kit (Zymo Research) to concentrate the ribodepleted samples, to carry out size selection and to digest any remaining DNA using DNase I treatment as described [206].



**Fig. 3.2** Removal of DNA-rRNA hybrids was performed in two consecutive steps using streptavidin-coated magnetic beads resuspended in 2x of 1x B&W buffer.

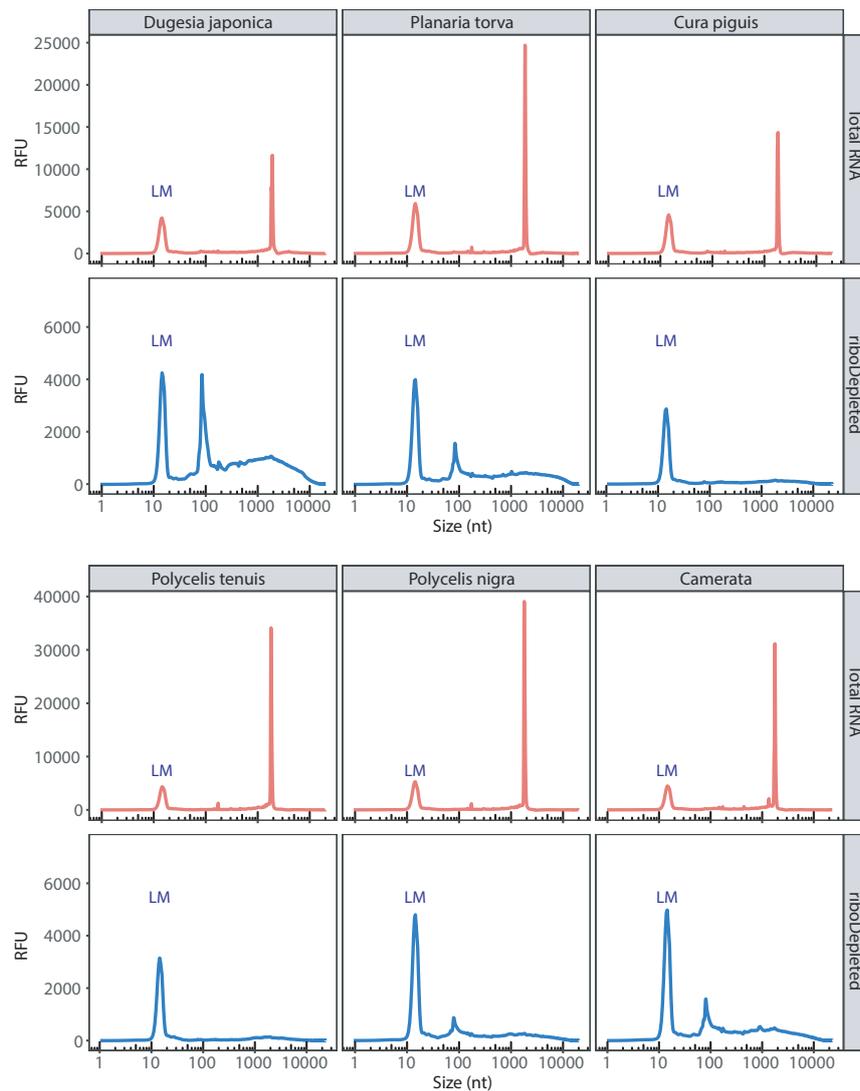
### 3.1.3 Ribosomal RNA depletion in planarian species related to *S. mediterranea*

Ribosomal DNA genes are among the most conserved sequences in all kingdoms of life. They are present in all organisms and are widely used for the construction of phylogenetic trees [197]. The latter is possible because of the low rate of nucleotide substitutions in rRNA sequences (about 1 - 2% substitutions occur per 50 million years based on bacterial 16S rRNA) [129]. The divergence of 18S rRNA sequence between different families of freshwater planarians lays in the range of 6 – 8%, while interspecies diversity does not exceed 4% [22]. Therefore, low rRNA divergence between taxa can be exploited for the design of universal probes for rRNA depletion in different organisms. To assess the specificity and universal applicability of our DNA probes, we depleted rRNA in flatworm species of the order Tricladida, all related to *S. mediterranea* (Fig. 3.3).



**Fig. 3.3** Phylogenetic tree showing the taxonomic position of the analyzed planarian species.

Total RNA separation profiles were analyzed before and after rRNA depletion of six planarian species from three different families. Two of these, *Dugesia japonica* and *Cura pinguis*, belong to the same family as *S. mediterranea*, the Dugesiiidae family. In addition, we examined three species from the family Planariidae (*Planaria torva*, *Polycelis nigra* and *Polycelis tenuis*) and one species from the genus *Camerata* of Uteriporidae (subfamily Uteriporinae). For all tested species our DNA probes proved efficient for the complete removal of rRNA, which migrated close to 2000 nt on all electropherograms (Fig. 3.4). The peak at about 100 nt in the rRNA-depleted samples represents a variety of small RNAs (5S and 5.8S rRNA, tRNAs, and other small RNA fragments), which escaped the size selection aimed at retaining only fragments longer than 200 nt. Taken together, the probes developed for *S. mediterranea* can be utilized for the removal of ribosomal RNA in a multitude of planarian species and may even be generally applicable to all studied planarian species.

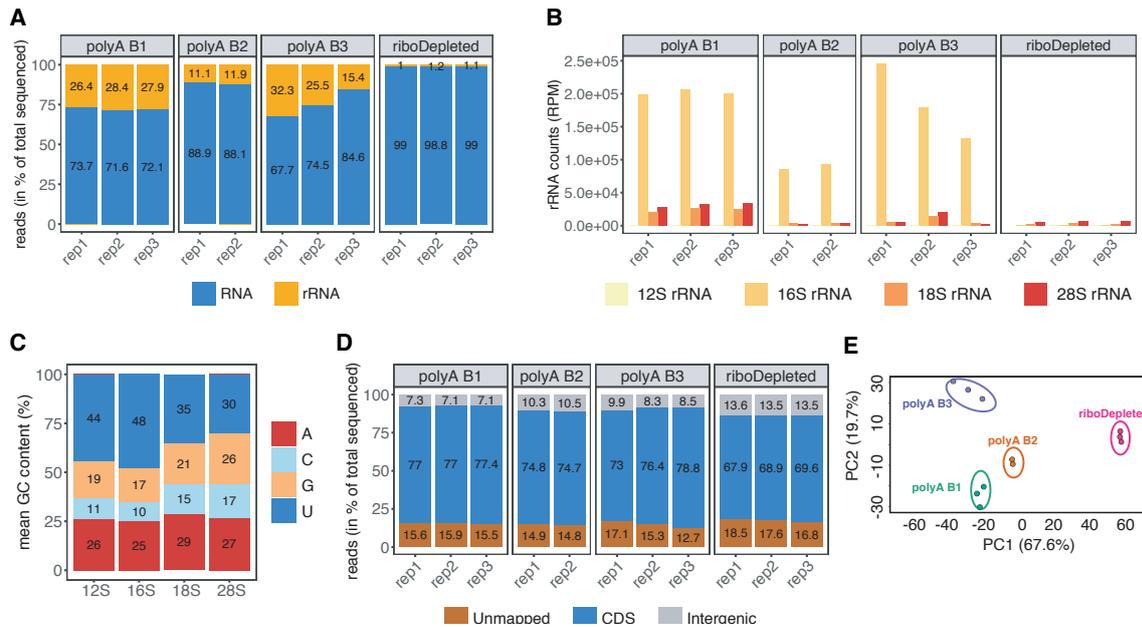


**Fig. 3.4** Total RNA separation profile before and after rRNA depletion. In all species analyzed, the 28S rRNA undergoes "gap deletion" maturation, which results in two co-migrating fragments. Both 28S fragments co-migrate with 18S rRNA, resulting in a single rRNA peak.

### 3.1.4 Comparison of RNA-seq libraries prepared using rRNA-depletion or poly(A) selection

To assess the efficiency of rRNA removal and the specificity of our DNA probes, we prepared and analyzed RNA-seq libraries from rRNA-depleted total RNA from *S. mediterranea*. Total RNA was extracted from 100,000 FACS-sorted planarian neoblasts, resulting in 70 – 100 ng of input RNA. RNA-seq libraries were prepared and sequenced as described [206] following 15 cycles of PCR amplification. The subsequent analysis of sequenced libraries confirmed

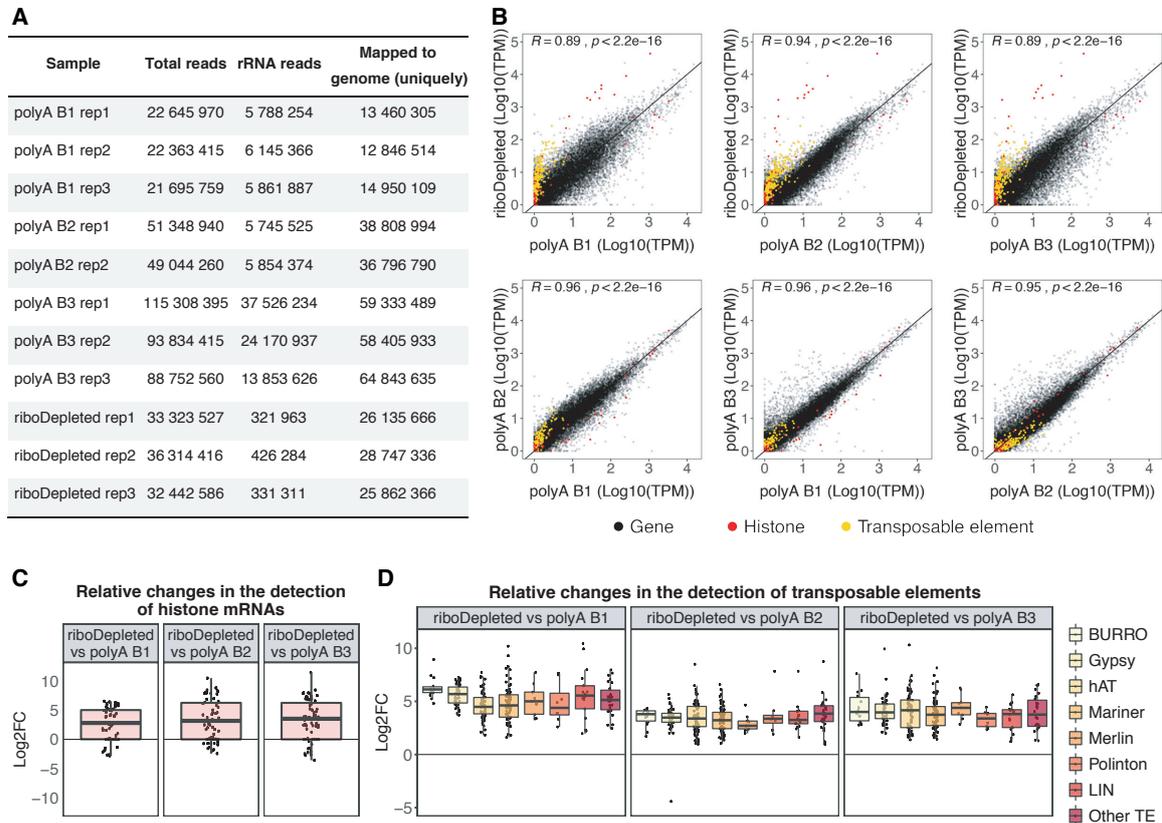
the efficient removal of rRNAs. Less than 2% of total sequenced reads constituted ribosomal RNA (Fig. 3.5A). Next, we compared our rRNA-depleted libraries with three publicly available planarian poly(A) enriched RNA-Seq datasets (poly(A) libraries) [34, 164, 120]. In case publicly available libraries were sequenced in paired-end mode, we analyzed only the first read of every pair to minimize the technical variation between libraries [196].



**Fig. 3.5** (A) Percentage of rRNAs reads in the sequenced libraries prepared from rRNA-depleted or poly(A)-enriched RNA. (B) rRNAs species remaining in the final sequenced libraries. (C) Nucleotide content of planarian rRNA. (D) Percentage of sequenced reads mapped to coding (CDS) and intergenic regions in the planarian genome. (E) Principal component analysis (PCA) biplot of log<sub>2</sub> expression data for coding genes reveals distinct clustering of all analyzed RNA-seq experiments.

As shown in Figure 3.5A, the ribodepleted libraries contained significantly less rRNA compared to all poly(A) enriched ones. Interestingly, the major rRNA species that remained after poly(A) selection was mitochondrial 16S rRNA (Fig. 3.5B). Although the planarian genome has a high A-T content (> 70%) [56], we could not attribute the overrepresentation of 16S rRNA in poly(A) libraries to a high frequency or longer stretches of A nucleotides as compared to other rRNA species (Fig. 3.5C). Moreover, using publicly available planarian poly(A)-position profiling by sequencing (3P-Seq) libraries [89], which allow the identification of 3'-ends of polyadenylated RNAs, no polyadenylation sites were detected in 16S rRNA. Therefore, we speculate that upon folding of 16S rRNA stretches of A nucleotides become exposed and facilitate the interaction with oligo-dT beads during transcript poly(A) selection.

We next assigned the analyzed datasets to the planarian genome. In ribodepleted libraries more than 13% of all mapped reads were assigned to intergenic regions, compared to 7% – 10.5% for poly(A)-enriched ones (Fig. 3.5D).



**Fig. 3.6** (A) Sequencing depth and number of reads mapped to the planarian genome in analyzed rRNA-depleted and poly(A)-enriched samples. (B) Comparison of gene expression in transcripts per million (TPM) between planarian rRNA-depleted and poly(A)-enriched (polyA) RNA-Seq data. The Pearson's correlation coefficient is indicated. (C) Increased representation of histone mRNAs in rRNA-depleted libraries. (D) Boxplot of log<sub>2</sub> fold differences in the expression values of transposable elements between rRNA-depleted and poly(A)-enriched libraries.

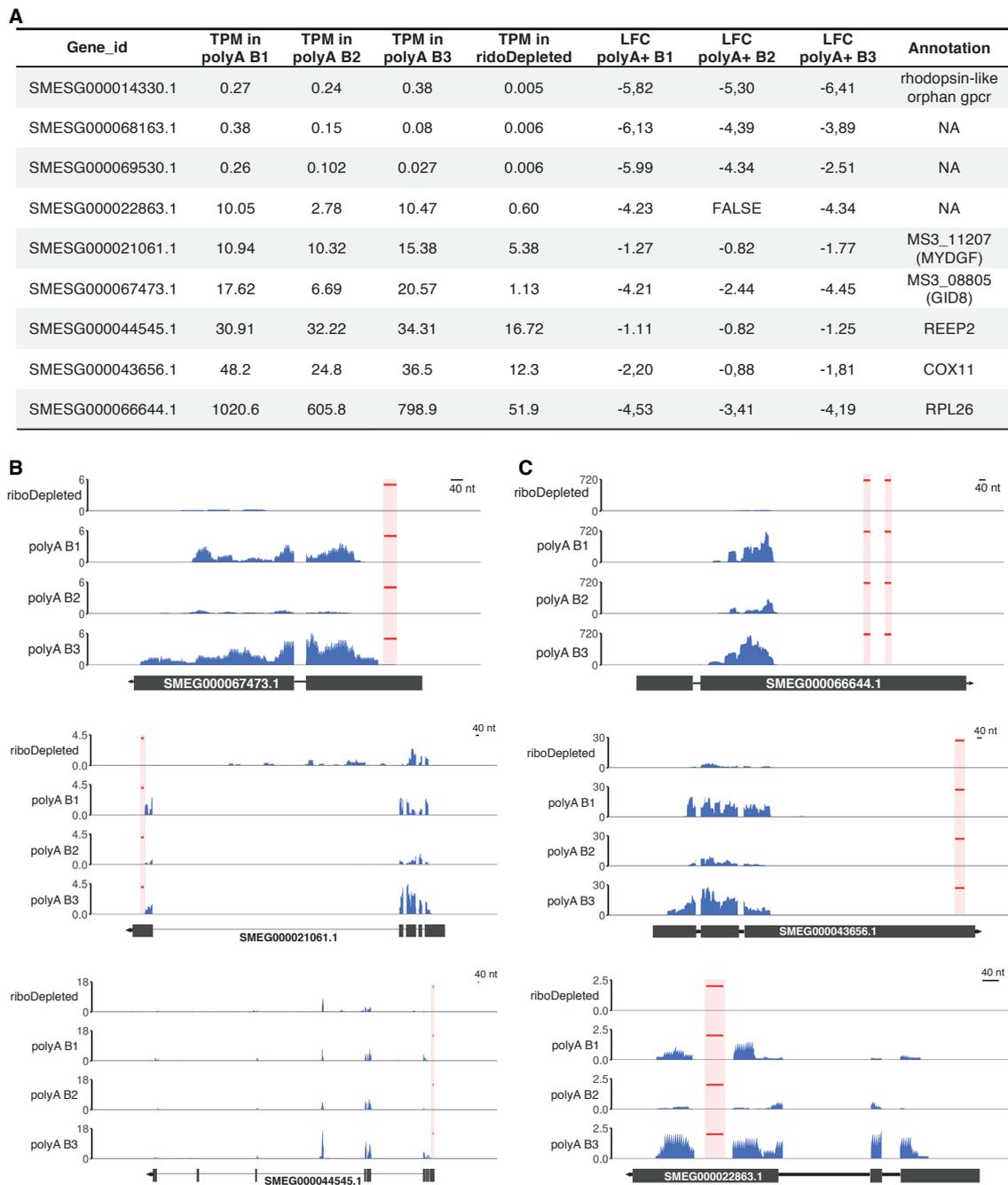
In addition, the percentage of unmapped reads was higher in rRNA-depleted libraries and constituted about 17.6%, which is on average 2.4% more than in poly(A) datasets. We speculate that for rRNA-depleted libraries the proportion of reads mapping to intergenic regions will increase in the future, once complete assemblies of the planarian genome are available. Currently, the planarian genome assembly consists of 481 scaffolds [56]. To detect gene expression variabilities between the analyzed libraries, we performed principal component analysis for the clustering of gene expression data. Although all poly(A) selected libraries were grouped closer together along the PC1 scale, all four analyzed datasets appeared

as separated clusters. This indicates considerable variation even amongst different batches of poly(A) libraries (Fig. 3.5E). One possible source of such variation might be the sequencing depth of the analyzed libraries, which varied considerably from 13 to 64 millions of mapped reads (Fig. 3.6A).

Next, to estimate the correlation between rRNA-depleted and poly(A) libraries, we calculated their Pearson correlation coefficients (Fig. 3.6B). We found the highest Pearson correlation between rRNA-depleted libraries and polyA B2 samples ( $R = 0.94$ ,  $p < 2.2 \times 10^{-16}$ ) (Fig. 3.6B). This could be due to their similar sequencing depth compared to the other polyA libraries. The transcripts, whose abundance was most significantly affected by poly(A) selection, were found to be histone mRNAs that are known to lack polyA tails (Fig. 3.6B, 3.6C) [116]. Their expression level appeared to be 8 – 10 log<sub>2</sub> fold higher in our rRNA-depleted libraries. Moreover, in the rRNA-depleted libraries we also detected significantly higher expression levels for transposable elements (Fig. 3.6B, 3.6D). Out of 316 planarian transposable element families [12], 254 were on average upregulated 5.2, 3.5 and 4.0 log<sub>2</sub> fold as compared to polyA B1, polyA B2 and polyA B3 libraries, respectively (Fig. 3.6D). Moreover, the rRNA-depleted libraries revealed that Burro elements, giant retroelements found in planarian genome [56], gypsy retrotransposons, hAT and Mariner/Tc1 DNA transposons are the most active transposable elements in planarian stem cells. Although some transposable elements are polyadenylated, long-terminal repeat elements (LTRs) lack poly(A)-tails [79]. This renders their detection in poly(A)-enriched sample non-quantitative.

### **3.1.5 Non-specific depletion of coding transcripts in ribodepleted libraries**

In using custom rRNA-depletion probes, our major concern was that the utilized probes would lead to unspecific co-depletion of planarian coding transcripts. To exclude this possibility, we first mapped our pool of 88 DNA probes in antisense orientation to the planarian transcriptome allowing up to 8 mismatches and gaps of up to 3 nt. This mapping strategy requires at least 75% of a DNA probe to anneal to its RNA target. It resulted in only 11 planarian genes to be potentially recognized by 20 DNA probes from our oligonucleotide pool. Next, we carried out a differential expression analysis of these 11 potentially targeted transcripts between the ribodepleted libraries and poly(A)-selected ones. The analysis revealed that 9 out of 11 potential targets were downregulated at least 1-fold in at least two poly(A) experiments (Fig. 3.7A).

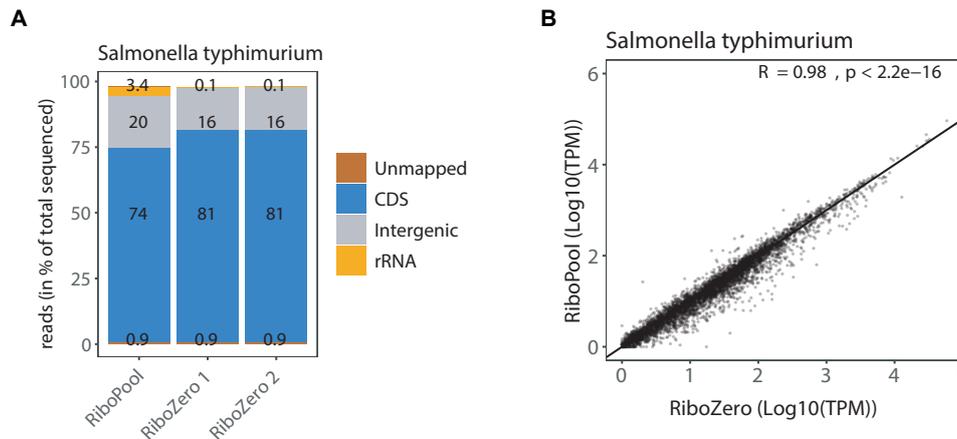


**Fig. 3.7** (A) Expression levels in TPM (transcripts per million) of nine transcripts targeted by probes utilized for ribodepletion. LFC stands for the log<sub>2</sub> fold difference in the expression level of individual transcripts between ribodepleted and poly(A) enriched libraires. (B) RNA-seq coverage profile for SMESG000067473.1, SMESG000021061.1 and SMESG000044545.1 in rRNA-depleted (riboDepleted) and poly(A) enriched (polyA B1, polyA B2, polyA B3) libraries. The location of antisense probes mapping to the transcripts is marked in red. (C) The same as in (B) for SMESG000066644.1, SMESG000043656.1 and SMESG000022863.1.

As the abundance of three transcripts (SMESG000014330.1 (rhodopsin-like orphan gpcr [159]), SMESG000068163.1 and SMESG000069530.1 (both without annotation)) was very low in all polyA libraries ( $<0.6$  transcripts per million (TPM)), we did not consider these any further. However, the remaining six transcripts were found to be significantly down-regulated in ribodepleted libraries. For three of these targeted genes (SMESG000067473.1, SMESG000021061.1 and SMESG000044545.1) the probes map in regions that display significant RNA-seq coverage (Fig. 3.7B). Therefore, their lower expression values in ribodepleted libraries is likely attributed to probe targeting. Intriguingly, for the remaining three targets (SMESG000066644.1, SMESG000043656.1 and SMESG000022863.1 annotated as *RPL26* (ribosomal protein L26), *COX11* (cytochrome c oxidase copper chaperone) and an unknown transcript, respectively) the probes were predicted to map to loci that do not exhibit RNA-seq coverage (Fig. 3.7C). The likely reason for this is inaccurate gene annotation. Alternatively, target regions might represent repetitive, multimapping sequences, which we excluded during read mapping. Taken together, our off-target analysis revealed that a maximum of 11 genes might be affected by our rRNA removal procedure - a very low number that underscores the specificity and efficiency of our depletion protocol.

### 3.1.6 Applicability of the described ribodepletion method to other organisms

To demonstrate the applicability of the developed rRNA workflow to other organisms, we employed our protocol to the depletion of ribosomal RNA from *Salmonella typhimurium* using a pool of organism-specific DNA probes (riboPOOL) developed by siTOOLS Biotech (Martinsried, Germany) (Fig. 3.8A). We compared the libraries resulting from the application of our newly developed procedure to the established rRNA depletion workflow that utilizes the Ribo-Zero rRNA Removal Kit (Bacteria) from Illumina. Removal of rRNA from a *S. typhimurium* sample using riboPOOL probes was as successful as a depletion reaction using Ribo-Zero, leaving as low as 3.4% rRNA in the final library (Fig. 3.8A). Moreover, an overall comparison of gene expression levels showed a high correlation (Pearson correlation  $R = 0.98$ ,  $p < 2.2e-16$ ) between riboPOOL depleted libraries and libraries prepared with the Ribo-Zero kit (Fig. 3.8B). Taken together, the rRNA depletion workflow described in this manuscript is robust and easily applicable to any bacterial and eukaryotic species of choice utilizing organism-specific probes.



**Fig. 3.8 (A)** Percentage of rRNA in sequenced libraries from *Salmonella typhimurium*. Libraries were prepared using our developed rRNA depletion workflow with organism-specific riboPOOL probes (siTOOLS Biotech) or the commercially available Ribo-Zero kit (Illumina). **(B)** Scatter plot comparing transcript abundance (TPM) between ribodepleted libraries using our developed workflow and the commercial Ribo-Zero kit. The Pearson's correlation coefficient is indicated.

## 3.2 Discussion part 2

For samples from typical model organisms, such as human, mouse and rat, there are numerous commercial kits available for the removal of rRNA, e.g. NEBNext from New England Biolabs, RiboGone from Takara and RiboCop from Lexogen. This also applies to typical gram-positive and gram-negative bacteria (MICROBExpress from ThermoFisher and Ribominus from Invitrogen). Moreover, these kits can be utilized with a certain degree of compatibility for the depletion of rRNA in organisms of distinct phylogenetic groups (e.g. RiboMinus Eukaryote Kit for RNA-Seq, Invitrogen). However, as the breadth of molecularly tractable organisms has increased in the past decade, the necessity to develop organism-specific rRNA depletion techniques has risen as well [31, 51, 158]. To date, custom protocols either use biotinylated antisense probes along with streptavidin-coated magnetic beads for rRNA removal or rely on the digestion of DNA-RNA hybrids with RNase H [88, 131, 104, 125].

In this study, we describe a novel rRNA depletion workflow for the planarian flatworm *S. mediterranea*. Our protocol is based on the hybridization of biotinylated DNA probes to planarian rRNA followed by the subsequent removal of the resulting rRNA-DNA hybrids by using streptavidin-labeled magnetic beads. We tested the efficiency and specificity of our protocol by depleting rRNA from total RNA of neoblasts, planarian adult stem cells. A comparative analysis between rRNA-depleted and poly(A)-selected libraries revealed that our protocol retains all information present in poly(A) selected libraries. Over and

above, we found ribodepleted libraries to contain additional information on histone mRNAs and transposable elements. The abundance of histone mRNAs in neoblasts is not unexpected, as planarian neoblasts are the only dividing cells in adult animals and thus require histones for packaging newly synthesized DNA [115, 143]. The high expression values of transposable elements likely reflects our ability to detect both non-poly(A) transcripts and degradation products of transposable elements generated by PIWI proteins loaded with transposon-specific piRNAs [170, 80]. Planarian PIWI proteins and their co-bound piRNAs are abundant in neoblasts and essential for planarian regeneration and animal homeostasis [145, 170, 80, 133]. Using our rRNA depletion protocol, we are now able to estimate the actual abundance of transposons and other repeats in planarians. This is important as these transcripts are generated from a large fraction of planarian genome (about 62% of the planarian genome comprise repeats and transposable elements) [56]. In addition, the planarian PIWI protein SMEDWI-3 is also involved in the degradation of multiple protein-coding transcripts in neoblasts (was shown in Chapter 2). Such mRNA degradation processes complicate the analysis of mRNA turnover using poly(A) enriched libraries, as these only represent mRNA steady-state levels. To study dynamic changes in mRNA levels is especially intriguing during neoblasts differentiation, as then the steady-state levels of numerous mRNAs are changing [174, 109]. Using our rRNA-depletion protocol, we can now determine whether mRNA expression changes are due to altered transcription rates or due to increased degradation. Taken together, rRNA-depleted RNA-seq libraries are particularly valuable for the investigation of piRNA pathway and RNA degradation processes, since they retain the dynamics inherent to cellular RNA metabolism. Furthermore, by successfully depleting rRNA from other freshwater triclad species, we could demonstrate the versatility of the DNA probes designed for *S. mediterranea*. Last, we validated the efficiency of the developed workflow by removal of rRNA in the gram-negative bacterium *S. typhimurium*. Therefore, the proposed workflow likely serves as an efficient and cost-effective method for rRNA depletion in any organism of interest.

# Chapter 4

## Conclusion

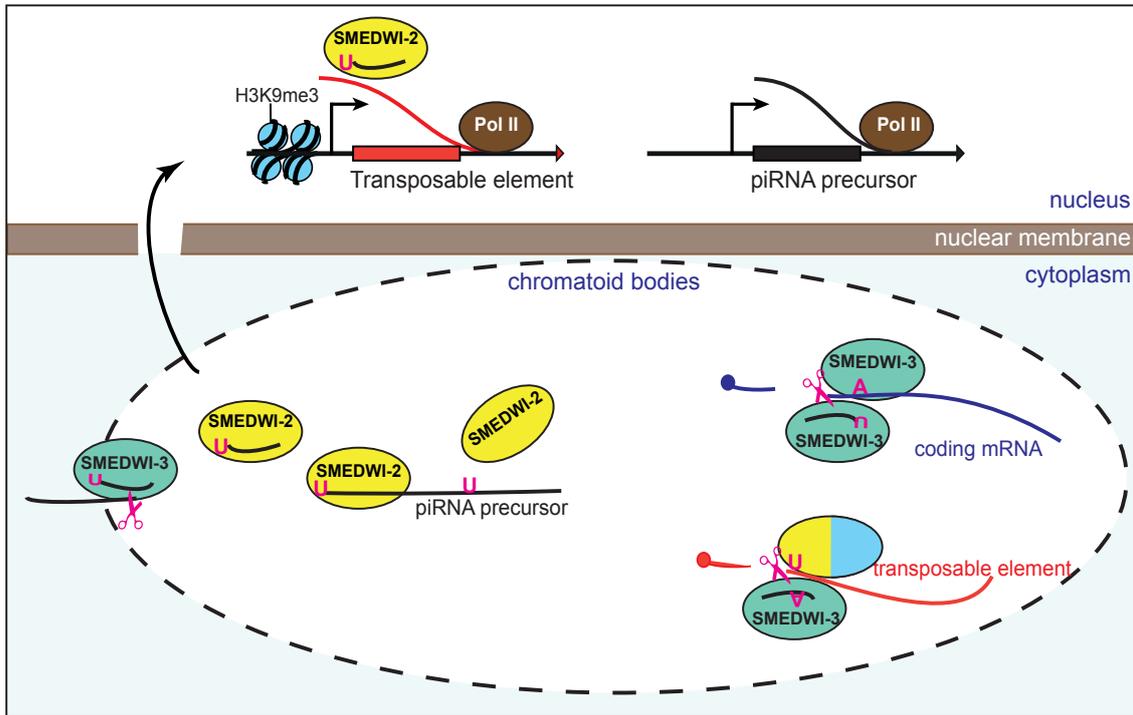
This study marks a significant step forward in understanding the role of piRNAs and PIWI proteins in planarian adult stem cells. The presented data reveal the general organization principles of the piRNA pathway in planarian flatworms and it uncovers the involvement of planarian piRNAs in mRNA surveillance.

The main findings of this study are summarized below (Fig. 4.1):

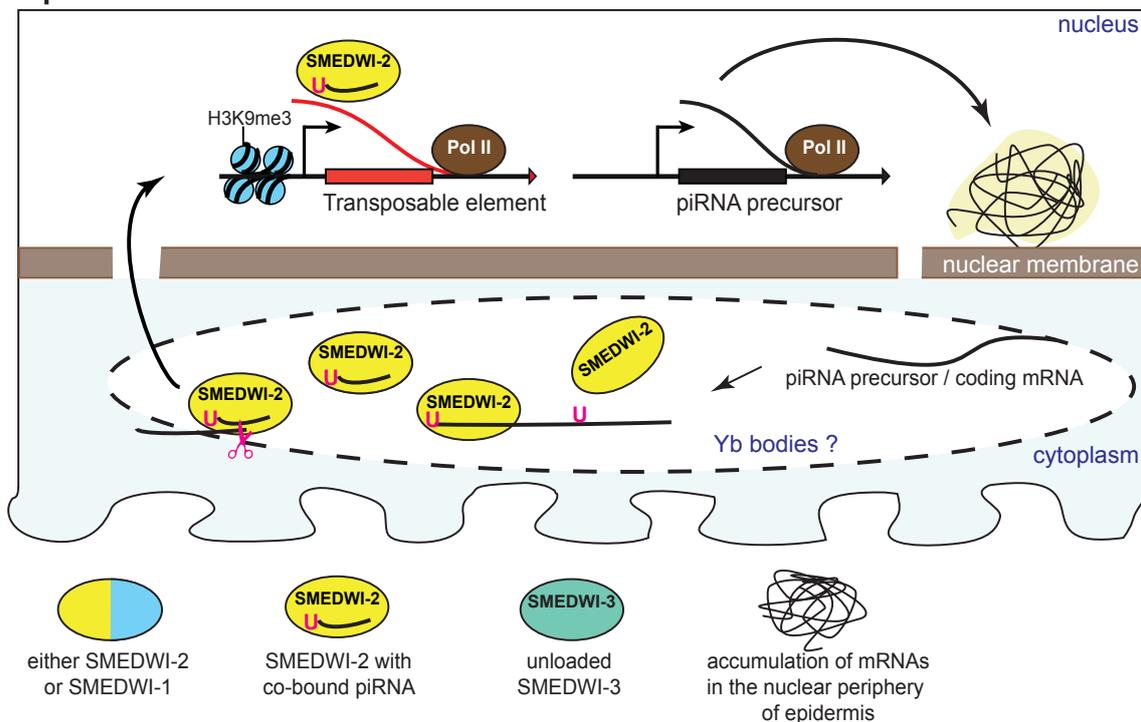
- The piRNA pathway regulates the expression of transposable elements at the level of transcription. In addition, post-transcriptional regulation is ensured by the degradation of TEs in a heterotypic ping-pong cycle operated predominantly by SMEDWI-2 and SMEDWI-3;
- SMEDWI-3 degrades a class of mRNA transcripts in a homotypic ping-pong cycle in adult stem cells;
- SMEDWI-3 also binds to another group of transcripts without triggering their degradation into piRNAs;
- The degree of complementarity between the piRNA guide and its target defines whether SMEDWI-3 binds or degrades targeted transcript;
- SMEDWI-3 ping-pong targets are present in nuclear foci in planarian epithelial cells along with the SMEDWI-2 protein. This indicates the presence of an active piRNA pathway in the planarian epidermis;
- Numerous systems biology techniques, such as Degradome-Seq, RNA-Seq, small RNA-Seq, HITS-CLIP and CLASH were successfully adapted to planarian flatworms;

- A ribosomal RNA depletion workflow was developed for planarian flatworms. It preserves the dynamic information of transcriptomes and facilitates the investigation of the piRNA pathway, in particular, and processes such as RNA degradation and decay, in general.

## Neoblasts



## Epidermis



**Fig. 4.1** Schematic representation of planarian piRNA pathway. In neoblasts piRNA pathway operates with the use of three SMEDWI proteins, ensuring the silencing of transposable elements and numerous mRNAs of coding genes. Epidermal piRNA pathway acts only by means of SMEDWI-2. The group 1 SMEDWI-3 ping-pong targets concentrate at the nuclear periphery of epidermal cells forming the foci reminiscent of the accumulation of piRNA precursor in nuclear structures of *Drosophila's* follicle cells.



# Chapter 5

## Experimental procedures

### 5.1 Experimental procedures of Results part 1

#### 5.1.1 Chemicals reagents

Reagents used in this study were purchased from Carl Roth GmbH & Co (Karlsruhe, Germany) unless specified otherwise in the text.

#### 5.1.2 Primer's sequences

All primer sequences used in the study can be found in Appendix A.

#### 5.1.3 Planarian culture

Asexual *Schmidtea mediterranea*, clonal line CIW4 [4], were maintained at 20°C in Montjuic solution (1.6 mM NaCl, 1.0 mM CaCl<sub>2</sub>, 1.0 mM MgSO<sub>4</sub>, 0.1 mM MgCl<sub>2</sub>, 0.1 mM KCl, 1.2 mM NaHCO<sub>3</sub>, pH 7.0) in the presence of 50 µg/ml gentamycin [23]. Animals were fed weekly with homogenized calf liver and starved at least 7 days prior to experiments.

#### 5.1.4 Antibodies production and purification

The anti-SMEDWI-1 antibody specific to peptide NEPEGPTETDQSLS [59] was a kind gift of Dr. K. Bartscherer (Hubrecht Institute, Utrecht, Netherlands).

The anti-SMEDWI-2 antibody was generated by immunizing rabbits with the peptide KKDEEGVEKEK (BioGenes, Berlin, Germany). Rabbit antiserum was used for all experiments.

For the production of a polyclonal anti-SMEDWI-3 antibody, the N-terminal 200 amino acids of SMEDWI-3 were fused to GST expressed in *Escherichia coli* strain *Rosetta DE3* and then used for immunization (BioGenes, Berlin, Germany). Obtained SMEDWI-3 antiserum was further purified against the antigen coupled to NHS-activated sepharose (GE Healthcare) according to manufacturer's instruction.

### 5.1.5 Expression of Strep-tagged SMEDWI

Full-length SMEDWI-1, SMEDWI-2 and SMEDWI-3, each carrying a Strep-SUMO-Star tag, were expressed in High Five insect cells. Following 72 hours of protein expression, cells were pelleted, snap-frozen in liquid nitrogen and homogenized in RIPA buffer (25 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1% NP-40, 0.5% Sodium deoxycholate, 0.1% SDS). Primers used for the cloning of full-length genes can be found in Appendix A.2.

### 5.1.6 Western blotting

For protein sample preparation, 3 - 5 snap-frozen worms (5 - 7 mm) were homogenized in 50  $\mu$ l lysis buffer (30 mM HEPES (pH 7.7) (Sigma, H989), 150 mM NaCl, 10 mM KCl, 4 mM MgCl<sub>2</sub>, 1 mM DTT, 0.5% Triton X-100, Complete EDTA-free protease inhibitor (Roche, 4693159001)) using a micropestle. The lysate was then cleared by centrifugation at 21,130 g for 15 min at 4°C. Protein concentration of the extracts was measured using Bradford assay. In total, 15 - 20  $\mu$ g of whole cell extract was loaded per lane and separated on 8% Bis-Tris gel before transfer onto nitrocellulose membrane. In total, 15 - 20  $\mu$ g of protein extract was loaded per lane. Antibodies for Western blots were used with the following dilutions:

rabbit polyclonal anti-SMEDWI-1	1:2,000;
rabbit antiserum anti-SMEDWI-2	1:500;
rabbit polyclonal anti-SMEDWI-3	1:4,000;
mouse monoclonal anti-Strep-tag (iba, 2-1207-001)	1:1,000;
mouse monoclonal anti- $\alpha$ -Tubulin (Sigma, T5168)	1:4,000;
goat anti-rabbit IgG-peroxidase (Sigma, A0545)	1:20,000;
goat anti-mouse IgG-peroxidase (Sigma, A9917)	1:20,000.

### 5.1.7 RNA interference

cDNA fragments for *smedwi-1* (SMESG000036375.1) (nucleotides (nts) 1579 - 2555), *smedwi-2* (SMESG000069984.1) (nt 81 - 975) and *smedwi-3* (SMESG000081970.1) (nt

1347 - 2332) were cloned into the pPR-T4P vector (kind gift of Dr. J. Rink, Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany) using SLIC-cloning [103]. *Unc-22* dsRNA from *C. elegans* comprising nt 6616 - 7358 was used as a control. dsRNA synthesis and RNAi feeding experiments were carried out as described [154] with the following modifications. In vitro transcription reaction was set in the presence of 8% PEG 8000. Synthesized dsRNA was DNase I treated (Roche, 04536282001) and sodium acetate/ethanol precipitated. The detailed protocol can be found in Appendix B.1. Primers used for cloning gene fragments into pPR-T4P vector and for in vitro transcription template preparation can be found in Appendix A.4 and Appendix A.3.

For *smedwi-2* RNAi experiments animals were fed twice (0 dpf (days post feeding) and 3 dpf) and sacrificed on day 7 post feeding. For *smedwi-1* and *smedwi-3* RNAi experiments animals were fed three times (0 dpf, 3 dpf and 6 dpf) and sacrificed on day 11 post feeding.

### 5.1.8 Flow cytometry

FACS sorting of X1 and Xins cells was performed by Dr. Elizabeth M. Duncan (University of Kentucky, Lexington, USA) as described [34].

### 5.1.9 Slide-section immunofluorescence

Slide-section immunofluorescence was performed by rinsing animals 3 times in planarian water, incubating in 5% N-acetyl-cysteine in 1x PBS for 5 min with light hand-rolling of tubes and then fixing in 4% PFA in 0.5x PBS overnight at 4°C with no rocking. Then dehydrated into ethanol by 30%, 50%, 70% dilutions in water and samples were stored for a minimum overnight and maximum 7 days at 4°C prior to embedding. Samples were automatically infiltrated with paraffin (Thermoscientific, Richard-Allan Type 9) using a Delta Pathos processor (Milestone Medical, Sorisole Italy). The automation included a 4 min rinse in 85% ethanol, 100% ethanol for 25 min at 65°C, 100% isopropanol for 35 min at 68°C, 2 changes of paraffin at 70°C each, 8 min and then 6 min (Tissue Infiltration Medium: Surgipath through Leica). Following a final step of 20 min in 65°C paraffin, the tissue was embedded in paraffin for transverse sectioning onto Superfrost Plus slides (Fisher Scientific) at 5 microns using a Leica RM2255 microtome (Leica Biosystems Inc. Buffalo Grove, IL). Slides were deparaffinized by incubating in 100% Xylene for 3 min (repeated twice), 100% ethanol for 3 minutes (repeated twice), 80% ethanol rinse and a water rinse. Slides were fixed with 4% PFA for 3 hours at room temperature (RT) and washed with PBS three times. Antigen retrieval was performed in citrate buffer heated to 95°C for 15 min,

cooled down for 20 min and then washed in TBST for 30 min. Slides were blocked with 4 drops of Background Buster (Innovex Biosciences) and incubated for 30 min at RT. Primary antibody was applied and incubated over night at 4°C. The following dilutions were used: anti-SMEDWI-1 at 1:200, anti-SMEDWI-3 at 1:4000 and anti-Y12 at 1:500. Slides were washed with TBST three times and secondary Alexafluor antibodies (647, 488) were applied at 1:500 at RT for 1 hour. Slides were washed three times with TBST and mounted using Prolong Gold Mounting media (Thermofisher). Sections were imaged using a Zeiss 710 with a Nikon 20x Plan-Apo 0.8 NA objective.

### 5.1.10 Whole-mount in situ hybridization

Whole-mount in situ hybridization were performed as described [34]. Investigated transcripts have the following gene identifiers: *ank1* (SMESG000076223.1), *dapk1* (SMESG000043474.1), *traf6* (SMESG000000294.1), *histone H2B* (SMESG000052758.1) [156].

### 5.1.11 Immunoprecipitation of SMEDWI proteins

Approximately 70 worms (7 - 10 mm) were collected, snap-frozen in liquid nitrogen and homogenized in 4 ml lysis buffer (30 mM HEPES (pH 7.7) (Sigma, H989), 150 mM NaCl, 10 mM KCl, 4 mM MgCl<sub>2</sub>, 1 mM DTT, 0.5% Triton X-100, Complete EDTA-free protease inhibitor (Roche, 4693159001)) with a Dounce homogenizer. The planarian lysate was cleared by centrifugation at 50,000 g for 30 min at 4°C, followed by lysate filtration through a 0.20 μm cellulose acetate syringe filter (LLG labware, 9.055501). The protein concentration of the resulted lysate was about 3 mg/ml. For each immunoprecipitation experiment 1 ml of cleared lysate was incubated with 10 μg of purified antibody or 30 μl of antiserum for 2 hours at 4°C with gentle rotation. Pre-immune serum was used as a negative control. Next, 50 μl pre-equilibrated Dynabeads Protein A (Invitrogen, 10002D) were added to each sample and the incubation was continued for an additional 2 hours at 4°C with gentle rotation. Beads were washed twice with low salt buffer (30 mM HEPES (pH 7.7), 150 mM NaCl, 10 mM KCl, 4 mM MgCl<sub>2</sub>, 5 mM EDTA, 1 mM DTT, 0.1% Triton X-100) and twice with high salt buffer (30 mM HEPES (pH 7.7), 300 mM NaCl, 10 mM KCl, 4 mM MgCl<sub>2</sub>, 5 mM EDTA, 1 mM DTT, 0.1% Triton X-100). Next, beads were transferred to a new tube and resuspended in 200 μl of Proteinase K buffer (200 mM Tris-HCl (pH 7.5) 300 mM NaCl, 25 mM EDTA, 1.5% SDS) containing 120 μg/ml Proteinase K (Roche, 03115879001). To extract piRNAs that co-immunoprecipitated with SMEDWI proteins, the beads were incubated with Proteinase K for 20 min at 42°C and 1000 rpm, followed

by phenol-chloroform extraction and ethanol precipitation of all isolated nucleic acids. For radioactive 5'-labeling the extracted nucleic acids were then dephosphorylated with calf intestine phosphatase (NEB, M0290S), isolated with peqGOLD Trifast (peqlab, 30-2010) and labeled with [ $\gamma$ - $^{32}\text{P}$ ]-ATP (PerkinElmer, NEG002A250UC) using T4 polynucleotide kinase (NEB, M0201S) according to the manufacturer's protocol. Labeled RNA was resolved on a 10% denaturing urea polyacrylamide gel. The detailed protocol can also be found in Appendix B.2.

### 5.1.12 Small RNA library preparation

Small RNA libraries were constructed as described [63] with some modifications. Briefly, total small RNA from 100,000 FACS-sorted X1 and Xins cells, or RNA co-immunoprecipitated by SMEDWI-1, SMEDWI-2, or SMEDWI-3 were used for small RNA cloning. Pre-adenylated 3'-adapters were ligated to the 3'-end of all RNAs by a truncated T4 RNA Ligase 2 (NEB, M0351S) in the presence of 10% PEG 8000 at 16°C overnight. The addition of PEG 8000 considerably improved the ligation efficiency of 3'-adapter to piRNAs. piRNAs carry 2'-O-methyl group on their 3'-end, which generally hampers adapter ligation. Next, T4 RNA Ligase 1 was used to add an RNA adapter to the 5'-end of all RNAs at 37°C for 1 hour. Reverse transcription of the resulted product was performed with SuperScript III (Invitrogen, 18080085) followed by PCR amplification, gel purification and size selection. Deep sequencing was performed on Illumina HiSeq1000 (single-end 50 nt) or Illumina NextSeq 500 (single-end 75 nt) platforms. The detailed protocol can be found in Appendix B.3. Primer and adapter sequences used for library preparation can be found in Appendix A.5.

### 5.1.13 Ribosomal RNA depletion

Ribosomal RNA depletion was performed as described in Chapter 3. The complete protocol can be found in Appendix B.4.

### 5.1.14 RNA library preparation

RNA libraries from 100,000 FACS-sorted planarian X1 and Xins cells were constructed as described [206]. Next generation sequencing was carried out on an Illumina Next-Seq 500 platform (paired-end 75 nt or single-end 75 nt mode).

### 5.1.15 SMEDWI-3 HITS-CLIP

SMEDWI-3 HITS-CLIP libraries were prepared as described [185] with the following modifications: Approximately 50 worms (7 - 10 mm) were dissociated in cold calcium- and magnesium-free buffer with 1% BSA (CMFB). Dissociated cells were then filtered, pelleted, and resuspended in cold CMFB buffer. Next, cells were irradiated at 254 nm using a Stratalinker 1800 (Stratagene) once with 400 mJ/cm<sup>2</sup> and then again after 30 sec at 200 mJ/cm<sup>2</sup> [124]. Then, cells were pelleted, snap-frozen in liquid nitrogen and resuspended in lysis buffer (30 mM HEPES (pH 7.7) (Sigma, H989), 150 mM NaCl, 10 mM KCl, 4 mM MgCl<sub>2</sub>, 1 mM DTT, 0.5% Triton X-100, 0.8 U/μl RNasin (Promega, N2515), Complete EDTA-free protease inhibitor (Roche, 03115879001)). After treatment with DNase I (Roche, 04536282001) the whole cell extract was centrifuged at 50,000 g for 30 min at 4°C. To immunoprecipitate SMEDWI-3 with crosslinked RNA, lysate was incubated with 18 μg of anti-SMEDWI-3 antibody for 2 hours at 4°C with gentle rotation, followed by incubation with 150 μl Dynabeads Protein A slurry (Invitrogen, 10002D) for another 2 hours at 4°C with gentle rotation. Next, beads were washed twice with low salt buffer (30 mM HEPES (pH 7.7), 150 mM NaCl, 10 mM KCl, 4 mM MgCl<sub>2</sub>, 5 mM EDTA, 1 mM DTT, 0.1% Triton X-100), twice with high salt buffer (30 mM HEPES (pH 7.7), 500 mM NaCl, 10 mM KCl, 4 mM MgCl<sub>2</sub>, 5 mM EDTA, 1 mM DTT, 0.1% Triton X-100) and twice with 1x PNK buffer (70 mM Tris-HCl (pH 7.5), 10 mM MgCl<sub>2</sub>, 5 mM DTT, 0.1% Triton). Following 3'-adapter ligation RNA-protein complexes were eluted from the beads, separated on an 8% Bis-Tris gel and transferred onto a nitrocellulose membrane (GE Healthcare, 10600012). RNA-protein complexes were then excised from the nitrocellulose membrane and crosslinked RNA was extracted using Proteinase K (Roche, 03115879001), followed by 5'-adapter ligation and DNase I treatment. Ligated RNA was then reverse transcribed and PCR-amplified in two steps as described [185]. Size-specific libraries were extracted excised from a 6% urea PAGE, collected using Spin-X filter tubes (Costar, CLS8163) and precipitated with ethanol. Prepared libraries were sequenced on an Illumina NextSeq 500 platform (single-end 75 nt). The detailed protocol can be found in Appendix B.5. Primer and adapter sequences can be found in Appendix A.6.

### 5.1.16 SMEDWI-3 CLASH

CLASH libraries were prepared as described for SMEDWI-3 HITS-CLIP with slight modifications [169]. After immunoprecipitation of crosslinked RNA-SMEDWI-3 complexes, samples were treated with RNase T1 (0.1 U in 500 μl of lysis buffer) for 5 min at 20°C. Next,

samples were washed once with low salt buffer, once with high salt buffer and three times with 1x PNK buffer. To phosphorylate the mRNA 5'-ends for chimera formation and discard possible DNA contaminants, sample were treated with DNase I and PNK 3'-phosphatase minus. Chimera ligation was conducted overnight in the presence of T4 RNA ligase 1 (1U/ $\mu$ l) and 10% PEG 8000. Next, 3'-ends of crosslinked transcripts were dephosphorylated with Antarctic phosphatase. Adapter ligation, library preparation and sequencing were carried out as described for SMEDWI-3 HITS-CLIP.

### **5.1.17 Degradome library preparation**

Degradome libraries (5'-monophosphate-dependent cloning of mRNA fragments) were prepared from 2.5  $\mu$ g total RNA of 1.8 million X1 FACS sorted cells as described [188]. Next generation sequencing was carried out on an Illumina Next-Seq 500 platform (single-end 75 nt mode). The detailed protocol can be found in Appendix B.6. Primer and adapter sequences can be found in Appendix A.7.

### **5.1.18 Processing of small RNA libraries**

Statistical and graphical analysis were performed with R/Bioconductor [48]. To process small RNA-seq data, 3'-adapters were trimmed using cutadapt [113] and filtered against planarian rRNA sequences with SortMeRNA [85]. Only sequences longer than 18 nt and shorter than 40 nt were kept for the following analysis. Processed reads were mapped against planarian genome [56] by using bowtie (1.1.2) [91] with the following settings “-v 0 -a -best -strata”. The resulting alignment files were converted into bed2 format with piPipes [61]. To account for multi-mapping reads, the total read count was divided by the number of locations that they map to. These weighted counts were used to characterize piRNAs mapping to genomic features with BEDTools (2.25.0) [140]. To compare the abundance of mRNA-derived piRNAs associated with SMEDWI-1, SMEDWI-2 and SMEDWI-3 and to compute the small RNA coverage of target genes, piRNAs were normalized to RPM sequences, and only uniquely mapping piRNAs were considered. The detailed pipeline can be found in Appendix C.1.

### **5.1.19 Calculation of the ping-pong signature**

Homotypic and heterotypic ping-pong signatures were calculated with piPipes [61].

### 5.1.20 Sequence logos and Venn diagrams

To generate sequence logos with seqLogo (1.46.0) piRNA sequences mapped to the reference genome were collapsed to avoid biases due to the presence of highly abundant piRNAs. Only mapped piRNAs that appear in at least two replicates were used to generate Venn diagrams with the eulerr package (4.1.0).

### 5.1.21 piRNA cluster prediction

proTRAC version 2.4.2 [149] was used to assign 405,725 sequence reads to 270 piRNA clusters, covering 3.8 million base pairs. ProTRAC was run with “-clsiz 1000, -pimax 40, -pimin 26, -pdens 0.09”.

### 5.1.22 Processing of the ChIP-Seq libraries

H3K4me3 ChIP-Seq reads were processed by Eric Ross (Stowers Institute for Medical Research, Kansas City, MO, USA) as following. Reads were aligned to the *Schmidtea mediterranea* genome with bowtie (version 1; parameters: -k 100 -strata) and formatted with samtools (version 1.8) [56, 102, 91, 156]. Peaks were called with macs2 (version: 2.1.1.20160309; parameters: -q 0.01) [205]. The R library DiffBind was used to identify peaks differentially bound between *smewi-2(RNAi)*, *smewi-3(RNAi)* and *unc-22 (C. elegans)* RNAi as control. BEDTools (version: v2.26.0) was used to calculate the overlap of differential peaks with annotated repeat families [140, 56, 156].

### 5.1.23 Processing of RNA-seq libraries

Reads after removal of 3'-adapters and quality filtering with Trimmomatic (0.36) [15] were trimmed to a length of 50 nt. Next, sequences mapped to planarian rRNAs were removed with SortMeRNA [85]. Reads were assigned to the reference genome or consensus transposable element sequences [12] in strand-specific mode and quantified with kallisto [17] with “-single -l 350 -s 30 -b 30” for single-end libraries and “-b 30” for paired-end libraries. Differential expression analysis of transposable elements upon RNAi knockdown was performed with edgeR (3.20.9) using the TMM normalization method [118, 147]. Differentially expressed planarian transcripts were identified by sleuth (0.29.0) [137]. The detailed pipeline can be found in Appendix C.2.

### 5.1.24 Processing of HITS-CLIP libraries

Before mapping reads to the planarian genome, 3'-adapters were removed with cutadapt, sequences mapping to planarian rRNA were discarded with SortMeRNA, and all duplicated reads were collapsed with the FASTX-Toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit](http://hannonlab.cshl.edu/fastx_toolkit)). Large and small CLIP reads from separately prepared upper and down band (Appendix B, Fig. B.8) libraries were pooled for their subsequent joint analysis. CLIP reads were mapped to the planarian genome using the bwa aligner with the following settings “bwa aln -n 0.06 -q 20” [101]. CLIP sites were determined employing CIMS analysis as implemented in the CTK tool kit [203, 168]. Briefly, sites with aligned CLIP reads carrying substitutions and deletions ( $FDR \leq 0.001$ ) that appeared in at least in two libraries and with a total number of overlapping unique tags  $k$  at the aligning position  $\geq 30$  were considered as SMEDWI-3 targeted CLIP regions.

For the annotation of CLIP regions to genomic features reads were mapped to the planarian genome using STAR [32] with the following settings: `-outFilterMatchNmin 15 -outFilterMatchNminOverLread 0.72`. Uniquely mapped CLIP libraries were annotated to transposable elements, piRNA clusters and coding genes with BEDTools [140]. The employed pipeline can be found in Appendix C.3.

### 5.1.25 Genic piRNA density across SMEDWI-3 HITS-CLIP target sites

To estimate the abundance of genic piRNAs across SMEDWI-3-targeted CLIP-sites, we calculated the density of the mapped SMEDWI-3 immunoprecipitated piRNAs per base pair of respective CLIP sites. Briefly, CLIP sites were defined as merged overlapping CLIP regions. Next, piRNAs were mapped to the CLIP sites with the following parameters “-v1 -a -best -strata”. Weighted counts for piRNAs mapping to CLIP regions in sense orientation were normalized to the length of the respective CLIP region (in bp).

### 5.1.26 Gene-set enrichment analysis (GAGE)

To analyze the coordinated differential expression of SMEDWI-3 transcripts, we carried out gene-set enrichment analysis with GAGE (2.28.2) [111]. Expression changes of the two groups of genes upon RNAi knockdown in X1 and Xins cells were assessed with edgeR [147].

### **5.1.27 Processing of Degradome-seq libraries**

Degradome libraries were processed using piPipes [61]. Only genes that had at least 5 unique pairs of 5' to 5' 10 nt overlap between degradome reads and antisense piRNAs were considered as degraded in the ping-pong cycle.

### **5.1.28 GO term enrichment analysis**

GO term enrichment analysis was conducted using topGO (2.30.1) with Fisher's exact test [2]. GO-term gene annotations were obtained from PlanMine [16, 157].

### **5.1.29 Identification of chimeric reads**

Chimeric reads were identified using the CLIP-chimaeric pipeline with default parameters [3, 184]. To calculate the base-pairing density for Group 1, Group 2 and Group 3 only chimeric reads with piRNAs part mapping within  $\pm 20$  bp from the midpoint of mRNAs fragments were considered.

### **5.1.30 Data availability**

All sequencing data have been deposited in the Gene Expression Omnibus (GEO), series GSE122199.

## **5.2 Experimental procedures of Results part 2**

### **5.2.1 Ribosomal RNA depletion**

Ribosomal RNA depletion was conducted as described in Chapter 3. To evaluate Fragment analyzer separation profiles, planarian total RNA (1000 ng for each sample) was subjected to rRNA depletion using varying concentrations of NaCl (0 mM, 50 mM, 250 mM, 500 mM) in the hybridization buffer.

### **5.2.2 Planarian rRNA-depleted RNA-Seq dataset**

Raw sequencing reads for planarian rRNA-depleted dataset were downloaded from the project GSE122199 (GSM3460490, GSM3460491, GSM3460492). The libraries were prepared and sequenced as described in method section 5.1.14.

### 5.2.3 Publicly available RNA-Seq datasets

Raw sequencing reads for all datasets were downloaded from the Sequence read archive (SRA). Planarian polyA B1 rep1, polyA B1 rep2, polyA B1 rep3 correspond to SRR2407875, SRR2407876, and SRR2407877, respectively, from the Bioproject PRJNA296017 (GEO: GSE73027) [34]. Planarian polyA B2 rep1, polyA B2 rep2 samples correspond to SRR4068859, SRR4068860 from the Bioproject PRJNA338115 [120]. Planarian polyA B3 rep1, polyA B3 rep2, polyA B3 rep3 correspond to SRR7070906, SRR7070907, SRR7070908, respectively, (PRJNA397855) [164]. Only first read of the pair was analyzed for polyA B2 and polyA B3 from Bioprojects PRJNA338115 and PRJNA397855.

### 5.2.4 *Salmonella typhimurium* SL1344 datasets

In total, four samples were sequenced for *Salmonella typhimurium* SL1344 by IMG Laboratory GmbH (Martinsried, Germany) on an Illumina NextSeq 500 platform (single-end, 75 bp). One sample represented untreated total RNA, two samples comprised RiboZero and one RiboPool-treated total RNA. Sequencing data are available at the NCBI Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>) under the accession number GSE132630.

### 5.2.5 Processing of RNA-Seq libraries

Planarian RNA-seq data were processed as follows: Reads after removal of 3'-adapters and quality filtering with Trimmomatic (0.36) [15] were trimmed to a length of 50 nt. For libraries sequenced in pair-end mode, only the first read of a pair was considered for the analysis. Next, sequences mapped to planarian rRNAs were removed with SortMeRNA [85]. Reads were assigned to the reference genome version SMESG.1 [56] or consensus transposable element sequences [12] in strand-specific mode. The abundance of transcripts was quantified with kallisto [17] using the settings: "--single -l 350 -s 30 -b 30". Differential gene expression analysis was performed with DeSeq2 [110]. To annotate RNA-Seq reads to coding regions (CDS), reads were mapped to the planarian genome using STAR [32] with the following settings: --quantMode TranscriptomeSam --outFilterMultimapNmax 1. RNA sequencing data from *Salmonella typhimurium* SL1344 were processed with READemption 0.4.3 using default parameters [43]. Sequenced reads were mapped to the RefSeq genome version NC\_016810.1 and plasmids NC\_017718.1, NC\_017719.1, NC\_017720.1.

### **5.2.6 Phylogenetic tree**

The phylogenetic tree was constructed using NCBI taxonomic names at phyloT (<https://phylot.biobyte.de>). The tree was visualized using the Interactive Tree of Life (iToL) tool [99].

### **5.2.7 Analysis of DNA probe specificity**

DNA probe sequences were mapped to the planarian transcriptome SMEST.1 [156] using the BURST aligner (v0.99.7LL; DB15) [1] with the following settings "-fr -i .80 -m FORAGE". Only sequences that mapped to genes in antisense orientation with no more than 8 mismatches were considered as potential probe targets.

# References

- [1] Al-Ghalith, Knights, G., and Dan, K. (2017). BURST enables optimal exhaustive DNA alignment for big data. <https://github.com/knights-lab/BURST>, 10.5281/zenodo.806850.
- [2] Alexa, A., Rahnenführer, J., and Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*, 22(13):1600–1607.
- [3] Alexiou, P., Maragkakis, M., Mourelatos, Z., and Vourekas, A. (2018). cCLIP-Seq: retrieval of chimeric reads from HITS-CLIP (CLIP-Seq) libraries. *Methods in molecular biology*, 1680:87–100.
- [4] Alvarado, A. S., Newmark, P. A., Robb, S. M. C., and Juste, R. (2002). The Schmidtea mediterranea database as a molecular resource for studying platyhelminthes, stem cells and regeneration. *Development*, 129(24):5659–5665.
- [5] Andersen, P. R., Tirian, L., Vunjak, M., and Brennecke, J. (2017). A heterochromatin-dependent transcription machinery drives piRNA expression. *Nature*, 549:54–59.
- [6] Aravin, A., Gaidatzis, D., Pfeffer, S., 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(7099):203–207.
- [7] Aravin, A. A., Naumova, N. M., Tulin, A. V., Vagin, V. V., Rozovsky, Y. M., and Gvozdev, V. A. (2001). Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Current Biology*, 11(13):1017–1027.
- [8] Aravin, A. A., Sachidanandam, R., Bourc’his, D., Schaefer, C., Pezic, D., Toth, K. F., Bestor, T., and Hannon, G. J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Molecular Cell*, 31(6):785–799.
- [9] 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(5825):744–747.
- [10] Arnold, C. P., Merryman, M. S., Harris-Arnold, A., McKinney, S. A., Seidel, C. W., Loethen, S., Proctor, K. N., Guo, L., and Sánchez Alvarado, A. (2016). Pathogenic shifts in endogenous microbiota impede tissue regeneration via distinct activation of TAK1/MKK/p38. *eLife*, 5:e16793.

- [11] Ashe, A., Sapetschnig, A., Weick, E.-M., Mitchell, J., Bagijn, M., Cording, A., Doebley, A.-L., Goldstein, L., Lehrbach, N., Le Pen, J., Pintacuda, G., Sakaguchi, A., Sarkies, P., Ahmed, S., and Miska, E. (2012). piRNAs can trigger a multigenerational epigenetic memory in the germline of *C. elegans*. *Cell*, 150(1):88–99.
- [12] Bao, W., Kojima, K. K., and Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, 6(11).
- [13] Barckmann, B., Pierson, S., Dufourt, J., Papin, C., Armenise, C., Port, F., Grentzinger, T., Chambeyron, S., Baronian, G., Desvignes, J.-P., Curk, T., and Simonelig, M. (2015). Aubergine iCLIP reveals piRNA-dependent decay of mRNAs involved in germ cell development in the early embryo. *Cell Reports*, 12(7):1205–1216.
- [14] Batki, J., Schnabl, J., Wang, J., Handler, D., Andreev, V. I., Stieger, C. E., Novatchkova, M., Lampersberger, L., Kauneckaite, K., Xie, W., Mechtler, K., Patel, D. J., and Brennecke, J. (2019). The nascent RNA binding complex SFiNX licenses piRNA-guided heterochromatin formation. *Nature Structural & Molecular Biology*, 26(8):720–731.
- [15] Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114–2120.
- [16] Brandl, H., Moon, H., Vila-Farré, M., Liu, S.-Y., Henry, I., and Rink, J. C. (2015). PlanMine - a mineable resource of planarian biology and biodiversity. *Nucleic Acids Research*, 44(D1):764–773.
- [17] Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, 34:525–527.
- [18] Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G. J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell*, 128(6):1089–1103.
- [19] 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(5906):1387–1392.
- [20] Brown, D. D. R. and Pearson, B. J. (2017). A Brain Unfixed: Unlimited Neurogenesis and Regeneration of the Adult Planarian Nervous System. *Frontiers in neuroscience*, 11:289.
- [21] Carranza, S., Baguña, J., and Riutort, M. (1999). Origin and Evolution of Paralogous rRNA Gene Clusters Within the Flatworm Family Dugesidae (Platyhelminthes, Tricladida). *Journal of Molecular Evolution*, 49(250–259).
- [22] Carranza, S., Giribet, G., Ribera, C., Baguña, and Riutort, M. (1996). Evidence that two types of 18S rDNA coexist in the genome of *Dugesia* (*Schmidtea*) *mediterranea* (Platyhelminthes, Turbellaria, Tricladida). *Molecular Biology and Evolution*, 13(6):824–832.
- [23] Cebrià, F. and Newmark, P. A. (2005). Planarian homologs of netrin and netrin receptor are required for proper regeneration of the central nervous system and the maintenance of nervous system architecture. *Development*, 132(16):3691–3703.

- [24] Chen, Y.-C. A., Stuwe, E., Luo, Y., Ninova, M., Le Thomas, A., Rozhavskaia, E., Li, S., Vempati, S., Laver, J. D., Patel, D. J., Smibert, C. A., Lipshitz, H. D., Toth, K. F., and Aravin, A. A. (2016). Cutoff suppresses RNA Polymerase II termination to ensure expression of piRNA precursors. *Molecular cell*, 63(1):97–109.
- [25] 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 & Development*, 12(23):3715–3727.
- [26] Czech, B. and Hannon, G. J. (2016). One loop to rule them all: the ping-pong cycle and piRNA-guided silencing. *Trends in Biochemical Sciences*, 41(4):324–337.
- [27] Czech, B., Munafò, M., Ciabrelli, F., Eastwood, E. L., Fabry, M. H., Kneuss, E., and Hannon, G. J. (2018). piRNA-Guided Genome Defense: From Biogenesis to Silencing. *Annual Review of Genetics*, 52(1):131–157.
- [28] De Fazio, S., Bartonicek, N., Di Giacomo, M., Abreu-Goodger, C., Sankar, A., Funaya, C., Antony, C., Moreira, P. N., Enright, A. J., and O’Carroll, D. (2011). The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature*, 480(7376):259–263.
- [29] Dennis, C., Brasslet, E., Sarkar, A., and Vaury, C. (2016). Export of piRNA precursors by EJC triggers assembly of cytoplasmic Yb-body in *Drosophila*. *Nature Communications*, 7(13739).
- [30] Dennis, C., Zanni, V., Brasslet, E., Eymery, A., Zhang, L., Mteirek, R., Jensen, S., Rong, Y. S., and Vaury, C. (2013). Dot COM, a nuclear transit center for the primary piRNA pathway in *Drosophila*. *PLOS ONE*, 8(9):1–6.
- [31] Dietrich, M. R., Ankeny, R. A., and Chen, P. M. (2014). Publication Trends in Model Organism Research. *Genetics*, 198(3):787–794.
- [32] Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2012). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1):15–21.
- [33] Draper, D. E. (2004). A guide to ions and RNA structure. *RNA*, 10(3):335–343.
- [34] Duncan, E., Chitsazan, A., Seidel, C., and Sánchez Alvarado, A. (2015). Set1 and MLL1/2 Target Distinct Sets of Functionally Different Genomic Loci In Vivo. *Cell Reports*, 13(12):2741 – 2755.
- [35] Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoepfner, D. J., Dash, C., Bazett-Jones, D. P., Le Grice, S., McKay, R. D. G., Buetow, K. H., Gingeras, T. R., Misteli, T., and Meshorer, E. (2008). Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell*, 2(5):437–447.
- [36] Elkayam, E., Kuhn, C.-D., Tocilj, A., Haase, A., Greene, E., Hannon, G., and Joshua-Tor, L. (2012). The Structure of Human Argonaute-2 in Complex with miR-20a. *Cell*, 150(1):100–110.

- [37] Elliott, S. A. and Sánchez Alvarado, A. (2013). The history and enduring contributions of planarians to the study of animal regeneration. *Wiley interdisciplinary reviews. Developmental biology*, 2(3):301–326.
- [38] ElMaghraby, M. F., Andersen, P. R., Pühringer, F., Hohmann, U., Meixner, K., Lendl, T., Tirian, L., and Brennecke, J. (2019). A heterochromatin-specific RNA export pathway facilitates piRNA production. *Cell*, 178(4):964–979.
- [39] Fabry, M. H., Ciabrelli, F., Munafo, M., Eastwood, E. L., Kneuss, E., Falciatori, I., Falconio, F. A., Hannon, G. J., and Czech, B. (2019). piRNA-guided co-transcriptional silencing coopts nuclear export factors. *eLife*, 8:e47999.
- [40] Fang, W., Wang, X., Bracht, J., Nowacki, M., and Landweber, L. (2012). Piwi-interacting RNAs protect DNA against loss during *Oxytricha* genome rearrangement. *Cell*, 151(6):1243–1255.
- [41] Fincher, C. T., Wurtzel, O., de Hoog, T., Kravarik, K. M., and Reddien, P. W. (2018). Cell type transcriptome atlas for the planarian *Schmidtea mediterranea*. *Science*, 360(6391).
- [42] Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 391:806–811.
- [43] Förstner, K. U., Vogel, J., and Sharma, C. M. (2014). READemption — a tool for the computational analysis of deep-sequencing–based transcriptome data. *Bioinformatics*, 30(23):3421–3423.
- [44] Friedländer, M. R., Adamidi, C., Han, T., Lebedeva, S., Isenbarger, T. A., Hirst, M., Marra, M., Nusbaum, C., Lee, W. L., Jenkin, J. C., Alvarado, A. S., Kim, J. K., and Rajewsky, N. (2009). High-resolution profiling and discovery of planarian small RNAs. *Proceedings of the National Academy of Sciences*, 106(28):11546–11551.
- [45] Gainetdinov, I., Colpan, C., Arif, A., Cecchini, K., and Zamore, P. D. (2018). A single mechanism of biogenesis, initiated and directed by PIWI proteins, explains piRNA production in most animals. *Molecular Cell*, 71(5):775–790.
- [46] Gan, H., Lin, X., Zhang, Z., Zhang, W., Liao, S., Wang, L., and Han, C. (2011). piRNA profiling during specific stages of mouse spermatogenesis. *RNA*, 17(7):1191–1203.
- [47] Garrey, S. M., Blech, M., Riffell, J. L., Hankins, J. S., Stickney, L. M., Diver, M., Hsu, Y.-H. R., Kunanithy, V., and Mackie, G. A. (2009). Substrate binding and active site residues in RNases E and G: role of the 5'-sensor. *Journal of Biological Chemistry*, 284(46):31843–31850.
- [48] Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J. Y., and Zhang, J. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome biology*, 5(10).

- [49] Ghildiyal, M. and Zamore, P. D. (2009). Small silencing RNAs: an expanding universe. *Nature Reviews Genetics*, 10:94–108.
- [50] Girard, A., Sachidanandam, R., Hannon, G. J., and Carmell, M. A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature*, 442(7099):199–202.
- [51] Goldstein, B. and King, N. (2016). The Future of Cell Biology: Emerging Model Organisms. *Trends in Cell Biology*, 26(11):818–824. Special Issue: Future of Cell Biology.
- [52] Goriaux, C., Dessel, S., Renaud, Y., Vaury, C., and Brasset, E. (2014). Transcriptional properties and splicing of the flamenco piRNA cluster. *EMBO reports*, 15(4):411–418.
- [53] Gou, L.-T., Dai, P., Yang, J.-H., Xue, Y., Hu, Y.-P., Zhou, Y., Kang, J.-Y., Wang, X., Li, H., Hua, M.-M., Zhao, S., Hu, S.-D., Wu, L.-G., Shi, H.-J., Li, Y., Fu, X.-D., Qu, L.-H., Wang, E.-D., and Liu, M.-F. (2014). Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis. *Cell Research*, 24(6):680–700.
- [54] Green, M. R. and Sambrook, J. (2012). *Molecular Cloning. A laboratory manual*, chapter 13, pages 953–960. 978-1-936113-41-5. Cold Spring Harbor Laboratory Press, 4th edition.
- [55] Grivna, S. T., Beyret, E., Wang, Z., and Lin, H. (2006). A novel class of small RNAs in mouse spermatogenic cells. *Genes & Development*, 20(13):1709–1714.
- [56] Grohme, M. A., Schloissnig, S., Rozanski, A., Pippel, M., Young, G. R., Winkler, S., Brandl, H., Henry, I., Dahl, A., Powell, S., Hiller, M., Myers, E., and Rink, J. C. (2018). The genome of *Schmidtea mediterranea* and the evolution of core cellular mechanisms. *Nature*, 554(7690):56–61.
- [57] Gunawardane, L. S., Saito, K., Nishida, K. M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M. C. (2007). A Slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science*, 315(5818):1587–1590.
- [58] Guo, L., Zhang, S., Rubinstein, B., Ross, E., and Alvarado, A. S. (2016). Widespread maintenance of genome heterozygosity in *Schmidtea mediterranea*. *Nature Ecology & Evolution*, 1(1):19.
- [59] Guo, T., Peters, A. H. F. M., and Newmark, P. A. (2006). A bruno-like gene is required for stem cell maintenance in planarians. *Developmental Cell*, 11(2):159–169.
- [60] Han, B. W., Wang, W., Li, C., Weng, Z., and Zamore, P. D. (2015). piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. *Science*, 348(6236):817–821.
- [61] Han, B. W., Wang, W., Zamore, P. D., and Weng, Z. (2014). piPipes: a set of pipelines for piRNA and transposon analysis via small RNA-seq, RNA-seq, degradome- and CAGE-seq, CHIP-seq and genomic DNA sequencing. *Bioinformatics*, 31(4):593–595.
- [62] Handler, D., Meixner, K., Pizka, M., Lauss, K., Schmied, C., Gruber, F. S., and Brennecke, J. (2013). The genetic makeup of the *Drosophila* piRNA pathway. *Molecular cell*, 50(5):762–777.

- [63] Hauptmann, J., Schraivogel, D., Bruckmann, A., Manickavel, S., Jakob, L., Eichner, N., Pfaff, J., Urban, M., Sprunck, S., Hafner, M., Tuschl, T., Deutzmann, R., and Meister, G. (2015). Biochemical isolation of Argonaute protein complexes by Ago-APP. *Proceedings of the National Academy of Sciences*, 112(38):11841–11845.
- [64] Hayashi, R., Schnabl, J., Handler, D., Mohn, F., Ameres, S. L., and Brennecke, J. (2016). Genetic and mechanistic diversity of piRNA 3'-end formation. *Nature*, 539:588–592.
- [65] Hayashi, T. and Agata, K. (2018). *A subtractive FACS method for isolation of planarian stem cells and neural cells*, pages 467–478. Springer New York.
- [66] Hayashi, T., Asami, M., Higuchi, S., Shibata, N., and Agata, K. (2006). Isolation of planarian X-ray-sensitive stem cells by fluorescence-activated cell sorting. *Development, Growth & Differentiation*, 48(6):371–380.
- [67] Helwak, A., Kudla, G., Dudnakova, T., and Tollervey, D. (2013). Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell*, 153(3):654–665.
- [68] Homolka, D., Pandey, R., Goriaux, C., Brasset, E., Vaury, C., Sachidanandam, R., Fauvarque, M.-O., and Pillai, R. (2015). PIWI slicing and RNA elements in precursors instruct directional primary piRNA biogenesis. *Cell Reports*, 12(3):418–428.
- [69] Houwing, S., Kamminga, L. M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Filippov, D. V., Blaser, H., Raz, E., Moens, C. B., Plasterk, R. H. A., Hannon, G. J., Draper, B. W., and Ketting, R. F. (2007). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in zebrafish. *Cell*, 129(1):69–82.
- [70] Howe, F. S., Fischl, H., Murray, S. C., and Mellor, J. (2017). Is H3K4me3 instructive for transcription activation? *BioEssays*, 39(1):e201600095.
- [71] Hyun, K., Jeon, J., Park, K., and Kim, J. (2017). Writing, erasing and reading histone lysine methylations. *Experimental & Molecular Medicine*, 49(4):324–324.
- [72] Ivankovic, M., Haneckova, R., Thommen, A., Grohme, M. A., Vila-Farré, M., Werner, S., and Rink, J. C. (2019). Model systems for regeneration: planarians. *Development*, 146(17).
- [73] Iwasaki, Y. W., Siomi, M. C., and Siomi, H. (2015). PIWI-interacting RNA: its biogenesis and functions. *Annual Review of Biochemistry*, 84(1):405–433.
- [74] Izumi, N., Shoji, K., Sakaguchi, Y., Honda, S., Kirino, Y., Suzuki, T., Katsuma, S., and Tomari, Y. (2016). Identification and functional analysis of the pre-piRNA 3' Trimmer in silkworms. *Cell*, 164(5):962–973.
- [75] Kang, H. and Alvarado, A. S. (2009). Flow cytometry methods for the study of cell-cycle parameters of planarian stem cells. *Developmental Dynamics*, 238(5):1111–1117.
- [76] Kashima, M., Kumagai, N., Agata, K., and Shibata, N. (2016). Heterogeneity of chromatoid bodies in adult pluripotent stem cells of planarian *Dugesia japonica*. *Development, Growth & Differentiation*, 58(2):225–237.

- [77] 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(3):1581–1590.
- [78] Kawaoka, S., Izumi, N., Katsuma, S., and Tomari, Y. (2011). 3' end formation of PIWI-interacting RNAs in vitro. *Molecular Cell*, 43(6):1015–1022.
- [79] Kazazian, H. H. (2004). Mobile Elements: Drivers of Genome Evolution. *Science*, 303(5664):1626–1632.
- [80] Kim, I. V., Duncan, E. M., Ross, E. J., Gorbovytska, V., Nowotarski, S. H., Elliott, S. A., Sánchez Alvarado, A., and Kuhn, C.-D. (2019). Planarians recruit piRNAs for mRNA turnover in adult stem cells. *Genes & Development*, page 10.1101/gad.322776.118.
- [81] Kim, K. W., Tang, N. H., Andrusiak, M. G., Wu, Z., Chisholm, A. D., and Jin, Y. (2018). A neuronal piRNA pathway inhibits axon regeneration in *C. elegans*. *Neuron*, 97(3):511–519.
- [82] Kiuchi, T., Koga, H., Kawamoto, M., Shoji, K., Sakai, H., Arai, Y., Ishihara, G., Kawaoka, S., Sugano, S., Shimada, T., Suzuki, Y., Suzuki, M. G., and Katsuma, S. (2014). A single female-specific piRNA is the primary determiner of sex in the silkworm. *Nature*, 509:633–636.
- [83] 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(6):1137–1149.
- [84] Kneuss, E., Munafò, M., Eastwood, E. L., Deumer, U.-S., Preall, J. B., Hannon, G. J., and Czech, B. (2019). Specialization of the *Drosophila* nuclear export family protein Nxf3 for piRNA precursor export. *Genes & Development*, page 10.1101/gad.328690.119.
- [85] Kopylova, E., Noé, L., and Touzet, H. (2012). SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, 28(24):3211–3217.
- [86] Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2018). miRBase: from microRNA sequences to function. *Nucleic Acids Research*, 47(1):155–162.
- [87] Kuhn, C.-D. and Joshua-Tor, L. (2013). Eukaryotic Argonautes come into focus. *Trends in Biochemical Sciences*, 38(5):263–271.
- [88] Kukutla, P., Steritz, M., and Xu, J. (2013). Depletion of ribosomal RNA for mosquito gut metagenomic RNA-seq. *Journal of visualized experiments : JoVE*, 74.
- [89] Lakshmanan, V., Bansal, D., Kulkarni, J., Poduval, D., Krishna, S., Sasidharan, V., Anand, P., Seshasayee, A., and Palakodeti, D. (2016). Genome-Wide Analysis of Polyadenylation Events in *Schmidtea mediterranea*. *G3: Genes, Genomes, Genetics*, 6(10):3035–3048.
- [90] Lambert, D., Leipply, D., Shiman, R., and Draper, D. E. (2009). The influence of monovalent cation size on the stability of RNA tertiary structures. *Journal of molecular biology*, 390(4):791–804.

- [91] 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 Biology*, 10(3).
- [92] 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 Research*, 19(10):1776–1785.
- [93] Lau, N. C., Seto, A. G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D. P., and Kingston, R. E. (2006). Characterization of the piRNA complex from rat testes. *Science*, 313(5785):363–367.
- [94] Le Thomas, A., Rogers, A. K., Webster, A., Marinov, G. K., Liao, S. E., Perkins, E. M., Hur, J. K., Aravin, A. A., and Tóth, K. F. (2013). Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes & Development*, 27(4):390–399.
- [95] Le Thomas, A., Tóth, K. F., and Aravin, A. A. (2014). To be or not to be a piRNA: genomic origin and processing of piRNAs. *Genome Biology*, 15(204).
- [96] Lee, H.-C., Gu, W., Shirayama, M., Youngman, E., Conte Jr., D., and Mello, C. (2012). *C. elegans* piRNAs mediate the genome-wide surveillance of germline transcripts. *Cell*, 150(1):78–87.
- [97] 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(5):843–854.
- [98] Lei, K., McKinney, S. A., Ross, E. J., Lee, H.-C., and Alvarado, A. S. (2019). Cultured pluripotent planarian stem cells retain potency and express proteins from exogenously introduced mRNAs. *bioRxiv*.
- [99] Letunic, I. and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research*, 47(W1):256–259.
- [100] Lewis, S. H., Quarles, K. A., Yang, Y., Tanguy, M., Frézal, L., Smith, S. A., Sharma, P. P., Cordaux, R., Gilbert, C., Giraud, I., Collins, D. H., Zamore, P. D., Miska, E. A., Sarkies, P., and Jiggins, F. M. (2018). Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. *Nature Ecology & Evolution*, 2(1):174–181.
- [101] Li, H. and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, 26(5):589–595.
- [102] Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup, . G. P. D. P. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16):2078–2079.
- [103] Li, M. Z. and Elledge, S. J. (2007). Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nature Methods*, 4(3):251 – 256.

- [104] Li, S.-K., Zhou, J.-W., Yim, A. K.-Y., Leung, A. K.-Y., Tsui, S. K.-W., Chan, T.-F., and Lau, T. C.-K. (2013a). Organism-Specific rRNA Capture System for Application in Next-Generation Sequencing. *PLOS ONE*, 8(9):e74286.
- [105] Li, X., Roy, C., Dong, X., Bolcun-Filas, E., Wang, J., Han, B., Xu, J., Moore, M., Schimenti, J., Weng, Z., and Zamore, P. (2013b). An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. *Molecular Cell*, 50(1):67–81.
- [106] Lim, A. K. and Kai, T. (2007). Unique germ-line organelle, nuage, functions to repress selfish genetic elements in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences*, 104(16):6714–6719.
- [107] Lim, R. S., Anand, A., Nishimiya-Fujisawa, C., Kobayashi, S., and Kai, T. (2014). Analysis of Hydra PIWI proteins and piRNAs uncover early evolutionary origins of the piRNA pathway. *Developmental Biology*, 386(1):237–251.
- [108] Lin, H. and Spradling, A. (1997). A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the *Drosophila* ovary. *Development*, 124(12):2463–2476.
- [109] Lou, C.-H., Shum, E. Y., and Wilkinson, M. F. (2015). RNA degradation drives stem cell differentiation. *The EMBO journal*, 34(12):1606–1608.
- [110] Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550.
- [111] Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D., and Woolf, P. J. (2009). GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics*, 10(161).
- [112] Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D., and Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research*, 47(W1):636–641.
- [113] Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):10–12.
- [114] Marzluff, W. F. and Duronio, R. J. (2002). Histone mRNA expression: multiple levels of cell cycle regulation and important developmental consequences. *Current Opinion in Cell Biology*, 14(6):692–699.
- [115] Marzluff, W. F. and Koreski, K. P. (2017). Birth and Death of Histone mRNAs. *Trends in Genetics*, 33(10):745–759.
- [116] Marzluff, W. F., Wagner, E. J., and Duronio, R. J. (2008). Metabolism and regulation of canonical histone mRNAs: Life without a poly(A) tail. *Nature Reviews Genetics*, pages 843–854.
- [117] Matsumoto, N., Nishimasu, H., Sakakibara, K., Nishida, K. M., Hirano, T., Ishitani, R., Siomi, H., Siomi, M. C., and Nureki, O. (2016). Crystal Structure of Silkworm PIWI-Clade Argonaute Siwi Bound to piRNA. *Cell*, 167(2):484–497.

- [118] McCarthy, D. J., Chen, Y., and Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, 40(10):4288–4297.
- [119] Meister, G. (2013). Argonaute proteins: functional insights and emerging roles. *Nature Reviews Genetics*, 14:447–459.
- [120] Mihaylova, Y., Abnave, P., Kao, D., Hughes, S., Lai, A., Jaber-Hijazi, F., Kosaka, N., and Aboobaker, A. A. (2018). Conservation of epigenetic regulation by the MLL3/4 tumour suppressor in planarian pluripotent stem cells. *Nature Communications*, 9(1):3633.
- [121] Mochizuki, K., Fine, N. A., Fujisawa, T., and Gorovsky, M. A. (2002). Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in Tetrahymena. *Cell*, 110(6):689–699.
- [122] Mohn, F., Handler, D., and Brennecke, J. (2015). piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. *Science*, 348(6236):812–817.
- [123] 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(6):1364–1379.
- [124] Moore, M. J., Zhang, C., Gantman, E. C., Mele, A., Darnell, J. C., and Darnell, R. B. (2014). Mapping Argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis [published correction appears in Nat Protoc. 2016 Mar;11(3):616]. *Nature protocols*, 9(2):263–293.
- [125] Morlan, J. D., Qu, K., and Sinicropi, D. V. (2012). Selective Depletion of rRNA Enables Whole Transcriptome Profiling of Archival Fixed Tissue. *PLOS ONE*, 7(8):e42882.
- [126] Murano, K., Iwasaki, Y. W., Ishizu, H., Mashiko, A., Shibuya, A., Kondo, S., Adachi, S., Suzuki, S., Saito, K., Natsume, T., Siomi, M. C., and Siomi, H. (2019). Nuclear RNA export factor variant initiates piRNA-guided co-transcriptional silencing. *The EMBO Journal*, 38(17):e102870.
- [127] Newmark, P. A. and Alvarado, A. S. (2002). Not your father’s planarian: a classic model enters the era of functional genomics. *Nature Reviews Genetics*, 3(3):210–219.
- [128] Nishimura, O., Hosoda, K., Kawaguchi, E., Yazawa, S., Hayashi, T., Inoue, T., Umesono, Y., and Agata, K. (2015). Unusually large number of mutations in asexually reproducing clonal planarian *Dugesia japonica*. *PLOS ONE*, 10(11):1–23.
- [129] Ochman, H., Elwyn, S., and Moran, N. A. (1999). Calibrating bacterial evolution. *Proceedings of the National Academy of Sciences*, 96(22):12638–12643.
- [130] Olivieri, D., Senti, K.-A., Subramanian, S., Sachidanandam, R., and Brennecke, J. (2012). The Cochaperone Shutdown Defines a Group of Biogenesis Factors Essential for All piRNA Populations in Drosophila. *Molecular Cell*, 47(6):954–969.

- [131] O’Neil, D., Glowatz, H., and Schlumpberger, M. (2013). Ribosomal RNA Depletion for Efficient Use of RNA-Seq Capacity. *Current Protocols in Molecular Biology*, 103(1):4.19.1–4.19.8.
- [132] Ozata, D. M., Gainetdinov, I., Zoch, A., O’Carroll, D., and Zamore, P. D. (2019). PIWI-interacting RNAs: small RNAs with big functions. *Nature Reviews Genetics*, 20(2):89–108.
- [133] Palakodeti, D., Smielewska, M., Lu, Y. C., Yeo, G. W., and Graveley, B. R. (2008). The PIWI proteins SMEDWI-2 and SMEDWI-3 are required for stem cell function and piRNA expression in planarians. *Rna*, 14(6):1174–1186.
- [134] Pearson, B. and Alvarado, A. S. (2008). Regeneration, stem cells, and the evolution of tumor suppression. *Cold Spring Harbor Symposia on Quantitative Biology*, 73:565–572.
- [135] Peiris, H. T., Hoyer, K. K., and Oviedo, N. J. (2014). Innate immune system and tissue regeneration in planarians: An area ripe for exploration. *Seminars in Immunology*, 26(4):295–302.
- [136] Pellettieri, J. (2019). Regenerative tissue remodeling in planarians – The mysteries of morphallaxis. *Seminars in Cell & Developmental Biology*, 87:13–21.
- [137] Pimentel, H., Bray, N. L., Puente, S., Melsted, P., and Pachter, L. (2017). Differential analysis of RNA-seq incorporating quantification uncertainty. *Nature Methods*, 14:687–690.
- [138] Preall, J. B., Czech, B., Guzzardo, P. M., Muerdter, F., and Hannon, G. J. (2012). Shutdown is a component of the *Drosophila* piRNA biogenesis machinery. *RNA*, 18(8):1446–1457.
- [139] Qi, H., Watanabe, T., Ku, H.-Y., Liu, N., Zhong, M., and Lin, H. (2011). The Yb body, a major site for Piwi-associated RNA biogenesis and a gateway for Piwi expression and transport to the nucleus in somatic cells. *Journal of Biological Chemistry*, 286(5):3789–3797.
- [140] Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842.
- [141] Rajasethupathy, P., Antonov, I., Sheridan, R., Frey, S., Sander, C., Tuschl, T., and Kandel, E. R. (2012). A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity. *Cell*, 149(3):693–707.
- [142] Rangan, P., Malone, C., Navarro, C., Newbold, S., Hayes, P., Sachidanandam, R., Hannon, G., and Lehmann, R. (2011). piRNA production requires heterochromatin formation in *Drosophila*. *Current Biology*, 21(16):1373–1379.
- [143] Reddien, P. W. (2013). Specialized progenitors and regeneration. *Development*, 140(5):951–957.
- [144] Reddien, P. W., Bermange, A. L., Murfitt, K. J., Jennings, J. R., and Alvarado, A. S. (2005a). Identification of genes needed for regeneration, stem cell function, and tissue homeostasis by systematic gene perturbation in planaria. *Developmental Cell*, 8(5):635–649.

- [145] Reddien, P. W., Oviedo, N. J., Jennings, J. R., Jenkin, J. C., and Sánchez Alvarado, A. (2005b). SMEDWI-2 is a PIWI-like protein that regulates planarian stem cells. *Science*, 310(5752):1327–1330.
- [146] Reuter, M., Berninger, P., Chuma, S., Shah, H., Hosokawa, M., Funaya, C., Antony, C., Sachidanandam, R., and Pillai, R. S. (2011). Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. *Nature*, 480(7376):264–267.
- [147] Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2009). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–140.
- [148] Rojas-Ríos, P. and Simonelig, M. (2018). piRNAs and PIWI proteins: regulators of gene expression in development and stem cells. *Development*, 145(17).
- [149] Rosenkranz, D. and Zischler, H. (2012). proTRAC - a software for probabilistic piRNA cluster detection, visualization and analysis. *BMC Bioinformatics*, 13(5).
- [150] Ross, R. J., Weiner, M. M., and Lin, H. (2014). PIWI proteins and PIWI-interacting RNAs in the soma. *Nature*, 505(7483):353–359.
- [151] Rouget, C., Papin, C., Boureux, A., Meunier, A.-C., Franco, B., Robine, N., Lai, E. C., Pelisson, A., and Simonelig, M. (2010). Maternal mRNA deadenylation and decay by the piRNA pathway in the early *Drosophila* embryo. *Nature*, 467:1128–1132.
- [152] Rouhana, L., Shibata, N., Nishimura, O., and Agata, K. (2010). Different requirements for conserved post-transcriptional regulators in planarian regeneration and stem cell maintenance. *Developmental Biology*, 341(2):429–443.
- [153] Rouhana, L., Vieira, A. P., Roberts-Galbraith, R. H., and Newmark, P. A. (2012). PRMT5 and the role of symmetrical dimethylarginine in chromatoid bodies of planarian stem cells. *Development*, 139(6):1083–1094.
- [154] Rouhana, L., Weiss, J. A., Forsthoefel, D. J., Lee, H., King, R. S., Inoue, T., Shibata, N., Agata, K., and Newmark, P. A. (2013). RNA interference by feeding in vitro–synthesized double-stranded RNA to planarians: Methodology and dynamics. *Developmental Dynamics*, 242(6):718–730.
- [155] Rouhana, L., Weiss, J. A., King, R. S., and Newmark, P. A. (2014). PIWI homologs mediate Histone H4 mRNA localization to planarian chromatoid bodies. *Development*, 141(13):2592–2601.
- [156] Rozanski, A., Moon, H., Brandl, H., Martín-Durán, J. M., Grohme, M. A., Hüttner, K., Bartscherer, K., Henry, I., and Rink, J. C. (2018a). PlanMine 3.0 - improvements to a mineable resource of flatworm biology and biodiversity. *Nucleic Acids Research*, 47(D1):812–820.
- [157] Rozanski, A., Moon, H., Brandl, H., Martín-Durán, J. M., Grohme, M. A., Hüttner, K., Bartscherer, K., Henry, I., and Rink, J. C. (2018b). PlanMine 3.0 - improvements to a mineable resource of flatworm biology and biodiversity. *Nucleic Acids Research*, 47(D1):812–820.

- [158] Russell, J. J., Theriot, J. A., Sood, P., Marshall, W. F., Landweber, L. F., Fritz-Laylin, L., Polka, J. K., Oliferenko, S., Gerbich, T., Gladfelter, A., Umen, J., Bezanilla, M., Lancaster, M. A., He, S., Gibson, M. C., Goldstein, B., Tanaka, E. M., Hu, C.-K., and Brunet, A. (2017). Non-model model organisms. *BMC Biology*, 15(1):55.
- [159] Saberi, A., Jamal, A., Beets, I., Schoofs, L., and Newmark, P. A. (2016). GPCRs Direct Germline Development and Somatic Gonad Function in Planarians. *PLOS Biology*, 14(5):e1002457.
- [160] Sahu, S., Dattani, A., and Aboobaker, A. A. (2017). Secrets from immortal worms: What can we learn about biological ageing from the planarian model system? *Seminars in Cell and Developmental Biology*, 70:108–121.
- [161] Saito, K., Ishizu, H., Komai, M., Kotani, H., Kawamura, Y., Nishida, K. M., Siomi, H., and Siomi, M. C. (2010). Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in *Drosophila*. *Genes & Development*, 24(22):2493–2498.
- [162] Saito, K., Nishida, K. M., Mori, T., Kawamura, Y., Miyoshi, K., Nagami, T., Siomi, H., and Siomi, M. C. (2006). Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes & Development*, 20(16):2214–2222.
- [163] Schirle, N. T., Sheu-Gruttadauria, J., and MacRae, I. J. (2014). Structural basis for microRNA targeting. *Science*, 346(6209):608–613.
- [164] Schmidt, D., Reuter, H., Hüttner, K., Ruhe, L., Rabert, F., Seebeck, F., Irimia, M., Solana, J., and Bartscherer, K. (2018). The Integrator complex regulates differential snRNA processing and fate of adult stem cells in the highly regenerative planarian *Schmidtea mediterranea*. *PLOS Genetics*, 14(12):1–28.
- [165] Schürmann, W. and Peter, R. (2001). Planarian cell culture: a comparative review of methods and an improved protocol for primary cultures of neoblasts. *Belgian Journal of Zoology*, 131:123–130.
- [166] Seth, M., Shirayama, M., Gu, W., Ishidate, T., Conte Jr., D., and Mello, C. (2013). The *C. elegans* CSR-1 Argonaute pathway counteracts epigenetic silencing to promote germline gene expression. *Developmental Cell*, 27(6):656–663.
- [167] Seth, M., Shirayama, M., Tang, W., Shen, E.-Z., Tu, S., Lee, H.-C., Weng, Z., and Mello, C. C. (2018). The coding regions of germline mRNAs confer sensitivity to Argonaute regulation in *C. elegans*. *Cell Reports*, 22(9):2254–2264.
- [168] Shah, A., Qian, Y., Weyn-Vanhentenryck, S. M., and Zhang, C. (2016). CLIP Tool Kit (CTK): a flexible and robust pipeline to analyze CLIP sequencing data. *Bioinformatics*, 33(4):566–567.
- [169] Shen, E.-Z., Chen, H., Ozturk, A. R., Tu, S., Shirayama, M., Tang, W., Ding, Y.-H., Dai, S.-Y., Weng, Z., and Mello, C. C. (2018). Identification of piRNA binding sites reveals the Argonaute regulatory landscape of the *C. elegans* germline. *Cell*, 172(5):937–951.

- [170] Shibata, N., Kashima, M., Ishiko, T., Nishimura, O., Rouhana, L., Misaki, K., Yone-mura, S., Saito, K., Siomi, H., Siomi, M., and Agata, K. (2016). Inheritance of a nuclear PIWI from pluripotent stem cells by somatic descendants ensures differentiation by silencing transposons in planarian. *Developmental Cell*, 37(3):226–237.
- [171] Shirayama, M., Seth, M., Lee, H.-C., Gu, W., Ishidate, T., Conte Jr., D., and Mello, C. (2012). piRNAs initiate an epigenetic memory of nonself RNA in the *C. elegans* germline. *Cell*, 150(1):65–77.
- [172] 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(5):964–980.
- [173] Siomi, M. C., Mannen, T., and Siomi, H. (2010). How does the royal family of Tudor rule the PIWI-interacting RNA pathway? *Genes & Development*, 24(7):636–646.
- [174] Solana, J., Gamberi, C., Mihaylova, Y., Grosswendt, S., Chen, C., Lasko, P., Rajewsky, N., and Aboobaker, A. A. (2013). The CCR4-NOT Complex Mediates Deadenylation and Degradation of Stem Cell mRNAs and Promotes Planarian Stem Cell Differentiation. *PLOS Genetics*, 9(12):e1004003.
- [175] Solana, J., Kao, D., Mihaylova, Y., Jaber-Hijazi, F., Malla, S., Wilson, R., and Aboobaker, A. (2012). Defining the molecular profile of planarian pluripotent stem cells using a combinatorial RNAseq, RNA interference and irradiation approach. *Genome biology*, 13(3):R19.
- [176] Solana, J., Lasko, P., and Romero, R. (2009). Spoltud-1 is a chromatoid body component required for planarian long-term stem cell self-renewal. *Developmental biology*, 328(2):410–421.
- [177] Soper, S. F. C., van der Heijden, G. W., Hardiman, T. C., Goodheart, M., Martin, S. L., 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(2):285–297.
- [178] Sun, S., Xie, H., Sun, Y., Song, J., and Li, Z. (2012). Molecular characterization of gap region in 28S rRNA molecules in brine shrimp *Artemia parthenogenetica* and planarian *Dugesia japonica*. *Biochemistry (Moscow)*, 77(4):411–417.
- [179] Teixeira, L. K. and Reed, S. I. (2013). Ubiquitin Ligases and Cell Cycle Control. *Annual Review of Biochemistry*, 82(1):387–414.
- [180] Théron, E., Dennis, C., Brassat, E., and Vaury, C. (2014). Distinct features of the piRNA pathway in somatic and germ cells: from piRNA cluster transcription to piRNA processing and amplification. *Mobile DNA*, 5(28).
- [181] Vagin, V. V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P. D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. *Science*, 313(5785):320–324.

- [182] van Keulen, H., Mertz, P. M., LoVerde, P. T., Shi, H., and Rekosh, D. M. (1991). Characterization of a 54-nucleotide gap region in the 28S rRNA gene of *Schistosoma mansoni*. *Molecular and Biochemical Parasitology*, 45(2):205–214.
- [183] van Wolfswinkel, J. C. (2014). Piwi and Potency: PIWI proteins in animal stem cells and regeneration. *Integrative and Comparative Biology*, 54(4):700–713.
- [184] Vourekas, A., Alexiou, P., Vrettos, N., Maragkakis, M., and Mourelatos, Z. (2016). Sequence-dependent but not sequence-specific piRNA adhesion traps mRNAs to the germ plasm. *Nature*, 531:390–394.
- [185] Vourekas, A. and Mourelatos, Z. (2014). *HITS-CLIP (CLIP-Seq) for mouse Piwi proteins*, pages 73–95. Humana Press, Totowa, NJ.
- [186] Vourekas, A., Zheng, Q., Alexiou, P., Maragkakis, M., Kirino, Y., Gregory, B. D., and Mourelatos, Z. (2012). Mili and Miwi target RNA repertoire reveals piRNA biogenesis and function of Miwi in spermiogenesis. *Nature Structural & Molecular Biology*, 19(8):773–781.
- [187] Wang, L., Dou, K., Moon, S., Tan, F. J., and Zhang, Z. Z. (2018). Hijacking oogenesis enables massive propagation of LINE and retroviral transposons. *Cell*, 174(5):1082–1094.
- [188] Wang, W., Yoshikawa, M., Han, B., Izumi, N., Tomari, Y., Weng, Z., and Zamore, P. (2014). The initial Uridine of primary piRNAs does not create the tenth Adenine that is the hallmark of secondary piRNAs. *Molecular Cell*, 56(5):708–716.
- [189] Ware, V. C., Renkawitz, R., and Gerbi, S. A. (1985). rRNA processing: removal of only nineteen bases at the gap between 28Sa and 28Sb rRNAs in *Sciara coprophila*. *Nucleic Acids Research*, 13(10):3581–3597.
- [190] Watanabe, T., Chuma, S., Yamamoto, Y., Kuramochi-Miyagawa, S., Totoki, Y., Toyoda, A., Hoki, Y., Fujiyama, A., Shibata, T., Sado, T., Noce, T., Nakano, T., Nakatsuji, N., Lin, H., and Sasaki, H. (2011). MitoPLD is a mitochondrial protein essential for nuage formation and piRNA biogenesis in the mouse germline. *Developmental Cell*, 20(3):364–375.
- [191] Watanabe, T., Takeda, A., Tsukiyama, T., Mise, K., Okuno, T., Sasaki, H., Minami, N., and Imai, H. (2006). Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes & Development*, 20(13):1732–1743.
- [192] Waymouth, C. (1970). Osmolality of mammalian blood and of media for culture of mammalian cells. *In Vitro*, 6(2):109–127.
- [193] Wedeles, C., Wu, M., and Claycomb, J. (2013). Protection of germline gene expression by the *C. elegans* Argonaute CSR-1. *Developmental Cell*, 27(6):664–671.
- [194] Wetmur, J. G. (1991). DNA Probes: Applications of the principles of nucleic acid hybridization. *Critical Reviews in Biochemistry and Molecular Biology*, 26(3-4):227–259.

- [195] Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell*, 75(5):855–862.
- [196] Williams, A. G., Thomas, S., Wyman, S. K., and Holloway, A. K. (2014). RNA-seq Data: Challenges in and Recommendations for Experimental Design and Analysis. *Current Protocols in Human Genetics*, 83(1):11.13.1–11.13.20.
- [197] Woese, C. R. (2000). Interpreting the universal phylogenetic tree. *Proceedings of the National Academy of Sciences*, 97(15):8392–8396.
- [198] Xiol, J., Spinelli, P., Laussmann, M., Homolka, D., Yang, Z., Cora, E., Couté, Y., Conn, S., Kadlec, J., Sachidanandam, R., Kaksonen, M., Cusack, S., Ephrussi, A., and Pillai, R. (2014). RNA clamping by Vasa assembles a piRNA amplifier complex on transposon transcripts. *Cell*, 157(7):1698–1711.
- [199] Yamanaka, S., Siomi, M. C., and Siomi, H. (2014). piRNA clusters and open chromatin structure. *Mobile DNA*, 5(22).
- [200] Yang, Z., Chen, K.-M., Pandey, R. R., Homolka, D., Reuter, M., Janeiro, B. K. R., Sachidanandam, R., Fauvarque, M.-O., McCarthy, A. A., and Pillai, R. S. (2016). PIWI slicing and EXD1 drive biogenesis of nuclear piRNAs from cytosolic targets of the mouse piRNA pathway. *Molecular Cell*, 61(1):138–152.
- [201] Zeng, A., Li, H., Guo, L., Gao, X., McKinney, S., Wang, Y., Yu, Z., Park, J., Semerad, C., Ross, E., Cheng, L. C., Davies, E., Lei, K., Wang, W., Perera, A., Hall, K., Peak, A., Box, A., and Sánchez Alvarado, A. (2018). Prospectively isolated tetraspanin+ neoblasts are adult pluripotent stem cells underlying planaria regeneration. *Cell*, 173(7):1593–1608.
- [202] Zeng, A., Li, Y. Q., Wang, C., Han, X. S., Li, G., Wang, J. Y., Li, D. S., Qin, Y. W., Shi, Y., Brewer, G., and Jing, Q. (2013). Heterochromatin protein 1 promotes self-renewal and triggers regenerative proliferation in adult stem cells. *Journal of Cell Biology*, 201(3):409–425.
- [203] Zhang, C. and Darnell, R. B. (2011). Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nature Biotechnology*, 29:607–614.
- [204] Zhang, D., Tu, S., Stubna, M., Wu, W.-S., Huang, W.-C., Weng, Z., and Lee, H.-C. (2018). The piRNA targeting rules and the resistance to piRNA silencing in endogenous genes. *Science*, 359(6375):587–592.
- [205] Zhang, Y., Liu, T., Meyer, C. A., Eickhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., and Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, 9(R137).
- [206] Zhang, Z., Theurkauf, W. E., Weng, Z., and Zamore, P. D. (2012). Strand-specific libraries for high throughput RNA sequencing (RNA-Seq) prepared without poly(A) selection. *Silence*, 3(9).
- [207] Zhang, Z., Wang, J., Schultz, N., Zhang, F., Parhad, S., Tu, S., Vreven, T., Zamore, P., Weng, Z., and Theurkauf, W. (2014). The HP1 homolog Rhino anchors a nuclear complex that suppresses piRNA precursor splicing. *Cell*, 157(6):1353–1363.

- 
- [208] Zhang, Z., Xu, J., Koppetsch, B., Wang, J., Tipping, C., Ma, S., Weng, Z., Theurkauf, W., and Zamore, P. (2011). Heterotypic piRNA ping-pong requires Qin, a protein with both E3 Ligase and Tudor domains. *Molecular Cell*, 44(4):572–584.
- [209] Zhu, S. J. and Pearson, B. J. (2018). Smed-myb-1 specifies early temporal identity during planarian epidermal differentiation. *Cell Reports*, 25(1):38–46.



# Appendix A

## Adapter and primer sequences

**Table A.1** Primer sequences used for cloning N-terminal 200 amino acids of SMEDWI-3 into pGEX vector. Part of primer corresponding to vector overhang is in lower-case.

Gene	Feature	Sequence
<i>pGEX</i>	Forward	TGACTGACTGACGATCTGCCTCGC
linearization	Reverse	TTCCGGGGATCCACGCGGAACC
<i>smedwi-3</i>	Forward	gttccgcgtggatccccggaaATGTCAGGAAGTAGTGG AATAGG TAGAGG
	Reverse	aggcagatcgtcagtcagtcaTTCTTTTATTTGAGTTGGATATTCA CGTC

**Table A.2** Primer sequences used for cloning Strep-tagged SMEDWI proteins into pFL vectors. Part of primer corresponding to vector overhang is in lower-case.

Gene	Feature	Sequence
<i>smedwi-1</i>	Forward	atgaccggtggccagcagatgggcATGGATTCAACTAATGTTACA AGTG
	Reverse	gtacttctcgacaagcttctactaCTGAGACACCAAGTAAAATTCAT AG
<i>smedwi-2</i>	Forward	attgaggctcacagagaacagattggtggaATGGAAGAAATCCCGGTG AAA
	Reverse	gtacttctcgacaagcttctactaCAGATAAAACAGGCGGTT
<i>smedwi-3</i>	Forward	attgaggctcacagagaacagattggtggaATGAGCGGTTCTAGCGGT AT
	Reverse	gtacttctcgacaagcttctactaCAAGTAGAACAGACGGTCAC

**Table A.3** Primer sequences used for cloning gene fragments into pPR-T4P vector. Part of a primer corresponding to vector overhang is in lower-case.

Gene	Feature	Sequence
<i>pPR-T4P</i>	Forward	CGGGTAGAATTGGCCGGGC
linearization	Reverse	CGGGATGGTAATGCCGCTAG
<i>smedwi-2</i>	Forward	ggctagcggcattaccatcccAGGTAAAATGGGAAATCGTG
	Reverse	taccgggcccggccaattctaccgTCCTGTGTATTTTGTAACAGC
<i>smedwi-3</i>	Forward	ggctagcggcattaccatcccGAACACGTATGGGCTACAA
	Reverse	taccgggcccggccaattctaccgCTTGTTCTTGCAAATGCGAT

---

<i>dapk1</i>	Forward	ggctagcggcattaccatcccgCAAGCAGTAATGGACATACAG
	Reverse	taccgggccccggccaattctaccgTGCACATTGGGATTTTCATTATC
<i>histone H2B</i>	Forward	ggctagcggcattaccatcccgATGGCTCCAAAAGCAAAAG
	Reverse	taccgggccccggccaattctaccgGGTAACAGCTTTGGTTCC
<i>traf6</i>	Forward	ggctagcggcattaccatcccgATGTCTGAAAGCGTTTGG
	Reverse	taccgggccccggccaattctaccgGGAAACGAACCGAAATTG
<i>ank1</i>	Forward	ggctagcggcattaccatcccgGCTGATCGGATCATTCAA
	Reverse	taccgggccccggccaattctaccgGTCAACAGGGTTTCTCC

---

**Table A.4** Primer sequences used for in vitro transcription template preparation. Part of the primer corresponding to T7 promoter overhang is in lower-case.

---

<b>Primer</b>	<b>Sequence</b>
T7+AA18 for T4P	ttgtaatacgaactcactataggagCCACCGGTTCCATGGCTAG
PR244 rev T4P	GGCCCAAGGGGTTATGTGG

---

### Next-generation sequencing library preparation

**Table A.5** Primer and adapter sequences used for small RNA-seq library preparation. The index part of the Primer 2 is in bold. Standard Illumina TruSeq indexes were used in the study. C7 stands for C7 amino linker modification at the 3' end. 5'(P) is a 5'-phosphate group.

Primer	Sequence
3'-adapter	5'(P)-TGGAATTCTCGGGTGCCAAGG-C7
5'RNA adapter	GUUCAGAGUUCUACAGUCCGACGAUC
RT Primer (RTP)	TGGAATTCTCGGGTGCCAAGG
Primer 1 (RP1)	AATGATACGGCGACCACCGAGATCTACACGTTTCAGAG TTCTACAGTCCGA
Primer 2 (=TruSeq index5 primer)	CAAGCAGAAGACGGCATA <b>CGAGATCACTGTGTGACT</b> GGAGTTCCTTGGCACCCGAGAATTCCA

**Table A.6** Primer and adapter sequences used for HITS-CLIP library preparation. The index part of the Primer 2 is in bold. Standard Illumina TruSeq indexes were used in the study. Inv.dT stands for inverted dT. 5'(P) is a 5'-phosphate group.

Primer	Sequence
adapter RA3(-P)	UGGAAUUCUCGGGUGCCAAGG-inv.dT
adapter RA3(+P)	5'(P)-UGGAAUUCUCGGGUGCCAAGG-inv.dT
RNA RA5 adapter	GUUCAGAGUUCUACAGUCCGACGAUC
RT Primer (RTP)	TGGAATTCTCGGGTGCCAAGG
DP5 primer	GTTTCAGAGTTCTACAGTCCGACGATC

---

Primer 1 (RP1)	AATGATACGGCGACCACCGAGATCTACACGTTTCAGAG TTCTACAGTCCGA
Primer 2 (=TruSeq in- dex5 primer)	CAAGCAGAAGACGGCATACGAGAT <b>CACTGTGTGACT</b> GGAGTTCCTTGGCACCCGAGAATTCCA

---

**Table A.7** Primer and adapter sequences used for Degradome-seq library preparation. The index part of the Primer 2 is in bold. Standard Illumina TruSeq indexes were used in the study.

---

<b>Primer</b>	<b>Sequence</b>
RNA RA5 adapter	GUUCAGAGUUCUACAGUCCGACGAUC
Deg-RT primer	GCACCCGAGAATTCCANNNNNNNN
Deg-PCR-1l	CTACACGTTTCAGAGTTCTACAGTCCGA
Deg-PCR-1r	GCCTTGGCACCCGAGAATTCCA
Primer 1 (RP1)	AATGATACGGCGACCACCGAGATCTACACGTTTCAGAG TTCTACAGTCCGA
Primer 2 (=TruSeq in- dex5 primer)	CAAGCAGAAGACGGCATACGAGAT <b>CACTGTGTGACT</b> GGAGTTCCTTGGCACCCGAGAATTCCA

---



# Appendix B

## Experimental protocols

### B.1 dsRNA synthesis for RNA interference

Day 1

1. Template preparation for dsRNA synthesis.

- Mix the following components to prepare master mix for 10 PCR reactions (50  $\mu$ l each):

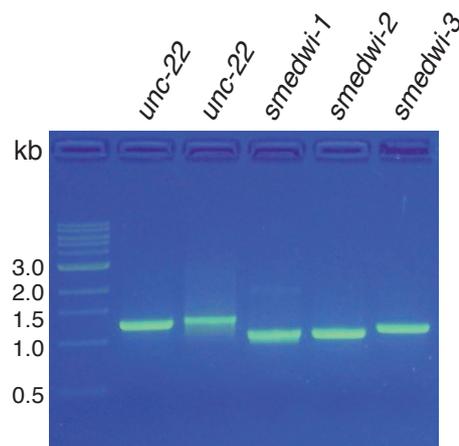
Volume ( $\mu$ l)	Reagent
457.4	H <sub>2</sub> O
55	10x ThermoPol Reaction buffer (NEB, B9004S)
11	10 $\mu$ M T7+AA18 for T4P
11	10 $\mu$ M PR244 rev T4P
11	10 mM dNTPs mix (NEB, N044S)
4.6	Taq DNA Polymerase (NEB, M0267S)
1 ng (per reaction)	pPR-T4P plasmid with a gene fragment

- PCR amplification program

Temperature	Time	Cycles
95°C	30''	
95°C	10''	
60°C	10''	20
68°C	1'40''	
68°C	3'	
4°C	∞	

## 2. PCR product purification

- Purify PCR product with QIAquick PCR Purification Kit (QIAGEN, 28104) according to manufacturer's instruction.
- Elute PCR product in 30  $\mu$ l of Elution buffer
- Control the quality and size of the amplified product on 0.8% agarose gel (Figure B.1).



**Fig. B.1** PCR product with flanking T7 RNA Polymerase promoter regions will serve as a template for in vitro transcription. Amplified products correspond to the following cDNA fragments: nt 6616 - 7358 for *unc-22*, nt 1579 - 2555 for *smedwi-1*, nt 81 - 975 for *smedwi-2* and nt 1347 - 2332 for *smedwi-3*.

## 3. In vitro transcription

- Mix the transcription reaction at RT to avoid the precipitation of DNA by spermidin, which is present in 10x IVT buffer (400 mM Tris-HCl (pH 8.0), 50 mM

DTT, 10 mM Spermidin, 0.1% (v/v) Triton X-100). However, all stock solutions should be kept on ice.

Volume ( $\mu$ l)	Reagent
9.5	H <sub>2</sub> O
10 (1 $\mu$ g)	Template (100 ng/ $\mu$ l)
12.5	20 mM rNTPs mix
2	500 mM MgCl <sub>2</sub>
10	40% PEG 8000
5	10x IVT buffer
1	T7 RNA Polymerase (no less than 5 mg/ml)

Incubate the reaction at 37°C for 4 hours or overnight at RT.

#### 4. DNase I treatment and dsRNA precipitation

- Add 2  $\mu$ l of DNase I (Roche, 04716728001) and 5.7  $\mu$ l of 10x DNase incubation buffer to the transcription reaction. Incubate the mixture for 15 min at 37°C and 800 rpm.
- Stop the reaction by adding 6.4  $\mu$ l of 500 mM EDTA (pH 8.0). Spin the mixture for 2 min at 21,130 g to precipitate inorganic pyrophosphate (white precipitate).
- Transfer the cleared reaction mix into new 1.5 ml low-binding tube. Add 7.1  $\mu$ l 3 M CH<sub>3</sub>COONa, 200  $\mu$ l 100% ethanol (2.5 volumes of the reaction mix). Incubate transcribed dsRNA overnight at -20°C.

#### Day 2

- Precipitate dsRNA by centrifugation at 21,130 g for 30 min at 4°C.
- Discard the supernatant. Add 800  $\mu$ l 75% ethanol and wash the pellets at 21,130 g for 5 min at 4°C .
- Carefully remove the supernatant and air-dry dsRNA pellets for 3 min.
- Resuspend dsRNA in 30 - 50  $\mu$ l of H<sub>2</sub>O incubating the pellet for 30 min at 37°C and 800 rpm.

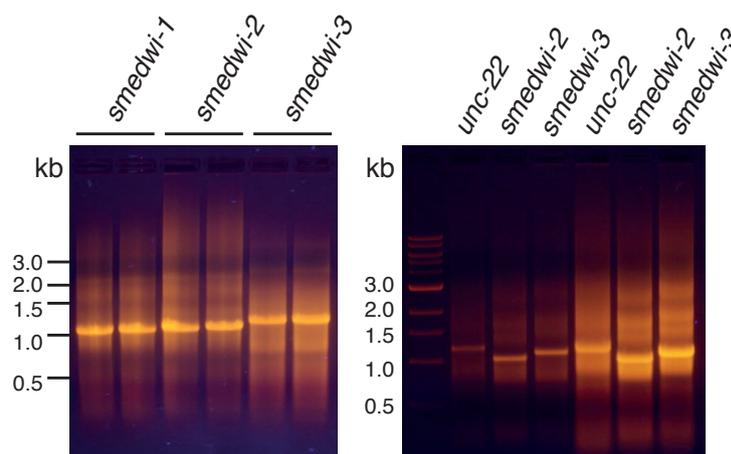
## 5. dsRNA annealing

- Transfer dsRNA into PCR-tubes and anneal two RNA strands using following PCR program:

90°C	3'0"
ramp to 70°C	0.1 °C/sec
70°C	3'0"
ramp to 50°C	0.1 °C/sec
50°C	3'0"
ramp to 25°C	0.1 °C/sec
25°C	5'0"

## 6. Examine in vitro transcribed dsRNA with agarose electrophoresis

- Mix 1  $\mu$ l of dsRNA with 20  $\mu$ l of 1x Formamide loading dye.  
2x Formamide loading dye: 95% Formamide (v/v), 18 mM EDTA (pH 8.0), 0.025% SDS, 0.025% Bromophenol blue (w/v), 0.025% Xylene cyanol (w/v).
- Prepare 0.8% agarose gel in 1x TAE buffer with 0.5  $\mu$ g/ml Ethidium Bromide to detect dsRNA. Run the gel at 130 V for 1 hour (Figure B.2).



**Fig. B.2** In vitro transcribed dsRNA.

## 7. RNAi food preparation

- Aliquot 75  $\mu$ l of 90% liver mixture (in 1x Monjuic solution) into PCR tubes.

- 
- Mix 25  $\mu\text{l}$  of 2  $\mu\text{g}/\mu\text{l}$  dsRNA (50  $\mu\text{g}$ ) with 2  $\mu\text{l}$  of red food dye (E124) to color RNAi food.
  - Add dsRNA/dye solution to 75  $\mu\text{l}$  of 90% liver mixture. Final concentration of dsRNA in the liver is 0.5  $\mu\text{g}/\mu\text{l}$ .
  - Feed worms with the RNAi food for 1 hour in the darkness.

## B.2 Immunoprecipitation of SMEDWI

### Day 1

#### 1. Lysate preparation

- To prepare worm lysate sufficient for 6 samples collect approximately 100 worms (7 - 10 mm), discard 1x Montjuic solution and snap-freeze them in liquid nitrogen.
- Transfer the frozen worms into Dounce homogenizer and add 4 ml of ice-cold lysis buffer (30 mM HEPES (pH 7.7) (Sigma, H989), 150 mM NaCl, 10 mM KCl, 4 mM MgCl<sub>2</sub>, 1 mM DTT, 0.5% Triton X-100, Complete EDTA-free protease inhibitor (Roche, 4693159001)). Dounce stroke the sample for 15 times.
- Clear planarian lysate by centrifugation at 50,000 g for 30 min at 4°C.

Optimal\*: After the centrifugation step filter planarian lysate through a 0.20  $\mu$ m cellulose acetate syringe filter (LLG labware, 9.055501) to remove lipids, which interfere with Bradford assay.

- Measure protein concentration in the lysate with Bradford assay. Expected concentration is  $\geq 4.5$  mg/ml.

#### 2. Indirect immunoprecipitation

- Aliquot 3 mg of planarian lysate into 1.5 ml low-binding tubes (Eppendorf, 0030108051).
- For each immunoprecipitation assay incubate the lysate with 10  $\mu$ g of purified antibody or 30  $\mu$ l of antiserum or pre-immune serum for 2 hours at 4°C with gentle rotation.
- To each sample add 50  $\mu$ l of washed twice in lysis buffer Protein A Dynabeads slurry (Invitrogene, 10002D). Continue incubation for another 2 hours.
- Place the tubes in the magnet stand for 1 min until the supernatant appears clear.
- Discard the supernatant, remove tube from the magnet stand and resuspend Dynabeads in low salt buffer.

Low salt buffer: 30 mM HEPES (pH 7.7), **150 mM NaCl**, 10 mM KCl, 4 mM MgCl<sub>2</sub>, 5 mM EDTA, 1 mM DTT, 0.1% Triton X-100

- Repeat the wash step one more time for a total of two washes with low salt buffer.
- Wash Dynabeads once with high salt buffer.

High salt buffer: 30 mM HEPES (pH 7.7), **300 mM NaCl**, 10 mM KCl, 4 mM MgCl<sub>2</sub>, 5 mM EDTA, 1 mM DTT, 0.1% Triton X-100

- Transfer washed beads into new 1.5 ml low-binding tube.
- Repeat the wash step with high salt buffer one more time for a total of two washes.
- Discard the supernatant and resuspend the beads in 200  $\mu$ l of Proteinase K buffer.

Proteinase K buffer: 200 mM Tris-HCl (pH 7.5) 300 mM NaCl, 25 mM EDTA, 1.5% SDS

- Add 2  $\mu$ l of Proteinase K (12 mg/ml) (Roche, 03115879001) to the final concentration of 120  $\mu$ g/ml. Incubate the beads with Proteinase K for 20 min at 42°C and 1000 rpm.

### 3. Phenol-Chloroform extraction

- Add 2 volumes (400  $\mu$ l) of phenol-chloroform-isoamyl alcohol (P/C/I) solution for RNA extraction (Roth, X985.1) and vortex for 1 min. Leave the mixture to stand for 5 min at RT. To separate the aqueous layer containing RNA, centrifuge the mixture at high speed (>16,000 g) for 5 min at RT.
- Collect the upper aqueous phase in a new low-binding tube.
- Add 1 volume (200  $\mu$ l) of 100% chloroform solution to the sample, vortex thoroughly for 1 - 2 min and centrifuge the mixture at high speed for 5 min at RT. This step removes traces of phenol in the collected aqueous layer.
- Collect the top aqueous solution and place into a new low-binding tube.
- Add 1  $\mu$ l of glycogene (20 mg/ml) (Roche, 10901393001) to each sample and vortex to mix the solution.
- Add 2.5 volumes (375  $\mu$ l) 100% ethanol. Mix solution well and precipitate overnight at -20 °C.

- Spin the solution at high speed for 30 min at 4 °C.
- Decant the supernatant and wash RNA pellet with 1 ml of 70% ethanol at high speed for 5 min at 4 °C.
- Carefully discard supernatant and shortly spin down the tube once again to collect the traces of ethanol.
- Remove residual ethanol and let the tube to air-dry with an open lid for 2 min.
- Resuspend RNA pellet in 17  $\mu$ l of H<sub>2</sub>O.

#### 4. RNA 5'-end dephosphorylation

- Mix the following reagents:

Volume ( $\mu$ l)	Reagent
17	immunoprecipitated RNA
2	10x CutSmart Buffer (NEB, B7204S)
1	CIP (NEB, M0290)

Incubate the mixture for 30 min at 37 °C and 800 rpm.

#### 5. Trifast extraction

- Add 500  $\mu$ l peqGOLD Trifast (peqlab, 30-2010) to the sample and mix thoroughly for 1 min. Leave the solution to stand for 5 min at RT.
- Add 100  $\mu$ l of 100% chloroform and vortex vigorously for 1 min. Incubate the mixture for 2 min at RT.
- Separate the aqueous layer containing RNA by centrifugation at high speed for 5 min at RT.
- Collect the top aqueous phase into new 1.5 ml low-binding tube.
- Add 1  $\mu$ l of glycogene and vortex to mix the solution.
- Add 250  $\mu$ l of 100% isopropanol. Mix solution well and precipitate RNA at RT for 10 min.
- Spin the solution at high speed for 15 min at RT.
- Carefully discard the supernatant and wash RNA pellet with 1 ml of 70% ethanol at high speed for 5 min at RT or 4 °C.

- Carefully discard the supernatant and shortly spin down the tube once again to collect the traces of ethanol.
- Remove residual ethanol and let the tube to air-dry with an open lid for 2 min.
- Resuspend RNA pellet in 16  $\mu$ l of H<sub>2</sub>O.

#### 6. RNA 5'-end labeling with [ $\gamma$ -<sup>32</sup>P]-ATP

- To label RNA 5'-end prepare the following mixture:

Volume ( $\mu$ l)	Reagent
16	dephosphorylated RNA
2	10x PNK buffer (NEB, B0201S)
1	[ $\gamma$ - <sup>32</sup> P]-ATP
1	T4 PNK (NEB, M0201L)

#### 7. Purification of labeled RNA with illustra Microspin G50 columns (GE Healthcaare, 27-5330-01) according to manufacturer's instructions:

- Re-suspend the resin in the column by vortexing.
- Loosen the cap one-quarter turn and twist off the bottom closure.
- Place the column in the supplied Collection tube. Spin down for 1 min at 735 g.
- Place the column into a fresh 1.5 ml low-binding tube and apply sample to the top-center of the resin, being careful not to disturb the resin bed.
- Spin for 2 minutes at 735 g. The purified sample is collected in the bottom of the 1.5 ml microcentrifuge tube.

#### 8. Denaturing RNA Urea-PAGE

- Mix labeled purified sample with 2x Formamide loading dye (95% Formamide (v/v), 18 mM EDTA (pH 8.0), 0.025% SDS, 0.025% Bromophenol blue (w/v), 0.025% Xylene cyanol (w/v)).
- Apply 10  $\mu$ l of immunoprecipitated RNA on 10% Urea gel. For whole RNA extract load  $\leq 1$  Bq/cm<sup>2</sup> of radioactive substance.
- Separate labeled RNA on 16 cm long 10% Urea gel at 21 mA for 1 hour 20 min.

- Expose the gel to a phosphoimager plate for a few hours or overnight, and scan the plate.

## B.3 TruSeq small RNA library preparation

### Adenylation of 3'-adapter

Day1

1. Prepare the following mix using 5' DNA Adenylation Kit (NEB, E2610S):

Volume ( $\mu$ l)	Reagent
2	3'-Adapter (100 $\mu$ M)
4	5' DNA Adenylation reaction buffer
26	H <sub>2</sub> O
4	ATP (1 mM)
4	Mth RNA ligase

Incubate the mixture for 1 hour at 65°C followed by enzyme inactivation for 5' at 85°C.

2. Phenol-Chloroform extraction

- Add to the mixture 160  $\mu$ l of H<sub>2</sub>O and 400  $\mu$ l of phenol-chloroform-isoamyl alcohol (P/C/I) solution for extraction of nucleic acids (Roth, 0038.1).
- Vortex for 1 min. Leave the mixture to stand for 5 min at RT. To separate the aqueous layer containing the DNA, centrifuge the mixture at high speed (>16,000 g) for 5 min at RT.
- Collect the upper aqueous phase in a new low-binding tube.
- Add 1 volume (200  $\mu$ l) of 100% chloroform solution to the sample, vortex thoroughly for 1 - 2 min and centrifuge the mixture at high speed for 5 min at RT. This step removes traces of phenol in the collected aqueous layer.
- Collect the top aqueous solution and place into a new low-binding tube.
- Add 1  $\mu$ l of glycogene (20 mg/ml) (Roche, 10901393001) to each sample and vortex to mix the solution.
- Add 2.5 volumes (375  $\mu$ l) 100% ethanol. Mix solution well and precipitate overnight at -20 °C.

**Day2**

- Spin the solution at high speed for 30 min at 4 °C.
- Decant the supernatant and wash DNA pellet with 1 ml of 70% ethanol at high speed for 5 min at 4 °C.
- Carefully discard the supernatant and shortly spin down the tube once again to collect the traces of ethanol.
- Remove residual ethanol and let the tube to air-dry with an open lid for 2 min.
- Resuspend DNA pellet in 20  $\mu$ l of H<sub>2</sub>O.
- Measure the concentration of adenylated adapter (ssDNA) with spectrophotometer. Adjust the total concentration of the adapter to 5  $\mu$ M (= 5 pmol/ $\mu$ l).

$$c \left[ \frac{\text{pmol}}{\mu\text{l}} \right] = \frac{c \left[ \frac{\text{ng}}{\mu\text{l}} \right] * 10^3}{7138 \left[ \frac{\text{g}}{\text{mol}} \right]};$$

$$\text{added } H_2O [\mu\text{l}] = \left[ \left( \frac{c \left[ \frac{\text{pmol}}{\mu\text{l}} \right]}{5} \right) * V_{Sol} [\mu\text{l}] \right] - V_{Sol} [\mu\text{l}]$$

MW of adenylated 3'-Adapter = 7138 g/mol

$V_{Sol}$  = Volume of adenylated 3'-Adapter (after concentration measurement)

**small RNA-seq library preparation****Day1**

## 1. Ligation of adenylated 3'-adapter

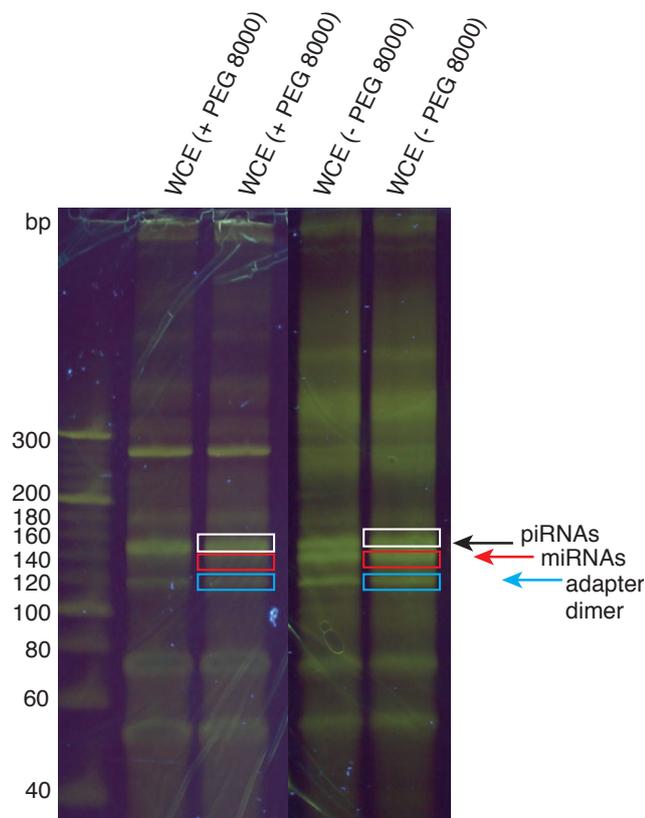
- Prepare the following reaction mix and pipette it into PCR tube:

Volume ( $\mu$ l)	Reagent
11	RNA IP (or 1 $\mu$ g of total RNA)
2	10x T4 RNA Ligase Reaction Buffer
4	50% PEG 8000
1	5 $\mu$ M adenylated 3'-adapter
Heat 30'' at 90°C	
2	T4 RNA Ligase 2, trunc. K227Q (NEB, M0351S)

- Incubate the mixture overnight ( $\approx 16$  hours) at  $16^{\circ}\text{C}$  (in PCR cycler).

**Day2**

- Incubate additional 1 hour at  $37^{\circ}\text{C}$ .
- Inactivate the enzyme for 10 min at  $65^{\circ}\text{C}$ .
- Spin down the reaction mixture and place on ice. The addition of PEG 8000 in the reaction mix considerably increases the ligation rate of 3'-adapter to piRNAs (Figure B.3).



**Fig. B.3** Small RNA-seq libraries prepared in the presence or absence of 10% PEG 8000 during 3'-adapter ligation. In the presence of PEG 8000 miRNAs band is not visible in the final library because of the high abundance of cloned piRNA species. Libraries -PEG 8000 had to be amplified deeper in order to see the piRNA band.

## 2. Ligation of RNA 5' adapter

- Add following components to the ligated RNA-3'adapter:

Volume ( $\mu$ l)	Reagent
20	RNA-3'adapter from the previous step
2	10x T4 RNA Ligase Reaction Buffer
0.4	100 mM ATP
3	100% DMSO
11.6	H <sub>2</sub> O
11.6	10 $\mu$ M 5' RNA adapter
Heat 30'' at 90°C	
2	T4 RNA Ligase 1 (NEB, M0204S)

Incubate the mixture for 1 hour at 37°C.

- Spin down the reaction mix and place on ice. Store the sample at -80°C.

### 3. cDNA Synthesis

- To reverse transcribe RNA add to the PCR tube:

Volume ( $\mu$ l)	Reagent
11	RNA from ligation reaction
1	100 $\mu$ M RT Primer (RTP primer)
4	5x first strand buffer
1.5	100 mM DTT
Incubate for 5 min at 65°C	
Place for >1 min on ice	
1	10 mM dNTP mix
0.5	100 mM DTT
1	SuperScript III RT (Invitrogene, 18080093)

– Vortex the mixture and spin down. Place into PCR cyclor.

- PCR program:

50°C	50'0"
85°C	5'0"
4°C	∞

#### 4. Pilot PCR amplification

- To check successful cloning and to estimate required number of cycles, prepare the following master mix:

Volume ( $\mu$ l)	Reagent
6.25	H <sub>2</sub> O
2	5x Phusion HF buffer
0.2	10 $\mu$ M Primer 1 (RP1)
0.2	10 $\mu$ M Primer 2 (=TruSeq index primer)
0.25	10 mM dNTP mix
0.1	Phusion Polymerase
+1	cDNA

- PCR amplification program. Try different number of cycles (12, 14 and 16 cycles):

Temperature	Time	Cycles
98°C	1'0"	
98°C	10"	
60°C	10"	16
72°C	30"	
72°C	3'	
4°C	∞	

#### 5. Analyze amplified products on 16 cm 6% PAGE.

- Prepare O'RangeRuler 20bp DNA Ladder (Thermo Scientific, SM1323) by mixing 8  $\mu$ l of the ladder with 2  $\mu$ l 5x Phusion HF buffer to adjust salt concentration in relation to PCR reaction.

- Add 2  $\mu$ l of 6x DNA loading dye (50% Glycerol in TE buffer, 0.25% Bromphenol Blue (w/v), 0.25% Xylene cyanol (w/v)) and 1  $\mu$ l of Urea Diluent to the PCR reaction and prepared ladder.
- Load samples onto 16 cm 6% PAGE and separate amplified products in 1x TBE buffer at 180 V for 3 hours.
- Stain the gel with SYBR Gold Nucleic Acid Gel Stain (Invitrogene, S11494) in 1x TBE for 5 min.

A band appearing at the size of 118 bp corresponds to adapter-adapter dimer. Construct with cloned planarian piRNAs ( $\approx$ 32 nts) migrates at 150 bp (Figure B.3).

### Day3

#### 6. Scale up PCR amplification

- Add to the PCR tube the following components:

Volume ( $\mu$ l)	Reagent
26.25	H <sub>2</sub> O
10	5x Phusion HF buffer
1	10 $\mu$ M Primer 1 (RP1)
1	10 $\mu$ M Primer 2 (=TruSeq index primer)
1.25	10 mM dNTP mix
0.5	Phusion Polymerase
+10	cDNA

- PCR amplification program:

Temperature	Time	Cycles
98°C	1'0"	
98°C	10"	
60°C	10"	16
72°C	30"	
72°C	3'	
4°C	∞	

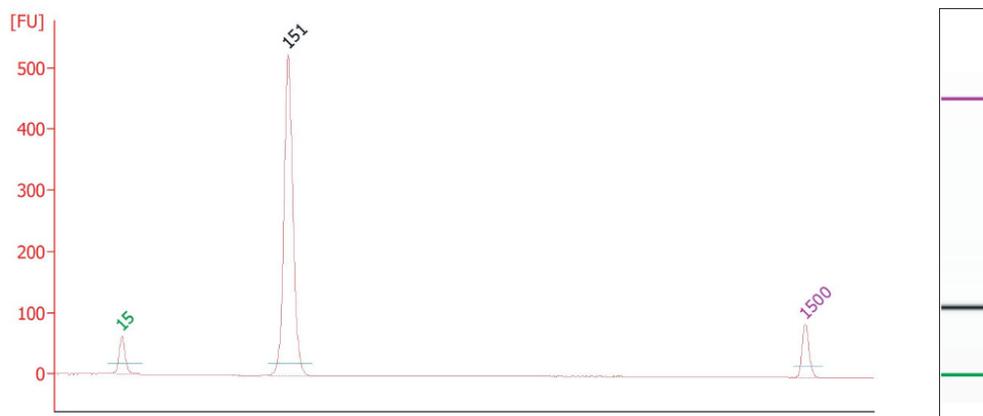
7. Separate amplified products on 16 cm 6% PAGE.

- Prepare O'RangeRuler 20bp DNA Ladder (Thermo Scientific, SM1323) by mixing 8  $\mu$ l of the ladder with 2  $\mu$ l 5x Phusion HF buffer to adjust salt concentration in relation to PCR reaction.
- Add 2  $\mu$ l of 6x DNA loading dye (50% Glycerol in TE buffer, 0.25% Bromphenol Blue (w/v), 0.25% Xylene cyanol (w/v)) and 1  $\mu$ l of Urea Diluent (Roth, 3047.1) to the PCR reaction and prepared ladder.
- Load samples onto 16 cm 6% PAGE splitting one reaction onto two lanes and separate amplified products in 1x TBE buffer at 180 V for 3 hours.
- Stain the gel with SYBR Gold Nucleic Acid Gel Stain (Invitrogen, S11494) in 1x TBE for 5 min.
- Cut out appropriate bands and place into 0.5 ml tubes punctured 2 - 3 times at the bottom with a gauge needle (0.8 mm).
- Place 0.5 ml tubes with gel pieces into 1.5 ml low-binding tubes and spin for 2 min at 16,000 g to shred gel into pieces.
- Discard 0.5 ml tubes.
- Add 300  $\mu$ l of elution buffer (300 mM NaCl, 2 mM EDTA (pH 8.0)) to each tube and incubate overnight at 25°C and 1000 rpm or for 2 hours at 37°C and 1,000 rpm.
- To separate eluate from gel pieces, pipette the solution and gel pieces into SpinX centrifuge tube filter (Costar, CLS8163). Spin the tubes at 16,000 g for 2 min.
- Transfer eluate into 1.5 ml low-binding tubes.
- Add 1  $\mu$ l of Glycogene, mix the solution by vortexing.

- To precipitate DNA add 2.5 volume of 100% ethanol and incubate overnight -20°C.

#### Day4

- Spin the solution at high speed (>16,000 g) for 30 min at 4°C.
- Decant the supernatant and wash DNA pellet with 1 ml of 70% ethanol at high speed for 5 min at 4 °C.
- Carefully discard the supernatant and shortly spin down the tube once again to collect the traces of ethanol.
- Remove residual ethanol and let the tube to air-dry with an open lid for 2 min.
- Resuspend DNA pellet in 12  $\mu$ l of H<sub>2</sub>O.
- Analyze small RNA-Seq libraries on the Bioanalyzer or Fragment analyzer (Figure B.4).



**Fig. B.4** Bioanalyzer electropherogram of prepared small RNA library. Peak at 151 bp corresponds to cloned piRNAs of 33 nt long.

## B.4 Planarian rRNA depletion

### 1. Total RNA input.

The following protocol can be used for the efficient depletion of rRNA from 100 - 1500 ng of total RNA. It is possible to scale up the procedure for the larger RNA input by increasing of the amount of biotinylated DNA probes and magnetic beads.

### 2. Hybridization of biotinylated DNA oligonucleotides (40-mers) antisense to rRNA

- To anneal biotinylated DNA 40-mers to rRNA prepare the following mix:

Volume ( $\mu$ l)	Reagent
10	hybridization buffer
10	RNA sample (1 $\mu$ g)
1	100 $\mu$ M biotinylated DNA 40-mers probes

Hybridization buffer: 20 mM Tris-HCl (pH 8.0), 1 M NaCl, 2 mM EDTA

- Gently mix the solution by pipetting and incubate at 68°C for 10 min.
  - Immediately put the tube at 37°C for 30 min.
- ### 3. Prepare magnetic beads Dynabeads MyOne Streptavidin C1 (Invitrogene, 65001) according to the manufacturer's instruction as following:
- Resuspend the beads in the vial. For each sample use 120  $\mu$ l (10  $\mu$ g/ $\mu$ l) of beads slurry.
  - Wash the beads twice with the equal volume (or at least 1 ml) of Solution A (0.1 M NaOH, 0.05 M NaCl, DEPC-treated). Add the washing buffer and incubate the mixture for 2 min. Next, place the tube on a magnet for 1 min and discard the supernatant.
  - Wash the beads once in Solution B (0.1 M NaCl, DEPC-treated).
  - Resuspend the beads in 2xB&W buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 2M NaCl) to a final concentration of 5  $\mu$ g/ $\mu$ l (twice original volume).
  - Separate the resuspended beads into 2 tubes. The beads from tube-1 will be used in the first round of rRNA depletion. Place the beads from tube-2 on a magnet for 1 min. Next, discard supernatant and resuspend in the equal volume of 1xB&W

(10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 1M NaCl). Final beads concentration is 5  $\mu\text{g}/\mu\text{l}$ . Tube-2 will be used in a second depletion step. Keep it at 37°C until required.

#### 4. Capture of RNA-DNA hybrids with magnetic beads

- Briefly centrifuge the tube with total RNA and probes. Add to the mixture the following:

Volume ( $\mu\text{l}$ )	Reagent
100	dilution buffer
120	magnetic beads (5 $\mu\text{g}/\mu\text{l}$ ) in 2xB&W (tube-1)
1	100 $\mu\text{M}$ biotinylated DNA 40-mers probes

Dilution buffer: 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 200 mM NaCl

- Resuspend by pipetting 10 times. Final concentration of NaCl is 1 M. Incubate the solution at 37°C for 15 min. Gently mix the sample occasionally by tapping.
- Place on magnet for 2 min. Carefully remove the supernatant and add additional 120  $\mu\text{l}$  of washed magnetic beads in 1xB&W (tube-2). Incubate the mixture at 37°C for 15 min with occasional gentle tapping.
- Place on a magnet for 2 min. Carefully remove the supernatant into the new tube and place it again on a magnet for another 1 min to remove the traces of magnetic beads from the sample.
- Collect the supernatant into a new tube.

#### 5. Sample concentration, size selection and DNase I treatment with RNA Clean & Concentrator-5 kit (Zymo Research, R1015) as described [206]. Briefly:

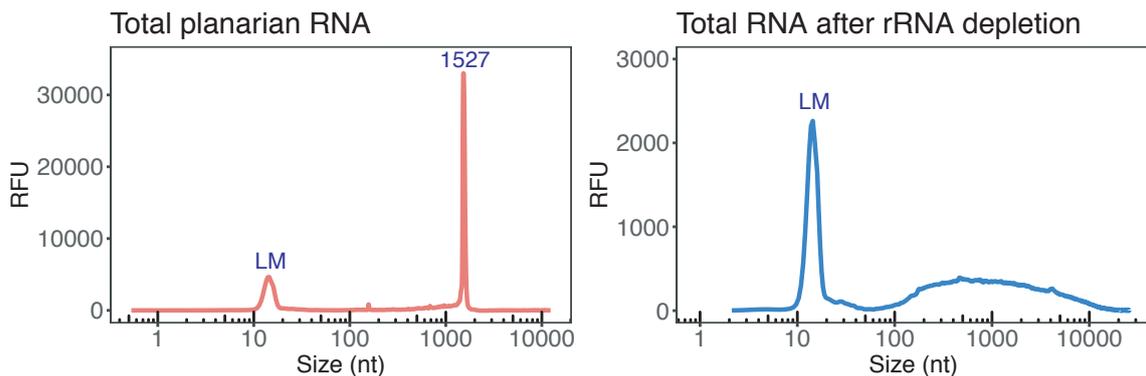
- Mix 180  $\mu\text{l}$  of RNA binding buffer with 180  $\mu\text{l}$  of 100% ethanol. Add this mixture to the 360  $\mu\text{l}$  of depleted RNA and mix.
- Transfer the buffer/ethanol/RNA mixture into a Zymo-Spin IC column. Centrifuge at 16,000 g for 1 min. Discard the flow-through.
- Add 400  $\mu\text{l}$  of RNA Wash buffer to the column, centrifuge at 16,000 g for 1 min. Discard the flow-through.

- To degrade contaminating DNA, mix the following reagents:

Volume ( $\mu\text{l}$ )	Reagent
1	DNase I (10 U/ $\mu\text{l}$ )
3	10x DNase I buffer
26	RNA Wash buffer

Apply 30  $\mu\text{l}$  of the mixture onto the column. Incubate the column at 37°C for 30 min. Centrifuge the column at 16,000 g for 1 min. Discard the flow-through.

- Add 400  $\mu\text{l}$  of RNA Prep buffer to the column, centrifuge at 16,000 g for 1 min. Discard the flow-through.
- Add 800  $\mu\text{l}$  RNA Wash buffer to the column, centrifuge at 16,000 g for 1 min. Discard the flow-through.
- Add 400  $\mu\text{l}$  RNA Wash buffer to the column, centrifuge at 16,000 g for 1 min. Discard the flow-through.
- Centrifuge the column at 16,000 g for 2 minutes.
- To elute the RNA, replace the collection tube with a new low-binding 1.5 ml tube, and then add 7  $\mu\text{l}$  of H<sub>2</sub>O to the column (minimum elution volume is 6  $\mu\text{l}$ ). Incubate at RT for 1 min. Centrifuge the column at 16,000 g for 1 min to collect RNA/flow-through (Figure B.5).



**Fig. B.5** Separation profile of planarian total RNA and rRNA depleted RNA. The large peak around 1527 nt corresponds to the 18S rRNAs and 28S rRNA, which was split into two parts during rRNA maturation.

## B.5 SMEDWI-3 HITS-CLIP

### Day 1

#### 1. Dissociate worms into single cells to achieve uniform UV-irradiation

- Pick 50 worms (7.5 - 10 mm) in a Petri dish.
- Rinse worms quickly in a small volume of cold 1x CMFB buffer. Remove as much liquid as possible. Keep the Petri dish on ice.

#### **10x CMF**

Component	For 100 ml solution	Final concentration
NaH <sub>2</sub> PO <sub>4</sub> × 2H <sub>2</sub> O	0.35 g	25.6 mM
NaCl (5 M)	2.85 ml	142.8 mM
KCl (2 M)	5.1 ml	102.1 mM
NaHCO <sub>3</sub>	0.79 g	94.2 mM
ddH <sub>2</sub> O	up to 100 ml	

**Glucose stock:** prepare solution of 120 mg glucose in 1 ml of H<sub>2</sub>O.

**1x CMFB:** 1xCMF; 0.1% BSA; 240 mg/L glucose; 15 mM HEPES-hemiNa (pH 7.7) in H<sub>2</sub>O.

**For 50 ml 1x CMFB:** 5 ml of 10x CMF; 0.5 g BSA; 100 μl of glucose stock (12 mg); 7.5 ml of 100 mM HEPES-hemiNa; 37.4 ml H<sub>2</sub>O.

- With a sharp blade dice the animals and collect them into 15 ml tube with 10 ml 1x CMFB.
- Rock on nutator for 20 min with occasional pipetting (every 2 min pipette up and down for 6 times with 1 ml pipette tip).
- Filter through 50 μm nylon cell strainer (CellTrics, 04-0042-2317) into 15 ml tube. Keep tubes on ice.
- Spin cells at 290 g for 5 min at 4°C, discard supernatant and reconstitute cells in 15 ml of 1x CMFB.
- Split reconstituted cells into two Petri dishes (7.5 ml each + 0.5 ml 1x CMFB) and irradiate with 254 nm UV-light in Stratalinker once with 400 mJ/cm<sup>2</sup> and once with 200 mJ/cm<sup>2</sup> with 30 sec interval in between to cool and mix cells on ice.

- Collect the cells and pellet them at 900 g for 3 min. Freeze in liquid nitrogen and keep at -80°C until required.

## Day 2

### 2. Immunoprecipitation

Use low-binding tubes for all following steps.

- Resuspend each pellet in 750  $\mu$ l of lysis buffer (30 mM HEPES (pH 7.7) (Sigma, H989), 150 mM NaCl, 10 mM KCl, 4 mM MgCl<sub>2</sub>, 1 mM DTT, 0.5% Triton X-100, 0.8 U/ $\mu$ l RNasin (Promega, N2515), Complete EDTA-free protease inhibitor (Roche, 03115879001)) and let it stand for 10 min on ice.
- Treat the lysate with 30 U (3  $\mu$ l) of DNase I per 750  $\mu$ l of lysate to digest genomic DNA and decrease viscosity of the solution. Incubate for 5 min at 37°C and 1,000 rpm.
- Ultracentrifuge lysate for 30 min at 50,000 g. Measure the concentration with Bradford assay.
- Add 18  $\mu$ g of anti-SMEDWI-3 AB per sample (3 mg of cleared lysate) or 30  $\mu$ l of pre-immune serum as a negative control. Incubated for 1 hour at 4°C with gentle rotation.
- Add 150  $\mu$ l of Dynabeads Protein A slurry (Invitrogen, 10002D) washed twice in lysis buffer to each tube. Incubate for another 2 hours rotating at 4°C.
- Wash the beads two times with low-salt buffer.  
Low salt buffer: 30 mM HEPES (pH 7.7), **150 mM NaCl**, 10 mM KCl, 4 mM MgCl<sub>2</sub>, 5 mM EDTA, 0.1% Triton X-100, 1 mM DTT
- Wash the beads one time with high-salt buffer.  
High-salt buffer: 30 mM HEPES (pH 7.7), **500 mM NaCl**, 10 mM KCl, 4 mM MgCl<sub>2</sub>, 5 mM EDTA, 0.1% Triton X-100, 1 mM DTT
- Trasfer washed beads into new 1.5 ml low-binding tube.
- Repeat the wash step with high salt buffer one more time for a total of two washes.
- Wash the beads two times with 1x PNK buffer.  
1x PNK buffer: 70 mM Tris-HCl (pH 7.5), 10 mM MgCl<sub>2</sub>, 5 mM DTT, 0.1% Triton X-100

### 3. RNA 3'-end dephosphorylation with Antarctic phosphatase (AP)

- Prepare the following master mix to resuspend washed beads:

Volume ( $\mu$ l)	Reagent
69	H <sub>2</sub> O
8	10x AP buffer
1	RNasin
2	AP enzyme (NEB, M0289S)

Add 80  $\mu$ l of the master mix to each sample. Incubate for 20 min at 37°C and 1,000 rpm.

- Wash beads with ice-cold buffer:
  - Three times with 1x PNK buffer
  - Once with 1x PNK buffer without DTT and Triton X-100 (70 mM Tris-HCl (pH 7.5), 10 mM MgCl<sub>2</sub>).

### 4. Radioactive labeling of 3'-adapter

- Radioactively label adapter beforehand as following:

Volume ( $\mu$ l)	Reagent
9.5	H <sub>2</sub> O
2.5	50 $\mu$ M RA3(-P)
25	[ $\gamma$ - <sup>32</sup> P]-ATP
5	10x PNK buffer
8	T4 PNK enzyme (NEB, M0201S)

- Incubate the mixture for 30 min at 37°C. Then add 1  $\mu$ l of 10 mM ATP and incubate the mixture for additional 5 min.
- Prepare Microspin G-25 columns (GE Healthcare, 27-5325-01). Resuspend the resin in a G-25 column by vortexing it upside down, and then break off the bottom seal and loosen the cap. Pre-centrifuge the column for 1 min at 735 g, and transfer the column to a fresh RNase-free microcentrifuge tube.

- Apply the phosphorylated adapter sample to the resin and spin the column for 2 min at 735 g at RT.
- Heat inactivate residual T4 PNK in the eluted labeled adapter RA3 for 20 min at 65°C.
- The labeled adapter can be used immediately or can be stored at -20°C until needed.

#### 5. 3'-adapter ligation

- Resuspended the beads in the following master mix:

Volume ( $\mu$ l)	Reagent
28	H <sub>2</sub> O
8	10x T4 RNA ligase buffer
2	100% DMSO
16	50% PEG 8000
8	10 mM ATP
1	RNasin
8	T4 RNA ligase 1 to 1U/ $\mu$ l (NEB, M0204S)
4	[ $\gamma$ - <sup>32</sup> P]-RA3 adapter

Incubate at 16°C for 1 hour with gentle rotation.

- Add additional 5  $\mu$ l of 100  $\mu$ M RA3(+P) and incubate overnight at 16°C with gentle rotation.

### Day 3

- Wash beads with ice-cold buffer:
  - One time Low-salt wash buffer
  - One time High-salt wash buffer
  - Transfer washed beads into new 1.5 ml low-binding tubes
  - Two times with 1x PNK buffer

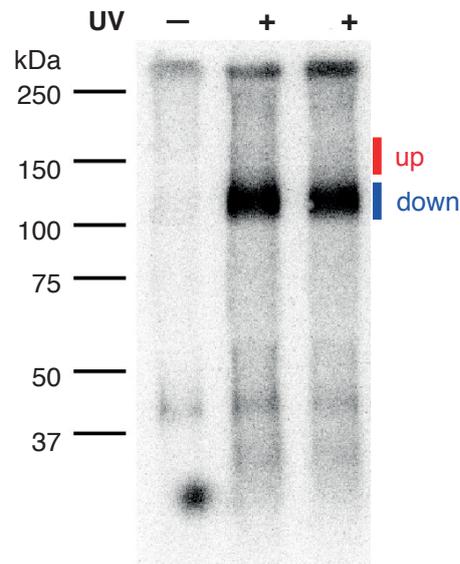
6. Treat the beads with T4 PNK to restore mono phosphate at the 5'-end of cross-linked RNA:

- Add 80  $\mu$ l of PNK mix to each tube:

Volume ( $\mu$ l)	Reagent
59	H <sub>2</sub> O
8	10x T4 PNK buffer
8	10 mM ATP
1	RNasin
0.5	DNAse I (15 U)
4	T4 PNK enzyme (NEB, M0201S)

Incubate for 30 min at 37°C and 1,000 rpm.

- Wash the beads with ice-cold buffer:
    - One time Low-salt wash buffer
    - One time High-salt wash buffer
    - Two times with 1x PNK buffer
  - After removing the final wash buffer, elute protein-RNA complexes with 20  $\mu$ l of 1x SDS loading buffer diluted in 1x PNK buffer.
  - Incubate the mixture for 12 min at 70°C and 1,000 rpm.  
6x SDS buffer: 300 mM Tris (pH 6.8), 12% SDS, 600 mM DTT, 0.6% Bromophenol Blue (w/v), 60% glycerol
  - Place the tube on magnetic stand and collect the supernatant containing RNA-protein complexes.
7. Separate RNA-protein complexes on 8% Bis-Tris PAGE Load samples on 8% Bis-Tris PAGE and run the gel for 55 min at 155 V in 1x high-molecular weight running buffer.  
5x High-molecular weight buffer: 250 mM MOPS, 250 mM Tris base, 5 mM EDTA (pH 8.0), 0.5% SDS
  8. Transfer RNA-protein complexes on nitrocellulose membrane for 1 hour at 90 V in 1x Western blot transfer buffer supplemented with 20% methanol.  
5x Western blot transfer buffer: 120 mM Tris base, 960 mM Glycine, 0.15% SDS
  9. After western blot transfer, rinse the membrane in RNase-free H<sub>2</sub>O, wrap the membrane in transparent plastic foil and expose for 1 hour to a phosphoimager plate (Figure B.6).



**Fig. B.6** SMEDWI-3 HITS-CLIP immunoprecipitated complexes. RNAs crosslinked to SMEDWI-3 were radiolabeled, separated on an 8% Bis-Tris gel and blotted onto a nitrocellulose membrane.

10. Cut out the radioactive signal (upper (7-10 kDa above main signal) and lower band separately) on a light-table.
11. Shred the membrane into 4 mm long pieces and place them in a new low-binding tube.
12. Proteinase K digestion
  - Prepare Proteinase K solution with a concentration 4 mg/ml in 1x PK buffer. Preincubate the solution for 30 min at 37°C to degrade possible present nucleases. 5x PK buffer: 500 mM Tris-HCl (pH 7.5), 300 mM NaCl, 50 mM EDTA
  - Add 200  $\mu$ l of Proteinase K solution to each tube containing membrane fragments. Incubate for 20 min at 37°C and 1,000 rpm.
  - Add 200  $\mu$ l of solution 7 M Urea in 1x PK buffer to each tube. Incubate for 20 min at 37°C and 1,000 rpm.
  - Add 400  $\mu$ l of phenol-chloroform-isoamyl alcohol (P/C/I) solution for RNA extraction and incubate for 20 min at 37°C and 1,000 rpm.
13. Phenol-Chloroform extraction
  - Spin tubes at high speed (>16,000 g) for 5 min at RT. Collect the upper aqueous phase.

- Add 400  $\mu$ l of 100% chloroform, vortex for 1 min, let the tube stand for 2 min, and centrifuge at high speed for 5 min. Collect the upper aqueous phase.
- Concentrate RNA with RNA Clean & Concentrator-5 kit (Zymo Research, R1015).
  - Add to the upper phase (360  $\mu$ l) 2 volumes of binding buffer (720  $\mu$ l).
  - Mix vigorously, separate into 2 tubes (540  $\mu$ l each tube).
  - Add 540  $\mu$ l of 100% ethanol to each tube. Mix and apply the solution into Zymo columns.
  - Spin the columns at 16,000 g for 30 sec. Discard the flow-through. Repeat the procedure until all RNA will be applied onto the columns.
  - Add 400  $\mu$ l of RNA Prep buffer and spin at 16,000 g for 1 min. Discard the flow-through.
  - Add 700  $\mu$ l of Wash buffer and spin at 16,000 g for 1 min. Discard the flow-through.
  - Add 400  $\mu$ l of Wash buffer and spin at 16,000 g for 1 min. Discard the flow-through.
  - Spin the columns for additional 2 min 16,000 g to remove the traces of Wash buffer.
  - For elution place the column into new low-binding tube and apply on top 7  $\mu$ l of H<sub>2</sub>O. Incubate the column for 1 min at RT and spin at 16,000 g for 1 min.

#### 14. Ligation of 5'-adapter

- Set the following reaction and pipette into PCR tubes:

Volume ( $\mu$ l)	Reagent
5.5	eluted from the membrane RNA
1.6	10x T4 RNA ligase buffer
1.5	10 mM ATP (final 1 mM)
0.5	RNasin
2.4	50% PEG 8000 (final 8%)
0.75	50% DMSO (final 2.5%)
3	DNase I (15 U)
1	100 $\mu$ M RNA RA5 adapter
4	T4 RNA ligase 1 (final 2 U/ $\mu$ l) (NEB, M0204S)

Incubate at 16°C overnight ( $\approx$ 16 hours).

#### Day 4

- Add to the mixture 34  $\mu$ l of H<sub>2</sub>O up to the final volume of 50  $\mu$ l and concentrate the sample with RNA clean & concentrate-5 extraction kit from Zymo Research.
  - Mix 50  $\mu$ l of the reaction mix with 100  $\mu$ l of Binding buffer. Mix vigorously.
  - Add 150  $\mu$ l of 100% ethanol to each tube. Mix the solution by vortexing and apply onto Zymo columns. Spin for 1 min at 16,000 g. Discard the flow-through.
  - Add 400  $\mu$ l RNA Prep buffer to each column, centrifuge at 16,000 g for 1 min. Discard the flow-through.
  - Add 700  $\mu$ l of Wash buffer and spin at 16,000 g for 1 min. Discard the flow-through.
  - Add 400  $\mu$ l of Wash buffer and spin at 16,000 g for 1 min. Discard the flow-through.
  - Spin the columns for additional 2 min 16,000 g to remove the traces of Wash buffer.
  - For elution place the column into new low-binding tube and apply on top 12  $\mu$ l of H<sub>2</sub>O. Incubate the column for 1 min at RT and spin at 16,000 g for 1 min.

#### 15. Reverse transcribe crosslinked RNA

- To reverse transcribe RNA add to the PCR tube:

Volume ( $\mu$ l)	Reagent
11	RNA from ligation reaction
1	10 $\mu$ M RT Primer (RTP primer)
4	5x first strand buffer
1.5	100 mM DTT
Incubate for 3 min at 65°C	
Place for >1 min on ice	
1	10 mM dNTP mix
0.5	100 mM DTT
1	SuperScript III RT (Invitrogene, 18080093)

– Vortex the mixture and spin down. Place into PCR cycler.

- PCR program:

25°C	5'0"
50°C	60'0"
70°C	15'0"
4°C	$\infty$

## 16. Pilot PCR amplification

- To check successful cloning and to estimate required number of cycles, prepare the following master mix:

Volume ( $\mu$ l)	Reagent
5.25	H <sub>2</sub> O
2	5x Phusion HF buffer
0.2	10 $\mu$ M RT Primer (RTP primer)
0.2	10 $\mu$ M DP5 primer
0.25	10 mM dNTP mix
0.1	Phusion Polymerase
+2	cDNA

- PCR amplification program. Try different number of cycles (18, 20 and 22 cycles):

Temperature	Time	Cycles
98°C	1'0"	
98°C	10"	
60°C	10"	20
72°C	30"	
72°C	3'	
4°C	$\infty$	

17. Analyze amplified products on 9 cm 6% PAGE.

- Prepare O'RangeRuler 20bp DNA Ladder (Thermo Scientific, SM1323) by mixing 2  $\mu$ l of the ladder with 8  $\mu$ l 1x Phusion HF buffer to adjust salt concentration in relation to PCR reaction.
- Add 2  $\mu$ l of 6x DNA loading dye (50% Glycerol in TE buffer, 0.25% Bromphenol Blue (w/v), 0.25% Xylene cyanol (w/v)) and 1  $\mu$ l of Urea Diluent to the PCR reaction and prepared ladder.
- Load samples onto 9 cm 6% PAGE and separate amplified products in 1x TBE buffer at 160 V for 45 min.
- Stain the gel with SYBR Gold Nucleic Acid Gel Stain (Invitrogene, S11494) in 1x TBE for 5 min.

A smear band corresponding to crosslinked RNA and piRNAs species should appear above 80 bp.

#### 18. Scale up PCR amplification

- Add to the PCR tube the following components:

Volume ( $\mu$ l)	Reagent
26.25	H <sub>2</sub> O
10	5x Phusion HF buffer
1	10 $\mu$ M RT Primer (RTP primer)
1	10 $\mu$ M DP5 primer
1.25	10 mM dNTP mix
0.5	Phusion Polymerase
+10	cDNA

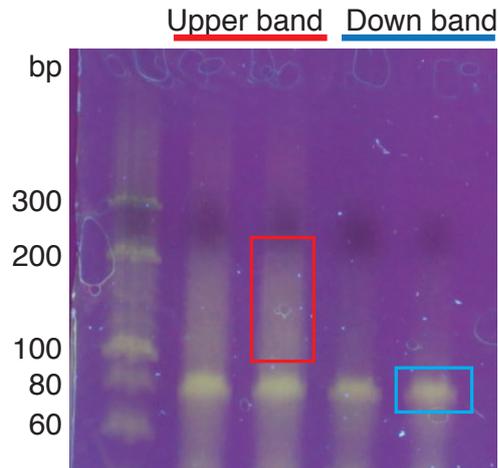
- PCR amplification program:

Temperature	Time	Cycles
98°C	1'0"	
98°C	10"	
60°C	10"	20
72°C	30"	
72°C	3'	
4°C	$\infty$	

#### 19. Separate amplified products on 9 cm 6% PAGE.

- Prepare O'RangeRuler 20bp DNA Ladder (Thermo Scientific, SM1323) by mixing 2  $\mu$ l of the ladder with 8  $\mu$ l 1x Phusion HF buffer to adjust salt concentration in relation to PCR reaction.
- Add 2  $\mu$ l of 6x DNA loading dye (50% Glycerol in TE buffer, 0.25% Bromphenol Blue (w/v), 0.25% Xylene cyanol (w/v)) and 1  $\mu$ l of Urea Diluent to the PCR reaction and prepared ladder.

- Load samples onto 9 cm 6% PAGE and separate amplified products in 1x TBE buffer at 160 V for 45 min.
- Stain the gel with SYBR Gold Nucleic Acid Gel Stain (Invitrogene, S11494) in 1x TBE for 5 min (Figure B.7).



**Fig. B.7** PAGE analysis of amplified PCR products from the excised upper or lower band of crosslinked RNA species with ligated adapters and reverse transcribed.

- Cut out appropriate bands and place into 0.5 ml tubes punctured 2 - 3 times at the bottom with a gauge needle (0.8 mm).
- Place 0.5 ml tubes with gel pieces into 1.5 ml low-binding tubes and spin for 2 min at 16,000g to shred gel into pieces.
- Discard 0.5 ml tubes.
- Add 400  $\mu$ l of elution buffer (300 mM NaCl, 2 mM EDTA (pH 8.0)) to each tube and incubate for 2 hours at 37°C and 1,000 rpm.
- To separate eluate from gel pieces, pipette the solution and gel pieces into SpinX centrifuge tube filter (Costar, CLS8163). Spin the tubes at 16,000 g for 2 min.
- Transfer eluate into 1.5 ml low-binding tubes.
- Add 1  $\mu$ l of Glycogene, mix the solution by vortexing.
- To precipitate DNA add 2.5 volume of 100% ethanol and incubate overnight -20°C.

### Day5

- Spin the solution at high speed for 30 min at 4°C.

- Decant the supernatant and wash DNA pellet with 1 ml of 70% ethanol at high speed for 5 min at 4 °C.
- Carefully discard the supernatant and shortly spin down the tube once again to collect the traces of ethanol.
- Remove residual ethanol and let the tube to air-dry with an open lid for 2 min.
- Resuspend DNA pellet in 10  $\mu$ l of H<sub>2</sub>O.

20. Add index primers. Pilot PCR might be necessary to determine the right number of cycle (see as above).

- Add to the PCR tube following components:

Volume ( $\mu$ l)	Reagent
26.25	H <sub>2</sub> O
10	5x Phusion HF buffer
1	10 $\mu$ M Primer 1 (RP1)
1	10 $\mu$ M Primer 2 (=TruSeq index primer)
1.25	10 mM dNTP mix
0.5	Phusion Polymerase
+10	cDNA

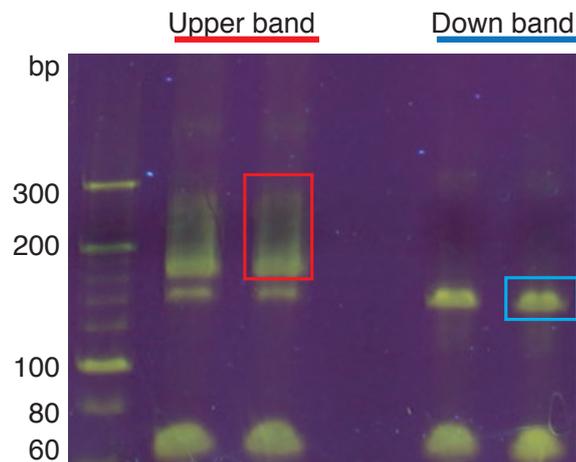
- PCR amplification program:

Temperature	Time	Cycles
98°C	1'0"	
98°C	10"	
60°C	10"	6
72°C	30"	
72°C	3'	
4°C	$\infty$	

21. Use 2.5 volume of Agencourt AMPure XP beads (BeckmanCoulter, A63880) to concentrate the PCR sample. Elute amplified DNA from the beads with 12  $\mu$ l of H<sub>2</sub>O.

## 22. Separate amplified products on 9 cm 6% PAGE.

- Prepare O'RangeRuler 20bp DNA Ladder by mixing 2  $\mu$ l of the ladder with 8  $\mu$ l 1x Phusion HF buffer to adjust salt concentration in relation to PCR reaction.
- Add 2  $\mu$ l of 6x DNA loading dye (50% Glycerol in TE buffer, 0.25% Bromphenol Blue (w/v), 0.25% Xylene cyanol (w/v)) and 1  $\mu$ l of Urea Diluent to the PCR reaction and prepared ladder.
- Load samples onto 9 cm 6% PAGE and separate amplified products in 1x TBE buffer at 160 V for 45 min.
- Stain the gel with SYBR Gold Nucleic Acid Gel Stain (Invitrogene, S11494) in 1x TBE for 5 min (Figure B.8).



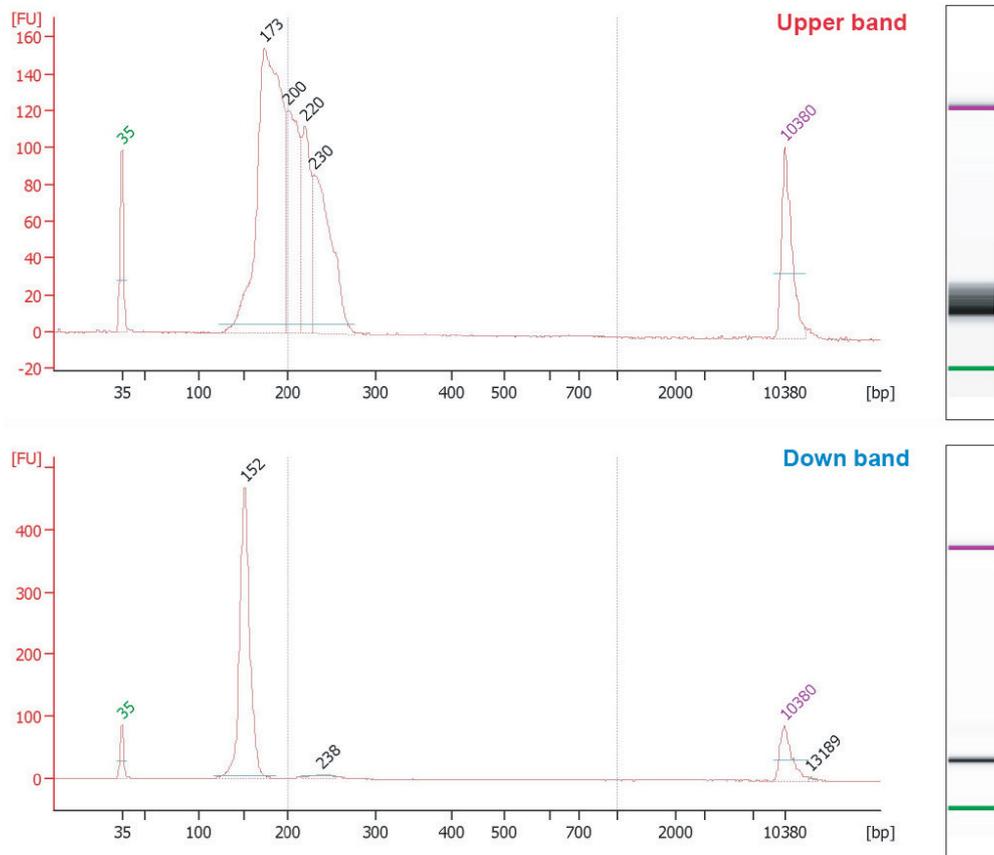
**Fig. B.8** PAGE analysis of HITS-CLIP libraries amplified from the excised upper or lower band.

- Cut out appropriate bands and place into 0.5 ml tubes punctured 2 - 3 times at the bottom with a gauge needle (0.8 mm).
- Place 0.5 ml tubes with gel pieces into 1.5 ml low-binding tubes and spin for 2 min at 16,000 g to shred gel into pieces.
- Discard 0.5 ml tubes.
- Add 400  $\mu$ l of elution buffer (300 mM NaCl, 2 mM EDTA (pH 8.0)) to each tube and incubate for 2 hours at 37°C and 1,000 rpm.
- To separate eluate from gel pieces, pipette the solution and gel pieces into SpinX centrifuge tube filter (Costar, CLS8163). Spin the tubes at 16,000 g for 2 min.
- Transfer eluate into 1.5 ml low-binding tubes.

- Add 1  $\mu$ l of Glycogene, mix the solution by vortexing.
- To precipitate DNA add 2.5 volume of 100% ethanol and incubate overnight  $-20^{\circ}\text{C}$ .

**Day6**

- Spin the solution at high speed for 30 min at  $4^{\circ}\text{C}$ .
  - Decant the supernatant and wash DNA pellet with 1 ml of 70% ethanol at high speed for 5 min at  $4^{\circ}\text{C}$ .
  - Carefully discard the supernatant and shortly spin down the tube once again to collect the traces of ethanol.
  - Remove residual ethanol and let the tube to air-dry with an open lid for 2 min.
  - Resuspend DNA pellet in 12  $\mu$ l of  $\text{H}_2\text{O}$ .
23. Analyze HITS-CLIP libraries on the Bioanalyzer or Fragment analyzer (Figure B.9).



**Fig. B.9** Bioanalyzer electropherogram of prepared HITS-CLIP libraries. The upper electropherogram corresponds to the library prepared from excised upper band with longer crosslinked RNA species. Below, electropherogram of HITS-CLIP library generated from the down band containing crosslinked piRNAs and shorter RNA species.

## B.6 Degradome-seq library preparation

### 1. Input total RNA

Total planarian RNA (2.5  $\mu\text{g}$ ) from FACS-sorted X1 cells was ribosomal RNA depleted as described above in Appendix B section B.4. RNA was eluted from Zymo Research columns with 12  $\mu\text{l}$  of  $\text{H}_2\text{O}$ .

### 2. Ligation of 5' adapter

- Prepare the following reaction mix:

Volume ( $\mu\text{l}$ )	Reagent
11	rRNA depleted RNA
2	5x 10x T4 RNA ligase buffer
2	10 mM ATP
1.8	100 mM DTT
1	25 $\mu\text{M}$ 5'-adapter (RA5)
2	T4 RNA ligase 1 (NEB, M0204S)

Incubate the mixture overnight at 16°C or for 2 hours at 37°C.

- To the reaction mix add 30  $\mu\text{l}$  of  $\text{H}_2\text{O}$  and purify RNA with RNA Clean & Concentrator-5 (Zymo Research, R1015) enriching for RNA  $\geq 200$  nt long according to manufacturer's instructions. Elute RNA with 11  $\mu\text{l}$  of  $\text{H}_2\text{O}$ .

### 3. cDNA synthesis

- To reverse transcribe RNA add to the PCR tube:

Volume ( $\mu$ l)	Reagent
10	RNA from ligation reaction
1	50 $\mu$ M Deg-RT primer
5	5x first strand buffer
Incubate for 3 min at 65°C	
Place for >1 min on ice	
5.25	H <sub>2</sub> O
1.5	10 mM dNTP mix
1.25	100 mM DTT
1	SuperScript III RT (Invitrogene, 18080093)

– Vortex the mixture and spin down. Place into PCR cyclers.

- PCR program:

25°C	5'0"
50°C	60'0"
70°C	15'0"
4°C	$\infty$

- Purify cDNA with 45  $\mu$ l (1.8 volume) Agencourt AMPure XP beads. Elute with 23  $\mu$ l of H<sub>2</sub>O.

#### 4. First PCR amplification

- Mix the following components:

Volume ( $\mu$ l)	Reagent
20	cDNA
16.25	H <sub>2</sub> O
10	5x Phusion HF buffer
1	10 $\mu$ M DEG PCR-1l primer
1	10 $\mu$ M DEG PCR-1r primer
1.25	10 mM dNTP mix
0.5	Phusion Polymerase

- PCR amplification program.

Try different number of cycles for a different input amount, organism or tissue:

Temperature	Time	Cycles
98°C	0'30"	
98°C	10"	
59°C	30"	2
72°C	12"	
98°C	10"	
68°C	10"	4
72°C	12"	
72°C	3'0"	
4°C	$\infty$	

- Purify PCR product with Agencourt AMPure XP beads with double size selection to keep only fragments of 200 - 400 bp.
  - Mix PCR product (50  $\mu$ l) with 0.6 volume (30  $\mu$ l) of Ampure beads. Incubate the mixture at for 5 min RT.
  - Place the tube on a magnet stand for 5 min to separate the beads from the supernatant.
  - Carefully transfer the supernatant into a new tube (**do not discard the supernatant**). Discard the beads, which now contain fragments larger 400 bp.

- Mix the supernatant from the previous step with additional 40  $\mu$ l of Ampure beads. Incubate 5 min at RT.
- Place the tube on a magnet stand for 5 min to separate the beads from the supernatant.
- Wash the beads twice with 70% ethanol as described [206] and elute DNA with 30  $\mu$ l H<sub>2</sub>O.

#### 5. Second PCR amplification

- Mix the following components:

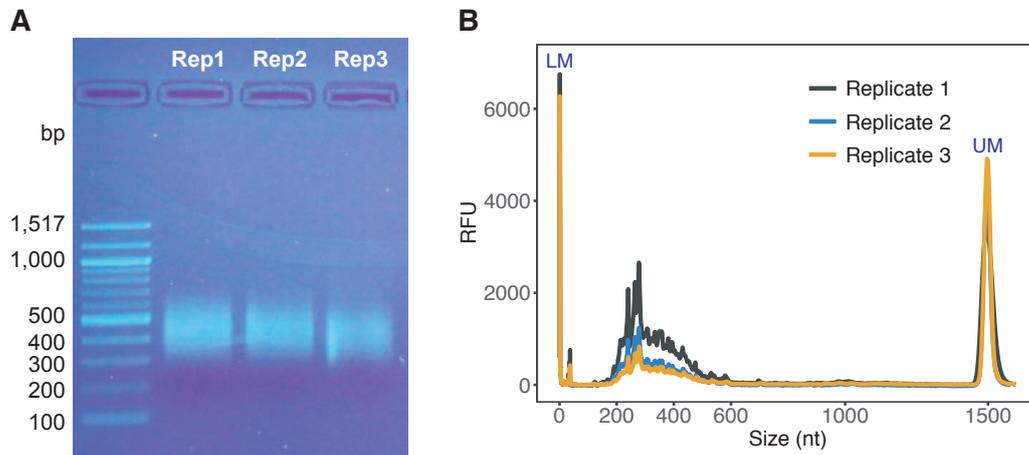
Volume ( $\mu$ l)	Reagent
30	eluted DNA
6.25	H <sub>2</sub> O
10	5x Phusion HF buffer
1	10 $\mu$ M Primer 1 (RP1)
1	10 $\mu$ M Primer 2 (=TruSeq index primer)
1.25	10 mM dNTP mix
0.5	Phusion Polymerase

- PCR amplification program.

Try different number of cycles for a different input amount, organism or tissue:

Temperature	Time	Cycles
98°C	0'30"	
98°C	10"	
68°C	30"	2
72°C	14"	
98°C	10"	6
72°C	14"	
72°C	3'0"	
4°C	$\infty$	

- Purify PCR product with one volume (50  $\mu$ l) Agencourt AMPure XP beads. Elute prepared Degradome-seq libraries with 12  $\mu$ l of H<sub>2</sub>O (Figure B.10).



**Fig. B.10** Separation profile of degradome-seq libraries.

(A) Prepared libraries were analyzed on 1.5% agarose gel. (B) Fragment analyzer electropherogram of degradome-seq libraries. Over amplified PCR duplicates appear as sharp peaks. LM stands for lower ladder with the length 1 nt. UM is an upper ladder with a size 1,500 nt.

# Appendix C

## Data processing pipelines

### C.1 small RNA-seq pipeline

The example pipeline shows the processing of small RNA libraries prepared from SMEDWI-3 immunoprecipitated piRNAs. SMEDWI-1 and -2 bound piRNAs were processed the same way.

```
# fastqc quality control
cd /Work1/iana/Data/Sm3wt/Raw_data
mkdir -p /Work1/iana/Data/Sm3wt/Analysis/1-fastqc
fastqc -o ../Analysis/1-fastqc *.fq.gz

# 3'-adapter clipping
mkdir -p /Work1/iana/Data/Sm3wt/Analysis/2-cutadapt

for f in *.fq.gz; do
output=../Analysis/2-cutadapt/${f%.*}_cutadapt.fq.gz
Log=../Analysis/2-cutadapt/${f%.*}.Log
cutadapt -a TGGAATTCTCGGGTGCCAAGG -m 18 -M 40 $f 1> $output 2>
    $Log
done

cd /Work1/iana/Data/Sm3wt/Analysis/2-cutadapt
mkdir -p report
fastqc -o './report' *.fq.gz

# rRNA filtering
gunzip *.fq.gz
mkdir -p ../3-sortmerna/rRNA

for f in *.fq; do
rRNA=../3-sortmerna/rRNA/${f%.*}rRNA
output=../3-sortmerna/${f%.*}sortmerna
sortmerna --ref /Work1/iana/Applications/sortmerna-2.1-linux-64/
    rRNA_databases/planarian_rRNA_full_sequences.fa,/Work1/iana/
    Applications/sortmerna-2.1-linux-64/index/
```

```

    planarian_rRNA_full_sequences --reads $f --fastx -a 8 --
    aligned $rRNA --other $output -v --log
done

cd ../3-sortmerna/rRNA

for f in *.log; do
Totalreads='grep "Total reads" $f | awk '{print $0}''
rRNA='grep "Total reads passing E-value
    threshold" $f | awk '{print $0}''
other='grep "Total reads failing E-value
    threshold" $f | awk '{print $0}''
echo -e "\n${f%.*}\t$Totalreads" >> ../
    sortmerna_statistic_output.txt
echo -e "${f%.*}\t$rRNA" >> ../sortmerna_statistic_output.txt
echo -e "${f%.*}\t$other" >> ../sortmerna_statistic_output.txt
done

gzip *.fq
cd ../

# Generate insert file
mkdir -p ../4-insert
for f in *.fq; do
piPipes_fastq_to_insert $f ../4-insert/${f/fq/insert}
done

gzip *fq

# Map the reads to genome with bowtie
cd ../4-insert
mkdir -p ../5-bowtie

bowtie-build /Work1/iana/Planarian_genome/final_dd_Smed_g4.fa /
    Work1/iana/Applications/bowtie-1.1.2/indexes/Smed_g4

for f in sm3wt{1,2,3}; do
DIR=/Work1/iana/Data/Sm3wt/Analysis/5-bowtie
bowtie -r -p 8 -v 0 -a --best --strata /Work1/iana/Applications/
    bowtie-1.1.2/indexes/Smed_g4 $f.insert -S $DIR/$f.sam 2>>
    $DIR/$f.bowtie.Log
samtools view -bS -@ 8 $DIR/$f.sam | samtools sort -@ 8 - > $DIR
    /$f.bam
rm $DIR/$f.sam
done

# Convert bam into bed2 file format
cd ../5-bowtie

for f in *bam; do
bedtools_piPipes_bamtobed -i $f > ${f%.*}.bed
piPipes_insertBed_to_bed2 ../4-insert/${f%.*}.insert ${f%.*}.
    bed > ${f%.*}.bed2
done

# Analyze piRNAs ping-pong signature

```

```

mkdir -p pingpong

for f in *bed2; do
awk $f 'length($7) > 25' > ${f%.*}_cutoff25.bed2
done

piPipes_local_ping_pong -u 37 -a sm3wt1_cutoff25.bed2 -b
    sm3wt2_cutoff25.bed2 -p 8 > ./pingpong/sm3wt1_sm3wt2.pp

# Length distribution of genome mapped unique piRNAs
mkdir -p Length
for f in *.bed2; do
awk $f '{print length($7)}' | sort -n | uniq -c > ./Length/${f%.*}_Length.txt
done

# Create fasta files to analyze nucleotide distribution and plot
    venn diagram
mkdir -p fasta

for f in sm3wt{1,2,3}; do
bedtools_piPipes getfasta -fi /Work1/iana/Planarian_genome/
    final_dd_Smed_g4.fa -bed $f.bed -fo ./fasta/${f}.pipipes.fa -s
    -name
done

# Calculate weighted counts of mapped piRNAs
for f in *cutoff25.bed2; do
awk 'BEGIN {OFS="\t"}; {print
    $0"\t"$4/$5}' $f > ${f%.*}.weightCounts.bed2
done

# Intersect mapped piRNAs with annotation files
mkdir -p ../6-bedtools/LINE
mkdir -p ../6-bedtools/DNA
mkdir -p ../6-bedtools/LTR
mkdir -p ../6-bedtools/Unk
mkdir -p ../6-bedtools/No_annot
mkdir -p ../6-bedtools/Cluster
mkdir -p ../6-bedtools/Exon

for f in *weightCounts.bed2; do
LINE=/Work1/iana/Planarian_genome/RepeatMaskerLINE.gtf
DNA=/Work1/iana/Planarian_genome/RepeatMaskerDNA.gtf
LTR=/Work1/iana/Planarian_genome/RepeatMaskerLTR.gtf
Unk=/Work1/iana/Planarian_genome/RepeatMaskerUnk.gtf
No_annot=/Work1/iana/Planarian_genome/genes_plus_TE.gtf
Cluster=/Work1/iana/Planarian_genome/pirna_clusters.gtf
Exon=/Work1/iana/Planarian_genome/smes_v2_hconf_SMESG.gtf
bedtools intersect -split -wo -f 0.5 -a $f -b $LINE > ../6-
    bedtools/LINE/${f%.*}.intersectLINE.bed
bedtools intersect -split -wo -f 0.5 -a $f -b $DNA > ../6-
    bedtools/DNA/${f%.*}.intersectDNA.bed
bedtools intersect -split -wo -f 0.5 -a $f -b $LTR > ../6-
    bedtools/LTR/${f%.*}.intersectLTR.bed

```

```

bedtools intersect -split -wo -f 0.5 -a $f -b $Unk > ../6-
  bedtools/Unk/${f%.*}.intersectUnk.bed
bedtools intersect -split -s -v -a $f -b $No_annot > ../6-
  bedtools/No_annot/${f%.*}.intersectNo_annot.bed
bedtools intersect -split -wo -f 0.5 -a $f -b $Cluster > ../6-
  bedtools/Cluster/${f%.*}.intersectCluster.bed
bedtools intersect -split -wo -f 0.5 -a $f -b $Exon > ../6-
  bedtools/Exon/${f%.*}.intersectExon.bed
done

cd ../6-bedtools
for f in sm3wt{1,2,3}.weightCounts; do
Unique_seq_sense='cat ./Exon/$f.intersectExon.bed | awk '$6==$15
  {print $0}' | awk '{sum+=$8} END {print sum}'
Unique_seq_antisense='cat ./Exon/$f.intersectExon.bed | awk '$6
  !=$15 {print $0}' | awk '{sum+=$8} END {print
  sum}'
Unique_seq_dna_sense='cat ./DNA/$f.intersectDNA.bed | awk '$6==
  $15 {print $0}' | awk '{sum+=$8} END {print
  sum}'
Unique_seq_dna_antisense='cat ./DNA/$f.intersectDNA.bed | awk '
  '$6!=$15 {print $0}' | awk '{sum+=$8} END {print
  sum}'
Unique_seq_line_sense='cat ./LINE/$f.intersectLINE.bed | awk '$6
  ==$15 {print $0}' | awk '{sum+=$8} END {print
  sum}'
Unique_seq_line_antisense='cat ./LINE/$f.intersectLINE.bed | awk
  '$6!=$15 {print $0}' | awk '{sum+=$8} END {print
  sum}'
Unique_seq_ltr_sense='cat ./LTR/$f.intersectLTR.bed | awk '$6==
  $15 {print $0}' | awk '{sum+=$8} END {print
  sum}'
Unique_seq_ltr_antisense='cat ./LTR/$f.intersectLTR.bed | awk '
  '$6!=$15 {print $0}' | awk '{sum+=$8} END {print
  sum}'
Unique_seq_Unk_sense='cat ./Unk/$f.bed.intersectUnk.bed | awk '
  '$6==$15 {print $0}' | awk '{sum+=$8} END {print
  sum}'
Unique_seq_Unk_antisense='cat ./Unk/$f.bed.intersectUnk.bed |
  awk '$6!=$15 {print $0}' | awk '{sum+=$8} END {print
  sum}'
Unique_seq_cluster_sense='cat ./Cluster/$f.intersectCluster.bed
  |awk '$6==$15 {print $0}' | awk '{sum+=$8} END {print
  sum}'
Unique_seq_cluster_antisense='cat ./Cluster/$f.intersectCluster.
  bed | awk '$6!=$15 {print $0}' | awk '{sum+=$8} END {print
  sum}'
No_annot_sense='cat ./No_annot/$f.intersectNo_annot.bed | awk '
  '{sum+=$8} END {print
  sum}'
echo -e "$f.intersectExon.bed\texon\t${Unique_seq_sense}\t${
  Unique_seq_antisense}" >> ./sm3ip_to_features_analysis.txt
echo -e "$f.intersectDNA.bed\tdna\t${Unique_seq_dna_sense}\t${
  Unique_seq_dna_antisense}" >> ./sm3ip_to_features_analysis.
  txt

```

```

echo -e "$f.intersectLINE.bed\tline\t${Unique_seq_line_sense}\t${
Unique_seq_line_antisense}" >> ./sm3ip_to_features_analysis.
txt
echo -e "$f.intersectLTR.bed\tltr\t${Unique_seq_ltr_sense}\t${
Unique_seq_ltr_antisense}" >> ./sm3ip_to_features_analysis.
txt
echo -e "$f.intersectUnk.bed\tUnk\t${Unique_seq_Unk_sense}\t${
Unique_seq_Unk_antisense}" >> ./sm3ip_to_features_analysis.
txt
echo -e "$f.intersectCluster.bed\tcluster\t${
Unique_seq_cluster_sense}\t${Unique_seq_cluster_antisense}"
>> ./sm3ip_to_features_analysis.txt
echo -e "$f.intersectNo_annot.bed\tNo_annot\t${No_annot_sense}\
tint(0)" >> ./sm3ip_to_features_analysis.txt
done

```

## C.2 RNA-seq pipeline

```

# RNA libraries quality control
cd /Work1/iana/Data/RNA_seq_Sm3RNAi/Raw_data
mkdir -p ../Analysis/1-fastqc
fastqc -o ../Analysis/1-fastqc *fq.gz

# Adapter removing and quality flitering
mkdir -p ../Analysis/2-trimmomatic

for f in *fq.gz; do
output=../Analysis/2-trimmomatic/${file%.*}_trimmomatic.fq.gz
Log=../Analysis/2-trimmomatic/${file%.*}_trimmomatic.Log"
java -jar /Work1/iana/Applications/Trimmomatic-0.36/trimmomatic
-0.36.jar SE -threads 8 -phred33 $f $output ILLUMINACLIP:"/
Work1/iana/Applications/Trimmomatic-0.36/adapters/TruSeq3-SE.
fa":2:20:10 SLIDINGWINDOW:5:20 MINLEN:18 &> $Log
done

cd ../Analysis/2-trimmomatic
mkdir -p report
fastqc -o ./report *trimmomatic.fq.gz

for f in *Log; do
reads='grep 'Input Reads:' $f | awk '{print $0}'
echo -e "${f%.*}:\t${reads}" >> ../trimmomatic_statistic_output
.txt
done

# rRNA filtering
mkdir -p ../3-sortmerna/rRNA
gunzip *.gz

for f in *.fq; do
rRNA=../3-sortmerna/rRNA/${f%_*}_rRNA
output=../3-sortmerna/${f%_*}_sortmerna

```

```

sortmerna --ref /Work1/iana/Applications/sortmerna-2.1-linux-64/
  rRNA_databases/planarian_rRNA_full_sequences.fa,/Work1/iana/
  Applications/sortmerna-2.1-linux-64/index/
  planarian_rRNA_full_sequences --reads $f --fastx -a 8 --
  aligned $rRNA --other $output -v --log
done

cd ../3-sortmerna/rRNA

for f in *.log; do
Totalreads='grep "Total reads" $f | awk '{print $0}';
rRNA='grep "Total reads passing E-value
  threshold" $f | awk '{print $0}';
other='grep "Total reads failing E-value
  threshold" $f | awk '{print $0}';
echo -e "\n${f%.*}\t$Totalreads" >> ../
  sortmerna_statistic_output.txt
echo -e "${f%.*}\t$rRNA" >> ../sortmerna_statistic_output.txt
echo -e "${f%.*}\t$other" >> ../sortmerna_statistic_output.txt
done

gzip *.fq
cd ../

# Assign filtered reads to planarian transcriptome and consensus
  transposable element sequences with kallisto
mkdir -p ../4-kallisto/rf

for f in *fq; do
output=../4-kallisto/rf/${f%_*}_sense"
Log=../4-kallisto/rf/${f%_*}.Log
kallisto quant -i /Work1/iana/Applications/kallisto/index/
  smes_v2_hconf_SMESG_flat+consensusTE -o $output --single -l
  350 -s 30 -b 30 --rf-stranded -t 8 $f &> $Log
done

```

### C.3 Processing of HITS-CLIP libraries

```

# HITS-CLIP libraries quality control
cd /Work1/iana/Data/CLIP/Raw_data
mkdir -p ../Analysis/1-fastqc
fastqc -o ../Analysis/1-fastqc *fq.gz

# 3'-adapter trimming
mkdir -p ../Analysis/2-cutadapt

for f in *fq.gz; do
output=../Analysis/2-cutadapt/${f%.*}_cutadapt.fq.gz
Log=../Analysis/2-cutadapt/${f%.*}.Log
cutadapt -a TGGAAATTCCTCGGGTGCCAAGG -m 18 $f 1> $output 2> $Log
done

cd ../Analysis/2-cutadapt
mkdir -p report

```

```

fastqc -o './report' *.fq.gz

for f in *.Log; do
WrittenReads='grep 'Reads written
  (passing filters):' $f | awk '{print $(NF-1)}'
echo -e "${f%%.*}:\t${WrittenReads}" >> ../
  cutadapt_statistic_output.txt
done

# Collapse PCR duplicates
mkdir -p ../3-collapsed
gunzip *.fq.gz

for f in *.fq; do
output=../3-collapsed/${f%%.*}_collapsed.fa
fastx_collapser -Q33 -i "$f" -o $output -v &> ../3-collapsed/${f
  %%.*}.Log
done

cd ../3-collapsed

for f in *.Log; do
WrittenReads='grep 'Output:' $f | awk '{print $0}'
echo -e "${f%%.*}:\t${WrittenReads}" >> ../
  collapser_statistic_output.txt
done

# Remove rRNA reads
mkdir -p ../4-sortmerna/rRNA

for f in *.fa; do
rRNA=../4-sortmerna/rRNA/${f%%_*}_rRNA
output=../4-sortmerna/${f%%_*}_sortmerna
sortmerna --ref /Work1/iana/Applications/sortmerna-2.1-linux-64/
  rRNA_databases/planarian_rRNA_full_sequences.fa,/Work1/iana/
  Applications/sortmerna-2.1-linux-64/index/
  planarian_rRNA_full_sequences --reads $f --fastx -a 8 --
  aligned $rRNA --other $output -v --log
done

cd ../4-sortmerna/rRNA

for f in *.log; do
Totalreads='grep "Total reads" $f | awk '{print $0}'
rRNA='grep "Total reads passing E-value
  threshold" $f | awk '{print $0}'
other='grep "Total reads failing E-value
  threshold" $f | awk '{print $0}'
echo -e "\n${f%%.*}\t${Totalreads}" >> ../
  sortmerna_statistic_output.txt
echo -e "${f%%.*}\t${rRNA}" >> ../sortmerna_statistic_output.txt
echo -e "${f%%.*}\t${other}" >> ../sortmerna_statistic_output.txt
done

cd ../

```

```

# Merge fasta files of down and upper band libraries

for f in *up_sortmerna.fa; do
sed -i '1~2 s/\/upper-/' $f
done

for f in *d_sortmerna.fa; do
sed -i '1~2 s/\/down-/' $f
done

mkdir -p ../5-combinedFA

for f rep{1,2,3,4,5,6,7,8,9}; do
cat $f_d_sortmerna.fa $f_up_sortmerna.fa > ../5-combinedFA/
    $f_merged.fa
done

cd ../5-combinedFA
gzip *fa

### CIMS analysis according to CTK pipeline (Shah et al. 2017)
###
# Map reads to the genome with bwa
mkdir -p ../6-bwa

bwa index -p /Work1/iana/Applications/bwa/smedbwa -a bwtsv '/
    Work1/iana/Planarian_genome/final_dd_Smed_g4.fa'

for f in *merged.fa.gz; do
bwa aln -t 8 -n 0.06 -q 20 /Work1/iana/Applications/bwa/smedbwa
    $f > ../6-bwa/${f%*_merged*}.sai
bwa samse /Work1/iana/Applications/bwa/smedbwa ../6-bwa/${f
    %*_merged*}.sai $f | gzip -c > "../6-bwa/${f%*_merged*}.sam.gz
"
done

# Parse alignment into bed
mkdir -p ../7-bed
cd ../6-bwa

for f in *.sam.gz; do
parseAlignment.pl -v --map-qual 1 --min-len 18 --mutation-file "
    ../7-bed/${f%.*}.mutation.txt" $f "../7-bed/${f%.*}.bed"
done

# Diagnostic step
mkdir -p ../7-bed/len-dist
cd ../7-bed
for f in *.bed; do
awk '{print $3-$2}' $f | sort -n | uniq -c | awk '{print
    $2"\t"$1}' > ./len-dist/${f%.*}.tag.uniq.len.dist.txt
done

# Number of mapped reads

for f in *.bed; do

```

```

wc -l $f &>> ${f%.*}.sign.peak.statistic.txt
cat *.sign.peak.statistic.txt > reads_statistic_output.txt
done

# Mutations in unique tags
for f in rep{1,2,3,4,5,6,7,8,9}; do
python2 /usr/local/ctk-master/joinWrapper.py $f.mutation.txt $f.
    bed 4 4 N $f.mutation.uniq.txt
done

for f in rep{1,2,3,4,5,6,7,8,9}; do
Total_mut='wc -l $f.mutation.txt'
Unique_mut='wc -l $f.mutation.uniq.txt'
echo -e "${Total_mut}\t${Unique_mut}" >> ./
    mut_uniq_statistic_output_forCIMS.txt
done

# Define mutation cites
mkdir -p ../8-CIMS/del
mkdir -p ../8-CIMS/ins
mkdir -p ../8-CIMS/sub

for f in *.mutation.uniq.txt; do
getMutationType.pl -t del $f ../8-CIMS/del/${f%.*}.del.bed
getMutationType.pl -t ins $f ../8-CIMS/ins/${f%.*}.ins.bed
getMutationType.pl -t sub --summary ../8-CIMS/${f%.*}.stat.txt
    $file ../8-CIMS/sub/${f%.*}.sub.bed
done

# Example of mutational analysis of substitutions
cd ../8-CIMS/sub

for f in *sub.bed; do
CIMS.pl -v -n 10 -p -c ./${f%.*}_cache_cluster --keep-cache --
    outp ${f%.*}.pos.stat.SUB.txt -v ../7-bed/${f%.*}.bed"
    $f "${f%.*}.sub.CIMS.txt"
done

mkdir -p FDR005
for f in *.CIMS.txt; do
awk -v prefix="rep1" '{if($9<=0.001){print
    $1"\t"$2"\t"$3"\t"prefix_"$4"\t"$5"\t"$6}}' $f | sort -k 9,9
    n -k 8,8nr -k 7,7n > ./FDR001/${f%.*}.p001.bed
done

# Intersect identified substitution sites with gene annotation
cd ./FDR001
mkdir -p intersect

for f in *.p001.bed; do
bedtools intersect -wa -wb -split -a $f -b /Work1/iana/
    Planarian_genome/smes_v2_hconf_SMESG.gtf > ./intersect/${f
    %.*}.intersect001.bed
done

```



## **(Eidesstattliche) Versicherungen und Erklärungen**

(§9 Satz 2 Nr. 3 PromO BayNAT)

*Hiermit versichere ich eidesstattlich, dass ich die Arbeit selbstständig verfasst und keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt habe (vgl. Art. 64 Abs. 1 Satz 6 BayHSchG).*

(§9 Satz 2 Nr. 3 PromO BayNAT)

*Hiermit erkläre ich, dass ich die Dissertation nicht bereits zur Erlangung eines akademischen Grades eingereicht habe und dass ich nicht bereits diese oder eine gleichartige Doktorprüfung endgültig nicht bestanden habe.*

(§9 Satz 2 Nr. 4 PromO BayNAT)

*Hiermit erkläre ich, dass ich Hilfe von gewerblichen Promotionsberatern bzw. -vermittlern oder ähnlichen Dienstleistern weder bisher in Anspruch genommen habe noch künftig in Anspruch nehmen werde.*

(§9 Satz 2 Nr. 7 PromO BayNAT)

*Hiermit erkläre ich mein Einverständnis, dass die elektronische Fassung meiner Dissertation unter Wahrung meiner Urheberrechte und des Datenschutzes einer gesonderten Überprüfung unterzogen werden kann.*

(§9 Satz 2 Nr. 8 PromO BayNAT)

*Hiermit erkläre ich mein Einverständnis, dass bei Verdacht wissenschaftlichen Fehlverhaltens Ermittlungen durch universitätsinterne Organe der wissenschaftlichen Selbstkontrolle stattfinden können.*

---

Ort,

Datum,

Unterschrift