Searching and characterizing novel 21U RNA biogenesis factors in *Caenorhabditis elegans*

Dissertation

Zur Erlangung des Grades

Doktor der Naturwissenschaften

Am Fachbereich Biologie

Der Johannes Gutenberg-Universität Mainz

Ricardo José Cordeiro Machado Rodrigues

geb. am 07/03/1988 in Lissabon, Portugal

Mainz, 2019

Dekan:

1. Berichterstatter:
2. Berichterstatter:

Tag der mündlichen Prüfung:

04/09/2019

À minha família. Àqueles que me apoiaram. Àqueles que amo.

To my family. To those who supported me. To those that I love.

Table of Contents

Summary

Germ cells are responsible for the transmission of genetic information across generations. Hence, genomic instability in these cells can come at the cost of individual fertility and the fitness of a species. In animals, the RNAi-like Piwi-pathway has evolved to counteract the activity of one of the biggest threats to genome stability, Transposable Elements (TEs). In this pathway, Piwi-clade argonaute proteins use Piwi-Interacting RNAs (piRNAs) to recognize and induce the silencing of genome parasites such as TEs.

The piRNAs of *Caenorhabditis elegans* are known as 21U RNAs. This organism has the astonishing repertoire of around 30,000 different 21U RNAs, which are loaded into the *C. elegans* Piwi argonaute PRG-1. This argonaute uses these 21 nucleotide long small RNAs to recognize and induce the silencing of genomic parasites. Recognition by PRG-1 induces silencing and initiates an epigenetic silencing memory that can be parentally perpetuated for an indefinite number of generations. This extraordinarily effective silencing is supported by both transcript and chromatin silencing.

The 21U RNAs are transcribed as precursors by RNA Polymerase II. These contain a 5'end Cap and flanking sequences that need to be removed before PRG-1 loading. The process by which this maturation is achieved is largely unknown, as is the molecular machinery involved. In this work, I search and find novel machinery required for 21U RNA processing.

I started by focusing on the 21U RNA biogenesis factor PID-1 (piRNA-induced silencing defective 1). I found that this protein is part of a larger, novel protein complex, I named PETISCO (PID-3, ERH-2, TOFU-6 and IFE-3 Small RNA COmplex), that is also required for 21U RNA biogenesis. This complex binds the 5' Caps of 21U RNA precursors via IFE-3, a eIF4E-family protein, typically required for mRNA translation in other organisms. Not only did I resolve the architecture of PETISCO, I also found that this complex is required for a second function, essential for embryonic development. This function is guided by the protein TOST-1 (Twenty One u AntagoniST 1), which I propose binds PETISCO interchangeably with PID-1. Together these two proteins independently guide the two functions of PETISCO: 21U RNA biogenesis and embryonic viability. Finally, I propose that the embryonic requirement of PETISCO is connected to Splice Leader snRNA and core histone mRNA regulation and discuss how this relation may have evolved.

Zusammenfassung

Keimzellen sind für die Weitergabe genetischer Informationen über Generationen hinweg verantwortlich. Daher kann die genomische Instabilität dieser Zellen die individuelle Fertilität und die Fitness einer Art beeinträchtigen. Bei Tieren hat sich der RNAi-ähnliche Piwi-Signalweg entwickelt, um der Aktivität mobiler genetischer Elemente (*Transposable Elements, TEs*), einer der größten Bedrohungen für die Genomstabilität, entgegenzuwirken. In diesem Signalweg verwenden Argonautenproteine, Proteine aus der Gruppe der Piwi-Proteine, Piwi-interagierende RNAs (piRNAs), um genomische Parasiten wie TEs zu erkennen und deren Stummschaltung zu induzieren.

Die piRNAs von *Caenorhabditis elegans* sind als 21U RNAs bekannt. Dieser Organismus hat das erstaunliche Repertoire von rund 30,000 verschiedenen 21U RNAs, die auf das *C. elegans* Piwi-Argonautenprotein PRG-1 beladen werden. Dieses Argonautenprotein verwendet die 21 Nucleotide langen, kleinen RNAs, um genomische Parasiten zu erkennen und deren Stummschaltung zu induzieren. Die Erkennung durch PRG-1 induziert die Stummschaltung und initiiert ein epigenetisches Gedächtnis, das für eine unbestimmte Anzahl von Generationen elterlich aufrechterhalten werden kann. Diese außerordentlich wirksame Stummschaltung wird sowohl auf der Transkriptions- als auch auf der Chromatinebene unterstützt.

Die 21U-RNAs werden von der RNA-Polymerase II als Vorläufer transkribiert. Diese enthalten eine 5'-Ende-Kappe und flankierende Sequenzen, die vor dem Laden von PRG-1 entfernt werden müssen. Der Prozess, durch den diese Reifung erreicht wird, ist weitgehend unbekannt, ebenso wie die beteiligten molekularen Maschinerien. In dieser Arbeit suche und finde ich neuartige Maschinerien, die für die 21U-RNA-Prozessierung erforderlich sind.

Ich konzentrierte mich zunächst auf den 21U-RNA-Biogenesefaktor PID-1 (piRNA-induced silencing defective 1). Ich entdeckte, dass dieses Protein Teil eines größeren, neuartigen Proteinkomplexes ist, den ich PETISCO (PID-3, ERH-2, TOFU-6 und IFE-3 Small RNA COmplex) nannte, und welcher auch für die 21U-RNA-Biogenese erforderlich ist. Dieser Komplex bindet die 5'-Kappen von 21U-RNA-Vorläufern über IFE-3, ein Protein der eIF4E-Familie, das typischerweise für die mRNA-Translation in anderen Organismen benötigt wird. Ich habe nicht nur die Architektur von PETISCO aufgelöst, sondern auch festgestellt, dass dieser Komplex für eine zweite Funktion erforderlich ist, die für die Embryonalentwicklung unerlässlich ist. Diese Funktion wird durch das Protein TOST-1 (Twenty One u AntagoniST 1) gesteuert. Diesbezüglich schlage ich ein Modell vor, in dem TOST-1 PETISCO austauschbar mit PID-1 bindet. Zusammen steuern diese beiden Proteine unabhängig voneinander die beiden Funktionen von PETISCO: 21U RNA-Biogenese und embryonale Lebensfähigkeit. Schließlich stelle ich zur Diskussion, dass der embryonale Bedarf von PETISCO mit der *Splice Leader* snRNA- und *Core-*

Histon-mRNA-Regulation zusammenhängen könnte und erörtere, wie sich diese Beziehung möglicherweise entwickelt hat.

List of abbreviations

AGO	Argonaute
ALG	Argonaute-like gene
At	*Arabidopsis thaliana*
bp	Base pairs
Cbg	*Caenorhabditis briggsae*
Cbr	*Caenorhabitis brunei*
Ce	*Caenorhabitis elegans*
Cj	*Caenorhabitis japonica*
Cre	*Caenorhabitis remanei*
CRISPR/Cas9	Clustered regularly interspaced short palindromic repeats/ CRISPR-associated protein 9
CSR	Chromosome segregation and RNAi defective
DIC	Differential interference contrast microscopy
DMA	Dimethyl arginine
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
dsRNA	Double-stranded RNA
eAGOs	Eukaryotic AGO
EGO	Enhancer of *glp*-one
Endo-siRNA	Endogenous siRNA
ERH	Enhancer of Rudimentary Homologue
Eri	Enhanced RNAi
Exo-RNAi	Exogenous RNAi
FDR	False discovery rate
GFP	Enhanced green fluorescent protein
H3K9me3	Histone H3 lysine 9 trimethylation
hAGO2	Human AGO2
HEN1	HUA enhancer 1
HENN	HEN-1 of nematode
HP1	Heterochromatin protein 1
HRDE	Heritable RNAi deficient
IBM	IFE binding motif

iCLIP	Individual-nucleotide resolution UV crosslinking and immunoprecipitation
IP	Immunoprecipitation
IP-LFQP	Immunoprecipitation followed by label free quantitative mass spectrometry
m7G	7-methylguanylate
Mel	Maternal effect lethal phenotype
MID	Middle
miRISC	microRNA RISC
miRNA	microRNA
mRNA	Messenger RNA
Mog	Masculinization of germline phenotype
Mrt	Mortal germline phenotype
MTase	Methyltransferase
MUT	Mutator
NRDE	Nuclear RNAi defective
nt(s)	Nucleotide(s)
pAGO	Prokaryotic Ago
PARN	Poly(A)-specific ribonuclease
PATCs	Periodic An/Tn clusters
PAZ	Piwi Argonaute and Zwille
PID	piRNA-induced silencing defective
piRNA	Piwi-interacting RNA
Piwi	P-element induced Wimpy testis
PRDE	piRNA-dependent silencing defective
PRG	Piwi related gene
PTGS	Post-transcriptional gene silencing
PTM	Post-translational modification
RDE	RNAi defective
rDNA	Ribosomal DNA
RdRP	RNA-dependent RNA Polymerase
RIP	RNA immunoprecipitation
RISC	RNA-induced silencing complex
RITS	RNA-induced transcriptional silencing
RNA Pol	RNA polymerase

RNAe	RNA-induced epigenetic silencing
RNAi	RNA interference
RNPs	Ribonucleoproteins
RPM	Reads per million mapped reads
RppH	RNA 5' Pyrophosphohydrolase
RRF	RNA-dependent RNA Polymerase family
RRM	RNA Recognition Motif
rRNA	Ribosomal RNA
RT-qPCR	Reverse transcriptase quantitative polymerase chain reaction
sgRNA	Single guide RNA
siRNA	Small interfering RNA
SL	Splice leader
smRNA	Small RNA
SNAPc	SnRNA-activating protein complex
SNPC	Small nuclear RNA activating complex
snRNAs	Small nuclear RNAs
ssRNA	Single-stranded RNA
TDR	Tudor
TE	Transposable element
TGS	Transcriptional gene silencing
TMG	2,2,7-trimethylguanylate
TOFU	Twenty one u-RNA biogenesis fouled up
TOST	Twenty one U antagonist
tRNA	Transfer RNA
USTC	Upstream sequence transcription complex
UTR	Untranslated region
WAGO	Worm Argonaute
WT	Wild type

1. Introduction

1.1. The emergence of RNA as an active molecule

The importance of nucleic acids as biomolecules was self-evident since their discovery, as ribonucleic Acid (RNA) and deoxyribonucleic acid (DNA) exist in virtually all known living material. Still, it was after the discovery of the DNA double helix (Watson and Crick, 1953) that the first models for the function of these molecules emerged. Crick follows this discovery with the central dogma of molecular biology (Crick, 1958) and RNA took a role as a transitory molecule between DNA and protein. The confirmation of the functions of mRNA (Brenner et al., 1961), tRNA (Hoagland et al., 1958) and rRNA (Palade, 1955), and the discovery of the *lac operon* by Jacob and Monod (Jacob and Monod, 1961) and later its regulatory system (Gilbert and Müller-Hill, 1966) further contributed to the more prominent role of DNA as an information centre and proteins as effectors and control factors of that information.

The idea of RNA as a transitory molecule started losing strength once other RNA functions were found: the small nuclear RNAs (snRNAs) were found to assemble into RNPs and be required for splicing (early reviewed in Dreyfuss et al., 1988); the discovery of the catalytic activity of RNA (Guerrier-Takada et al., 1983; Kruger et al., 1982) not only gave RNA a more prominent role in the cell machinery but it rose this molecule to the centre of biology as a possible primordial biomolecule. A molecule that can store information in its base pairing, work as catalyst and self-replicate, is a frontrunner candidate for the effector molecule of the first cell.

The RNA field has grown in parallel with the RNA species that have been found through time. The last few decades have revealed a myriad of RNA subspecies that impose their regulatory function by both transcriptional gene silencing (TGS) and/or post-transcriptional gene silencing (PTGS). The former is achieved by changing the amount of RNA produced from a particular gene, while the latter acts upon an already transcribed RNA. One of the first examples of TGS by RNA was the discovery of the silencing of the X chromosome by the coating of the Xist RNA (Lee et al., 1996). The first observers of RNA based PTGS was probably Lee et al., 1993. The authors discovered a small RNA (smRNA) of 21 to 22 nucleotides that originated from the LIN-4 *locus* of *Caenorhabditis elegans* and was complementary to the LIN-14 3'UTR and impaired its translation. The phenomenon of smRNAs and the fact that double stranded RNA (dsRNA) can interfere with gene expression was named RNA interference (RNAi) and was described by Fire et al., 1998. From that moment RNAi has both been the object of intense study and has revolutionized molecular biology as a tool in the study of the cell.

In this introduction I will shortly introduce RNA interference pathways that are present across most eukaryotes. I will contextualize the discovery of RNAi and how argonautes are central proteins in this process. Finally, I will focus on *C. elegans*, the animal model of this thesis,

in which the RNAi pathways have undergone rich branching and specialization. At the end I will shortly introduce trans-splicing in *C. elegans* as this will be relevant in other sections of this thesis.

1.2. RNA interference

RNA interference is a process by which a smRNA works as a cofactor to an argonaute protein (AGO) that uses this smRNA as a sequence complementary guide to find the RNA it should target. The size spectrum of this smRNA ranges between 19-33 nucleotides and changes according to the AGO protein it interacts with and the RNAi pathway of which the AGO is part (Tolia and Joshua-Tor, 2007). The process was unknowingly first observed upon introducing overexpression transgenes into petunias, which led to the silencing of the RNA of the endogenous *locus* in this organism (van der Krol et al., 1990; Napoli et al., 1990). More observations followed, suggesting that the effect was the action of antisense RNA targeting the gene (Baulcombe, 1996). It was in the work of Lee et al., 1993 that it was first proposed that the interference was due to the annealing of two complementary RNA molecules. This would inhibit translation according to the stoichiometric ratio of the two molecules, a hypothesis that would come to be contradicted (Fire et al., 1998). Although these effects were first observed in plants it would only be the works of Andrew Fire and Craig Mello on RNAi in *C. elegans* that would start to unravel the molecular mechanisms behind RNAi.

As Fire et al., 1998 introduced dsRNA in *C. elegans* they observed that small dsRNA sequences would lead to the full silencing of complementary RNA. This silencing was also inherited through several generations. The observed generational effect contradicted the hypothesis that silencing was due to translation inhibition via RNA-RNA annealing in a stoichiometric manner, and likely there was a full machinery and mechanism dedicated to this process. Since that work this mechanism has been observed across eukaryotes, with few exceptions (Drinnenberg et al., 2009).

Multiple subspecies of smRNA have been found amongst eukaryotes. As they vary in function it is simpler to divide these molecules based on their origin. The original studies on RNA interference looked into the effects of dsRNA. In those observations, relatively long dsRNA molecules (~300-1000nts) were injected in worms (Fire et al., 1998). Conversely, the smRNA cofactor used by the AGO sizes up to 33 nucleotides. The size difference implies that the dsRNA molecule needs to be cleaved into smaller sized molecules that can interact with the AGO. The catalyst of this reaction is the RNA endonuclease Dicer. Dicer is an RNAse III domain containing protein that has the ability of recognizing and cleaving dsRNA into smaller dsRNA molecules that are then downstream loaded into AGOs (Bernstein et al., 2001; MacRae et al., 2008). smRNAs originated from dsRNA are thus categorized as Dicer-dependent (Fig. I1), of which two classes are known: microRNAs and short interfering RNAs. Other smRNAs originate

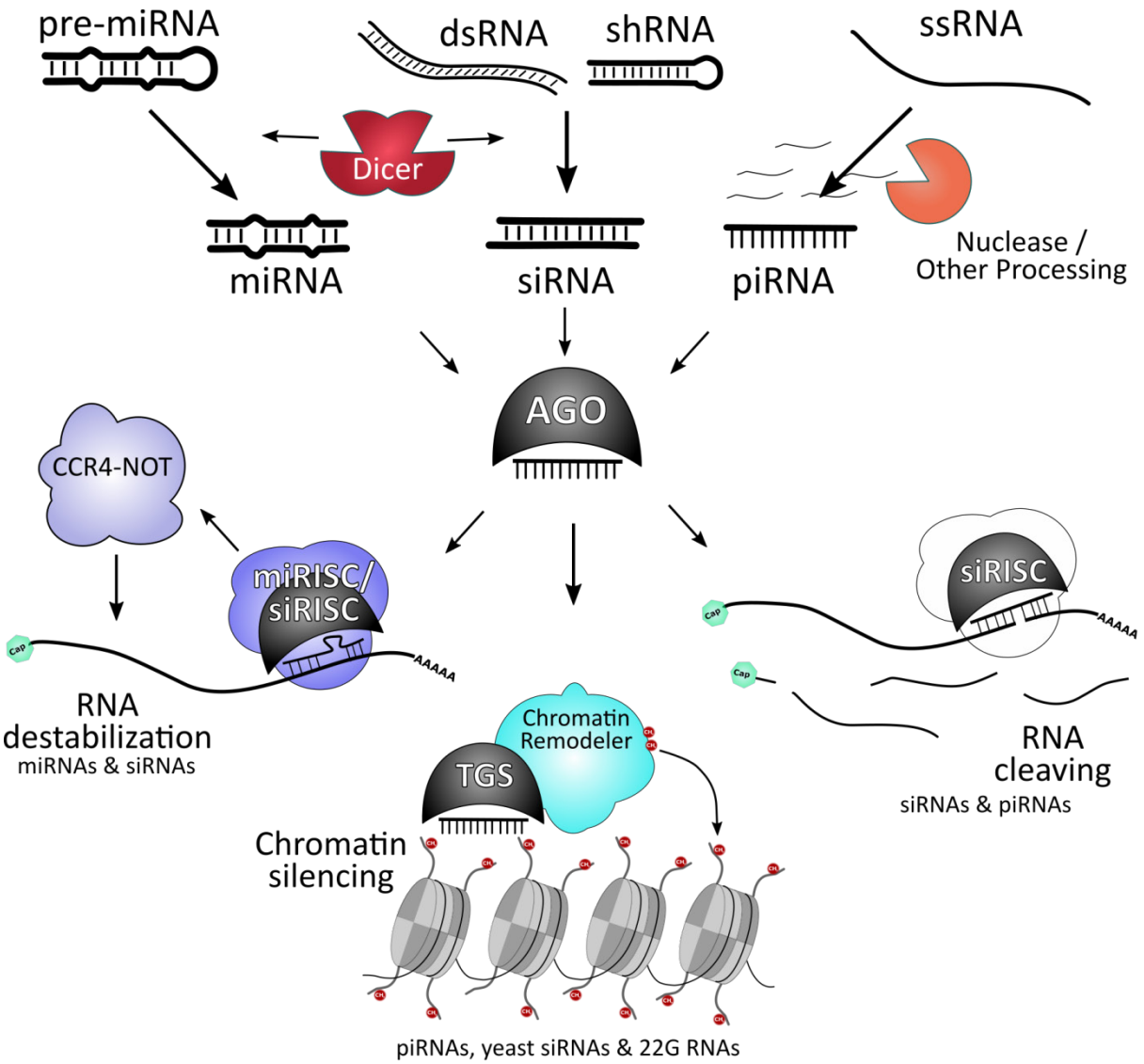

Fig I1 – Schematic representing the biogenesis and effects of RNAi pathways
Dicer cleaves pre-miRNAs into miRNAs and short hairpin RNAs (shRNA) and dsRNAs into siRNAs. piRNAs are produced from ssRNA and are Dicer independent. The smRNAs are loaded into an AGO protein that uses the smRNA as a sequence template to find targets and induce their silencing. The silencing can be induced by: RNA destabilization, e.g. miRNAs and siRNAs with incomplete sequence complementarity; RNA cleaving, e.g. siRNAs and piRNAs; or transcriptional gene silencing by chromatin remodelling, e.g. piRNAs and yeast siRNAs. Each silencing is induced by the recruitment of AGO co-factors.

from single stranded RNA (ssRNA) molecules that are processed and loaded into AGO proteins. As they do not need Dicer cleavage in their processing they are classified as Dicer-independent (Fig. I1). The most studied class of Dicer-independent smRNAs is likely to be PIWI-interacting RNAs (piRNAs), a smRNA species so named for their ability to interact with the PIWI clade of AGO proteins.

1.3. Argonautes at the centre of RNAi

The first argonaute protein to be discovered was found in a forward genetic screen for cell elongation in *Arabidopsis thaliana* (Bohmert et al., 1998). The elongation of the leaves in early development of *ago1* mutants reminded the authors of cephalopod tentacles and so they named

the gene after the octopus species *Argonauta argo*. Since the discovery of this first argonaute, examples of this protein superfamily have been found amongst bacteria, archaea and most eukaryotes (Swarts et al., 2014a).

The Argonaute protein superfamily has a wide distribution among species, with their origin diverging in each domain of life. The current evidence shows that the last common ancestor of archaea and eukaryotes was likely to both have an AGO protein and the associated RNAi machinery, revealing the ancestry of this protein family and pathway (Shabalina and Koonin, 2008). On the other hand, bacterial AGOs are likely the result of horizontal gene transfer, evident from the lack of correlation between the phylogeny of these AGOs and the speciation of bacteria (Makarova et al., 2009).

The structure of argonautes was first solved using prokaryote argonautes (pAGOs) (Rashid et al., 2007; Song et al., 2004). These attempts failed to include the smRNA cofactor in the structure, but managed to give precious insight into how an AGO could interact with nucleic acids. It was later found that many pAGOs preferably bind a single stranded DNA (ssDNA) cofactor rather than an ssRNA (Ma et al., 2005; Swarts et al., 2014b, 2015; Yuan et al., 2005). Little information is currently available about the functions of RNA/DNA interference in prokaryotes, but multiple functions have been proposed for pAGOs, including being part of systems that protect the host against foreign genetic invaders, such as phages and transposons (Lisitskaya et al., 2018).

AGOs are more widespread amongst eukaryotes, but the number of different argonautes found in a given species does vary greatly. Some species have undergone large multiplications of AGOs in their genome (e.g. *C. elegans* has 27 AGOs), while others seem to have lost these genes all together (e.g. *S. cerevisiae*)(Drinnenberg et al., 2009; Yigit et al., 2006). Still, eukaryotic argonautes (eAGOs) have central roles in RNAi-based mechanisms, where they use a small ssRNA as a cofactor, named guide RNA, to identify complementary RNA sequences and, typically, induce their silencing (Fig. I1). Currently, four AGO clades have been identified in eukaryotes (Swarts et al., 2014a): the AGO clade, a class of Argonautes that interacts with microRNAs and short interfering RNAs and is more closely related to AtAGO1; the PIWI clade, which is closely related to *Drosophila*'s P-element Induced Wimpy testis (PIWI) AGO; the *Trypanosoma* clade, exclusive to trypanosomatids; and Worm argonautes (WAGOs), a class that is specific to nematodes and performs multiple functions (Fig. I2A). As examples of the AGO clade can be found in most eukaryotic *phyla,* these are largely accepted as the ancestor argonaute clade from which the others have branched (Shabalina and Koonin, 2008). In animals, the earliest branching event was the PIWI clade (Grimson et al., 2008) which is present in most animals and is exclusive to the animal kingdom.

1.3.1. The argonaute structure shapes its smRNA partner

RNA interference is a classification that includes multiple pathways that use a smRNA species combined with an AGO. Each pathway is categorized according to the intervenient AGO and the specific smRNA species. In truth, the nature of the smRNA and the function of the pathway are determined by the AGO itself: the structure of the AGO and its binding pocket can determine the size, the 5'end nucleotide and chemical modifications of the smRNA, while the AGO specific interactors determine the overall action of a specific RNAi pathway.

Eukaryotic Argonautes exhibit a quite stable overall structure, harbouring four domains from N- to C-terminal end: the N-terminal Domain, the PAZ (PIWI-Argonaute-Zwille) domain, the MID (middle) domain and the PIWI domain (Swarts et al., 2014a) (Fig. I2B). The PIWI domain harbours the nuclease activity, also known as slicer activity, of the AGO and has an RNAse H fold with a catalytic amino acid tetrad DED(H/D) (Parker et al., 2004; Rivas et al., 2005; Song et al., 2004). AGOs with the full catalytic motif have the ability to cleave the target RNA and induce the silencing of a transcript in this manner (Liu et al., 2004). *In vitro* experiments also show the N-terminal domain as important for target cleavage, but the exact function of domain in this process is still unclear (Hauptmann et al., 2013). Otherwise, this domain is required for unwinding RNA duplexes bound to the AGOs (Kwak and Tomari, 2012).

The PAZ domain interacts with the 3' end of the smRNA and protects the smRNA from exonucleolytic degradation (Lingel et al., 2003; Song et al., 2003; Yan et al., 2003). Upon extensive 3' end pairing of the smRNA to the target RNA, the PAZ domain releases the 3' end. This allows complementary base pairing and conformational changes for target cleavage (Wang et al., 2009). This conformational change exposes the 3'end of the smRNA to modifications, such a uridylation, that promote the degradation of the smRNA cofactor by exonucleases (Ameres et al., 2010). Some PAZ domains also confer RNA modification selectivity. Although this domain normally interacts with the 2'-OH of the smRNA, some smRNAs are protected from uridylation by methylation of this site (2'-O-CH_3) (Ameres et al., 2010). This is catalysed by HEN1-methyltransferases (Billi et al., 2012; Horwich et al., 2007; Kamminga et al., 2010, 2012; Montgomery et al., 2012) and AGOs that interact with methylated smRNA species have a higher affinity for these modified RNAs (Simon et al., 2011).

The MID domain forms a pocket in which several of its amino acids interact with the 5' end nucleotide of the smRNA cofactor (Boland et al., 2010; Elkayam et al., 2012; Frank et al., 2010; Schirle and MacRae, 2012). This domain normally has a selectivity to interact with the 5'end PO_4 of the smRNA (Boland et al., 2010; Schirle and MacRae, 2012). However, certain RNA species have different 5' ends, such as the 22G RNAs in *C. elegans,* which carry a 5' end triphosphate (Pak and Fire, 2007; Ruby et al., 2006). It is not yet known how the MID domains

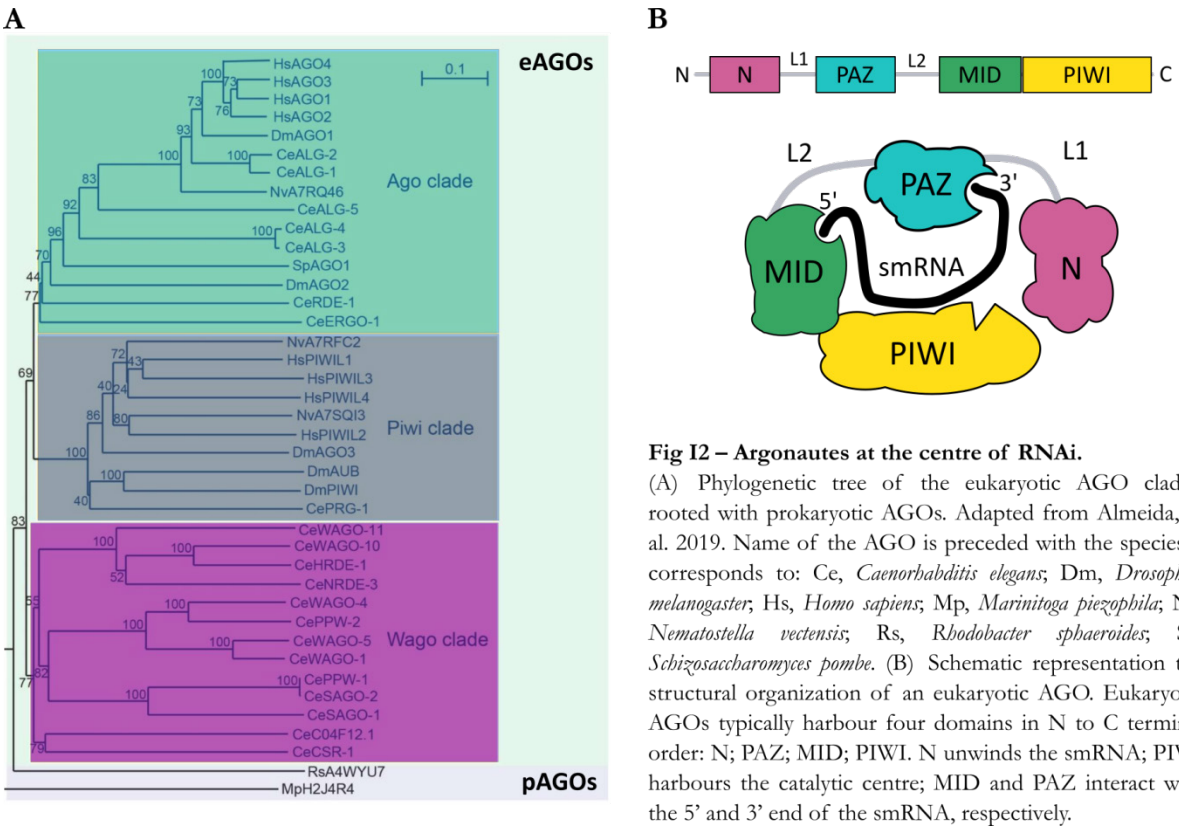

Fig I2 – Argonautes at the centre of RNAi.
(A) Phylogenetic tree of the eukaryotic AGO clades, rooted with prokaryotic AGOs. Adapted from Almeida, et al. 2019. Name of the AGO is preceded with the species it corresponds to: Ce, *Caenorhabditis elegans*; Dm, *Drosophila melanogaster*; Hs, *Homo sapiens*; Mp, *Marinitoga piezophila*; Nv, *Nematostella vectensis*; Rs, *Rhodobacter sphaeroides*; Sp, *Schizosaccharomyces pombe*. (B) Schematic representation the structural organization of an eukaryotic AGO. Eukaryotic AGOs typically harbour four domains in N to C terminal order: N; PAZ; MID; PIWI. N unwinds the smRNA; PIWI harbours the catalytic centre; MID and PAZ interact with the 5' and 3' end of the smRNA, respectively.

of 22G RNA-interacting AGOs accommodate this modification, but the fact that these can bind a 5'PPP end reveals the remarkable plasticity of these domains. Moreover, the MID domains, apart from conferring 5'end modification selection, can also provide 5'end nucleotide binding selectivity. This is achieved by a five amino acid sequence in this domain named the selectivity loop (Frank et al., 2010). Thus, the AGO selects its smRNA and shapes the smRNA population it interacts with.

The cleaving ability of an AGO does not strictly determine its function. One can find catalytically active and inactive AGOs that induce silencing of their targets (Bagijn et al., 2012; De Fazio et al., 2011; Liu et al., 2004; MacRae et al., 2008; Reuter et al., 2011). Also the same smRNA species can have different effects depending on the argonaute it interacts with, e.g. nematode 22G RNAs can induce nuclear or cytoplasmic silencing depending on the interacting AGO (Buckley et al., 2012; Yigit et al., 2006). This means that albeit AGOs are at the centre of the RNAi machinery, these proteins do not only act alone, but typically recruit other machinery that complements their activity (Fig. I1). Thus, the actions of each AGO cannot be solely described by their structure or enzymatic activity, but rather the context of their action. This context is established by the AGO-interacting proteins and their connections to secondary pathways that determine the processes downstream of target recognition by these proteins.

1.4. microRNAs

The microRNA (miRNA) pathway is a Dicer-dependent RNAi related pathway whose majority of small RNAs originate from precursor RNA molecules, the primary microRNAs (pri-miRNAs). Pri-miRNAs are encoded in the genome as single non-coding RNA genes and are transcribed by RNA Polymerase II (RNA Pol II). Upon transcription, the pri-miRNA folds on itself creating a dsRNA stem-loop structure intercalated with mismatches and non-paired bulges which are recognized by a specific processing machinery, the "microprocessor". The microprocessor is constituted by two proteins: Drosha and Pasha, an RNAse III type enzyme and its cofactor, respectively. Together, they cleave the pri-miRNA stem loop leaving this structure intact. At this stage the stem loop is named pre-miRNA and is actively exported from the nucleus by Exportin 5 and RAN GTPases (Ha and Kim, 2014). More rarely, pre-miRNAs can originate from introns and are named mirtrons. In place of the microprocessor, these are processed by the splicing machinery and form the same stem loop and bulged structures that are recognized and exported from the nucleus as the other pre-miRNAs.

Cytoplasmic pre-miRNAs are recognized by the Dicer/TRBP complex that digests these stem loops into 21-23 nucleotide long dsRNA molecules. This is followed by loading of the small RNA into the argonaute protein of the microRNA-Induced Silencing Complex (miRISC). Only one of the strands is loaded into the AGO, named miRNA, while the other strand, named miRNA*, is usually degraded. The choice of which strand is loaded is determined by its thermodynamic stability (Khvorova et al., 2003; Schwarz et al., 2003). Once loaded, miRISC identifies its targets by base pair complementarity between transcripts and the miRNA. Interestingly, in animals exact complementarity is not required and mismatches between the two sequences are allowed (Bartel, 2018). Specifically, perfect pairing is usually required from positions 2 to 8 of the miRNA, named seed sequence, while the 3'end is more permissive to mismatches. In fact, the seed sequence is often conserved throughout species and miRNA families can be classified according to this sequence. miRISC generally targets the 3' untranslated region (UTR) of mRNAs, although there are reported cases of 5'UTR and coding region targeting, these have less silencing efficiency (Bartel, 2018). The silencing is executed by inhibiting translation or promoting destabilization of the mRNA by deadenylation, rather than target slicing (Eulalio et al., 2008; Guo et al., 2010; Hendrickson et al., 2009) (Fig. I1).

The physiological functions of miRNAs are broad as they regulate gene expression in multiple pathways. Mutations in certain miRNAs or miRNA families can lead to severe deleterious effects in development, and disease. On the other hand, no function has been found for the majority of miRNAs, even for some that are highly conserved (Alvarez-Saavedra and Horvitz, 2010; Miska et al., 2007). Thus, the current perspective on miRNA function is that these

perform the post-transcriptional fine tuning of gene(s) that they target and their function only becomes evident under conditions that stress these systems (Bartel, 2018; Brenner et al., 2010). This tuning is achieved by a mix of factors including miRNA concentration and the miRNA 3'end complementarity to its target (Brancati and Großhans, 2018).

1.5. Short interfering RNAs

Short interfering RNAs (siRNAs) is a class that comprises the remaining non-miRNA Dicer-dependent RNAi pathways (Fig. I1). The origin of siRNAs is usually a long dsRNA that can be exogenous (exo-siRNAs), such as dsRNA viral genomes, or endogenous to the cell (endo-siRNAs). In similarity to miRNAs, the dsRNA is digested by Dicer and then loaded into the AGO of the short interfering RNA Induced Silencing Complex (siRISC). siRISC transcript targeting, typically requires perfect pairing between the smRNA and the target RNA. In these cases target cleaving is achieved by the AGO. Still, siRNA-induced PTGS can be achieved by non-perfect pairing, which in large follows the rules of miRISC and leads to inhibition of translation (reviewed in Carthew and Sontheimer, 2009). siRNAs have been attributed different functions and I will give a short introduction on the physiological function of these smRNAs.

Endo-siRNAs vary in function depending on each species; while in *Drosophila* and mammals these RNAs are targeting and silencing Transposable Elements (TEs), in *Schizosaccharomyces pombe* they are required for the establishment of heterochromatic domains. Murine siRNAs originate from repetitive sequences and *loci* that are transcribed from both the plus and minus strands in germline cells (Tam et al., 2008; Watanabe et al., 2006, 2008). These transcripts form dsRNA molecules that are recognized and cleaved by Dicer. Opposite to murine siRNAs, *Drosophila* siRNAs are somatic, but their origin can also be *loci* transcribed from both strands, or short RNA hairpins, which are processed by Dicer. These are loaded to AGO2 in siRISC and also target TEs (Chung et al., 2008; Czech et al., 2008; Ghildiyal et al., 2008; Kawamura et al., 2008; Okamura et al., 2008).

In fission yeast, centromere heterochromatic assembly is dependent on siRNAs (Fig. I3A). There, the RNA-induced transcriptional silencing complex (RITS) contains the only AGO orthologue of *S. pombe*, Ago1, and uses its loaded siRNAs to recognize pericentromeric regions and induce heterochromatin assembly and silencing (Volpe, 2002). The genomic regions flanking the centromeres are highly repetitive and are transcribed from both strands (Verdel and Moazed, 2005). The dsRNA generated are recognized and cleaved by Dicer and loaded into Ago1 in the RITS complex (Volpe, 2002). The RITS complex then has the ability to recognize the same transcripts from which the smRNA cofactor was generated at their transcription site (Motamedi et al., 2004). This complex recruits the Clr4-Rik1-Cul4 (CLRC) complex that catalyses H3K9me3

deposition in the repetitive DNA sites, followed by the deposition of Swi6, a H3K9me3-interacting HP1 family protein (Bühler et al., 2006; Kato et al., 2005; Motamedi et al., 2004; Noma et al., 2004). Finally, Swi6 recruits an RNA-dependent RNA Polymerase (RDRP) containing complex that uses the local repetitive sequences as a template to generate more dsRNAs and amplify the RNAi response (Verdel and Moazed, 2005; Verdel et al., 2004). Thus, this siRNA pathway serves as a self-reinforcing loop and acts both as PTGS and TGS.

1.6. PIWI-interacting RNAs

PIWI-interacting RNAs (piRNAs) represent a class of smRNAs that is Dicer-independent and animal specific. Their name derives from the fact that these RNA species specifically interact with AGOs of the PIWI clade. Typically, the expression of these RNAs and AGOs is restricted to the germline, where they survey and protect the germline genome from foreign invaders, such as TEs. Germ cells are responsible for the continuation of a species, and any unsolicited genetic insertion can lead to sterility or severely impair descendants. Thus, TE surveillance takes a frontstage in these cells (Aravin et al., 2006; Batista et al., 2008; Brennecke et al., 2007; Das et al., 2008; Houwing et al., 2007; Ozata et al., 2019).

The general machinery of the Piwi pathway is similar amongst animals, such as the use of a PIWI protein and the fact that the smRNA originates from a ssRNA molecule and is thus Dicer-independent. Still, there are many variations between animals and I will briefly describe two well studied models for this pathway: *Drosophila* and mouse (Fig. 3B and C). *C. elegans* also harbours a Piwi pathway, but as the central object of this thesis, it will be described in more detail further ahead in a dedicated section.

1.6.1. The Drosophila Piwi pathway

The *Drosophila* Piwi pathway, is essential for this organism, as it is evidenced by the severe deleterious effects on mutants for components of this pathway (Czech et al., 2018). *Drosophila* harbours three PIWI proteins, which all take a part in the pathway: Piwi, AGO3 and Aubergine (Aub). AGO3 and Aub pair with each other and are the main players in the so-called ping-pong amplification cycle in which transcripts are silenced. Piwi shuttles between the nucleus and the cytoplasm and mainly acts in TGS (Akkouche et al., 2017; Brennecke et al., 2007).

Mature piRNAs are 24-30 nucleotides long, but they originate from long ssRNA molecules, the piRNA precursors, that are processed in multiple steps. PiRNA precursors are transcribed by RNA Pol II from genomic clusters harbouring repetitive sequences such as TEs (Brennecke et al., 2007). The clustering of these sequences is likely due to the strategy employed by this species to recognize new foreign targets for silencing. An active TE creates copies of itself and moves around the genome. Once a copy of an invading agent has the (un)fortunate luck of

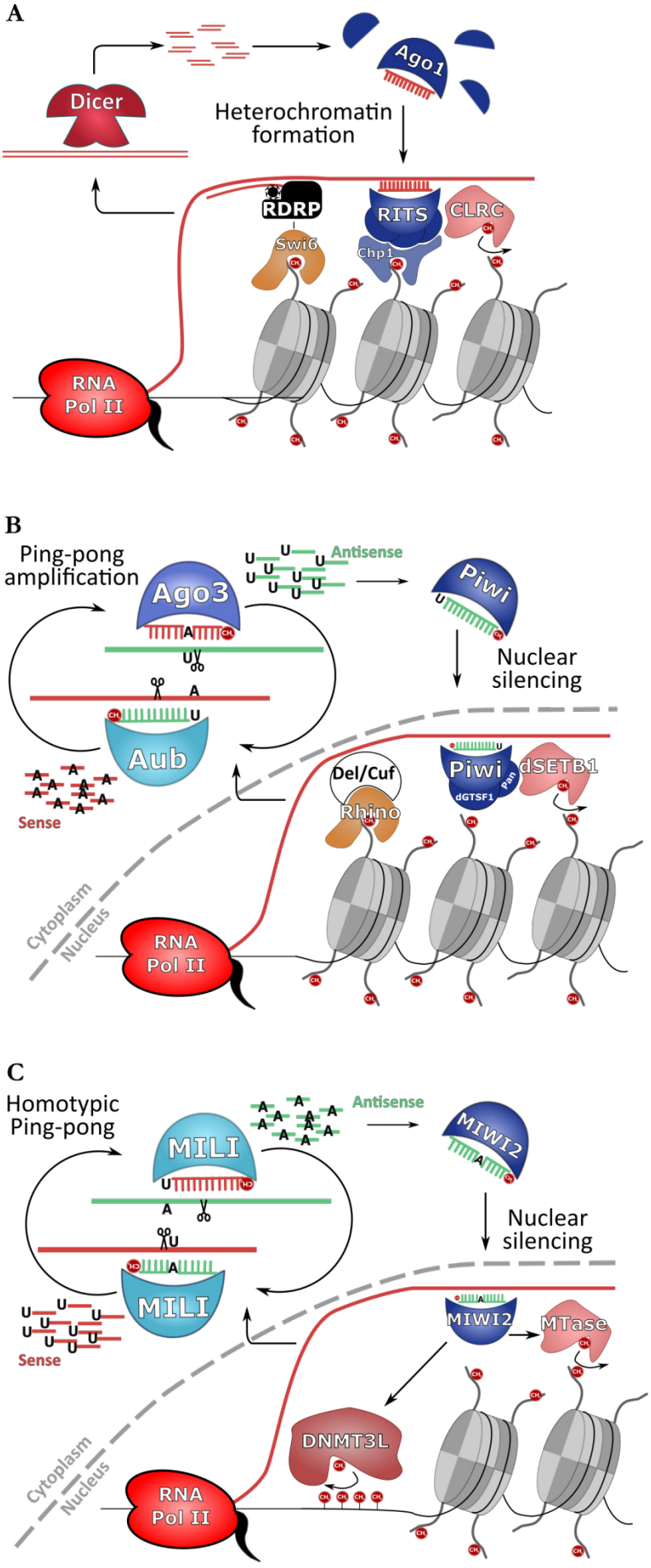

Figure I3 - Detailed schematics of piRNA and yeast siRNA pathways.
(A) The *S. pombe* siRNA pathway. In *S. pombe* dsRNAs are cleaved by dicer and siRNAs are loaded into Ago1/RITS. RITS recognizes the heterochromatic transcripts and recruits machinery for heterochromatin formation (CLCR). Heterochromatin is bond by Swi6 that recruits RDRP making new dsRNAs, creating a self-reinforcing loop.
(B) The *Drosophila* piRNA pathway. piRNA precursor transcripts are transcribed by RNA Pol II. In the ping-pong amplification loop, these are recognized and cleaved by Aub, creating sense piRNAs that are loaded in AGO3. AGO3 recognizes and cleaves antisense long transcripts creating antisense piRNAs that are loaded into Aub, continuing the loop, or Piwi. The latter is transported to the nucleus and induces chromatin silencing. The ping-pong signature is created by the 5' end U bias.
(C) The murine pre-pachytene piRNA pathway. Like *Drosophila* piRNA precursor transcripts are transcribed by RNA Pol II. In the ping-pong amplification both antisense and sense piRNAs are created by MILI. Part of the antisense piRNAs created by MILI is loaded into MIWI2, which is transported into the nucleus and induces chromatin silencing. In both (B) and (C) piRNAs are methylated (red circles)

falling within one of these clusters, it is identified as a foreign sequence and is directed into piRNA production. The clusters are located within or in the boundary of heterochromatic regions, such as pericentromeric and telomeric regions, and can be classified as uni-strand or dual-strand clusters. The latter have served as the preferred model to study piRNA biogenesis in *Drosophila* (Czech et al., 2018; Guzzardo et al., 2013; Yamashiro and Siomi, 2018), and I will focus on these.

Dual-strand clusters are marked by the presence of the HP1 homologue Rhino and the repressive histone mark H3K9me3 (Klattenhoff et al., 2009). Rhino allows for the transcription of these heterochromatic and normally repressed regions by recruiting specialized machinery. This protein is in complex with Cutoff and Deadlock and licenses the heterochromatin transcription by recruiting Moonshiner and TRF2 (Andersen et al., 2017; Klattenhoff et al., 2009; Mohn et al., 2014; Zhang et al., 2014). These are specialized paralogues of general transcription factors TFIIA and TBP, that are able to initiate transcription in a promoter-independent manner (Andersen et al., 2017). These resulting transcripts are recognized by UAP56, a DEAD-box helicase, that ensures the export of these precursor into the cytoplasm by specialized nuclear export factors (Batki et al., 2019; ElMaghraby et al., 2019; Fabry et al., 2019; Zhang et al., 2012b, 2014). Interestingly, loaded Piwi is the protein that licenses these dual-clusters for piRNA precursor production in adult ovaries (Akkouche et al., 2017; Bartel, 2018), creating a self-reinforcing loop in resemblance to *S. pombe* siRNAs (Fig. I3A).

PiRNAs are processed in the cytoplasm, particularly, this process is believed to occur in nuage: perinuclear granules to which most of the piRNA machinery is localized (Lim and Kai, 2007). The long transcripts are processed into piRNAs by either the primary processing pathway, and are named primary piRNAs, or the ping-pong amplification cycle (Czech and Hannon, 2016; Czech et al., 2018). In the former, the endonuclease Zucchini cleaves the transcripts generating smRNAs with free 5' and 3' ends (Han et al., 2015; Ipsaro et al., 2012; Mohn et al., 2015; Nishimasu et al., 2012). After the 3'end is fine-tuned by the exonuclease Nibbler, the smRNA is loaded into a PIWI protein and 2'-O-methylated at its 3'end by HEN1, becoming a fully matured piRNA (Horwich et al., 2007; Wang et al., 2016). The loaded PIWI protein then can initiate the secondary piRNA production.

Secondary piRNAs are produced via the ping-pong amplification cycle (Fig. I3B). In this cycle two PIWI argonautes degrade TE transcripts and use their degradation products to load new PIWI proteins and amplify the silencing response. The 5' end generated by the slicing of one PIWI on a target transcript is then used by another Piwi argonaute to generate a sense strand piRNA (Brennecke et al., 2007; Gunawardane et al., 2007). The 3'end is then again tuned by the Zucchini/Nibbler duo. Piwi binds preferentially to smRNAs whose 5'ends were catalysed by

Zucchini. These are have an uracil bias at their 5'end, which creates the signature of the ping-pong cycle: a 5'end U in antisense piRNAs and an A at position 10 of sense piRNAs, where the transcript is cleaved (Gunawardane et al., 2007). In the case of *Drosophila* the antisense piRNA binding PIWIs are Piwi and Aub while AGO3 mainly binds sense piRNAs. The two antisense PIWI proteins have specialized functions: Aub is cytoplasmic and performs the ping-pong cycle, while Piwi is nuclear and induces TGS (Czech and Hannon, 2016; Czech et al., 2018; Ozata et al., 2019).

1.6.2. The murine Piwi pathway

In many ways the murine Piwi pathway resembles that of *Drosophila*; it involves three PIWI proteins (MILI, MIWI and MIWI2), a cytoplasmic and a nuclear phase, a ping-pong signature and the same 2'-O-methylation of the piRNAs (Kirino and Mourelatos, 2007a, 2007b). On the other hand, in mouse these PIWI proteins are expressed at different developmental stages, and this pathway is solely required for male fertility (Carmell et al., 2007; Deng and Lin, 2002; Kuramochi-Miyagawa et al., 2004). The loss of this pathway in female gonads is a mouse-specific adaptation, as other mammals express the required machinery in female gonads (Roovers et al., 2015).

MILI is expressed throughout the life of germ cells (Aravin et al., 2006), it preferentially binds sense piRNAs and is responsible for the ping-pong amplification cycle. To be precise, MILI has been proposed to act with itself in what is called a homotypic ping-pong cycle, where this AGO is both the slicer and receiver of new piRNAs (De Fazio et al., 2011). MIWI2 is co-expressed with MILI from embryogenesis to birth, although it requires MILI cleavage to be loaded (De Fazio et al., 2011), it does not seem to take part in this cycle (Fig. I3C). MIWI2, in resemblance to *Drosophila* Piwi, is loaded with piRNAs and transported to the nucleus, where it targets TE sequences and induces the DNA methylation and silencing of these sequences (Aravin et al., 2008; Carmell et al., 2007). At these early developmental stages piRNAs originate from the processing of repetitive sequence clusters and are named pre-pachytene piRNAs.

MIWI is only expressed in adult testes on the onset of the pachytene stage of meiosis (Aravin et al., 2008). At these stages MILI and MIWI can be loaded with either sense or anti-sense transcripts, but the origin of the piRNAs differs. Instead of originating from TE elements the precursor transcripts originate from intergenic loci that are activated by the A-MYB transcription factor (Li et al., 2013). Due to their origin, pachytene piRNAs no longer target TEs, but are suggested to regulate gene expression in spermatogenesis (Ozata et al., 2019).

1.6.3. piRNAs go beyond TE silencing

The piRNA pathway targets TEs in most of the animals it has been studied in, and it is thought to have evolved in the context of this function (Parhad and Theurkauf, 2019). Still, as a

gene regulatory tool it is not surprising that some species have used this pathway for other purposes. The mouse pachytene piRNAs are suggested to take part in spermatogenesis (Simonelig, 2014; Zhao et al., 2013) and other functions have been reported for this pathway. In the germline, the piRNA pathway has been reported to be used for mRNA anchoring and mRNA stabilization in *Drosophila* embryos (Rojas-Ríos and Simonelig, 2018). Also piRNAs have been reported to act on sex determination in both flies and nematodes (Kiuchi et al., 2014; Tang et al., 2018) and in planarians and cnidarians they have been proposed to play central roles in regeneration (Rojas-Ríos and Simonelig, 2018; van Wolfswinkel, 2014).

Somatic functions have also been reported for piRNAs. In mosquitos, the piRNA pathway has been expanded to combat viral infections (Miesen et al., 2015; Morazzani et al.,

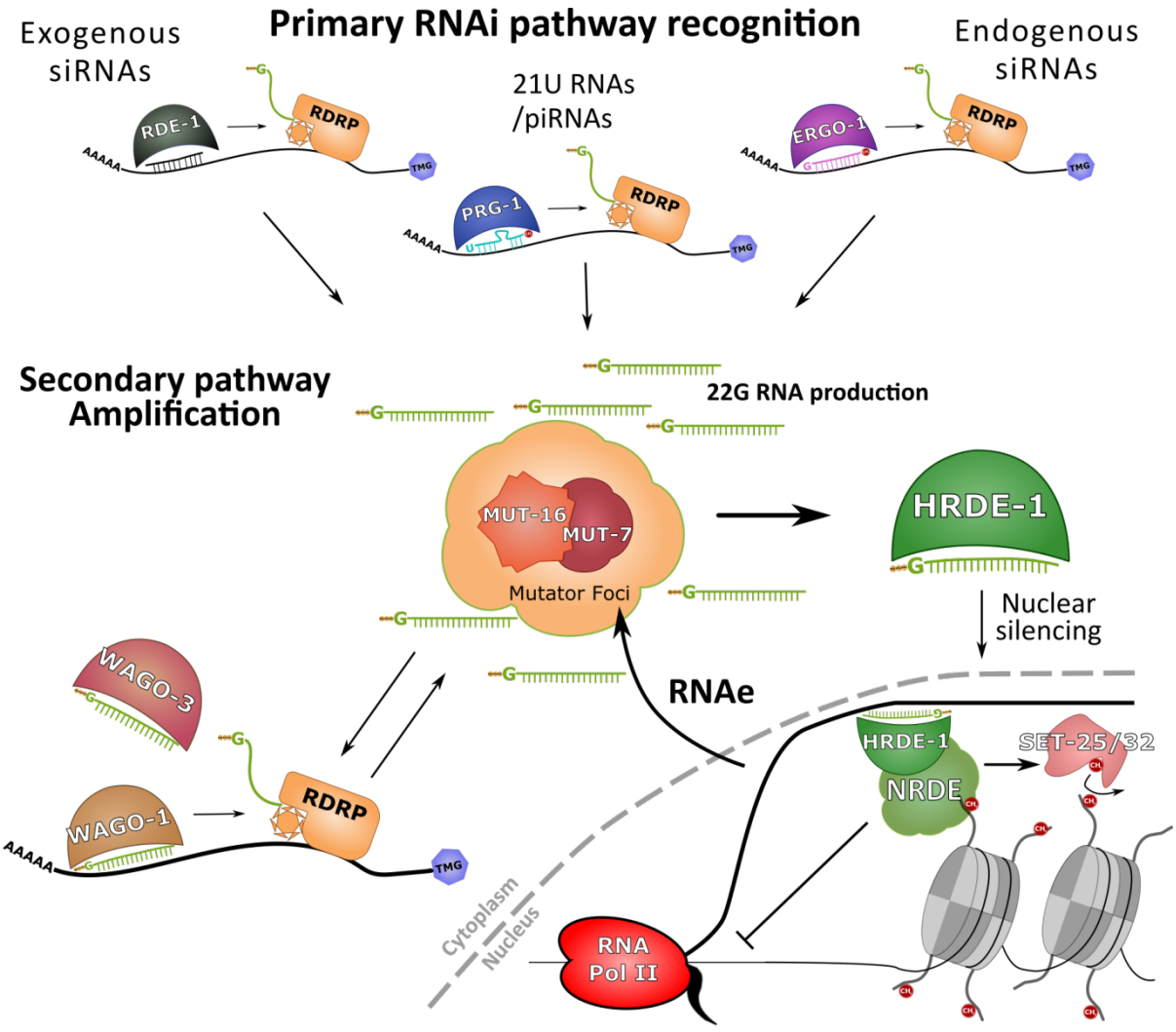

Figure I4 – Schematic representation of primary and secondary RNAi pathways of *C. elegans*.
Transcripts are targeted for silencing by primary pathways, including the exogenous siRNAs, endogenous siRNAs and 21U RNAs, the piRNAs of *C. elegans*. All of these induce silencing by activating 22G RNA production via the Mutator Complex. The RNAi response is amplified and 22G RNAs are loaded into WAGO proteins. These silence transcripts in the cytoplasm (WAGO-1/WAGO-3) or are transported to the nucleus (HRDE-1/NRDE-3 in the soma), where they induce chromatin remodelling via the NRDE complex. The 22G RNA pathway can feed itself, i.e. target recognition by WAGOs induces the production of more 22G RNAs. If the primary recognition is made by 21U RNAs it can induce chromatin changes that lead to a transgenerational positive feedback loop, silencing those transcripts indefinitely. The latter process is named RNAe for RNA-induced epigenetic silencing.

2012; Schnettler et al., 2013). In neurons these small RNAs have been reported to be required for axon regeneration (Kim et al., 2018) and behavioural adaptation (Moore et al., 2019; Rajasethupathy et al., 2012), but the number of studies showing these effects is rather limited and the molecular mechanisms behind them often unresolved. Thus, the somatic functions of piRNAs are still heavily debated subject in the field. Nevertheless, multiple examples for novel functions of piRNAs have been reported and this growing field will be interesting to follow.

1.7. RNAi pathways in *C. elegans*

The first RNAi pathway was described in *C. elegans* (Fire et al., 1998), but it was unknown to the authors that this model is rather unique when it comes to RNA interference. The genome of these nematodes encodes 27 AGOs, including 18 that belong to the worm specific AGO clade (WAGO) (Yigit et al., 2006). The large number of AGOs is combined with smRNA species and pathways that are unique to nematodes, including 21U RNAs, 22G RNAs and 26G RNAs. These act sequentially creating complex gene regulation networks based on RNAi (Fig. I4). Some pathways do not silence recognized transcripts *per se*, but rather induce the silencing of these transcripts via a downstream pathway. The former type is classified as a primary pathway, while the latter type are classified as secondary. I will introduce these pathways and how they are connected. As miRNAs are essentially between this species and other animals (reviewed in Ambros and Ruvkun, 2018), they will not be included in the following description.

1.7.1. Exogenous siRNAs

In most animals the siRNA pathway uses the miRNA machinery to perpetuate silencing. Interestingly, *C. elegans* has specialized machinery to acquire exogenous dsRNA and to induce internal silencing (Braukmann et al., 2017). The silencing by these molecules is also spread through the worm body and even embryos (Marré et al., 2016).

The exogenous dsRNA molecule is recognized by RDE-4, which recruits the worm Dicer (DCR-1) and the WAGO argonaute RDE-1 (Parker et al., 2006; Tabara et al., 2002). DCR-1 cleaves the dsRNA and the smRNA duplexes are loaded into RDE-1 (Ketting, 2001; Tabara et al., 1999). RDE-1 then cleaves the passenger strand releasing itself to find target transcripts. This AGO has the ability to cleave target transcripts, but does not require this activity to silence them (Steiner et al., 2009). Instead, it recruits proteins RDE-10, RDE-11 and RDE-12 to drive these transcripts into a secondary siRNA pathway (Shirayama et al., 2014; Yang et al., 2012, 2014; Zhang et al., 2012a). The secondary siRNA pathway corresponds to the 22G RNA pathway, which I will describe later ahead.

1.7.2. 26G RNAs, the endogenous siRNAs

Besides the exo-siRNAs, *C. elegans* contains a second class of siRNAs, the 26G RNAs. These are 26nts long and have 5'end bias for a guanine, thus their name (Ruby et al., 2006). 26G RNAs are produced by the ERI complex, which includes the RDRP RRF-3 and DCR-1 (Duchaine et al., 2006; Thivierge et al., 2012).

The ERI complex exists as a pre-complex that is brought together by the zinc finger protein GTSF-1, ensuring effective 26G RNA production (Almeida et al., 2018). In the current model the ERI complex is recruited to a target transcript in an unknown manner, and RRF-3 synthesizes an antisense strand creating a dsRNA molecule (Blumenfeld and Jose, 2016). This turns the transcript into a DCR-1 substrate that is cleaved into a 26nt long smRNA (Welker et al., 2010). The newly formed 26G RNA can be loaded into two distinct RNAi pathways: ALG-3/ALG-4 or ERGO-1 (Almeida et al., 2018, 2019a; Conine et al., 2010, 2013; Han et al., 2009).

1.7.2.1. ALG-3/ALG-4 26G RNAs

ALG-3 and ALG-4 are two *C. elegans* paralogues of the AGO clade of argonautes that are expressed only during spermatogenesis. There, these assume redundant functions and only double mutants present a phenotypes (Conine et al., 2010). The absence of these proteins lead to High Incidence of Males (Him) and sterility (Conine et al., 2010; Gent et al., 2009; Pavelec et al., 2009). The Him phenotype is attributed to missegregation of the X chromosome, while sterility is due to the lack of sperm activation. Interestingly, although these 26G RNAs target spermatogenesis specific genes, rather than inducing their degradation, they seem to promote spermatogenesis and transmit a paternal memory of gene expression (Conine et al., 2010, 2013). They have been proposed to lead to downstream action by the CSR-1 pathway (described below), but the exact action of the ALG-3/4 pathway is not yet understood.

1.7.2.2. ERGO-1 26G RNAs

ERGO-1, like ALG-3/-4, belongs to the AGO clade of argonautes, but early divergence from the other members of this clade has made it difficult to classify (Almeida et al., 2019b; Yigit et al., 2006). These differences may be due to the type of smRNA it interacts with, which are 3'end 2'-O-methylated by HENN-1 (Billi et al., 2012; Kamminga et al., 2012; Montgomery et al., 2012; Vasale et al., 2010) and thus require a distinct structure, permissive towards this interaction. The activity of the HENN-1 methyltransferase is required for stable 26G RNA populations of this pathway (Billi et al., 2012; Kamminga et al., 2012; Montgomery et al., 2012).

The ERGO-1 branch of 26G RNAs is expressed from the oogenic germline and to the second larval stage (L2) of *C. elegans* (Billi et al., 2012; Vasale et al., 2010) and targets a wide range of transcripts for silencing. These sequences include non-germline specific pseudogenes, long non-coding RNAs (lncRNAs) and other non-conserved repeat-rich sequences. The silencing is

induced via the recruitment of the 22G RNA pathway, making this a primary RNAi pathway (Fischer et al., 2011; Gent et al., 2010; Han et al., 2009; Newman et al., 2018; Vasale et al., 2010; Zhou et al., 2014). The silencing capacity of 26G RNAs has the ability of being transmitted both maternally and from the germline to the soma (Almeida et al., 2019a). Together with the fact that these smRNAs are methylated and that they target repetitive sequences, the ERGO-1 pathway draws parallel to the piRNA pathway. Hence, it has been suggested that in *C. elegans,* there is no single piRNA pathway but that ERGO-1 together with the PRG-1/21U RNA pathway both constitute the piRNA-like pathways of this nematode (Almeida et al., 2019b).

1.7.3. 21U RNAs

The genome of *C. elegans* encodes only one functional PIWI clade AGO, named Piwi Related Gene 1 (PRG-1). This argonaute was found to interact with 21U RNAs and these are only stable in the presence of PRG-1 (Batista et al., 2008; Das et al., 2008; Wang and Reinke, 2008). Thus, this smRNA species has generally been considered the piRNAs of *C. elegans*. Their name is due to their strong 5'end uracil bias and their 21nt long length (Ruby et al., 2006). Not only do these smRNAs bind a PIWI clade AGO, they are also germ cell specific (Batista et al., 2008; Das et al., 2008; Wang and Reinke, 2008), 2'-O-methylated by HENN-1 (Billi et al., 2012; Kamminga et al., 2012; Montgomery et al., 2012) and are responsible for the recognition and silencing of foreign sequences (Ashe et al., 2012; Das et al., 2008; de Albuquerque et al., 2015; Lee et al., 2012a; Luteijn et al., 2012; Shirayama et al., 2012), as piRNAs in other species (Aravin et al., 2006; Brennecke et al., 2007; Girard et al., 2006).

The similarities between the 21U RNA pathway and piRNAs are not extended to 21U RNA biogenesis, mode of action or phenotypes. In contrast with Piwi pathways in other animals, ablation of the 21U RNA pathway does not lead to full sterility or major transposon activity. Instead, deletion of *prg-1* leads to modest male fertility issues (Wang and Reinke, 2008) and only one class of transposons, the Tc3, is targeted directly and is reactivated in these mutants (Das et al., 2008). The absence of this pathway also registers transgenerational fertility loss, the so-called Mortal Germline (Mrt) phenotype, but this is an epigenetic effect and not TE associated (Simon et al., 2014). These differences are probably associated with the fact that rather than directly disrupting the transcripts it recognizes, the PRG-1 pathway induces silencing via a self-sustainable 22G RNA pathway (Ashe et al., 2012; Bagijn et al., 2012; Lee et al., 2012a; Luteijn et al., 2012; Shirayama et al., 2012), in similarity to other silencing RNAi pathways in *C. elegans* (Billi, 2014). I will next describe in detail our current knowledge of the different steps of 21U RNA biology.

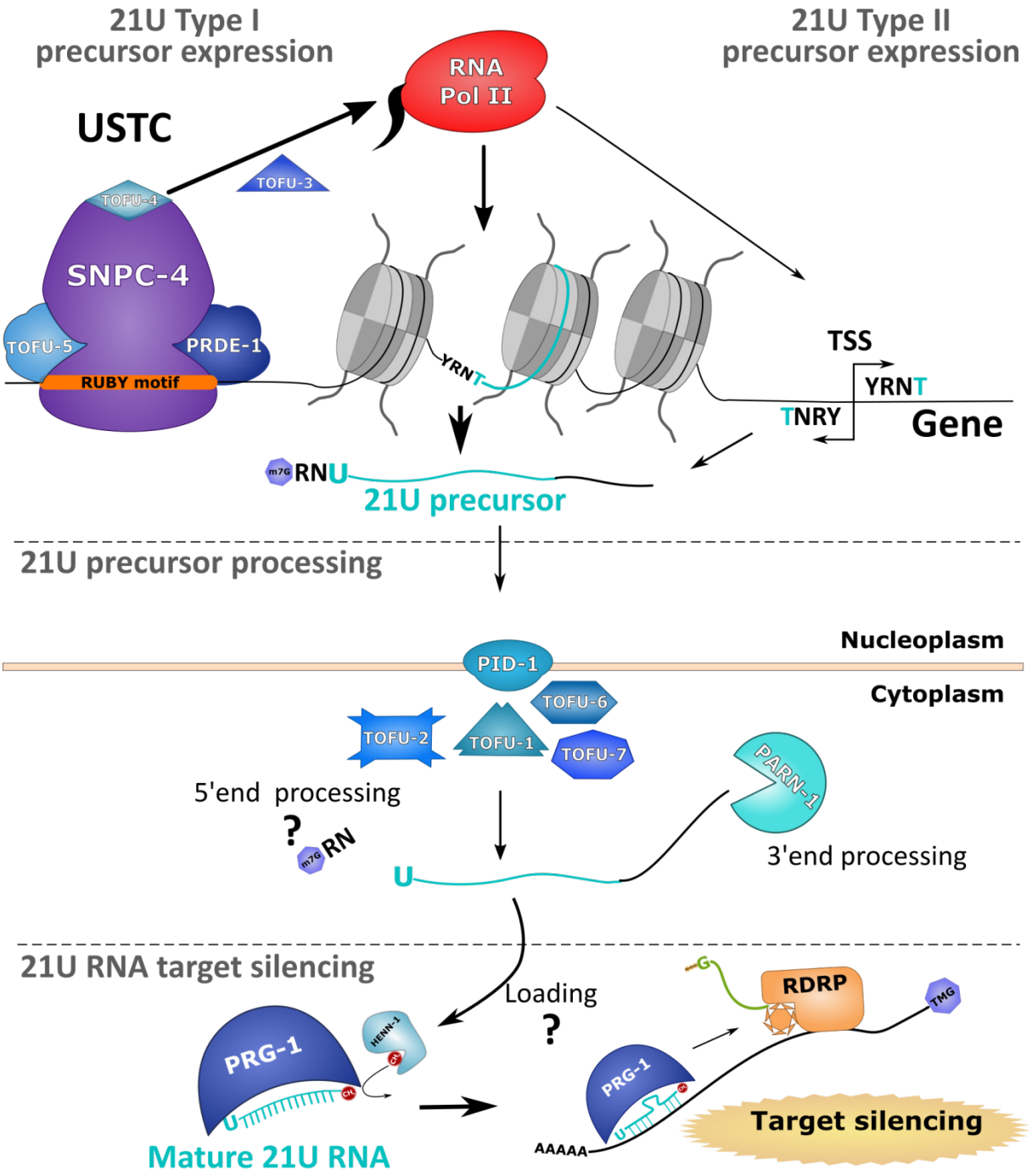

Figure I5 - Schematic representation of 21U RNA biosynthesis.
C. elegans type I 21U RNAs are transcribed as roughly 26 nucleotide long precursors by RNA polymerase II. This transcription is dependent on the upstream sequence transcription complex (UTSC) that includes TOFU-4/-5, PRDE-1 and SNPC-4. The UTSC recognizes the Ruby motif upstream of each individual 21U RNA locus. Type II precursors are less abundant and are also transcribed by RNA Pol II. These derive from transcription start sites (TSS) that start with the same YRNT motif as type I 21U RNAs. Once expressed the 21U RNA precursor has to be both exported to the cytoplasm and processed into a 21 nucleotide long RNA containing a Uracil at the 5'. TOFU factors and PID-1 are known to be required for this process but the molecular mechanism is unknown. The 5'end processing molecular mechanism remains elusive but the 3' end shortening is catalysed by the exonuclease PARN-1. After 3' end shortening, the methyltransferase HENN-1 2'-O-methylates the 21U RNA. PRG-1 uses the 21U RNA for target recognition but does not require perfect pairing. Induces silencing via the 22G RNA pathway.

1.7.3.1. The origin of 21U RNAs

C. elegans contains an impressive repertoire of 21U RNAs, with around 30,000 currently identified (Batista et al., 2008; Das et al., 2008; Gu et al., 2012). These are first produced in the form of precursors, which can be classified as Type I and Type II, in accordance to their origin (Gu et al., 2012).

Type I 21U RNA precursors constitute the bulk of the produced RNA precursors and PRG-1-bound 21U RNAs (Batista et al., 2008; Gu et al., 2012). These are transcribed by RNA Pol II and exist as distinct *loci* in the *C. elegans* genome (Cecere et al., 2012; Ruby et al., 2006). Each *locus* has an upstream motif, also named Ruby motif, that is recognized by specific transcription factors that license 21U RNA precursor transcription (Billi et al., 2013; Kasper et al., 2014; Ruby et al., 2006; Weick et al., 2014; Weng et al., 2019). The Ruby motif confers each of the 21U RNA genes the ability to function as a single transcriptional unit (Fig. I5). As such, it restricts expression of 21U RNAs to germ cells and determines male/female germline enrichments of each *locus* (Billi et al., 2013). Interestingly, 21U RNA genes can be found throughout the genome, but the vast majority of 21U RNA *loci* is located within two 3Mb clusters on chromosome IV (Ruby et al., 2006), an architecture that is believed to facilitate their transcription (Beltran et al., 2019; Kasper et al., 2014; Weick et al., 2014; Weng et al., 2019).

The transcription machinery of Type I 21U RNA precursors has been co-opted from that of snRNA expression (Beltran et al., 2019; Kasper et al., 2014; Weng et al., 2019). Specifically, the transcription factor SNPC-4, the worm orthologue of SNAPc, uses 21U RNA expression specific adaptors to recruit RNA Pol II (Fig. I5). These adaptors include PRDE-1, TOFU-4 and TOFU-5, which together with SNPC-4 form the so-called Upstream Sequence Transcription Complex (USTC) that recognizes the Ruby motif and can even be visualized in the chromosome IV clusters (Kasper et al., 2014; Weick et al., 2014; Weng et al., 2019). The transcription start site (TSS) is located two nucleotides upstream of the 5'U of the mature 21U RNA, and follows a YRNT motif (T corresponds to the 5'end U) required for effective transcription (Billi et al., 2013; Gu et al., 2012; Ruby et al., 2006). The termination of 21U RNA precursor transcription is suggested to be induced by promoter pausing and cancelled elongation by RNA Pol II, permitting the transcription of these short transcripts (Beltran et al., 2019). This type of termination is not site-specific and adds an uncertain amount of nucleotides to the 3'end termini of the 21U RNA precursor, which in the end is 23 to 28 nucleotides long and capped (Gu et al., 2012).

The Type II 21U RNA precursors have the same general structure as the ones from Type I and only differ on their origin (Fig. I5). These are derived from small capped RNAs that are bi-directionally transcribed by RNA Pol II at TSS of coding genes (Gu et al., 2012). When TSS sites

contain YRNT motifs in their premises, abortive transcripts caused by RNA Pol II pausing will have the same characteristics as other 21U RNA precursors and the 21U RNA pathway uses these by-products to increase the repertoire of total 21U RNAs (Beltran et al., 2019; Gu et al., 2012).

1.7.3.2. The maturation of 21U RNA precursors

Once a 21U RNA precursor is expressed it should be transported to the cytoplasm, where PRG-1 is present (Wang and Reinke, 2008). The mechanism by which this takes place is not yet known, but mechanisms involving the protein PiRNA Induced silencing Defective 1 (PID-1) have been proposed (de Albuquerque et al., 2014; Zeng et al., 2019; this study). This protein is required for 21U RNA processing and contains putative nuclear import and export signals (de Albuquerque et al., 2014), making it a front runner candidate for this process.

The mature 21U RNA has a 5'end PO_4, is 21nts long and is 3'end 2'-O-methylated, hence, during processing it loses the cap and two nucleotides at the 5'end and a 3'end extension (Fig. I5). The exact mechanism by which the 5'end is processed in these smRNAs is still unknown. On the other hand, the 3'end extension is processed by the exonuclease PARN-1 (Fig. I5). This step is not required for a functional PRG-1 pathway, as this AGO can still harbour 3'end extended 21U RNAs, albeit its activity becomes less efficient (Tang et al., 2016a). PARN-1 trimming is followed by HENN-1 mediated 2'-O-methylation of the 21U RNA. Unlike other Piwi pathways, this process is not essential, but its absence decreases 21U RNA stability and targeting efficiency (Billi et al., 2012; Horwich et al., 2007; Kamminga et al., 2010, 2012; Kirino and Mourelatos, 2007a, 2007b; Montgomery et al., 2012). Other proteins, named Twenty One u Fouled Up (TOFU) 1 and 2, were found in a reverse genetic screen to be required for the processing of 21U RNAs (Goh et al., 2014). The authors draw the conclusion that these are processing 21U RNAs due to precursor accumulation in strains depleted for these factors. Still, further studies are required to determine the exact function of these proteins in the 21U RNA pathway.

1.7.3.3. PRG-1 target recognition and silencing

The loading of PRG-1 is an unknown process, but once loaded, PRG-1 uses the 21U RNA to find and induce the silencing of a target RNA. Once a target is recognized by PRG-1, it initiates a secondary silencing response that is both strong and stable. E.g. a transgene targeted for silencing by PRG-1 can be silenced for an indefinite number of generations and this silencing is maintained even upon removal of the PRG-1 pathway (Ashe et al., 2012; Lee et al., 2012a; Luteijn et al., 2012; Shirayama et al., 2012). This silencing state is named RNA-induced epigenetic silencing (RNAe), and in this process small RNAs are inherited and maintain an epigenetic memory of silenced transcripts (Fig. I4). Interestingly, RNAe can only be initiated by 21U RNA

recognition of a target, but once the RNAe state is achieved, it becomes independent of the 21U RNA pathway and is maintained by the 22G RNA pathway (Ashe et al., 2012; Luteijn et al., 2012; Shirayama et al., 2012).

The epigenetic memory is propagated by the 22G RNA pathway but it is guided by the PRG-1 pathway. In laboratory conditions one can genetically remove the 22G RNA silencing epigenetic memory. This loss does not stop 22G RNA activity, but rather 22G RNAs become erratic and start targeting endogenous transcripts that should be expressed. In the presence of the PRG-1 pathway this miss-targeting does not happen. This is avoided because PRG-1 assigns the proper targets to the 22G RNAs pathway. Thus, although the PRG-1 pathway does not maintain silencing memory, it is required to re-initiate it (de Albuquerque et al., 2015; Phillips et al., 2015).

The silencing of a target does not require the slicing activity of PRG-1, as catalytically inactive PRG-1 is still able to induce target silencing (Bagijn et al., 2012). Interestingly, it also does not require full sequence complementarity between the 21U RNA and the target transcript (Bagijn et al., 2012; Shen et al., 2018; Zhang et al., 2018). In similarity to the miRNA targeting rules (see miRNAs), target recognition by PRG-1 only requires perfect pairing in the positions 2 to 8 of the 21U RNA that work as a seed sequence (Shen et al., 2018; Zhang et al., 2018). Together with the variety of available 21U RNA sequences, this pathway has virtually the ability to recognize any sequence (Bagijn et al., 2012). The current model is that the tremendous silencing potential of the PRG-1 pathway is countered by CSR-1 pathway, which allows, stimulates or licenses gene activity in the *C. elegans* germline (described below).

1.7.4. 22G RNAs

Secondary siRNAs, or 22G RNAs, are an integration node of other RNAi silencing pathways in *C. elegans* (Fig. I4). Their production is triggered by a primary RNAi pathway, such as the ERGO-1, PRG-1 and RDE-1 pathways. Target recognition by a primary pathway results in an amplified response by the 22G RNA pathway (Bagijn et al., 2012; Pak and Fire, 2007; Sijen et al., 2007; Vasale et al., 2010). Their name, like other *C. elegans* smRNA species, derives from their 5' end G bias and 22nt length. Besides these features, 22G RNAs are also characterized by the presence of a triphosphate group at their 5'end (Pak and Fire, 2007). The triphosphate is a mark left by RDRPs during 22G RNA synthesis. The RDRP uses the target transcripts as templates and produces single 22nt long anti-sense transcripts (Fig.I4). These are not cleaved by Dicer, but directly loaded into WAGOs, thus leaving the triphosphate of the first nucleotide intact (Gent et al., 2010; Maniar and Fire, 2011; Pak and Fire, 2007; Pak et al., 2012; Ruby et al., 2006; Vasale et al., 2010).

The genome of *C. elegans* encodes four RDRPs, of which two are capable of producing 22G RNAs: RRF-1 and EGO-1. It also encodes for multiple WAGOs able to interact with these

smRNAs. RRF-1 produces 22G RNAs in response to RDE-1, ERGO-1 and PRG-1, but it has been reported that EGO-1 can redundantly assume this function (Bagijn et al., 2012; Gent et al., 2010; Gu et al., 2009; Sijen et al., 2007; Vasale et al., 2010). Independently of which primary pathway has triggered the 22G RNA pathway, their production is dependent on the Mutator (MUT) complex (Fig. I4)(Ketting et al., 1999; Phillips et al., 2012, 2014). Mutants for this complex are unable to silence TEs and their transposition leads to the occurrence random mutations, thus its name (Collins et al., 1987; Ketting et al., 1999; Vastenhouw et al., 2003).

The silencing activity of 22G RNAs can act at both TGS and PTGS, depending on which WAGO is involved. WAGO-1 and WAGO-4 are cytoplasmic and drive PTGS (Gu et al., 2009; Wan et al., 2018; Xu et al., 2018; Yigit et al., 2006), while nuclear WAGOs HRDE-1 and NRDE-1 drive the TGS in the germline and the soma, respectively (Buckley et al., 2012; Guang et al., 2008). TGS is established with the assistance of other Nuclear RNAi Deficient (NRDE) proteins, that recruit chromatin remodelling machinery and silence the target *locus* (Ashe et al., 2012; Buckley et al., 2012; Guang et al., 2008; Luteijn et al., 2012; Shirayama et al., 2012; Xu et al., 2018). This silencing pathway can become self-sustainable, but the process by which 22G RNAs generate more 22G RNAs against a target is yet unknown. Nonetheless, such 22G RNA driven 22G RNA production drives the PRG-1-triggered RNAe and a smRNA based transgenerational epigenetic memory (Ashe et al., 2012; Buckley et al., 2012; de Albuquerque et al., 2015; Guang et al., 2008; Luteijn et al., 2012; Phillips et al., 2015; Shirayama et al., 2012). Interestingly, recent studies have shown that this epigenetic memory can be triggered by environmental stimuli and increases the adaptability of descendants (Juang et al., 2013; Klosin et al., 2017; Moore et al., 2019; Posner et al., 2019; Rechavi et al., 2014). This is a striking discovery that demonstrates the existence for potential heritable adaptations without genetic modifications.

1.7.5. The CSR-1 pathway and PATCs as antagonists to PRG-1 silencing activity

CSR-1 is the only WAGO of *C. elegans* that is required for viability (Claycomb et al., 2009; Yigit et al., 2006). This WAGO is also loaded with 22G RNAs, but these differ from other 22G RNAs as they are mutator complex independent and are produced solely by EGO-1 (Claycomb et al., 2009; Maniar and Fire, 2011; Phillips et al., 2014; van Wolfswinkel et al., 2009). This pathway has been associated with different functions, including: core histone mRNA 3'end processing (Avgousti et al., 2012), the fine tuning of embryonic transcripts (Gerson-Gurwitz et al., 2016), parental memory and expression licensing of transcripts (Conine et al., 2013; Lee et al., 2012a; Seth et al., 2013, 2018; Wedeles et al., 2013). It is known that worms require a catalytically intact CSR-1 for survival (Gerson-Gurwitz et al., 2016), but its precise mode of action of this pathway and how it is required for viability are not yet understood. Currently, the most

prominent hypothesis for the action of this pathway is that it creates an RNAi-based parental memory and that this determine endogenous versus non-endogenous transcripts in *C. elegans*.

The parental memory hypothesis arises from the fact that, instead of gene silencing, this pathway targets germline transcripts that are expressed (Lee et al., 2012a; Seth et al., 2018). The PRG-1 pathway can virtually recognize and induce the silencing of any transcript (Bagijn et al., 2012; Seth et al., 2018; Zhang et al., 2018). Hence, it has been proposed that CSR-1, rather than silencing, provides a parental memory of endogenous transcripts and counters the silencing actions of the PRG-1 pathway (Claycomb et al., 2009; Conine et al., 2013; Lee et al., 2012a; Seth et al., 2013, 2018; Wedeles et al., 2013). The current model predicts that there is a "tug of war" between the PRG-1 and the CSR-1 pathways and this will determine licensing or silencing of a transcript (de Albuquerque et al., 2015; Lee et al., 2012a; Phillips et al., 2015; Seth et al., 2013, 2018; Wedeles et al., 2013).

Recently, a second feature has been discovered in sequences that further counteracts the silencing activity of PRG-1: Periodic A_n/T_n Clusters (PATCs). These are short AT-rich sequences localized in the non-coding parts of a transcript that positively regulate gene expression (Frøkjær-Jensen et al., 2016). Adding these sequences to a normally silenced sequence, such as a GFP transgene, can be sufficient to license its expression (Frøkjær-Jensen et al., 2016). The mechanism by which PATCs transmit a licensing information is not yet known, but a correlation between their presence and expression of endogenous genes in the germline has been found (Zhang et al., 2018).

1.8. The PZM is a biomolecular condensate that is part of the RNAi silencing process

The importance of biomolecular condensates in the form of liquid-liquid phase separation is a field of growing importance in RNA biology. Although not exclusively, the RNA world is particularly rich in examples of membraneless granules (Banani et al., 2017; Buchan, 2014; Morimoto et al., 2013; Seydoux, 2018). These types of structures are established by protein-protein, protein-RNA and RNA-RNA interactions that increase the local concentration of reaction components and reaction efficiency (Banani et al., 2017; Seydoux, 2018). The purpose of these structures varies from proposed functions in mRNA degradation or storage, such as P bodies and stress granules (Decker and Parker, 2012) to germ cell determinants (Bontems et al., 2009; Raz, 2003; Roovers et al., 2018). Across animal species the piRNA pathway is associated with perinuclear granules and TE silencing is believed to be active there (Brennecke et al., 2007; Houwing et al., 2007; Huang et al., 2011; Kamminga et al., 2010; Siomi et al., 2011; Voronina, 2013). *C. elegans* has different RNAi pathways that are connected to the PRG-1 and TE silencing. Still, each of these has also been proposed to act within perinuclear granules.

The first *C. elegans* perinuclear granules to be observed were the so called P granules, which got their name from the fact that they localize to the posterior P cell in early embryos (Strome and Wood, 1982). Their formation requires the presence of RNA helicases PGL-1 and PGL-3 (Hanazawa et al., 2011; Updike and Strome, 2009) and these granules harbour multiple AGOs, including PRG-1, CSR-1 and WAGO-1 (Batista et al., 2008; Claycomb et al., 2009; Gu et al., 2009; Wang and Reinke, 2008; Yigit et al., 2006). P granules have been found to reside on nuclear pores and have been proposed to be hubs for mRNA surveillance (Pitt et al., 2000; Sheth et al., 2010).

Interestingly, other perinuclear granules have been found to harbour RNAi machinery in *C. elegans*. The imaging of the mutator complex has shown that MUT proteins localize to perinuclear granules that are typically close to, but do not overlap with P granules. These were named mutator foci and loss of these granules leads to complete loss of the RNAi silencing pathway, demonstrating the importance of these structures (Phillips et al., 2012). Importantly, P granules are not affected by loss of mutator foci or vice-versa (Phillips et al., 2012). In recent work, a third perinuclear granule has been found, the Z granule. Specifically, this structure harbours WAGO-4 and its partner ZNFX-1, two proteins required for the inheritance of RNAi (Ishidate et al., 2018; Wan et al., 2018; Xu et al., 2018). Wan et al., 2018 find that the Z granule resides between the mutator foci and the P granule and they name this overall structure the PZM. The authors suggest a model where the contents of each of these granules communicate with each other. Thus, the multiple steps of RNAi are condensed in one location just outside the nuclear pore, allowing a swift and effective silencing response against a targeted sequence.

1.9. Trans-splicing in *C. elegans*

Along with the interesting specializations of its RNAi pathways, *C. elegans* harbours an additional RNA-related process of which little is known about: trans-splicing. The majority of mRNAs of *C. elegans* (70%) contain a 22 nucleotide sequence at their 5'UTR that is not transcribed from the gene *locus* (Allen et al., 2011; Gu et al., 2012). This sequence is donated by trans-splicing of one of two splice leader (SL) RNAs, SL1 or SL2. The genome of *C. elegans* contains operons that harbour about 15% of protein coding genes. The SL2 RNA is only trans-spliced into the downstream genes of polycistronic pre-mRNAs (Allen et al., 2011). Otherwise, no other correlation is known between trans-spliced or non-trans-spliced gene classes in this organism. Moreover, although this is a common strategy for gene expression amongst eukaryotes, the adaptive significance of these mechanisms remains unclear (Blumenthal, 2004; Danks et al., 2015; Douris et al., 2010; Zaslaver et al., 2011). Yet, the presence of a SL sequence at the 5'end of a given mRNA has been shown to improve its translation efficiency (Yang et al., 2017).

SL RNAs are believed to be derived from spliceosome U snRNAs and exist in the form of snRNPs, which carry Sm proteins and a 2,2,7-trimethylguanylate (TMG) Cap (Blumenthal, 2005; Hastings, 2005; Philippe et al., 2017). The transcription of SL RNAs also resembles that of other snRNAs. These genes contain proximal sequence elements (PSE) (Thomas et al., 1990), that promote RNA Pol II transcription of SL genes (Maroney et al., 1990). SL1 and SL2 genes differ in organization. The latter are spread throughout the genome, in resemblance of U snRNAs (Huang and Hirsh, 1989). On the other hand, the 100 SL1 RNA genes are clustered in chromosome V in 1kb repeats, intercalated with 5S rRNAs (Krause and Hirsh, 1987), which are transcribed in opposite orientation to SL1 genes by RNA Pol III.

The splice reaction is catalysed by the spliceosome snRNPs (Hannon et al., 1991) and it requires the presence of a signal sequence at 5'end of the mRNA transcript (Conrad et al., 1991, 1993, 1995). The signalling sequence resembles a splice acceptor site with no functional 5' splice donor upstream of it. This codes the substitution of the 5'UTR, named outron, for the SL RNA which becomes the splicing donor. Interestingly, the transesterification reaction leaves behind a Y-shape intron and a TMG capped mRNA that can be translated (Nilsen, 1993).

1.10. Aim of this thesis

Recognizing and controlling foreign invading sequences is a process of great importance for the perpetuation of a species. In animals, RNA interference mechanisms, such as the 21U RNA/Piwi pathway, are often the weapon of choice in genome defence against these invaders. In this work, we aimed at further understanding the molecular processes behind the 21U RNA pathway. Specifically, we use the nematode *Caenorhabditis elegans* as a model to understand the biosynthetic processes of the small RNA cofactor. The PID-1 protein is used as a start step to search for unknown biosynthesis machinery and we identify a novel protein complex at the centre of 21U RNA processing. Interestingly, we find that rather than specialized machinery, this nematode has co-opted pre-existing machinery into 21U RNA biogenesis. This novel protein complex sits in the crossroads between the 21U RNA pathway and essential cell processes that include snRNAs and histone mRNAs. This serves as an example of how a cell can use different toolkits for novel functions, in this case genome defence.

2. Materials and Methods

Parts of the text and figures included in this chapter and the Results chapter were published in the following scientific paper:

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

Individual contributions to the work mentioned in this thesis are listed in this chapter.

2.1. *Caenorhabditis elegans* genetics and culture

Caenorhabditis elegans strains were cultured according to standard laboratory conditions (Brenner, 1974) unless otherwise stated. Animals for IP-MS were grown at 20ºC in OP50 high density plates (Schweinsberg and Grant, 2013) for two generations and synchronized and plated on standard plates for the generation before harvest, lest indicated otherwise. Bristol N2 strain was used as reference wild type strain. Strain list below.

2.1.1. *Caenorhabditis elegans* strain list

Strain reference	Genotype	Usage
EG7833	*oxTi559[Peft-3::tdTomato::H2B::unc-54 3'UTR + Cbr-unc-119] I; unc-119(ed3) III*	CRISPR/Cas9
EG7893	*oxTi615[Peft-3::tdTomato::H2B::unc-54 3'UTR + Cbr-unc-119]; unc-119(ed3) III*	CRISPR/Cas9
EG8897	*unc-119(ed3) III; oxTi947[Peft-3::GFP::2xNLS::tbb-2 3'UTR + Cbr-unc-119] V*	CRISPR/Cas9
HT1593	*unc-119(ed3) III*	miniMos Transgenes
KK359	*tofu-6(it20); unc-4(e120)/mnC1 dpy-10(e128) unc-52(e444) II.*	
N2	Wild Type	NGS; IP-LFQP and RIP
QA137	*tofu-6(yt2) II; ytEx100*	
RFK114/SX1287	*mjIs145[Pmex-5::EGFP::his-58::21UR-1(sense)::tbb-2(3'UTR)]II*	Embryo Microscopy
RFK180	*mjIs144[Pmex-5::egfp::his-58::21UR-1_as::tbb-2(3'UTR)]; pid-1(xf14) II*	
RFK182	*pid-1(xf35) II*	NGS
RFK183	*pid-1(xf36) II*	
RFK184	*mjSi22[Pmex-5::mCherry::his-58::21UR-1_as::tbb-2(3'UTR)] I; pid-1(xf35) II*	Microscopy
RFK514	*unc-119(ed3) III; ife-3(xf101); oxTi947[Peft-3::GFP::2xNLS::tbb-2 3'UTR + Cbr-unc-119]/ oxTi664[Peft-3::TdTomato::H2B::unc-54 3'UTR + Cbr-unc-119] V*	
RFK515	*unc-119(ed3) III; ife-3(xf102); oxTi947[Peft-3::GFP::2xNLS::tbb-2 3'UTR + Cbr-unc-119]/ oxTi664[Peft-3::TdTomato::H2B::unc-54 3'UTR + Cbr-unc-119] V*	NGS
RFK523	*pid-3(tm2417) I/hT2[bli-4(e937) let-?(q782) qIs48](I;III).*	NGS
RFK625	*unc-119(ed3) III; xfIs137[Ppid-3::pid-3::mCherry::Myc::pid-3(3'UTR); Cbr-unc-119] II.*	Microscopy
RFK647	*pid-1(xf14); mjIs144[Pmex-5::egfp::his-58::21UR-1_as::tbb-2(3'UTR)] II; xfIs117[Ppid-1::pid-1::mCherry::V5::pid-1(3'UTR); Cbr-unc-119] V*	RNAi essay
RFK679	*pid-3(tm2417); xfIs136[Ppid-3::pid-3::mCherry::Myc::pid-3 (3'UTR); Cbr-unc-119] I*	IP-LFQP and RIP
RFK684	*xfIs123[Ptofu-6::tofu-6::GFP::HA::tofu-6(3'UTR); Cbr-unc-119] V*	Microscopy
RFK696	*xfIs121[Pife-3::3xFLAG::mCherry::ife-3::ife-3(3'UTR) + Cbr-unc-119] II; ife-3(xf101); oxTi947[Peft-3::GFP::2xNLS::tbb-2 3'UTR + Cbr-unc-119] V*	IP-LFQP and RIP
RFK697	*xfIs121[Pife-3::3xFLAG::mCherry::ife-3::ife-3(3'UTR) + Cbr-unc-119] II; xfIs123 [Ptofu-6::tofu-6::GFP::HA::tofu-6(3'UTR) + Cbr-unc-119] V*	Microscopy
RFK700	*xfIs136[Ppid-3::pid-3::mCherry::Myc::pid-3(3'UTR); + Cbr-unc-119] I;xfIs123 [Ptofu-6::tofu-6::GFP::HA::tofu-6(3'UTR) + Cbr-unc-119] V*	Microscopy
RFK701	*xfIs136[Ppid-3::pid-3::mCherry::Myc::pid-3(3'UTR); + Cbr-unc-119] I; pid-1(xf35) II*	IP-LFQP
RFK703	*bnIs1[Ppie-1::GFP::pgl-1 + unc-119(+)], xfIs136[Ppid-3::pid-3::mCherry::Myc::pid-3(3'UTR)+ Cbr-unc-119] I*	Microscopy
RFK721	*tofu-6(it20), unc-4(e120) II; xfIs123[Ptofu-6::tofu-6::GFP::HA::tofu-6(3'UTR)+ Cbr-unc-119] V*	IP-LFQP
RFK742	*xfIs167[Perh-2::erh-2::EGFP::OLLAS::erh-2 (3'UTR) + Cbr-unc-119] I; erh-2(xf168), oxTi615[eft-3p::tdTomato::H2B::unc-54 3'UTR + Cbr-unc-119]; unc-119(ed3) III*	IP-LFQP and Microscopy
RFK810	*erh-2(xf168); oxTi615[Peft-3::tdTomato::H2B::unc-54 3'UTR + Cbr-unc-119] III/ qC1[dpy-19(e1259) glp-1(q339) qIs26] III*	NGS
RFK861	*tost-1(xf191); oxTi615[Peft-3::tdTomato::H2B::unc-54 3'UTR + Cbr-unc-119] III*	

RFK874	*pid-3(xf149), oxTi559[Peft-3::tdTomato::H2B::unc-54 3'UTR + Cbr-unc-119] III/ hT2[bli-4(e937) let-?(q782) qIs48](I;III)*	
RFK875	*pid-3(xf153), oxTi559[Peft-3::tdTomato::H2B::unc-54 3'UTR + Cbr-unc-119] III/ hT2[bli-4(e937) let-?(q782) qIs48](I;III)*	
RFK876	*pid-3(xf151), oxTi559[Peft-3::tdTomato::H2B::unc-54 3'UTR + Cbr-unc-119] III/ hT2[bli-4(e937) let-?(q782) qIs48](I;III)*	
RFK905	*tost-1(xf194), oxTi615[Peft-3::tdTomato::H2B::unc-54 3'UTR + Cbr-unc-119] III/ qC1[dpy-19(e1259) glp-1(q339) qIs26] III*	NGS
RFK912	*tost-1(xf196), oxTi615[Peft-3::tdTomato::H2B::unc-54 3'UTR + Cbr-unc-119] III*	Temperature Shift

2.2. Mutant generation with CRISPR/Cas9 system

Mutant alleles were generated as described in Friedland et al., 2013. gRNAs were selected under the criteria: NGG PAM site, highest GC content and specificity according to CRISPRdirect (Naito et al., 2015) and Zhang Lab's http://crispr.mit.edu. Two to three gRNA, singularly cloned into Addgene plasmid #46169, were injected (35ng/µl) together with Addgene plasmid #46168 (50ng/µl) and co-injection marker pRR83 (5ng/µl) into adult worms (specific strains above). F1 worms positive for pharynx GFP expression were isolated, allowed reproduction, lysed in single worm lysis buffer (5 mM KCl, 2,5 mM MgCl2, 10 mM Tris pH8,3, 0,45% IGEPAL, 0,45% Tween-20, 0,01% gelatin) and genotyped for mutations using NEB Taq DNA Polymerase (M0273X) according to manufacturer's instructions. Isolated mutants were outcrossed at least two times before balancing. CRISPR/Cas9 guide RNA sequences below:

Target	CRISPR/Cas9 Guide RNA
ife-3	GCCTCCGTGCCGGGATTCGA
ife-3	GACACCCCCTCCAGAATCGC
ife-3	GAGCCCAGCGATTCTGGAGG
pid-3	GAAAATGGTTGCCCATCAGA
pid-3	GTGGAAGAATGTGCACGACG
pid-3	GGCGGATTTCAAGTCGAAAT
erh-2	GTGAGAATTATTATGTTTAA
erh-2	GAGCAGCTGATTTCTTGGAA
erh-2	GAAGATCATCATAGAAACAT
tost-1	GATAGTTCTGAAACATAACC
tost-1	GAGCTTCTTCTCATCAGTAG
tost-1	GATGGCAGTAGTCATTCTGA

2.3. miniMos transgene insertion and mapping

Random miniMos insertions were made through injection of unc-119(ed3) carrying worms. *C. briggsae* unc-119 was used as a selection marker. Injections and mapping were made in accordance to Frøkjær-Jensen et al., 2014.

2.4. Embryonic arrest and transgene complementation tests

Embryonic arrest stage was determined by single picking wildtype and mutant gravid individuals, bleaching and synchronizing them in M9 buffer for 16h. Embryos and larvae were then imaged with wide-field microscopy.

Individuals carrying mutant alleles or carrying both mutant allele and corresponding miniMos transgene were singled at L4 larvae (n>5) stage and allowed to self-fertilize and reproduce for 24h. At this time point progenitors were removed from the plate and embryos/larvae counted. After 48h the number of larvae in each plate was counted and the proportion of arrested embryos for each progenitor was determined.

2.5. Microscopy

Wide-field fluorescence microscopy images were obtained using a Leica DM6000B and Confocal microscopy images were acquired with a Leica TCS SP5. Images were processed using Leica LAS software, ImageJ and Adobe Photoshop.

2.5.1. Immunostaining

Adult worms were dissected in Egg Buffer (25mM HEPES pH7,4, 118mM NaCl, 48mM KCl, 2mM EDTA, 0,5mM EGTA) with 1%(v/v) Tween20 and fixed 5 minutes by adding 1:1 Egg Buffer+2% formaldehyde followed by a wash step in Egg Buffer. Cuticle was then removed by Freeze cracking (Duerr, 2013). An extra fixation step of 1 minute in -20ºC Methanol preceded three washes in PBS (137mM NaCl, 2,7mM KCl, 10mM Na_2HPO_4, 2mM KH_2PO_4, pH7,5) with 0,5%(v/v) Tween20 (PBST). After 1 hour in blocking buffer (PBST+10% Bovine Serum) samples were co-stained overnight at 4ºC with 1:200 dilutions of RFP-Booster_Atto647N (Chromotek, rba647n-10) and GFP-Booster_Atto488 (Chromotek, gba488-10). Staining was followed with multiple PBST washes and samples were mounted in ProLong Gold Antifade Mountant (ThermoFisher Scientific, P36930).

2.5.2. Live imaging

RNAi treated RFK114 adult worms were dissected and mounted in Egg Buffer (25mM HEPES pH7,4, 118mM NaCl, 48mM KCl, 2mM EDTA, 0,5mM EGTA) and directly imaged in a Leica DM6000B wide-field fluorescence microscope.

2.6. Yeast two hybrid

Two-hybrid assays were performed in the haploid strain PJ69-4a and the pGAD and pGBD plasmid series as described previously in James et al., 1996. Cell pinning was performed with Rotor HAD (Singer Instruments, ROT-001).

2.7. RNAi experiments

HT115(DE3) bacteria carrying Timmons and Fire L4440 RNAi feeding vector (Timmons and Fire, 1998) were grown over 10 hours and seeded directly onto RNAi plates (standard NGM; 1mM IPTG and 50μg/mL ampicillin). HT115(DE3) with empty L4440 or carrying *pid-3* or *tost-1* targeting RNA were taken from the Ahringer RNAi library (Kamath et al., 2003). Remaining vectors were made by inserting cDNA of its corresponding gene into L4440 and then retransformed into HT115(DE3).

In all RNAi treatments animals were synchronized at L1 larvae and seeded into RNAi plates containing induced bacteria. For RT-qPCR, worms were imaged in adulthood and harvested for RNA collection. Experiment was repeated three independent times.

2.8. RNA isolation and treatments

C. elegans were collected off plate and washed with M9 buffer(22mM KH_2PO_4, 42mM Na_2HPO_4, 85mM NaCl, 1mM $MgSO_4$) followed by a wash with ultrapure water and lysis in Worm Lysis Buffer (0,2M NaCl, 0,1M Tris pH8,5, 50mM EDTA, 0,5% SDS) with 1mg/mL Proteinase K (Sigma-Aldrich, P2308) for 30 minutes at 65°C. After pelleting and removing debris, three volumes of TRIzol LS (ThermoFisher Scientific, 10296-028) were added to sample and RNA precipitation was carried out according to producer's instructions with the aid of Phase lock Gel – Heavy tubes (QuantaBio, 2302830). Eluted RNA samples were depleted of DNA using TURBO DNA-free Kit (Ambion, AM1907).

RNA Immunoprecipitation samples (see below) were obtained by adding TRIzol LS directly to IP beads after washes. The consecutive isolation steps follow the above described process.

2.8.1. RNA RppH and CIP-RppH treatment

For RppH treatments a portion of an RNA sample was collected and treated with RNA 5' Pyrophosphohydrolase (RppH) (NEB, M0356) for the purpose of removing 5' Cap structures (Almeida et al., 2019c). Each was treated in ThermoPol Buffer (20 mM Tris.Cl, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0,1% Triton X-100, pH8.8, NEB, B9005) for 1 hour at 37°C with 10 units of RppH. Reaction was stopped by adding EDTA to 10 mM and heating to 65°C for 5 minutes. RNA was purified by ethanol precipitation. CIP-RppH samples were treated in CutSmart Buffer (NEB B7204S) with Calf Intestinal Phosphatase (NEB, M0290L) at

20U/10µg RNA and 0.5U/µl SUPERase.In RNase Inhibitor (Ambion, AM2696) for 1h at 37°C. After TRIzol re-purification samples were treated with RppH as described above.

2.9. RT-qPCR

Cultured worms and RNA samples were isolated as described above. Reverse transcription for each sample was performed with 500ng of total RNA using ProtoScript First Strand cDNA Synthesis Kit (NEB, E6300) and Oligo $d(T)_{23}VN$. For histone and trans-splicing qPCR Random Primer Mix was used instead. qPCR 10µl reactions were set up with iTaq Universal SYBR Green Supermix (Bio-Rad, 1725121), 500mM primer concentration and a volume ratio of 1/5 cDNA. PCR cycles and measurements were made in an Applied Biosystems ViiA7 Real Time PCR System (ThermoFisher Scientific). Cycling conditions were made according to iTaq manufacture's recommendations: Standard run, temperature increments of 1,6°C/s; 95°C for 30 seconds, 40 cycles of 95°C for 15 seconds and 60°C for 1 minute; melt curve calculation: 15 seconds at 95°C, 1 minute at 60°C, temperature increments of 0,05°C/s to 95°C and hold for 15 seconds. Technical duplicates and biological triplicates were used. ΔΔCT method was used as an analysis method (Schmittgen and Livak, 2008). *pmp-3* was used as a normalization factor (Hoogewijs et al., 2008). Error bars represent the standard deviation of three biological replicates. Used primers are listed below.

Target	Sequence
GFP_Fw	ATGGTGTTCAATGCTTCTCG
GFP_Rev	TGACTTCAGCACGTGTCTTGT
pmp-3_control_Fw	GTTCCCGTGTTCATCACTCAT
pmp-3_control_Rev	ACACCGTCGAGAAGCTGTAGA
his-66_Fw	CAAGCAAGTTCACCCAGACA
his-66_Rev	TCTCCTGGGAGAATCAAACG
his-65_Fw	GTCGGTCGTCTTCACCGTAT
his-65_Rev	AGCTTGTTGAGCTCCTCGTC
SL1_Fw	GGTTTAATTACCCAAGTTTGAG
pmp3_5prime_Fw	CCCTCGACTTCCTCTTCTTT
pmp3_5prime_Rev	AAACGTGTCAGCATCCTTGT

2.10. *tost-1(xf196)* temperature shift assays

2.10.1. Viable progeny quantification

RFK912 worms were cultured at 15°C on standard plates. At the start of experiment they were selected and singled into standard plates at L4 larvae stage. After overnight (O/N) culture at 15°C or 25°C individual worms were transferred into a new plate and shifted to corresponding

temperature together with the plate of O/N egg lay. Every 2 hours individuals were transferred into a new plate. Eggs were counted in each of these plates on the day of egg lay and two days after larvae were counted for survival assay. As control we include RFK912 worms which underwent the same treatment except with no temperature shift and transferred into new plates every 4 hours.

2.10.2. Small RNA sequencing

Synchronized RFK912 and N2 worms were cultured at 15ºC on standard plates. At gravid adult stage plates were shifted to 25ºC and collected after 4h and 12h and RNA was isolated as indicated above. 0h indicates non-temperature shifted plates.

2.11. Small RNA Library preparation and sequencing

Performed by the IMB genomics core facility

NGS library prep was performed with NEXTflex Small RNA-Seq Kit V3 following Step A to Step G of Bioo Scientific`s standard protocol (V16.06). Libraries starting amount and PCR cycles can be consulted in the table below. Amplified libraries were purified by running an 8% TBE gel and size-selected for 18 – 40nt. Libraries were profiled in a High Sensitivity DNA on a 2100 Bioanalyzer (Agilent technologies) and quantified using the Qubit dsDNA HS Assay Kit, in a Qubit 2.0 Fluorometer (Life technologies). Samples of each individual experiment were pooled in equimolar ratio. Sequences were deposited at SRA, submission number PRJNA503945.

Experiment	Sample (each 3x)	Starting Material	PCR cycles	Equipment	Run type
pid-3	*pid-3(tm2417)*	1000ng	15	NextSeq 500 Flowcell	Highoutput 75-cycle-kit, SR for 1x 83 cycles plus 7 cycles for the index read
	pid-3(tm2417)/+	1000ng	15	NextSeq 500 Flowcell	Highoutput 75-cycle-kit, SR for 1x 83 cycles plus 7 cycles for the index read
ife-3	*ife-3(xf102)*	500ng	15	NextSeq 500/550 Flowcell	SR for 1x 75 cycles plus 7 cycles for the index read
	N2	500ng	15	NextSeq 500/550 Flowcell	SR for 1x 75 cycles plus 7 cycles for the index read
erh-2	*erh-2(xf168)*	500ng	15	NextSeq	SR for 1x 75 cycles plus 7

				500/550 Flowcell	cycles for the index read
	erh-2(xf168)/+	500ng	15	NextSeq 500/550 Flowcell	SR for 1x 75 cycles plus 7 cycles for the index read
tost-1	*tost-1(xf194)*	2000ng	12	NextSeq 500 Flowcell	Midoutput 150-cycle-kit, PE for 2x 75 cycles plus 16 cycles for the index read
	pid-1(xf35)	2000ng	12	NextSeq 500 Flowcell	Midoutput 150-cycle-kit, PE for 2x 75 cycles plus 16 cycles for the index read
	N2	2000ng	12	NextSeq 500 Flowcell	Midoutput 150-cycle-kit, PE for 2x 75 cycles plus 16 cycles for the index read

2.12. Biochemistry

2.12.1. Worm preparation

Synchronized non-gravid adult worms were collected off plate and washed with M9 buffer followed by a wash with ultrapure water. Pellets were frozen with liquid nitrogen and kept at -80°C until usage.

2.12.2. Lysate preparation

Worm Pellets were thawed on ice and mixed 1:1 with 2x Lysis Buffer (20 mM Tris.Cl, 300 mM NaCl, 1 mM EDTA, 1%(v/v) IGEPAL CO-630, pH7,5) with 2x protease inhibitors (cOmplete Mini, EDTA-free, Roche, 11836170001). Bioruptor Plus (Diagenode) sonicator was used to lyse worms (10 cycles 30/30 seconds, 4°C, high energy) and debris removed by spinning. Lysate protein concentration was determined with Pierce BCA Protein Assay Kit (ThermoFisher Scientific, 23225).

2.12.3. Immunoprecipitations

Lysates were diluted in 1x Lysis Buffer+ 1x Protease inhibitors to a final concentration of 1,5 mg of protein/mL and a total of 0,75 mg of protein was used per IP. At this step input samples were collected into 2x NuPAGE LDS Sample buffer (Life Technologies, NP0007)+ 200 mM DTT and boiled for 10 minutes. Anti-mCherry IPs were performed with RFP-Trap_MA beads (Chromotek, rtma-20) and anti-GFP IPs with GFP-Trap_MA beads (Chromotek, gtma-20), in both cases 25 µl of bead slurry was used and samples were rotated at 4°C for 2 hours.

Subsequent washes were made with Wash Buffer (10 mM Tris.Cl, 150 mM NaCl, 0,5 mM EDTA, pH7,5)+ Protease inhibitors in accordance with Chromotek protocols. Washed beads were resuspended in 2x NuPAGE LDS Sample Buffer + 200 mM DTT and boiled for 10 minutes, making the samples ready for loading.

2.12.4. RNAse treated immunoprecipitations

These followed the above described protocol with an additional RNAse A/T1 Mix (ThermoFisher Scientific, EN0551) treatment step. After lysate dilution samples were divided in two (Control and +RNAse) and 20 µL of RNAse A/T1 mix was added per 1 mL of +RNase sample. Control and +RNase samples were rotated for 20 minutes at 4ºC and followed by the described IP protocol.

2.12.5. Stringent washes treated immunoprecipitations

These followed the immunoprecipitation standard protocol with the only difference being the Wash Buffer used. Depending on the IP-LFQP Wash Buffer contained 150mM, 350mM or 500mM of NaCl as indicated in graphs.

2.12.6. Western blot

Inputs and IP samples were loaded into 4-12% gradient gels (ThermoFisher, NP0321BOX) and run with 1x NuPAGE MES SDS Running Buffer (ThermoFisher, NP0002). Transfer to an Immobilon PVDF, 0,45 µm membrane (Merck Millipore, IPVH00010) was executed with 1x NuPAGE Transfer Buffer (ThermoFisher Scientific, NP0006) 20%(v/v) Methanol. Membrane was probed with rabbit anti-PID-1 Q5941 (de Albuquerque et al., 2014) and detected with Amersham ECL Select Western Blotting Detection Reagent (GE Healthcare, RPN2235). Background recognition by anti-PID-1 ab is used as loading control.

2.12.7. Endogenous PID-1 Immunoprecipitations

Performed by Dr. Bruno F.M. de Albuquerque at the Rene F. Ketting Lab

200 µL of synchronized adult worms were resuspended in 500 □L of IP lysis buffer (25 mM Tris pH7,5, 150 mM NaCl, 1,5 mM $MgCl_2$, 1 mM DTT, 0,1% Triton X-100, complemented with 2x protease inhibitor) and sonicated at 4ºC for 10 cycles of 30/30 seconds, high intensity using a Bioruptor Plus (Diagenode). Cell debris was removed via spinning and 30 µL of washed Dynabeads Protein G (Life Technologies, 1004D) and 10 µL of anti-PID-1 antibody (Q5941) was added to the lysates and incubated under rotation for 3 hours at 4ºC. The beads were then washed 3x 5 minutes in wash buffer (25 mM Tris pH7,5, 150 mM NaCl, 1,5 mM $MgCl_2$, 1 mM DTT, complemented with 2x protease inhibitor) and resuspended in 30 µL of NuPAGE LDS buffer.

2.12.8.Mass Spectrometry

Performed by Sabrina Dietz at the Falk Butter Lab

Samples were separated on a 4–12% gradient Bis-Tris gel (ThermoFisher, NP0321) in MOPS SDS Running Buffer (ThermoFisher, NP0001) at 180 V for 10 minutes, afterward separately processed by in-gel digest (Kappei et al., 2013; Shevchenko et al., 2006) and desalted using a C18 StageTip (Rappsilber et al., 2007). The digested peptides were separated on a 25cm reverse-phase capillary (75µm inner diameter) packed with Reprosil C18 material (Dr. Maisch). Separation of the peptides was done with the EASYnLC 1000 system (Thermo) along a 2 hour gradient increasing from 2 to 40% Buffer B. For PID-1 IPs the gradient was shortened to 90 minutes. Measurement was done on a Q Exactive Plus mass spectrometer (Thermo) operated with a Top10 data-dependent MS/MS acquisition method per full scan (Bluhm et al., 2016). Measurements were processed with MaxQuant version 1.5.2.8 (Cox and Mann, 2008) using the wormbase protein fasta database (version WS265) and standard settings except LFQ quantitation and match between runs were activated. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD011500.

2.12.9.Size exclusion chromatography-Western blot

Performed in collaboration with Dr. Christian Renz at the Helle Ulrich Lab

Extract without tagged PETISCO in Fig.R5H refers to 3xFLAG::RRF-3; *rrf-3(pk1426)* adult extracts. These extract fractions were refurbished from Almeida et al., 2018. Lysates for TOFU-6::GFP::HA (RFK721) and ERH-2::GFP::OLLAS (RFK742) were prepared as described above. 4,5mg of total protein were separated on a Superose 6, 10/300 GL size exclusion column (GE Healthcare, 17517201), using a NGC Quest system (BioRad) and samples were collected as described in Almeida et al., 2018. Amicon Ultra 3kDa cutoff filter units (Merck-Millipore, UFC500324) were used to 13x concentrate each fraction and 25% of sample was used for SDS-PAGE/Western blot. TOFU-6::GFP::HA, ERH-2::GFP::OLLAS and PID-1 proteins were detected in independent extracts with antibodies anti-HA (Sigma-Aldrich, SAB4300603-100UG), anti-GFP (Santa Cruz Biotechnology, sc-9996) and anti-PID-1 (Q5941), respectively.

2.13. RIPseq

2.13.1.Lysate preparation

Worm Pellets were thawed on ice and mixed 1:1 with 2x Lysis Buffer (20 mM Tris.Cl, 300 mM NaCl, 1 mM EDTA, 1%(v/v) IGEPAL CO-630, pH7,5) with 2x protease inhibitors (cOmplete Mini, EDTA-free, Roche, 11836170001) and 2x SUPERase.In RNase Inhibitor (Ambion, AM2696). Bioruptor Plus (Diagenode) sonicator was used to lyse worms (10 cycles

30/30 seconds, high energy). Lysate protein concentration was determined with Pierce BCA Protein Assay Kit (ThermoFisher Scientific, 23225). Lysates were diluted in 1x Lysis Buffer+ 1x Protease inhibitors+ 1x SUPERase.In RNase Inhibitor to a final concentration of 1,5 mg of protein/mL and a total of 2,1 mg of protein was used per IP. Each lysate was cleared with 225 µL of Binding Control magnetic agarose beads (Chromotek, bmab-20) for 1 hour at 4°C.

2.13.2.Immunoprecipitation

Quadruplicate anti-mCherry RIPs were performed with RFP-Trap_MA beads (Chromotek, rtma-20). 75 µl of bead slurry per sample blocked for 1 hour with Blocking Buffer [2% (w/v) BSA, 2,5 mg/mL tRNA from *E.coli* MRE 600 (SigmaAldrich, 10109541001), 10 mM Tris.Cl, 150 mM NaCl, 0,5 mM EDTA, pH7,5] and washed with Wash Buffer (10 mM Tris.Cl, 150 mM NaCl, 0,5 mM EDTA, pH7,5). Inputs from cleared lysates were taken and mixed 3:1 with TRIzol. 75 µl of blocked bead slurry was added to the remaining cleared lysate and samples were rotated at 4°C for 2 hours. Subsequent washes were made with Wash Buffer+ Protease inhibitors. Washed beads were resuspended in 100 µL of Nuclease free water and immediately mixed with 400 µL of TRIzol.

2.13.3.Library preparation and Sequencing

Performed by the IMB Genomics Core Facility

NGS library prep was performed with NEXTflex Small RNA-Seq Kit V3 following Step A to Step G of Bioo Scientific`s standard protocol (V16.06). Libraries were prepared with a starting amount of 100 ng and amplified in 18 PCR cycles. Amplified libraries were purified by running an 8% TBE gel and size-selected for 18 – 40nt. Libraries were profiled in a High Sensitivity DNA on a 2100 Bioanalyzer (Agilent technologies) and quantified using the Qubit dsDNA HS Assay Kit, in a Qubit 2.0 Fluorometer (Life technologies). Total amount of samples were divided in two pools. Each pool was mixed in equimolar ratio and sequenced on a NextSeq 500/550 Flowcell, SR for 1x 75 cycles plus 7 cycles for the index read.

2.14. Bioinformatic analysis

2.14.1.Alignments, domain structure predictions

Protein alignments were performed with either ClustalO (Sievers et al., 2011) for two sequences or MUSCLE (Edgar, 2004b, 2004a) for multiple sequence alignments (MSA). Representation of the alignments was also made with ESPript 3.0 (Robert and Gouet, 2014).

Structure predictions were made with both with MPI Bioinformatics toolkit HHpred (Alva et al., 2016) and I-TASSER (Yang et al., 2015). HHpred domain predictions were performed using local alignments with HHblits uniclust30_2018_08 local as MSA generation method and maximal number of 8 and E-value threshold of $1x10^{-3}$ for generation steps. I-

TASSER was performed using hAGO2 RCSB Protein Data Bank ID: 4F3T (Schirle and MacRae, 2012) as a guide for I-TASSER modelling.

2.14.2.smRNA sequencing analysis

Performed by Dr. Antonio Miguel de Jesus Domingues at the Ketting Lab

Raw reads were first processed to remove adapters with v1.9 cutadapt (https://cutadapt.readthedocs.io/en/stable/) (seqtk trimfq -L 50 | cutadapt -a TGGAATTCTCGGGTGCCAAGG -O 5 -m 26 -M 48), followed by removal of reads containing low quality calls with the FASTX-Toolkit v0.0.14 (fastq_quality_filter -q 20 -p 100 -Q 33). The information of the read sequence and the 5' and 3' random UMIs (NNNN-RNA sequence-NNNN) was then used to collapse reads with identical sequences, including that of the UMIs, using a command-line script. UMIs were then removed (seqtk trimfq -b 4 -e 4), and reads shorter than 15 nucleotides were filtered out (seqtk seq -L 15) before mapping against the *C. elegans* genome (WBcel235, ensembl) with bowtie v0.12.8 (Langmead et al., 2009) (-q –sam –phred33-quals –tryhard –best –strata -v 0 -M 1). Coverage tracks were generated with Bedtools 2.25.0 (Quinlan and Hall 2010) (genomeCoverageBed -bg -split -scale) to summarize genomic read coverage, and the bigwigs created with bedGraphToBigWig. Normalization was done to total mapped reads. For visualization, the alignments of different replicated for the same sample were merged with bamtools-2.3.0 merge (Barnett et al., 2011). For the RIP-seq experiments, merged alignments were further processed to create $\log_2$(IP/input) normalized tracks using DeepTools (Ramírez et al., 2016) (bigwigCompare –binSize 1 –ratio log2).

Reads mapping to annotated features in the custom GTF were counted with htseq-count v0.9.0 (Anders et al., 2015) (htseq-count -f bam -m intersection-nonempty ~~s reverse)~~. To identify RNA-bound to the complex in RIPseq, we used DESeq2 (Love et al. 2014) with the formula '~replicate+condition' in which each IP is being compared to the corresponding input control. Replicates are paired as they are generated from the same biological sample. All the samples in the dataset were included in construction of the DESeq2 object, in order to estimate the dispersion more robustly.

For the differential expression of histone, 21U and miRNAs we performed pairwise differential gene expression estimation with DESeq2 (Love et al., 2014), using the counts for all conditions in the construction of the DESeq2 object for a more robust dispersion estimate. For visualization, genes belonging to a particular biotype (piRNA, miRNA, histones) were extracted from the final results table and their mean expression versus log2 Fold-Change shown in an MA-plot.

The number of reads mapping to different RNA classes was estimated with a combination of a custom Python script to select reads by size and nucleotide bias, available at

https://github.com/adomingues/filterReads/blob/master/filterReads/filterSmallRNAclasses.py, and bedtools intersect to match reads with annotated features. 21U RNAs were defined as reads with 18-40 bases mapping sense (intersectBed -s -f 0.85) to an annotated 21U RNA *locus* for type I, or the type II 21U-RNA identified in Gu et al., 2012, supplementary table 3B, following conversion of the coordinates from WS215 to WBcel235 with crossMap (Zhao et al., 2014). 21U-RNA precursors are reads that map to the -2 position with respect to the mature 5' end of annotated 21U-RNAs and are at least 23 nucleotides in length. 26G RNAs are 26 nucleotide long reads mapping antisense to protein coding genes, pseudogenes and lincRNA. For 26G RNAs a minimum overlap of 1 base was required (intersectBed default). miRNAs were defined as reads mapping sense to annotated miRNAs (intersectBed -s -f 1.0). Genomic locations were extracted from a custom gtf (genes + transposons) using the biotype information.

Metagene profiles were created with DeepTools. Read coverage was summarized with computeMatrix scale-regions –metagene –missingDataAsZero -b 50 -a 50 –regionBodyLength 98 –binSize 1 –averageTypeBins mean. As SL genes are multicopy genes, the setting –averageTypeBins was set to "sum". Using the "mean" of SL sequences did not alter the profile obtained (data no shown). The final metagene figure was created with plotProfile –plotType lines –perGroup.

2.14.3. Evolutionary analysis of PETISCO

Performed by Dr. Peter Sarkies at the Peter Sarkies Lab

Predicted proteomes of *Caenorhabditis* species were downloaded from Caenorhabditis.org. Selected other species were downloaded from Wormbase parasite (parasite.wormbase.org - WBPS5). The proteome of *C. elegans* WS235 was used as the test file for reciprocal blastp searches against all other species, recording only the best hit. The bit score of the best blast hit was extracted for the PETISCO complex, with ERH-1, PRDE-1 and PRG-1 included for comparison. In the case that no best reciprocal blast hit was found, the score was given as 0. Scores were then normalized by dividing by the score from blasting *C. elegans* against itself, to control for the different lengths of the proteins. In the case that no hit was found, genomic nucleotide sequence was downloaded and Exonerate was used to search for potential unannotated orthologues; identified hits were then reconstructed from the predicted protein sequence output from Exonerate and then tested using best reciprocal blastp as above. The heatmap.2 function in gplots within R was used to generate the heatmap. The order of the species in the heatmap was estimated by calculating the mean blastp score for the entire *C. elegans* proteome.

3. Results

Parts of the figures included in this chapter were published in the following scientific paper:

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

Individual contributions to the results here mentioned are listed under the Material and Methods.

3.1. Mass spectrometry screen identifies RNA machinery as PID-1 cofactors

pid-1 (PiRNA Induced silencing-Defective 1) has been previously described as essential for 21U RNA processing (de Albuquerque et al., 2014). This relatively small protein (19kDa) has no assigned domains, leaving little clues to its molecular activity within the 21U RNA pathway. The size and lack of discernible domains led us to consider it unlikely that this protein acts by itself. We therefore aimed at understanding PID-1 activity via identification of its potential interactors.

In order to identify PID-1 interactors we performed a series of immunoprecipitations (IP) targeting the endogenous PID-1 protein, followed by label-free quantitative proteomics (LFQP). We used a polyclonal antibody to perform quadruplicate IPs in wild-type (WT) worms and strains carrying previously described mutations of the *pid-1* gene as a control (de Albuquerque et al., 2014). PID-1 cofactors should only be enriched in WT worms as unspecific Ab binding enrichments should be cleared as background in our comparison with the mutant strains. Under these conditions we found nine proteins to be systematically co-precipitating with PID-1 (Fig. R1), including four factors with known connections to RNA biology: IFE-3, TOFU-6, Y23H5A.3 and F35G12.11 (Goh et al., 2014; Jankowska-Anyszka et al., 1998; Sugiyama et al., 2016). Interestingly, TOFU-6 has previously been found in an RNAi screen as part of the 21U RNA biogenesis machinery (Goh et al., 2014). In the same study Y23H5A.3 has also been found to be a 21U RNA biogenesis candidate. However, this gene was not included in the authors' validation short list, and hence its role in 21U RNA biogenesis remained uncertain. In high throughput yeast two hybrid screens, TOFU-6 has also been reported to interact with both IFE-3 and Y23H5A.3 (Boxem et al., 2008; Simonis et al., 2009), increasing our confidence of a

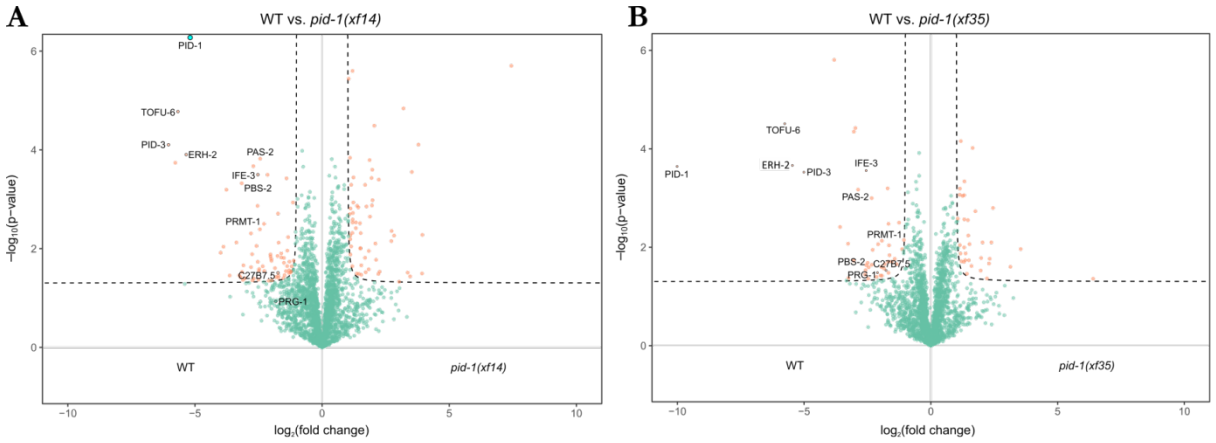

Figure R1 - PID-1 has a systematic set of interactors.
Volcano plot representing label-free proteomic quantification of PID-1 IPs from non-gravid adult extracts. IPs were performed and analysed in quadruplicates. The x-axis represents the median fold enrichment of individual proteins in wild type (WT) versus (A) *pid-1(xf14)* or (B) *pid-1(xf35)* mutant strains. y-axis indicates -Log_{10}(p-value) of observed enrichments. Dashed lines represent thresholds at p=0.05 and 2-fold enrichment. Blue data points represent values out of scale. Red and Green data points represent values above and below threshold, respectively.

partnership between these proteins.

The remaining four proteins included the 20S proteasome subunits PBS-2 and PAS-2, glycine decarboxylase 1 (GLDC-1) and protein arginine methyltransferase 1 (PRMT-1). The object of the present study is the 21U RNA biogenesis. Although we cannot exclude a role for PBS-2, PAS-2, GLDC-1 and PRMT-1 on this pathway, the ontology of the remaining factors brings them closer to our object. Thus, in the interest of not diluting our resources we focused our efforts on the RNA related cofactors of PID-1.

3.2. Biochemical nature of PID-1 interactors predicts RNA related functions

Some of the RNA related interactors of PID-1, e.g. IFE-3 and TOFU-6, have already been named from previous work. Y23H5A.3 and F35G12.11 have yet to be assigned a name. Thus, we have named the first PiRNA Induced silence-Defective 3 (PID-3) for its association with PID-1. T21C9.4 and F35G12.11 are the two enhancer of rudimentary orthologues present in *C. elegans*. We respectively named them enhancer of rudimentary homologue (ERH) 1 and 2, in line with other enhancer of rudimentary orthologues. ERH-2 was named as such for being the more distant of the two from the human Erh1. For each PID-1 interactor we find relevant data pointing towards an action in RNA biology:

3.2.1. *IFE-3*

IFE-3 is one of the five *C. elegans* orthologues of eIF4E, a protein superfamily that is known to interact with the 5' Cap of RNA. Specifically, IFE-3 selectively binds 7-methylguanylate (m7G)(Keiper et al., 2000; Miyoshi et al., 2002), an integral part of the RNA 5' end Cap. It is a 28kDa protein with a single eIF4E superfamily domain (Fig. R2A), and is the closest C. elegans orthologue to human eIF4E1.

3.2.2. *ERH-2*

ERH-2 is a 13kDa protein with a single domain (Fig. R2A) only found within this protein superfamily (Arai et al., 2005). To our knowledge there are no studies on either of the Erh worm orthologues. Still, the *Schizosaccharomyces pombe* Enhancer of Rudimentary orthologue has been associated with nuclear RNA decay (Sugiyama et al., 2016) and this protein superfamily has been recurrently linked to RNA metabolism (Weng et al., 2012; Wojcik et al., 1994).

3.2.3. *TOFU-6*

BLAST analysis (Altschul, 1997) of these factors fails to find orthologues of PID-3 or TOFU-6 outside of the *Caenorhabditis* genus. In an attempt to find possible orthologues and protein domains undetectable by the BLAST tool, we analysed these two proteins using the MPI

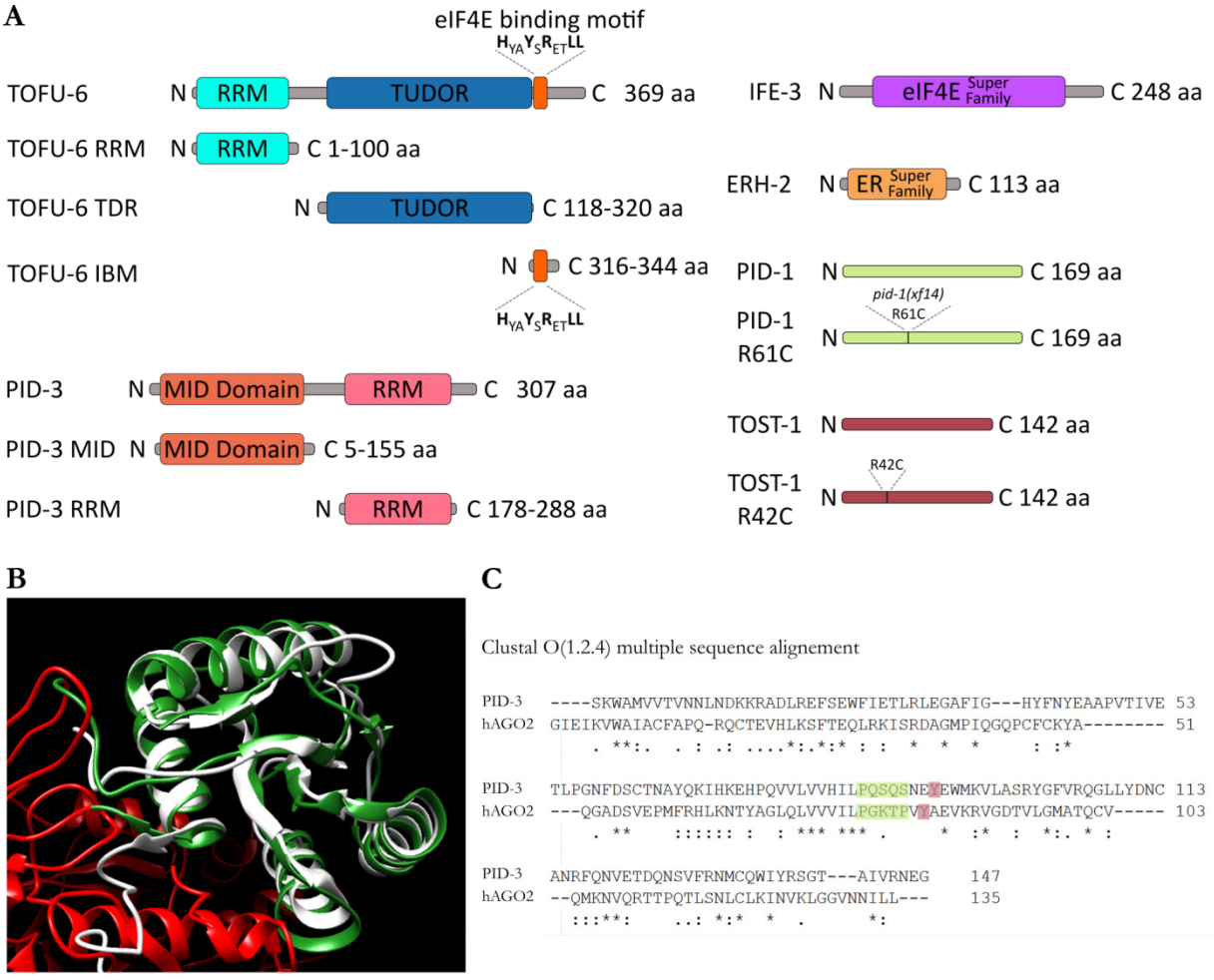

Figure R2 - Protein domains of PID-1 interactors. (A) Schematic representation of domain composition of PID-1 interactors. Note that the same scheme applies to for domains used in the Y2H grid. In that experiment proteins were C-terminally fused to GAL4 activation or binding domains. (B) Structural alignment between the PID-3 MID domain and hAGO2. PID-3 is in white, hAGO2 in red and hAGO2 MID domain in green. (C) ClustaO alignment of PID-3 and hAGO2 MID domains. Note that there is very little sequence similarity including the nucleotide selectivity loop (green) and the 5'end PO_4 interacting tyrosines (red) that do not align.

Bioinformatics Toolkit's HHpred (Zimmermann et al., 2018). HHpred uses pairwise comparison of hidden Markov models (HMM) to find remote homology between protein sequences or structures. Interestingly, this tool not only confirms the previously reported presence of an RNA Recognition Motif (RRM) and a TUDOR Domain (TDR) (E-value $1.2x10^{-9}$ and $8.6x10^{-23}$, respectively) in TOFU-6 (Goh et al., 2014), but also identifies a possible eIF4E binding motif at the C-terminal portion of this protein (E-value 8.1) (Fig. R2A).

3.2.4. PID-3

We find that PID-3, like TOFU-6, contains an RRM domain followed by a domain with high resemblance to the human AGO2 MID domain (E-value $2,4x10^{-4}$ and $7,0x10^{-4}$, respectively) (Fig. R2A). MID domains are found in AGO proteins and are known to play a key role in the interaction between the 5'PO_4 of the sRNA and the AGO (Boland et al., 2010; Frank et al., 2010; Schirle and MacRae, 2012). We further pursued this lead by performing an I-TASSER structure alignment of this domain with the previously published hAGO2 structure (Schirle and MacRae,

2012; Yang et al., 2015). The I-TASSER algorithm gave the PID-3 MID domain structure a confidence score of -0.03 in a lowest to highest confidence range of -5 to 2. Although we find little sequence homology between these two proteins, there is a high resemblance between the predicted structure of the PID-3 MID domain and the MID domain of hAGO2 (Fig. R2B and C). Thus, we hypothesize that PID-3 may interact with the 5' PO_4 of RNAs (see section 1.3.1.).

3.3. PETISCO proteins localize to the P-granules in the adult gonad

Determining the subcellular localization of a protein is essential to understand its function as it provides us context for its action. We created transgenes for each of the PID-1 interactors carrying fluorescent and epitope tags in order to investigate their protein expression and subcellular localization (Fig. R3A). Each transgene was created using miniMos (Frøkjær-Jensen et al., 2014), a technique for stable expression of randomly inserted transgenes based on the *Drosophila* transposon Mos1. In all cases the endogenous promoter and 3'UTR were used in order to replicate the original gene expression pattern (Fig. R3A). We tested the ability of these transgenes to replicate the function of wild type alleles by genetic complementation. We find in all cases that embryonic lethality of corresponding mutants (discussed below) is fully complemented by the transgenic alleles (Fig. R3B and C).

C. elegans glp-4 mutants lack germline tissue when cultured under stringent conditions (Beanan and Strome, 1992). The lack of detection of a protein in this strain serves as an indirect demonstration that it is expressed only in the worm germline. This approach was previously used to demonstrate germline specificity for both PID-1 and IFE-3 (de Albuquerque et al., 2014; Amiri et al., 2001). In accordance, we observed that the expression of all the different fusion

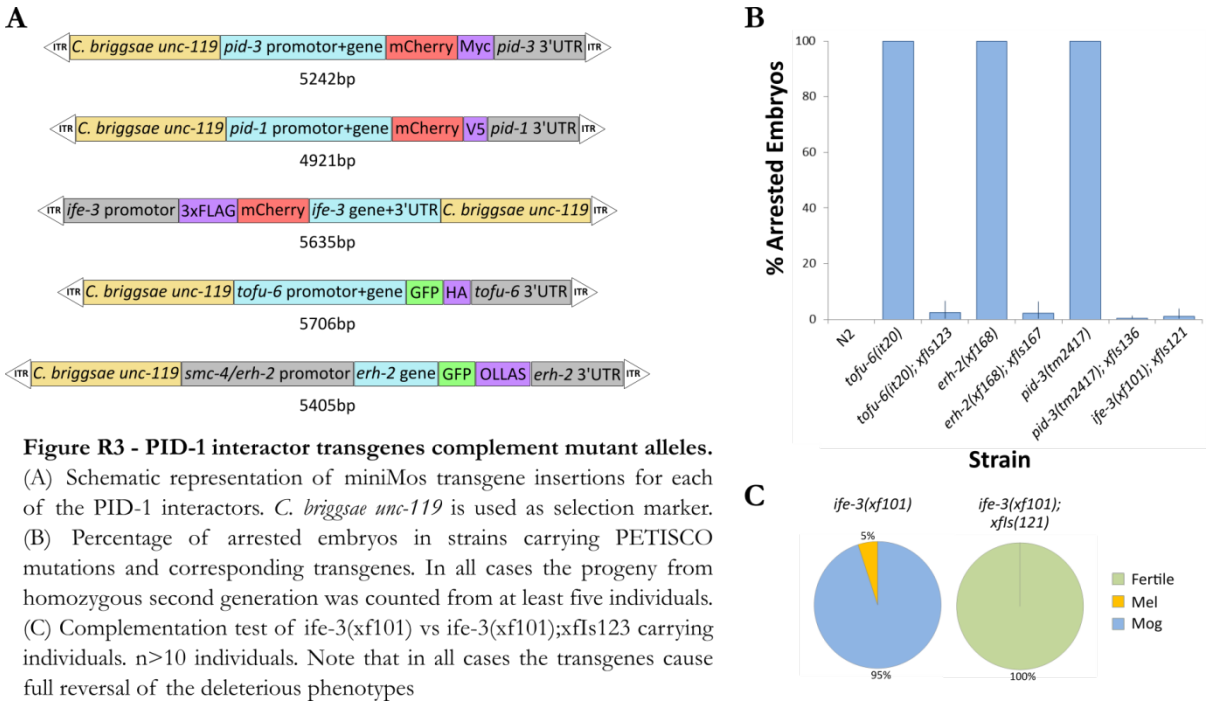

Figure R3 - PID-1 interactor transgenes complement mutant alleles. (A) Schematic representation of miniMos transgene insertions for each of the PID-1 interactors. *C. briggsae unc-119* is used as selection marker. (B) Percentage of arrested embryos in strains carrying PETISCO mutations and corresponding transgenes. In all cases the progeny from homozygous second generation was counted from at least five individuals. (C) Complementation test of ife-3(xf101) vs ife-3(xf101);xfIs123 carrying individuals. n>10 individuals. Note that in all cases the transgenes cause full reversal of the deleterious phenotypes

proteins was restricted to the germline in adult animals (Fig. R4). In addition, all proteins, except PID-1 were also found to be expressed in embryos.

3.3.1. PID-1 and IFE-3 localize to P granules

We find PID-1 localising to perinuclear granules in the distal adult gonad (Fig. R4C). This localization becomes diffused after the gonadal groove on the onset of oogenesis and expression is progressively lost through the oocyte maturation stages and is absent in early embryos (Fig. R4D). In contrast, IFE-3 protein is present throughout the adult germline and embryos, but still localising to perinuclear foci (Fig. R4). In embryos these foci are restricted to the posterior P cell (Fig. R4E and F), which gives rise germ cell lineage (Sulston and Horvitz, 1977; Sulston et al., 1983).

The *C. elegans* germline contains multiple types of perinuclear granules (see section 1.8), known to flank one another in a granular macrostructure named PZM (Wan et al., 2018). The distribution of the IFE-3 positive foci has a striking resemblance to the distribution of P granules, which are part of this macrostructure. We tested if the IFE-3 and PID-1 positive foci corresponded to P granules by crossing our transgenes with a P granule marker transgene: PGL-1::GFP (Cheeks et al., 2004). We found the majority of PID-1 and IFE-3 positive foci to localize with PGL-1::GFP positive granules or immediately around them (Fig. R4B).

We conclude from these observations that PID-1 resides at P granules at the distal adult gonad only, while IFE-3 resides at these sites throughout development. Still, we may not exclude a fraction of these proteins to reside to the remaining PZM as we do not find a complete overlap between PGL-1 and our proteins.

3.3.2. PID-3, ERH-2 and TOFU-6 follow the same expression pattern

We find the expression of PID-3, ERH-2 and TOFU-6 proteins to follow the same distribution pattern along the worm body: strong expression of these proteins in both the adult germline and embryos, including localization in perinuclear granules in the distal adult gonad (Fig. R4). As the distribution of these granules resembled the PID-1 positive granules, we further tested the localization of these granules in relation to PGL-1::GFP and each other.

We find that these proteins localize together at the same subcellular structure, the P granule (Fig. R4B). Like PID-1, we find that this co-localization is not a complete overlap and may reflect a partial overlap with other PZM granules. Nonetheless their co-localization further supports our IP-LFQP data and that these are functionally interacting proteins.

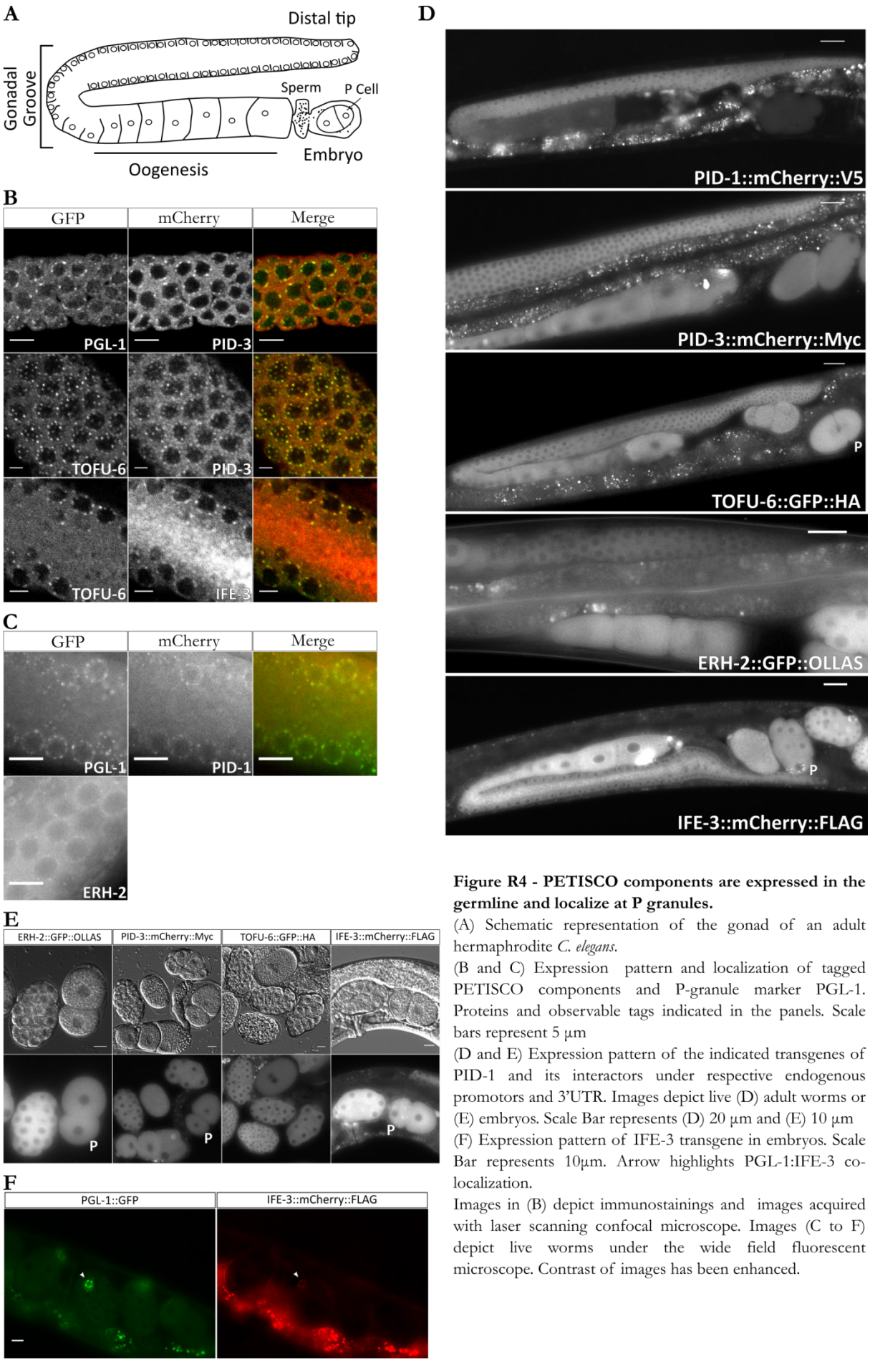

Figure R4 - PETISCO components are expressed in the germline and localize at P granules.

(A) Schematic representation of the gonad of an adult hermaphrodite *C. elegans*.

(B and C) Expression pattern and localization of tagged PETISCO components and P-granule marker PGL-1. Proteins and observable tags indicated in the panels. Scale bars represent 5 µm

(D and E) Expression pattern of the indicated transgenes of PID-1 and its interactors under respective endogenous promotors and 3'UTR. Images depict live (D) adult worms or (E) embryos. Scale Bar represents (D) 20 µm and (E) 10 µm

(F) Expression pattern of IFE-3 transgene in embryos. Scale Bar represents 10µm. Arrow highlights PGL-1:IFE-3 co-localization.

Images in (B) depict immunostainings and images acquired with laser scanning confocal microscope. Images (C to F) depict live worms under the wide field fluorescent microscope. Contrast of images has been enhanced.

3.4. PID-1 interactors form a protein complex named PETISCO

To understand the association between PID-1 and its interactors we performed IP-LFQP experiments with each of our proteins of interest (Fig. R5). As the access to antibodies targeting *C. elegans* proteins is rather limited, we used the transgenic strains rescuing endogenous null mutations (Fig. R3), mentioned in the section above, as baits for these experiments. Extracts of each of these strains were used in combination with mCherry or GFP-targeting nanobodies to perform IP-LFQP in quadruplicates. In each case the same quadruplicate experiment was performed using wild type animal extracts as control.

Strikingly, independently of which PID-1 interactor we used as bait to perform the IP-LFQP, we found PID-3, ERH-2, TOFU-6 and IFE-3 to be co-precipitating (Fig. R5A-E). The interaction between PID-1 and PID-3 was further confirmed with IP-WB (Fig. R5F). We tested if PID-1 was required for the co-precipitation of the remaining factors by repeating the PID-3 IP-LFQP experiment in *pid-1(xf35)* mutant extracts (Fig. R5G). We find that the PID-1 interactors are still co-precipitating, and thus conclude that these interactions are independent of PID-1.

Co-IP experiments are in general not suitable to claim whether interacting proteins form stable complexes, as co-precipitation can also result from rather transient interactions. However, the consistent significant enrichment of all these proteins in each of the others' IP-LFQP experiments does suggest that these proteins form a concrete multi-protein complex. We explored the size of this complex by performing size exclusion chromatography of total worm extracts and probing for PID-1, ERH-2 and TOFU-6 (Fig. R5H). The elution profiles of these three proteins resemble each other closely, consistent with the idea that they form a stable complex. The approximate molecular weight at which they elute is approximately 400kDa. As the molecular weights of the potential interactors add up to 117kDa (152kDa with added epitopes), these results also suggest that this complex exists at least as a dimer. Interestingly, ERH proteins from human and *S. pombe* have indeed been shown to occur as dimers. These data strongly support that the identified PID-1 interactors form novel protein complex and we have named it PID-3, ERH-2, TOFU-6, IFE-3 small RNA Complex or PETISCO.

We note that PID-1 is not enriched in all of these IPs (Fig. R5C and D). Although we find this protein to have significant enrichments in IFE-3 and PID-3 IPs, when using ERH-2 and TOFU-6 as bait we find that PID-1 remains below our pre-determined significance threshold. To note PID-1 remains close to the border of significance in the TOFU-6 IP-LFQP experiment (*P* value=0.06). Still, we find that PID-1 and ERH-2 to be direct interactors (see below) and believe this PID-1 absence to be a false negative result. This outcome may be an artefact caused by the usage of the same IP affinity tag: GFP.

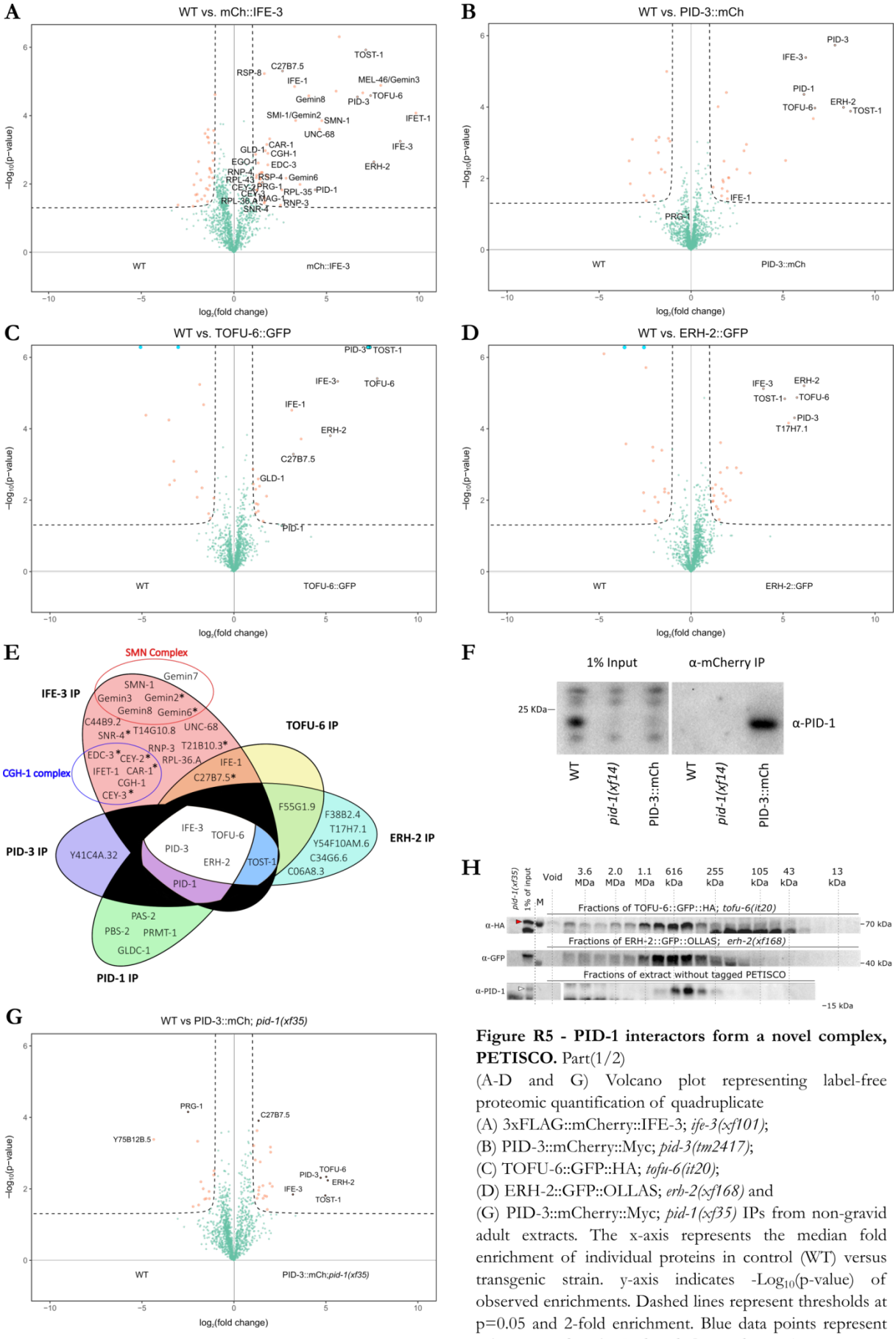

Figure R5 - PID-1 interactors form a novel complex, PETISCO. Part(1/2)

(A-D and G) Volcano plot representing label-free proteomic quantification of quadruplicate

(A) 3xFLAG::mCherry::IFE-3; *ife-3(xf101)*;

(B) PID-3::mCherry::Myc; *pid-3(tm2417)*;

(C) TOFU-6::GFP::HA; *tofu-6(it20)*;

(D) ERH-2::GFP::OLLAS; *erh-2(xf168)* and

(G) PID-3::mCherry::Myc; *pid-1(xf35)* IPs from non-gravid adult extracts. The x-axis represents the median fold enrichment of individual proteins in control (WT) versus transgenic strain. y-axis indicates -Log_{10}(p-value) of observed enrichments. Dashed lines represent thresholds at p=0.05 and 2-fold enrichment. Blue data points represent values out of scale. Red and Green data points represent above and below threshold respectively. End of part (1/2)

Interestingly, we find all elements of PETISCO to interact with the protein C35D10.13, which we have named TOST-1 (see below) (Fig. R5A-F). This is a 16.2kDa protein of unknown function which we do not find in our PID-1 IP-LFQP data. This may be an additional PETISCO subunit we previously failed to detect, or a PETISCO interactor outside of the PID-1 context. Therefore we included this protein in our downstream analysis.

3.4.1. *IFE-3 is a partner of additional protein complexes*

In all of the performed IP-LFQP datasets we find IFE-3 to have the largest set of interactors (Fig. R5A and F). In addition to PETISCO components, IFE-3 interacts with multiple gemins and core components of the SMN complex. The SMN complex is a complex required for the assembly of snRNPs (Battle et al., 2006a), possibly indicating a role of IFE-3 in this pathway.

As IFE-3 is an eIF4E orthologue one would expect to find eIF4G within its interaction network (Peter et al., 2015), which is not the case. Possibly, IFE-3 may not be performing translation initiation functions as its paralogues. Most translated transcripts in *C. elegans* are trans-spliced and carry a 2,2,7-trimethylguanylate (TMG) Cap that is not recognized by IFE-3 (Keiper et al., 2000; Miyoshi et al., 2002; Nilsen, 1993), hence it may be that IFE-3 has acquired a non-translation related function in *C. elegans*. Interestingly, we do detect IFET-1, and to a lesser extent the CGH-1 complex, as a close interactors of IFE-3. IFET-1 is the *C. elegans* orthologue of eIF4E nuclear import factor 1 (EIF4ENIF1), a protein that negatively regulates translation by competing with eIF4E for 5' end Cap binding (Andrei et al., 2005). Additionally this protein is known to partner with the CGH-1 complex, in order to protect and inhibit the translation of maternally inherited mRNAs (Boag et al., 2008). Conceivably, IFE-3 may be part of a machinery that negatively regulates the translation of m7G capped transcripts.

3.5. PETISCO is a stable protein complex in adult worms

As we have determined the components of PETISCO, we aimed at drawing a better picture of this complex by examining the molecular binding partners of each of its elements. In our first approach we attempted to solve the binding interactions of PETISCO using IP-LFQP. In this experiment, the same TOFU-6 IP-LFQP experiments were performed, but now the washing steps of the IPs were executed under high salt concentration ([NaCl] 150mM increased

Figure R5 - PID-1 interactors form a novel complex, PETISCO. Part(2/2)
(E) Venn Diagram summarizing significant interactions in PETISCO protein IPs. *represents protein found significantly enriched in only one experiment. For convenience the CGH-1 and the SMN complex components are highlighted. (F) mCherry pull-down of wild type (WT), *pid-1(xf14)* mutant and PID-3::mCherry::Myc; *pid-3(tm2417)* (PID-3::mCh) carrying worms. IPs were performed in non-gravid adult extracts. Membrane was probed for endogenous PID-1. Background recognition by the anti-PID-1 antibody is used as loading control. (H) Size-exclusion chromatography of non-gravid adult worm extracts. See methods for more details on strains. Fractions were collected and probed for HA, GFP or PID-1. Approximate molecular weight of fractions indicated in figures. Red arrow indicates full length TOFU-6::GFP and white arrow PID-1. *pid-1(xf35)* extract is used as probing control.

up to 500mM). Under these conditions we expected that only direct binding partners or interactions separated by a single degree would remain enriched in the IP-LFQP assay readout. Strikingly, we found that the PETISCO protein enrichments remained intact under these conditions, showing that the interactions within PETISCO are very robust or at least not salt-sensitive (Fig. R6A).

PETISCO proteins are linked to RNA biology, hence we investigated the participation of RNA in the establishment PETISCO protein interactions. We tested this by exposing worm extracts to a 30 minute RNaseA/T1 treatment followed by the described standard IP-LFQP experiment (Fig. R6B and C). IP-LFQP of either PID-3 or IFE-3 under these conditions revealed little change in interacting partners (Fig. R6B and C). Only, the loss of PID-1 as an interacting partner of IFE-3 was observed (Fig. R6B). These data suggest that PETISCO likely exists as a stable protein complex in adult worms. Furthermore, this complex is maintained via robust protein-protein binding and these interactions are unlikely to be mediated by RNA molecules. Yet, we cannot exclude that an RNA bound may be shielded from RNase activity.

3.6. The architecture of PETISCO subunit interactions

The robustness of PETISCO interactions consequently made it difficult to determine within the complex which proteins are binding to one another. As *in vivo* these proteins are systematically co-precipitating, we turned to an external system to determine these interactions: yeast two-hybrid (Y2H). In this system we expose proteins to each other in pairs and score for protein-protein interaction according to the proliferation of the host *Saccharomyces cerevisiae* auxotrophic strain. The transfected proteins are recombinantly fused to the Gal4 transcription factor binding or activation domain and positive interactions lead to the expression of genes which compensate for auxotrophy of histidine and adenine. We made a Y2H array and exposed it to two different selective media to test both interaction pairs and corresponding binding strength. This assay allowed us to map interaction pairs (Fig. R7).

Under standard stringency plates (TRP^- LEU^- HIS^-) we find that IFE-3 binds directly to TOFU-6 (Fig. R7A). TOFU-6 additionally binds PID-3 which in turn is a binder of ERH-2 (Fig. R7A). Remarkably, despite its modest size, ERH-2 is the component of PETISCO with the largest array of interactions: it is not only able to bind PID-3 but also PID-1 and TOST-1 (Fig. R7A). Furthermore, under the same conditions, we find evidence for dimerization of ERH-2 (Fig. R7A), although this interaction is ablated under higher stringency conditions (TRP^- LEU^- HIS^- ADE^-). Moreover, we find the interactions in PETISCO to remain unaltered under high stringency conditions, with the exception of the interaction between TOFU-6 and IFE-3, which according to this assay, stands as the weakest link within PETISCO (Fig. R7A).

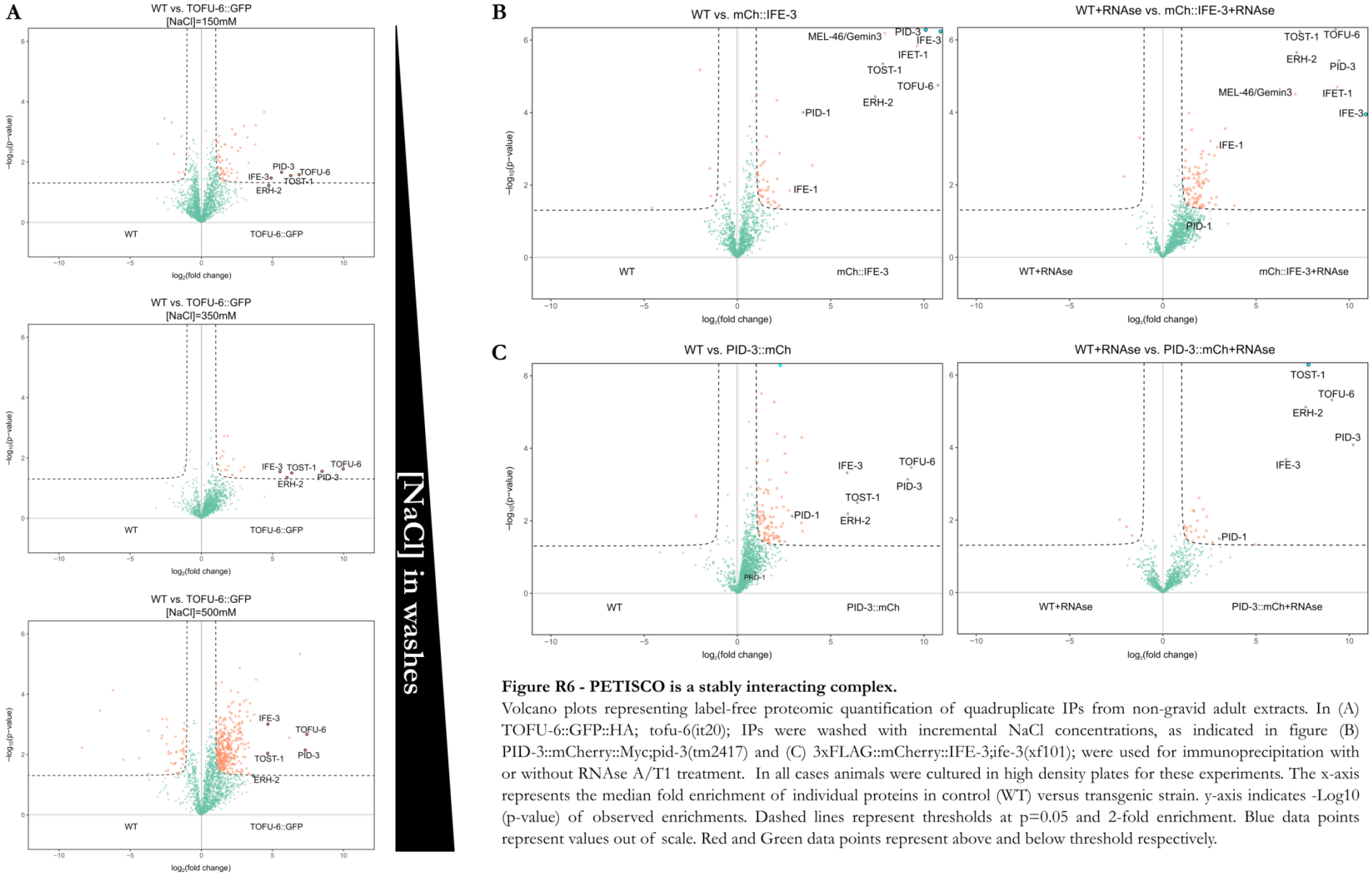

Figure R6 - PETISCO is a stably interacting complex.

Volcano plots representing label-free proteomic quantification of quadruplicate IPs from non-gravid adult extracts. In (A) TOFU-6::GFP::HA; tofu-6(it20); IPs were washed with incremental NaCl concentrations, as indicated in figure (B) PID-3::mCherry::Myc;pid-3(tm2417) and (C) 3xFLAG::mCherry::IFE-3;ife-3(xf101); were used for immunoprecipitation with or without RNAse A/T1 treatment. In all cases animals were cultured in high density plates for these experiments. The x-axis represents the median fold enrichment of individual proteins in control (WT) versus transgenic strain. y-axis indicates -Log10 (p-value) of observed enrichments. Dashed lines represent thresholds at p=0.05 and 2-fold enrichment. Blue data points represent values out of scale. Red and Green data points represent above and below threshold respectively.

In an effort to increase our understanding of PETISCO we increased the resolution in a second Y2H assay. We recapitulated each individual interaction found in the original Y2H array using single domain constructs of each of the PETISCO proteins (Fig. R7B-D).

3.6.1. TOFU-6 and IFE-3 interact via an eIF4E binding motif

As mentioned, we found IFE-3 and TOFU-6 to be interacting partners. While IFE-3 is constituted by a single domain specific to the eIF4E family, we were able to identify two different protein domains (RRM and TDR) and a C-terminal putative amino acid motif (HxxYxRxxLL) for eIF4E (IFE) binding (Gosselin et al., 2013) in TOFU-6 (Fig. R2A). As we recapitulate the original TOFU-6:IFE-3 interaction we find that this binding is perpetuated by the putative IFE binding motif (IBM) of TOFU-6 in both standard and high stringency conditions (Fig. R7B). Interestingly, we note that this motif alone shows a higher affinity to IFE-3 then within the full TOFU-6 context, where high stringency ablates this interaction (Fig. R7B). This could imply that the IBM may be modulated by other TOFU-6 domains. Alternatively, it could be an artefact of the Y2H system. For instance, TOFU-6 may not fold well inside the yeast nucleus, or expression levels of the short IBM construct may be significantly higher than that of full length TOFU-6.

3.6.2. PID-3 interaction with TOFU-6 and ERH-2 is mediated by its RRM domain

The TOFU-6:PID-3 interaction is supported by their corresponding RRM domains (Fig. R7C). The expression of an individual RRM domain of TOFU-6 or PID-3 is sufficient to establish binding with the full length PID-3 or TOFU-6, respectively. Likewise, the RRM domains of TOFU-6 and PID-3 are capable of establishing an interaction between them (Fig. R7C). These results are reproducible in both standard and high stringency conditions. Furthermore, the function of the PID-3 RRM domain is not restricted to its interaction with TOFU-6. We find that this domain also mediates the interaction between PID-3 and ERH-2 (Fig. R7D), a robust interaction that remains stable under high stringency conditions.

TOFU-6 and ERH-2 both interact with the PID-3 RRM domain (Fig. R7C and D). One could therefore consider that these interactions may be mutually exclusive. This could be addressed in a Y2H context by expressing free nuclear PID-3 RRM in a manner that would bridge the two Gal4 domains bound to ERH-2 and TOFU-6. In a simultaneous bound the Y2H strain should be able to grow in selective media. However, the fact that each of these proteins is enriched in the IP-LFQP assay of one another, strongly suggests that these proteins coexist in the same complex. Consequently, TOFU-6 and ERH-2 are likely to interact with different sections of the PID-3 RRM domain.

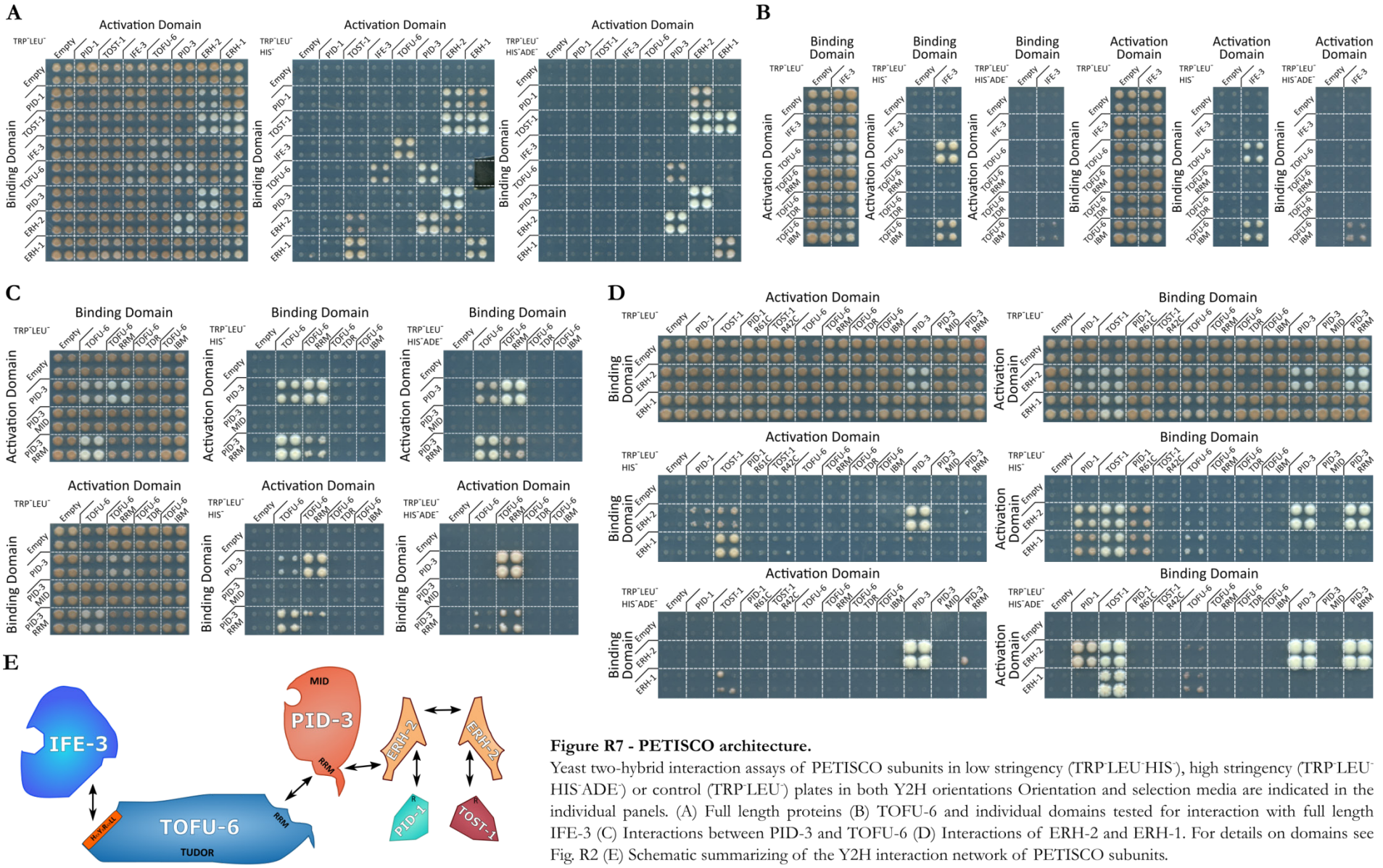

Figure R7 - PETISCO architecture.

Yeast two-hybrid interaction assays of PETISCO subunits in low stringency (TRP⁻LEU⁻HIS⁻), high stringency (TRP⁻LEU⁻HIS⁻ADE⁻) or control (TRP⁻LEU⁻) plates in both Y2H orientations Orientation and selection media are indicated in the individual panels. (A) Full length proteins (B) TOFU-6 and individual domains tested for interaction with full length IFE-3 (C) Interactions between PID-3 and TOFU-6 (D) Interactions of ERH-2 and ERH-1. For details on domains see Fig. R2 (E) Schematic summarizing of the Y2H interaction network of PETISCO subunits.

As of yet, no interaction has been detected for the Tudor domain of TOFU-6. Tudor domains often bind symmetrically dimethylated arginines (sDMAs). This modification may be absent from the TOFU-6 target for lack of the proper PTM signal in *S. cerevisiae,* and thus this interaction remained undetected in our Y2H assay. Possibly, TOFU-6 may yet interact with a different PETISCO protein or this domain may serve as a docking site for another protein, thus far not identified *in vivo*. Nonetheless, altogether these data draw a clear picture of the architecture of PETISCO and further support its existence as a robust protein complex.

3.7. PETISCO is required for 21U RNA induced silencing

PETISCO is a robust protein complex that is the main binding partner of PID-1, a protein essential for 21U RNA biogenesis. Thus, we speculated that this complex plays a role in this pathway.

In order to address this, we performed RNAi depletion of each PETISCO subunit in a *C. elegans* strain carrying a transgenic 21U RNA sensor (21U sensor) that is specifically recognised and silenced by the 21U RNA pathway (Bagijn et al., 2012). Individuals with a defective 21U RNA pathway are unable to silence the 21U sensor, leading to the expression of an H2B::GFP fusion protein that accumulates in the germline nuclei (Fig. R8A). Transgenes targeted by the 21U RNA pathway tendentially undergo RNAe (see sections 1.7.3 and 1.7.4), rendering them

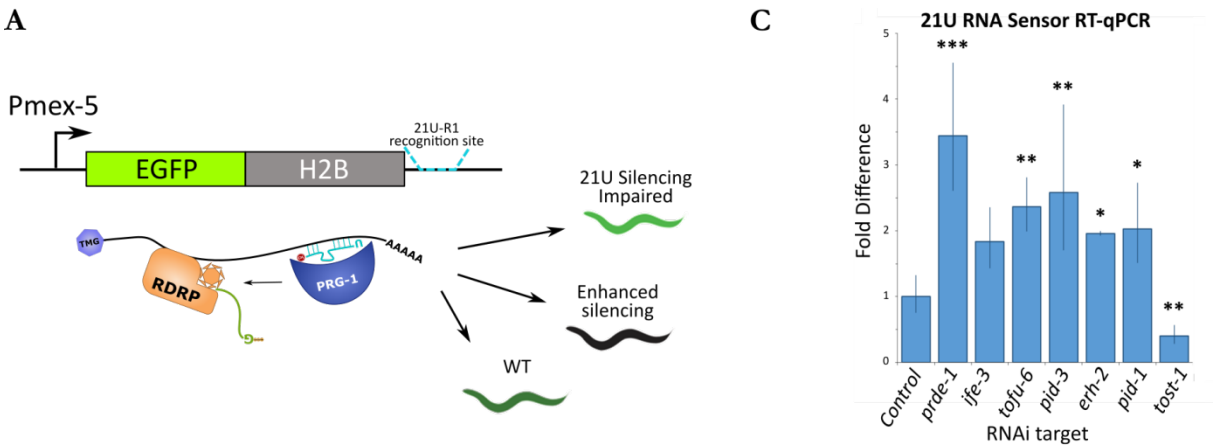

Figure R8 - PETISCO is required for 21U RNA induced silencing.
(A) Schematic representation of the 21U RNA sensor transgene activity in worms with a sensitized background. (B) Wide field fluorescent microscopy of adult hermaphrodites carrying 21U sensor. Worms were subjected to RNAi via feeding (targets indicated in figure) from L1 larval stage to adulthood. Empty RNAi Vector serves as negative control. Scale Bar represents 10 µm. (C) Quantitative RT-PCR of 21U sensor transgene in adult worm populations of (B). Values are obtained from experimental triplicates and normalized to pmp-3 mRNA levels. Significance of sample vs. control was tested with Dunnett's post hoc test: ***p-value<0.001; **p-value<0.01; *p-value<0.05. Error bars represent standard deviation.

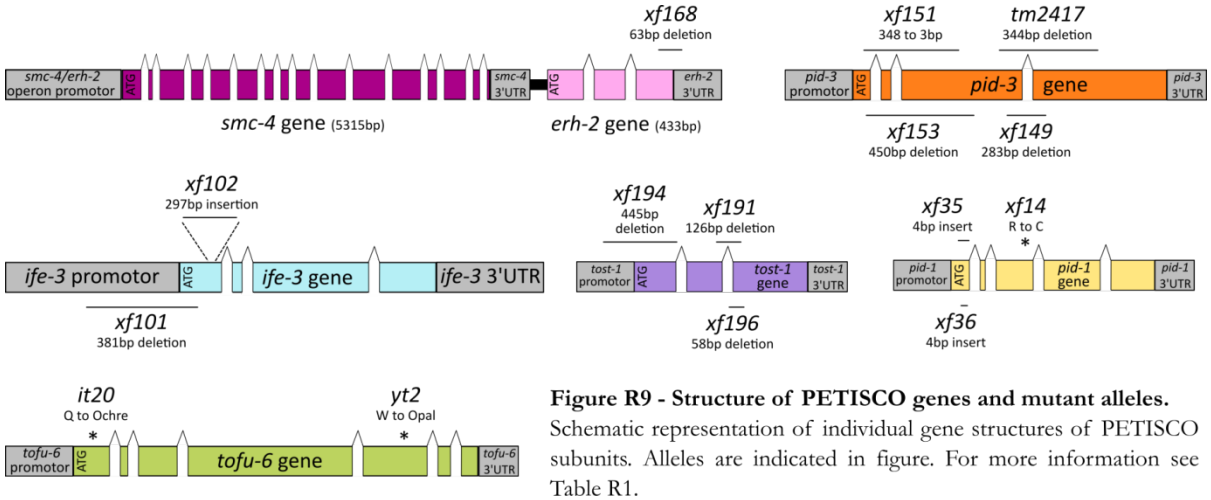

Figure R9 - Structure of PETISCO genes and mutant alleles. Schematic representation of individual gene structures of PETISCO subunits. Alleles are indicated in figure. For more information see Table R1.

insensitive to loss of 21U RNA activity. To prevent this effect, we created a strain carrying the 21U sensor combined in a *pid-1(xf14)* loss of function background, carried by a PID-1::mCherry*(xfIs117)* transgene. We found that this strain is unable to fully compensate the defective endogenous locus, maintaining the sensor in a partially silenced non-RNAe state. In addition, this also allows us to probe for enhanced 21U RNA function, as it leads to a fully silenced 21U sensor. Animals of this sensitized strain were exposed to RNAi, via feeding, against *pid-1*, *tost-1,* and each of the PETISCO genes. As positive and negative controls RNAi against *prde-1* and an empty RNAi vector were used respectively. Effects on the 21U sensor were scored with both microscopy and RT-qPCR (Fig. R8B-C).

Gene	Allele	Mutation Type	21U RNA Presence	Terminal Phenotypes	Ratio (%)
ife-3	*xf101*	Start Codon Loss	n.d.	Mog/Mel	95/5
	xf102	Frameshift	-	Mog/Mel	n.d.
	RNAi	n.a.	n.d.	Mog/Mel	47/53#
pid-3	*tm2417*	Frameshift	--	Mel	100
	xf149	Frameshift	n.d.	Mel	100
	xf151	Inframe Deletion	n.d.	Mel	100
	xf153	Frameshift	n.d.	Mel	100
tofu-6	*it20*	Nonsense	n.d.	Mel	100
	yt2	Nonsense	n.d.	Mel	100
	RNAi	n.a.	--*	Mel	n.d.
tost-1	*xf191*	Frameshift	n.d.	Viable	n.d.
	xf194	Start Codon Loss	+	Mel	100
	xf196	Splice Site Loss	n.d.	(TS) Mel	100 (25ºC)
pid-1	*xf14*	Missense	--	n.d.	n.d.
	xf35	Frameshift	--	Mog	<1
	xf36	Frameshift	--	n.d.	n.d.
erh-2	*xf168*	Stop Codon Loss	--	Mel	100

Table R1 - PETISCO subunit mutants display maternal effect lethality.
Summary of phenotypes observed in the various mutants described and used in this study. "n.a." not applicable; "n.d." not determined; "-" mild 21U RNA defect; "--" severe 21U RNA defect; "+" no 21U RNA defect; "(TS)" temperature sensitive; "#" counts are for gonadal arms (n=38) due to mixed phenotypes in individuals; "*" according to Goh et al. (2014)

PID-3, ERH-2, IFE-3 and TOFU-6 knock down all activated the 21U sensor transgene (Fig. R8B). Similar to the effect of *pid-1* knock down, the depletion of the PETISCO components led to a 2-fold increase in the expression of the 21U sensor, while the depletion of *prde-1*, required for 21U RNA precursor expression, led to a 3.5-fold expression increase compared to the empty vector (Fig. R8C). This data shows that PETISCO, like PID-1, is required for 21U RNA induced silencing.

Interestingly, we find that *tost-1* knock down led to the opposite effect on the 21U sensor. Depletion of TOST-1 led

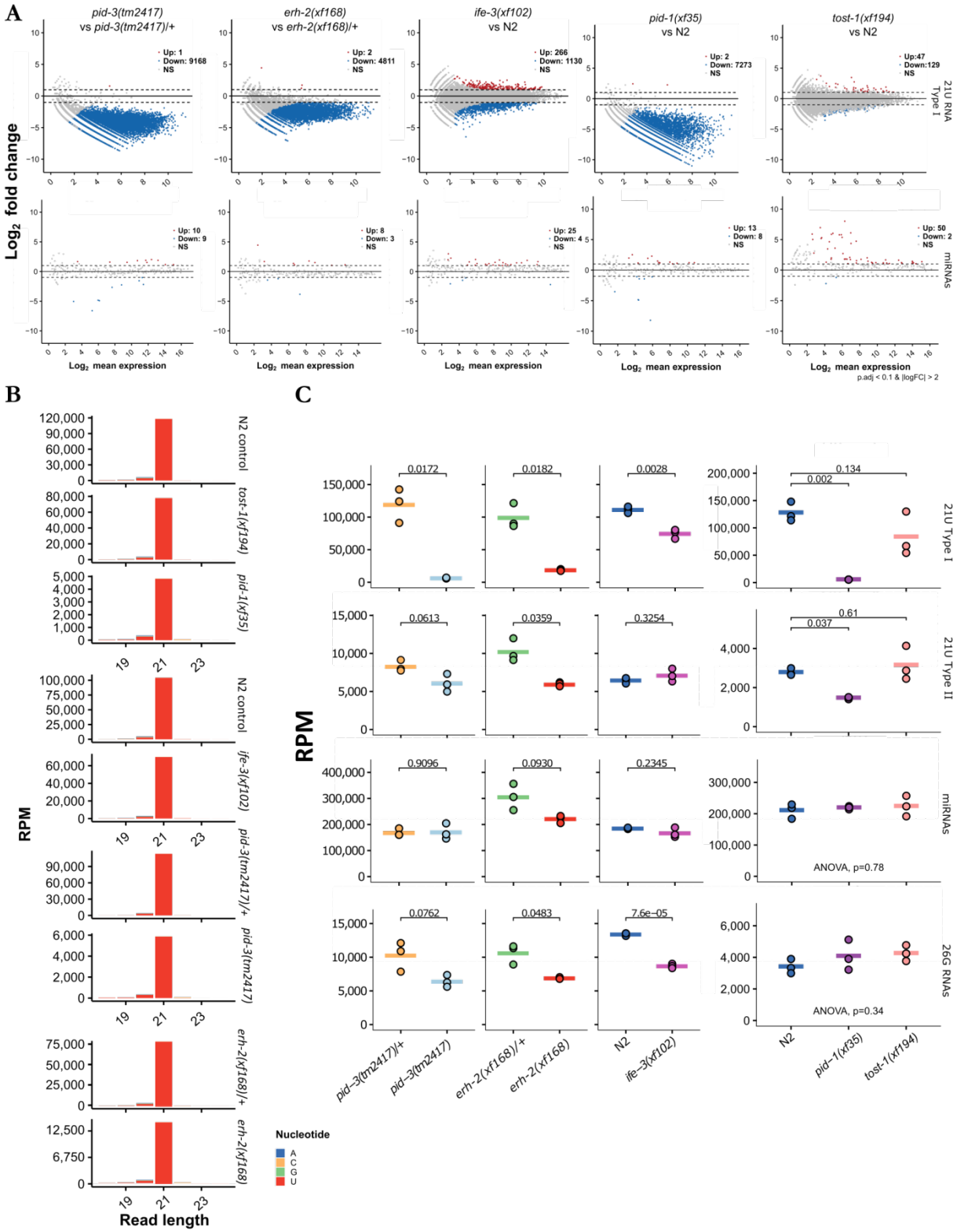

Figure R10 - Small RNA sequencing of PETISCO mutants.
pid-3(tm2417), erh-2(xf168) and *ife-3(xf102)* samples and controls consist of non-gravid adults, while *pid-1(xf35)* and *tost-1(xf194)* samples and control are gravid adult individuals. (A) Pairwise differential gene expression of type I 21U RNAs and miRNAs of all sequenced mutants vs. respective control. Mutants are indicated in the panels. Red and blue dots indicate upregulated and downregulated transcripts, respectively. (B) Length distribution and nucleotide bias of mature type I 21U RNAs that remain in PETISCO mutants. Graphs represent the average of the three replicates of (A). Nucleotide bias is represented as colour code. (C) Global levels of 21U RNAs type I, type II, miRNAs and 26G RNAs of samples of (A). Values are in reads per million mapped reads (RPM). Each dot represents a replicate of three and horizontal bar represents the total mean. Significance tested either with Student's t-test or Dunnett's post hoc test when appropriate. Respective mutant, small RNA population and statistical p-values are indicated in the figure.

to the decrease of the 21U sensor expression to 0.4-fold of empty vector treated animals. This result suggests that this protein is acting as an antagonist to PID-1 and PETISCO activity in the 21U RNA induced silencing, hence we named it TOST-1, for Twenty One U Antagonist 1.

3.8. Mutations in PETISCO lead to 21U RNA absence

We further pursued the effects of PETISCO on the 21U RNA pathway by performing small RNAseq on null mutants for the PETISCO genes. As not all of these were available in public databases we generated mutants for each of these using the CRISPR/Cas9 system (Fig. R9 and Table R1). *tofu-6* has previously been shown to be required for 21U RNA biogenesis (Goh et al., 2014), therefore we focused our efforts on the remaining elements. After confirmation of the mutant alleles, each sequencing experiment was performed in triplicate and normalized to total library depth, together with its respective control strain.

In synchrony with the reactivation of the 21U sensor, we found *ife-3(xf102)*, *erh-2(xf168)* and *pid-3(tm2417)* to have a significant decrease (Student's t-test, *P* value ≤0.05) in their total 21U RNA population (Fig. R10A and C). As in *pid-1(xf35)* mutants, the few remaining 21U RNAs are indistinguishable in length and 5' end U bias (Fig. R10B). On the other hand, these mutant strains have only modest to no alterations in their miRNA or type II 21U RNA populations (Fig. R10A and C), excluding a small RNA-wide effect. We note that the severity of 21U RNA depletion is

substantially higher in *erh-2(xf168)* and *pid-3(tm2417)* than in *ife-3(xf102)*. This discrepancy may be a reflection of the function of IFE-3 in the complex, which may be redundant with one of the other orthologues of eIF4E. Indeed, we find that IFE-1 is enriched in some of our IP-LFQP assays (Fig. R5E). These data show that PETISCO is required for the presence of 21U RNAs.

TOST-1 depletion enhanced the silencing of the 21U sensor (Fig. R8). Yet, we found a mild, bellow significance drop in the total Type I 21U RNA levels of *tost-1(xf194)* versus N2 worms (Fig. R10C). Thus, the enhanced silencing of the sensor might be a secondary effect of TOST-1 absence. Nevertheless, this data shows a contrast between TOST-1 activity and that of PETISCO and PID-1 in the 21U RNA pathway. Conversely, we found some miRNAs to be upregulated in *tost-1(xf194)* individuals (Fig. R10A). This indicates that TOST-1 may selectively regulate some miRNAs or this may be a secondary *tost-1(xf194)* effect.

We found that the less abundant 26G RNA species have their total count modestly reduced upon deletion of *ife-3*, *pid-3* or *erh-2* (Fig. R10C). These findings indicate a possible connection of PETISCO to general 26G RNA biogenesis or specifically one of its subpopulations (see section 1.7.2.). Yet, we found no clear differences in the known subpopulations of these small RNAs (data not shown). Given that the effect on 26G RNAs is

much weaker than that on 21U RNAs, we consider it likely that these observations result from indirect effects.

3.9. PETISCO is required for embryonic development

When mutating PETISCO genes, we were surprised to find that mutant individuals registered severe deleterious effects, unlike other known 21U RNA mutants. We observed two phenotypes for homozygous *ife-3(xf102)* individuals: Masculinization Of Germline (Mog) and Maternal Effect Lethal (Mel). In the former phenotype, *C. elegans* hermaphrodites develop only sperm and are sterile, whereas in the latter, mutant adults are fertile but their progeny suffer from developmental arrests and do not hatch. Under standard culture conditions, we found Mog (95%) to be more common than Mel (5%) in *ife-3* individuals (Fig. R3C). In agreement with our findings, these phenotypes have been described in previous work on *ife-3* (Mangio et al., 2015), although a conclusive explanation for these effects remained elusive. It is noteworthy that, although we do not find Mel *pid-1(xf35)* mutants, we do find at low frequency (<1%) *pid-1(xf35)* Mog individuals (Fig. R11A and Table R1), which may be related to the *ife-3(xf102)* phenotype.

In contrast with *ife-3*, *pid-3(tm2417)* and *erh-2(xf168)* are not Mog, but are exclusively Mel. 100% of the descendants of *pid-3(tm2417)* and *erh-2(xf168)* hermaphrodites arrest at approximately 100 cell stage, before gastrulation (Fig. R11B). *pid-3(tm2417)* and *erh-2(xf168)* strains are capable of generating fertile male individuals, but the embryonic arrest cannot be complemented by cross-breeding with a WT paternal line (data not shown). Thus, these are

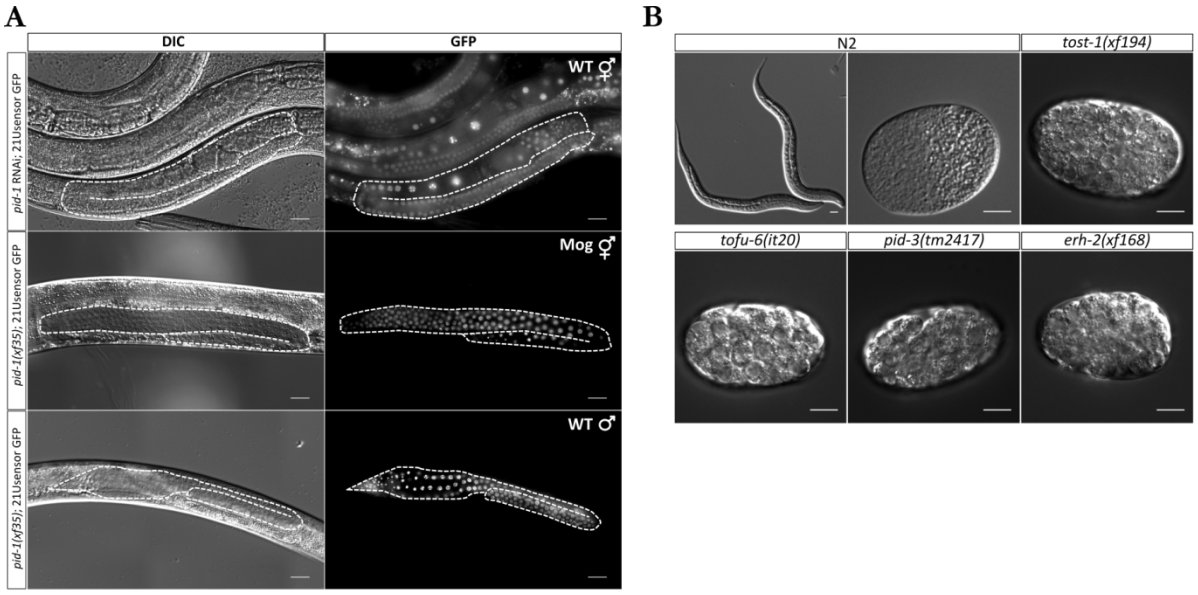

Figure R11 - Terminal phenotypes of PETISCO mutants.
(A) Fluorescent microscopy of the germline of 21U sensor GFP of *pid-1(xf35)*; 21U sensor worms. *pid-1(xf35)* hermaphrodites display a low frequency (<1%) Mog phenotype (second row). Bottom row a male germline of the same strain. Top row a hermaphrodite germline of a pid-1 knock-down. Scale Bars represent 20 μm. Contrast of images has been enhanced.
(B) Embryos laid by wild-type (N2) and PETISCO homozygous mutant mothers. Images are taken one day after egg lay. Mutant embryos arrest between 64-128 cell stage, while WT embryos are hatched (left N2 image). Occasional non-hatching wild-type embryos (right N2 image) do not show the characteristic stage of the PETISCO mutants. Scale bar represents 10 μm.

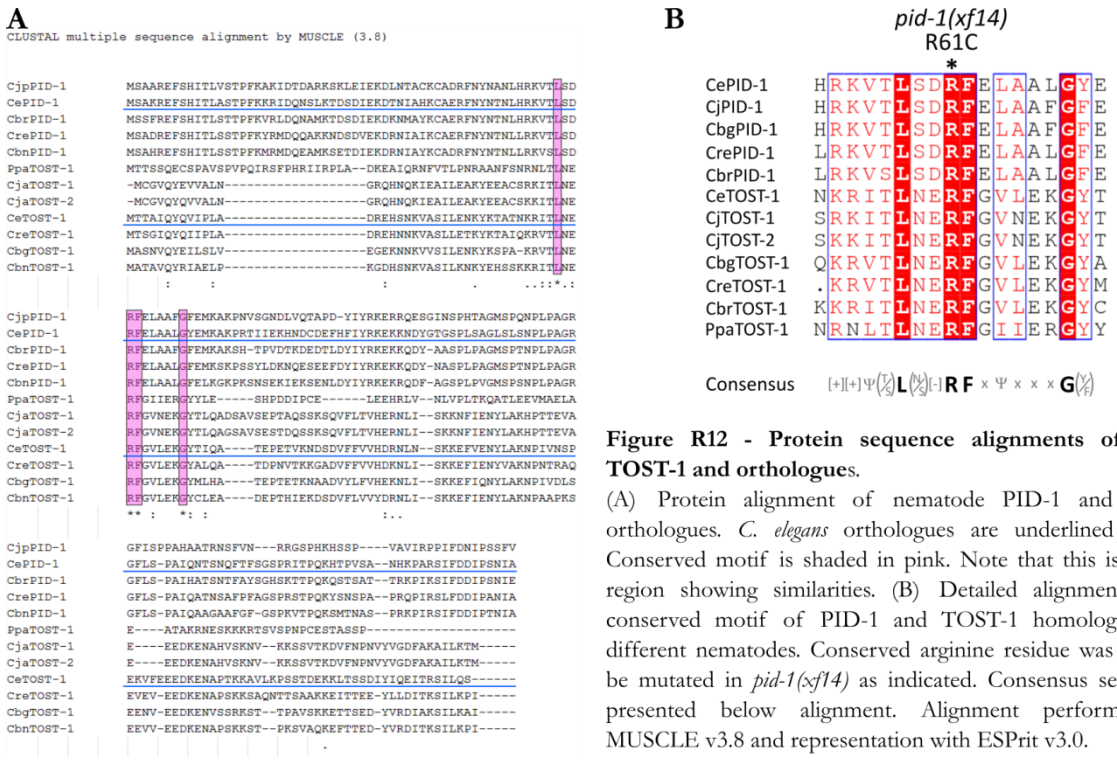

Figure R12 - Protein sequence alignments of PID-1, TOST-1 and orthologues.
(A) Protein alignment of nematode PID-1 and TOST-1 orthologues. *C. elegans* orthologues are underlined in blue. Conserved motif is shaded in pink. Note that this is the only region showing similarities. (B) Detailed alignment of the conserved motif of PID-1 and TOST-1 homologues from different nematodes. Conserved arginine residue was found to be mutated in *pid-1(xf14)* as indicated. Consensus sequence is presented below alignment. Alignment performed with MUSCLE v3.8 and representation with ESPrit v3.0.

neither sperm nor gene dosage defects, but a fully penetrant Mel phenotype. This observation is in accordance with the phenotype already described for *tofu-6* mutants (Minasaki and Streit, 2007), a gene originally named *mel-47*.

Interestingly, we found *tost-1(xf194)* hermaphrodites to also be 100% Mel and their descendants to arrest at the same stage as *pid-3(tm2417)* and *erh-2(xf168)* (Fig. R11B and Table R1). This is a strong indicator that TOST-1 shares an essential function with PETISCO, for which PID-1 is not required.

3.10. PID-1 and TOST-1 define distinct functions of PETISCO

Mutants for the 21U RNA pathway, such as *prg-1* and *pid-1*, have not been associated with severe viability phenotypes, although, they have been acknowledged to have a transgenerational sterility phenotype (Simon et al., 2014). This phenomenon known as mortal germline (Mrt) and is a phenotype where in each generation animals progressively lose their fertility to the point of sterility. The Mel phenotype caused by PETISCO ablation is a single generation fully penetrant phenotype, contrasting with the previously described phenotypes for the 21U RNA pathway. PID-1 interacts with PETISCO but its ablation does not lead to a Mel phenotype. On the other hand, the PETISCO interactor TOST-1 is not required for the 21U RNA pathway, but *tost-1(xf194)* mutants are Mel. We find that both of these proteins are directly binding ERH-2 (Fig. R7D). Additionally, TOST-1 is systematically enriched in PETISCO protein IP-LFQP assays (Fig. R5) while we do not find it when using PID-1 as bait (Fig. R1). These data lead us to speculate that TOST-1 and PID-1 may be interacting with ERH-2 in a mutually exclusive

manner, defining two versions of PETISCO. We probed into how these interactions may be established by looking at the biochemical nature of TOST-1 and PID-1.

3.10.1.A shared amino acid motif mediates ERH-2 binding by PID-1 and TOST-1

In comparing TOST-1 and PID-1 we find that, either than size and the lack of predictable protein domains, these proteins share little resemblance, and are unlikely to be evolutionarily related (Fig. R12A). However, by complementing this comparison with other PID-1 and TOST-1 orthologues we find a shared and conserved amino acid motif: $_{[+][+]}\Psi_{(T/S)}\mathbf{L}_{(N/S)[-]}\mathbf{RF}_{x}\Psi_{xxx}\mathbf{G}_{(Y/F)}$ (Fig. R12B). Remarkably, the conserved arginine contained within this consensus motif was previously identified to be essential for PID-1 activity: the first *pid-1* allele to be identified, *pid-1(xf14)*, was an R61C missense mutation that led to complete the loss of PID-1 protein (de Albuquerque et al., 2014).

We asked if the mentioned arginine could be responsible for the PID-1 interaction with PETISCO via ERH-2. To address this, we performed a Y2H assay in which we recapitulated the *pid-1(xf14)* allele and investigated how it interfered with its ability to bind ERH-2 (Fig. R7D). For comparison, we included in this assay the ERH-1, the second enhancer of rudimentary orthologue of *C. elegans*. We find that PID-1(R61C) significantly loses its ability to interact with ERH-2, and is only able to grow colonies under standard stringency (Fig. R7D). This result shows that this amino acid plays a key role in the interaction between PID-1 and ERH-2.

We further tested if this motif is responsible for the TOST-1:ERH-2 protein interaction and recreated the same Y2H experiment with TOST-1 and TOST-1(R42C). The latter is a TOST-1 variant carrying a mutation analogous to that of PID-1(R61C) (Fig. R2A). Interestingly, this missense mutation has a severe effect on the interaction between TOST-1 and ERH-2, and no interaction was observed in either of the tested stringency conditions (Fig. R7D). Thus, we conclude that PID-1 and TOST-1 interact with PETISCO via ERH-2, through a conserved amino acid motif. Surprisingly, we obtained the same results for interactions between PID-1 and TOST-1 variants and ERH-1 (Fig. R7D). This result indicates that this amino acid motif may be a common motif for ERH protein binding.

3.10.2.PID-1 and TOST-1 bind PETISCO and drive its function independently

The use of the same amino acid motif by PID-1 and TOST-1 suggests that these share the same binding pocket of ERH-2. Previously, we mentioned the absence of TOST-1 in our PID-1 IP-LFQP data (Fig. R1), supporting a model where these proteins do not co-exist within PETISCO. Moreover, PID-1 is required for 21U RNA biogenesis and TOST-1 for embryonic viability, while PETISCO subunits are required for both these functions. Hence, we conclude

that PID-1 and TOST-1 bind PETISCO via ERH-2 and guide it through its two different functions: 21U RNA biogenesis and embryogenesis.

3.11. TOST-1 is required in late oogenesis and early embryogenesis

Within the pool of generated *tost-1* we found that *tost-1(xf194)* does not have a significant impact in 21U RNAs displaying a fully penetrant Mel phenotype. Interestingly, we found a second allele, *xf196*, showing a different outcome. *tost-1(xf196)* harbours a 58bp deletion which includes the splice acceptor site of the third exon of *tost-1* (Fig. R9 and Table R1). This results in either a frameshift mutation or a truncation of the second half of this protein. In contrast to *tost-1(xf194)*, *tost-1(xf196)* has a temperature dependent phenotype: individuals are fully penetrant Mel when cultured at restrictive temperatures (25ºC) and viable under low stringency culture conditions (15ºC). Furthermore, the viability of this allele indicates that the critical part of TOST-1 activity is contained within the N terminal intact part of this protein. The ERH-2 interacting motif remains intact in *tost-1(xf196)*, in accordance with our observation that is likely essential for TOST-1 activity. Moreover, this temperature sensitive allele gave us the ability to induce TOST-1 inactivity, and the Mel phenotype, at will, by shifting animals from permissive to restrictive temperatures. Thus, we used this allele to determine the stage by which this protein is required for embryonic viability.

We tested this by culturing L4 larvae of *tost-1(xf196)* at both permissive and restrictive temperatures overnight, shifting the gravid adults from permissive to restrictive temperatures and vice-versa, and finally accounting for their progeny viability for a total of ten hours. This experiment allowed us to both determine the amount of time needed to induce the Mel phenotype and the time required to reverse it, defining a developmental window for the essential function of TOST-1 and PETISCO.

Animals shifted from 25ºC to 15ºC are able to start producing viable offspring after eight hours at the permissive temperature (Fig. R13A). The reversibility of the Mel phenotype indicates that there is no structural malformation of the gonad. Yet, the recovery time is significantly longer than the predicted time it would take from fertilization to egg-laying (circa 300 minutes), indicating that viable embryos were both fertilized and raised at 15ºC. This data suggests that TOST-1 activity is required already at the adult gonad before fertilization.

In the opposite experiment, individuals transferred from 15ºC to 25ºC, revealed a clear loss of viability within just two hours of exposure (Fig. R13A). This time window indicates that these embryos were residing at the uterus at the time of the temperature shift (Altun and Hall, 2009). Embryos residing in the uterus are fertilized and their development is independent of the gonad. Thus, the presence of active TOST-1 is required in both the parental gonad and the embryo. Interestingly, we observed that late stage embryos shifted to restrictive temperature, tend

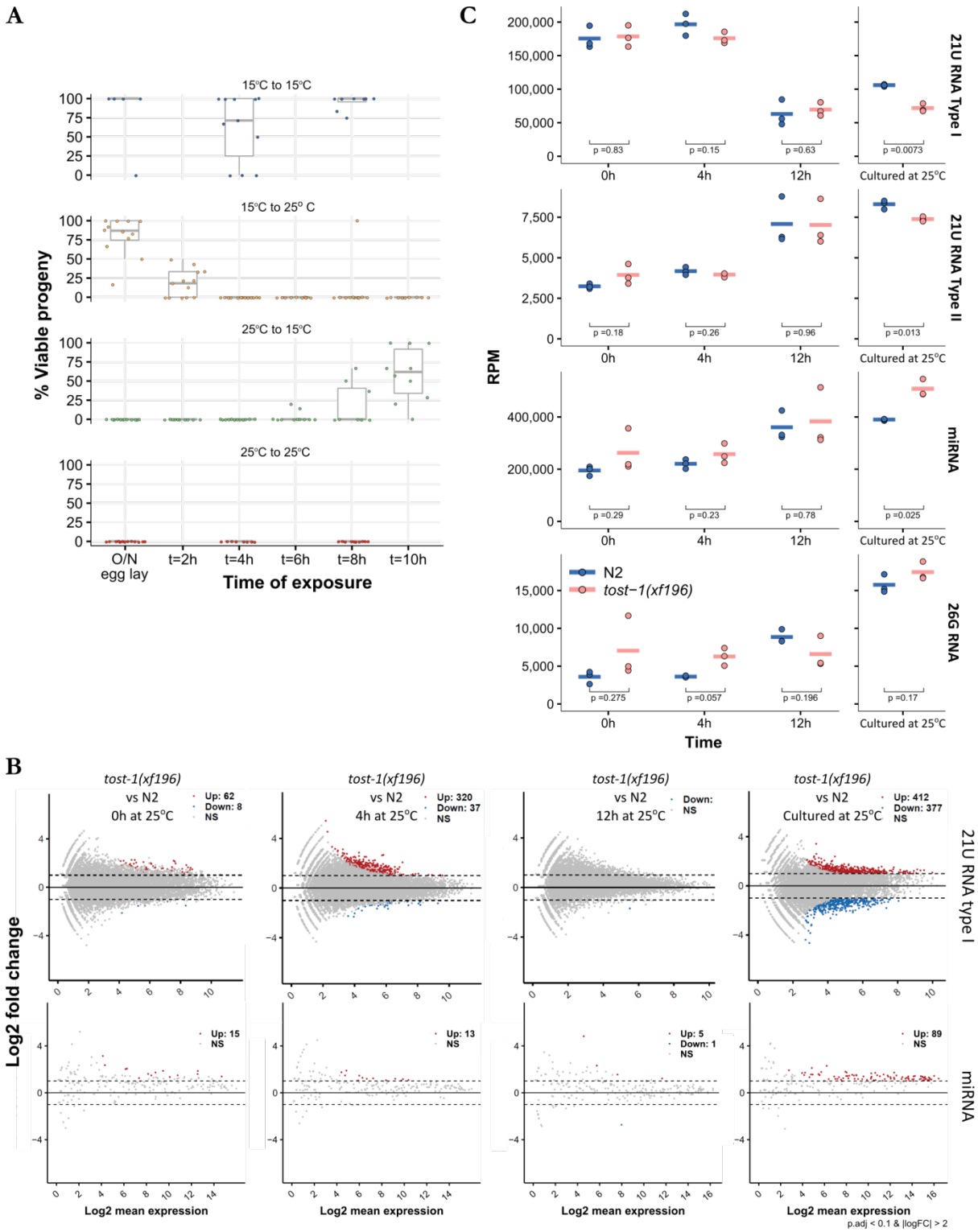

Figure R13 - Temperature shift assays of the temperature sensitive allele *tost-1(xf196)*.

(A) L4 larvae were grown overnight (O/N) at 15ºC or 25ºC and shifted to 25ºC or 15ºC the next morning as adults. O/N egg lay was shifted to corresponding temperature in parallel. Individuals were changed to new plates every 2h and progeny counted at egg stage and L2 larvae. Each point represents the progeny of an individual worm. (B and C) smRNA sequencing of temperature shifted *tost-1(xf196)* populations. (B) Pairwise differential gene expression of type I 21U RNAs and miRNAs in *tost-1(xf196)* vs. wildtype (N2) control. Samples consist of synchronized gravid adults exposed to stringent temperatures (25ºC) for 0h, 4h, 12h. Otherwise synchronized L1 larvae were cultured at 25ºC and collected as gravid adults. Red and blue dots indicate upregulated and downregulated transcripts respectively.

(C) Global levels of 21U RNAs type I, type II, miRNAs and 26G RNAs of samples of (B). X-axis represents the time of exposure or fully cultured at 25ºC. Values are in reads per million mapped reads (RPM). Each dot represents a sequenced replicate and horizontal bar represents the total mean. Significance tested with Student's t-test. N2 WT in blue and *tost-1(xf196)* in pink. Small RNA populations and p-values are indicated in the figure.

to escape embryonic arrest and grow into adulthood. We conclude from these data that TOST-1, and PETISCO, are specifically required during late oogenesis to early embryogenesis.

3.11.1. tost-1 mutations do not affect small RNAs directly

tost-1(xf196) as an inducible allele is able to minimize secondary effects of long term exposure to TOST-1 ablation. Thus, we further used *tost-1(xf196)* to validate the smRNAseq results obtained with *tost-1(xf194)*. smRNAseq was performed in triplicate N2 and *tost-1(xf196)* gravid adult populations. These were either cultured for 0h, 4h and 12h in stringent temperatures or animals were cultured completely at 25ºC starting their first larval stage (L1) (Fig. R13B and C).

We observed that *tost-1(xf196)* animals fully cultured at 25ºC showed a slight decline in the total amount of 21U RNAs and a small upregulation of some miRNAs (Fig. R13B and C). These results resemble those we obtained for *tost-1(xf194)* (Fig. R10 A and C). Still, we found no changes in any small RNA class between *tost-1(xf196)* and N2 in temperature shifted populations. *tost-1(xf196)* embryos are irreversibly arrested within 4h of exposure to stringent temperatures (Fig. R13A). As we do not find small RNA defects within this timeframe, we conclude that these are not the cause for the embryonic arrest and that the observed small RNA changes are likely an indirect result of the long-term absence of TOST-1.

3.12. PID-1 is required for 21U RNA precursor stability

PID-1 is required for 21U RNA precursor processing and the presence of 21U RNAs, making PETISCO an ideal candidate for a 21U RNA precursor processing platform. Hence, we asked if 21U RNA precursors are affected by the loss of PID-1. 21U RNA precursors are a lowly abundant class of small RNAs. In order to focus on 21U RNA precursors multiple steps were required: RNA samples were treated with Calf Intestinal Phosphatase (CIP), followed by 5' RNA Pyrophosphohydrolase (RppH). The first treatment prevented the cloning of the 5' ends of RNAs, while the second hydrolysed the RNA 5' Caps into 5' PO_4, restricting library cloning to capped RNAs (Almeida et al., 2019c; Gu et al., 2012). We then restricted our bioinformatic analysis to sequencing reads longer than 22 nucleotides, starting at the position -2 of 21U loci (Goh et al., 2014), ensuring the focus on 21U RNA precursors.

Triplicate RNA samples of *pid-1(xf35)* and N2 populations were treated and sequenced accordingly. We found that there is a significant reduction of the total 21U RNA precursor population in *pid-1(xf35)* mutants (Fig. R14A). This shows that PID-1 is required for the stability of 21U RNA precursors, possibly through binding of these molecules to PETISCO.

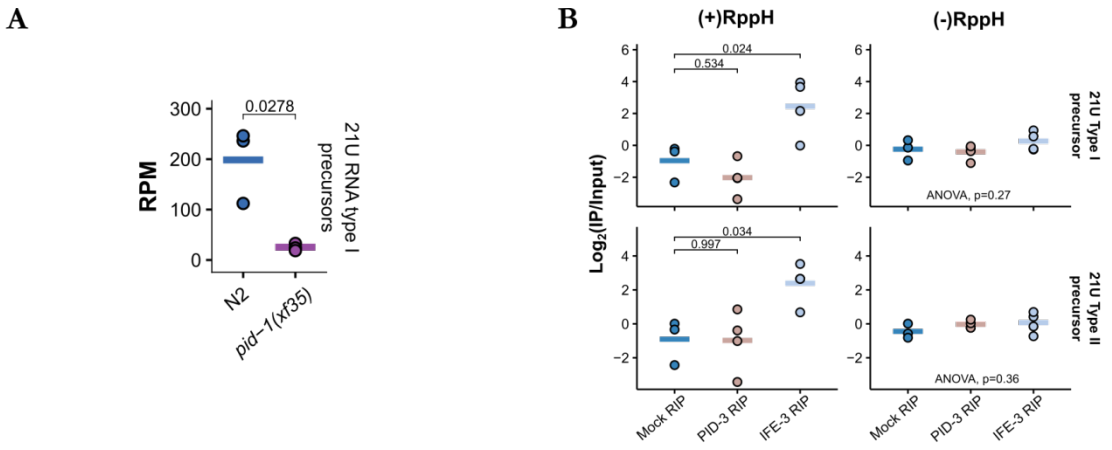

Figure R14 - PETISCO interacts 21U RNA precursors.
In all cases individual data points of three independent replicates are shown and horizontal bar represents the mean.
(A) Global levels of type I 21U RNA precursors in wild type (N2) and *pid-1(xf35)* non gravid adults. Values obtained from CIP-RppH treated smRNA libraries. RPM stands for reads per million mapped reads. Significance was tested with Student's t-test and p-value is indicated in graph.
(B) Fold enrichments of 21U RNA precursors in Mock (N2), IFE-3 and PID-3 RIPs over paired inputs in non-gravid adult worms. Left column displays RppH treated and right column non-treated samples. Significance of sample vs. Mock RIP tested with Dunnett's post hoc test and p-values are indicated in the graph.

3.13. PETISCO interacts with 21U RNA precursors

We investigated whether PETISCO has the ability to interact with 21U RNA precursors by performing RNA immunoprecipitation followed by small RNA sequencing (RIPseq) of IFE-3 and PID-3. The previously mentioned strains carrying *ife-3(xf102)* or *pid-3(tm2417)* and complementing mCherry tagged transgenes were used to perform RIPseq experiments. In order to increase our confidence in the RNA interactors, each RIPseq was performed in quadruplicate. A mock RIPseq experiment in a non-transgenic WT strain was used as control. RNA isolated from the RIP eluates and corresponding lysate RNA input samples were sequenced in order to detect specific enrichments in the RIP. To detect the 5' Capped 21U RNA precursors, half of each RNA sample was treated with RppH before library preparation. In this protocol we avoided the CIP treatment due to the limited RNA sample obtained. To note, after bioinformatic analysis, we found one of the mock replicates to differ significantly from the remaining samples and this sample was excluded.

Mock and PID-3 RIPseq samples did not show any significant enrichment for either 21U RNAs or 21U RNA precursors in IP/Input in either treatment (Fig. R14B). On the other hand, while IFE-3 IPs do not show a significant enrichment in untreated libraries, we find a clear IP/Input enrichment (4-fold) for both type I and type II 21U RNA precursors in RppH treated samples (Fig. R14B). These data show that IFE-3 interacts with capped 21U RNA precursors, supporting PETISCO as an integral part of the 21U RNA processing machinery.

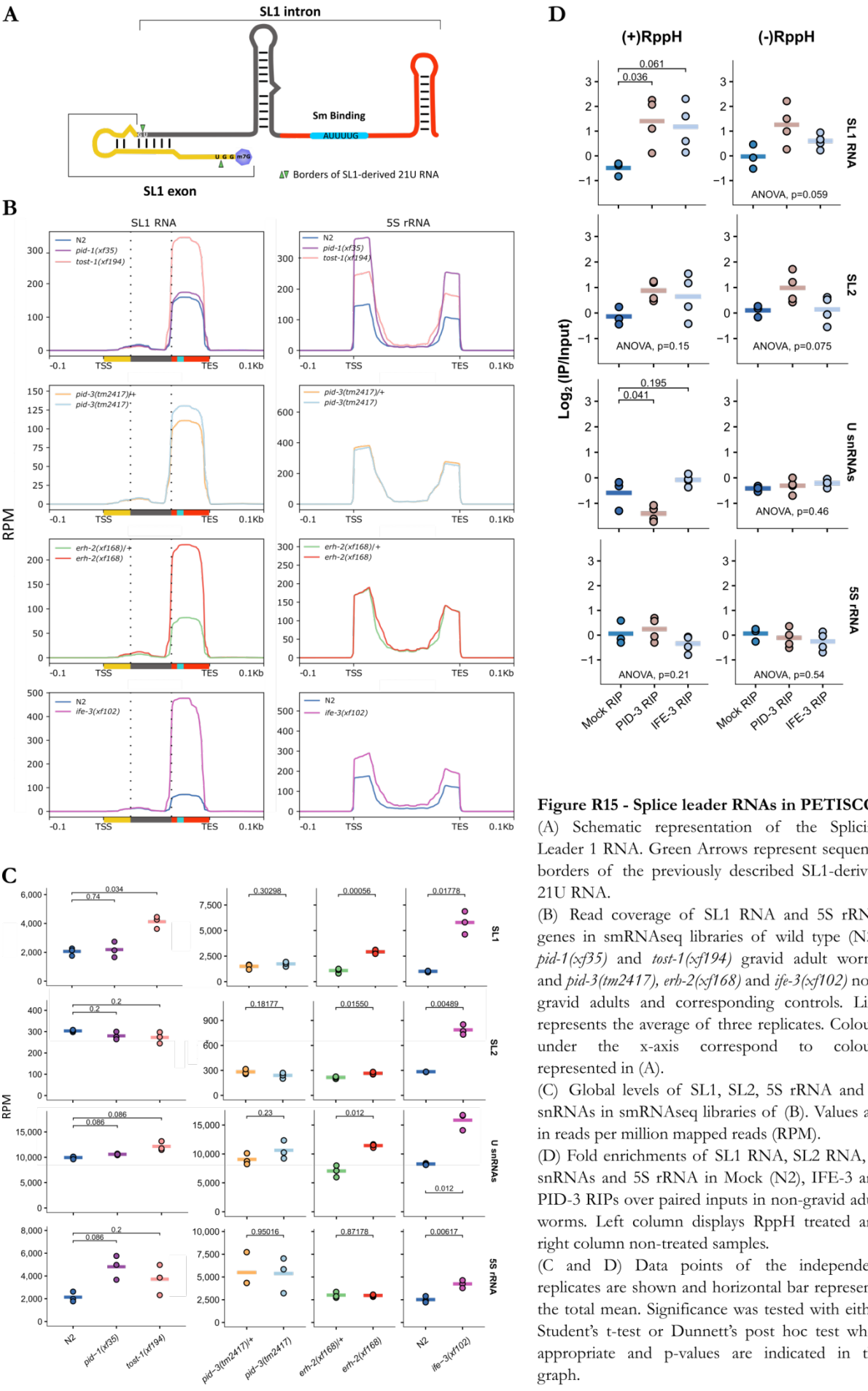

Figure R15 - Splice leader RNAs in PETISCO.
(A) Schematic representation of the Splicing Leader 1 RNA. Green Arrows represent sequence borders of the previously described SL1-derived 21U RNA.
(B) Read coverage of SL1 RNA and 5S rRNA genes in smRNAseq libraries of wild type (N2), *pid-1(xf35)* and *tost-1(xf194)* gravid adult worms and *pid-3(tm2417)*, *erh-2(xf168)* and *ife-3(xf102)* non-gravid adults and corresponding controls. Line represents the average of three replicates. Colours under the x-axis correspond to colours represented in (A).
(C) Global levels of SL1, SL2, 5S rRNA and U snRNAs in smRNAseq libraries of (B). Values are in reads per million mapped reads (RPM).
(D) Fold enrichments of SL1 RNA, SL2 RNA, U snRNAs and 5S rRNA in Mock (N2), IFE-3 and PID-3 RIPs over paired inputs in non-gravid adult worms. Left column displays RppH treated and right column non-treated samples.
(C and D) Data points of the independent replicates are shown and horizontal bar represents the total mean. Significance was tested with either Student's t-test or Dunnett's post hoc test when appropriate and p-values are indicated in the graph.

3.14. PETISCO interacts with SL1 RNA and is required for its homeostasis

C. elegans does not require 21U RNAs for embryonic survival. As PETISCO is required for embryonic development, its function should be broader than just 21U RNA biogenesis. Hence, we queried if the essential function of this complex could be related to additional RNA populations. Null mutations of *ife-3* lead to an interchangeable Mel/Mog phenotype, a phenotype also observable in *mel-46* mutants (Minasaki et al., 2009), the *C. elegans* orthologue of Gemin3. We have identified MEL-46 and other components of the SMN complex, as interactors of IFE-3 (Fig. R5). This complex is a major processor of snRNAs into snRNPs and it has been recently shown that 21U RNA transcription shares major components with the snRNA transcription machinery (Kasper et al., 2014; Weng et al., 2019). We therefore hypothesized that these two pathways may share further machinery, and that PETISCO may be acting on snRNA homeostasis.

3.14.1. PETISCO and tost-1 mutants accumulate SL1 snRNAs

We probed into snRNAs using our small RNA sequencing data. Strikingly, we found a subclass of snRNAs, splice leader sequences, to be affected in some of the mutants. In both *erh-2(xf168)* and *ife-3(xf102)* we found a 3 and 5-fold accumulation, respectively, of a 3'end fragment of SL1 RNA (Fig. R15B), the major splice leader sequence of *C. elegans*. Contrarily, we found no significant differences in splicing snRNAs or 5S rRNA, a structural ribosomal RNA transcribed from the same cluster as the SL1 RNA (Fig. R15B). *ife-3(xf102)* mutants further show a comparable accumulation for the SL2 RNA (Fig. R15C), albeit this observation is supported by an overall low read count. Importantly, the same SL1 RNA fragment has a comparable accumulation in *tost-1(xf194)* individuals, but not in *pid-1(xf35)* mutants (Fig. R15B and C). This pattern is the reverse to that of 21U RNA defects found on these two mutants (Fig. R10 A and

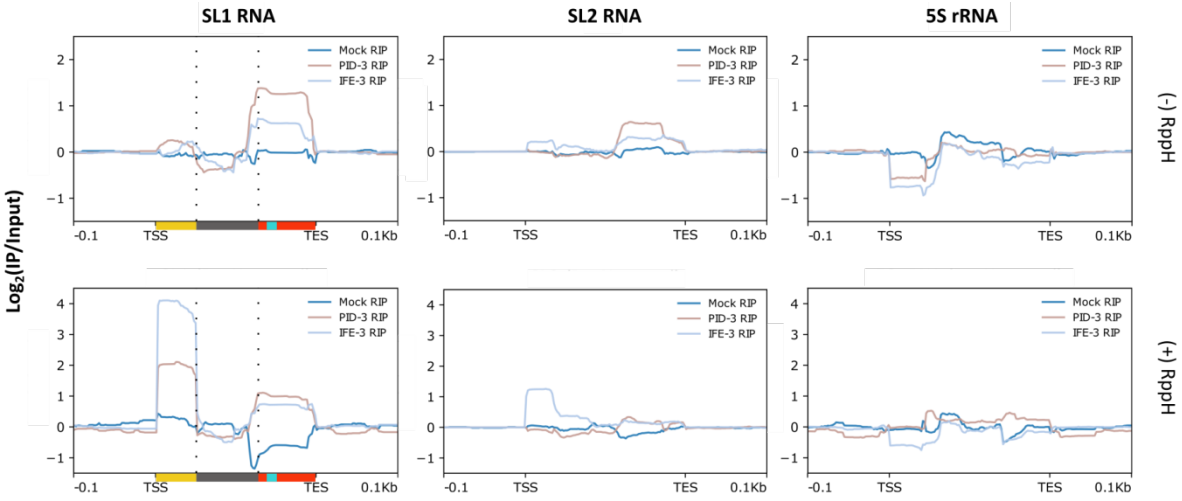

Figure R16 - Splice leader RNA sequences interacting with PETISCO.
Coverage profile of fold enrichments of SL1 RNA, SL2 RNA and 5S rRNA in Mock (N2), IFE-3 and PID-3 RIPs normalized to paired inputs in non-gravid adult worms. Lower row displays RppH treated and top row non-treated samples. Colours under the x-axis correspond to colours represented in Figure R15A.

C). Thus, these data support two different actions of this complex driven by PID-1 and TOST-1.

The SL1 accumulation may be due to PETISCO-related SL1 processing in which the absence of the complex leads to the accumulation of this RNA. On the other hand *pid-3(tm2417)* mutants do not show this accumulation. We speculate that the accumulation phenotype may require SL1 RNA binding, and that PID-3, being at the centre of PETISCO, may be essential for such binding. The absence of IFE-3, ERH-2 or TOST-1, the more peripheral elements of the complex, may allow binding and consequent SL1 RNA accumulation, possibly due to disturbed processing.

3.14.2. PETISCO binds 5' Capped SL1 snRNAs

To further test the link between PETISCO and the SL1 RNA we analysed our RIPseq data for both IFE-3 and PID-3. In accordance to our predictions, we found a small increase for SL1 RNA mapping reads in both PID-3 and, to a less extent, IFE-3 RIPs in comparison to inputs (Fig. R15D). These differences were intensified by RppH treatment (Fig. R15D). Adversely, mock RIPs showed either no enrichments for SL1 RNA or a depletion for this RNA species in RppH treated samples. This result strongly suggests that SL1 RNA species interact with PETISCO and that these carry a 5' end Cap.

We find the same pattern of enrichment for SL2 RNA, although to a lesser extent (Fig. R15D). This may be due to the lower amount of total reads we find for SL2 RNA. Interestingly, we find no differences between mock, PID-3 and IFE-3 RIPs for other snRNAs or 5S rRNA, further supporting a specific interaction between PETISCO and SL1/2 RNA.

The library preparation protocol used in this study is optimized for small RNA sequences (see section 2.11.), meaning that we are unable to capture full length splice leader RNA transcripts in significant amounts. In order to understand the SL1 RNA fragments we are observing, we plotted the RIPseq:SL1 mapping reads over the SL1 gene (Fig. R15D).

We found that our results are derived from two fragments: a 5' end fragment, only present in RppH treated libraries, and a 3' end fragment present in both treatments (Fig. R16). Since it is present only under the RppH treatment, we conclude that the 5' end fragment is capped. This capped fragment is enriched approximately 16-fold in IFE-3 and 4-fold in PID-3 compared to the mock RIP (Fig. R16). This suggests that IFE-3 interacts with the capped 5'end of SL1 RNA. The 3' end fragment corresponds to the same fragment we found accumulated in PETISCO and *tost-1* mutants (Fig. R15B and R16) and has comparable enrichments in both IFE-3 and PID-3 RIPs. We interpret these results as PETISCO binding the full length 5' Capped SL1 RNA. The absence of the middle fragment of SL1 may be due to degradation by nucleases during the RIP procedure. Altogether, we find that SL1 RNA and PETISCO are binding partners and PETISCO depletion affects SL1 RNA homeostasis.

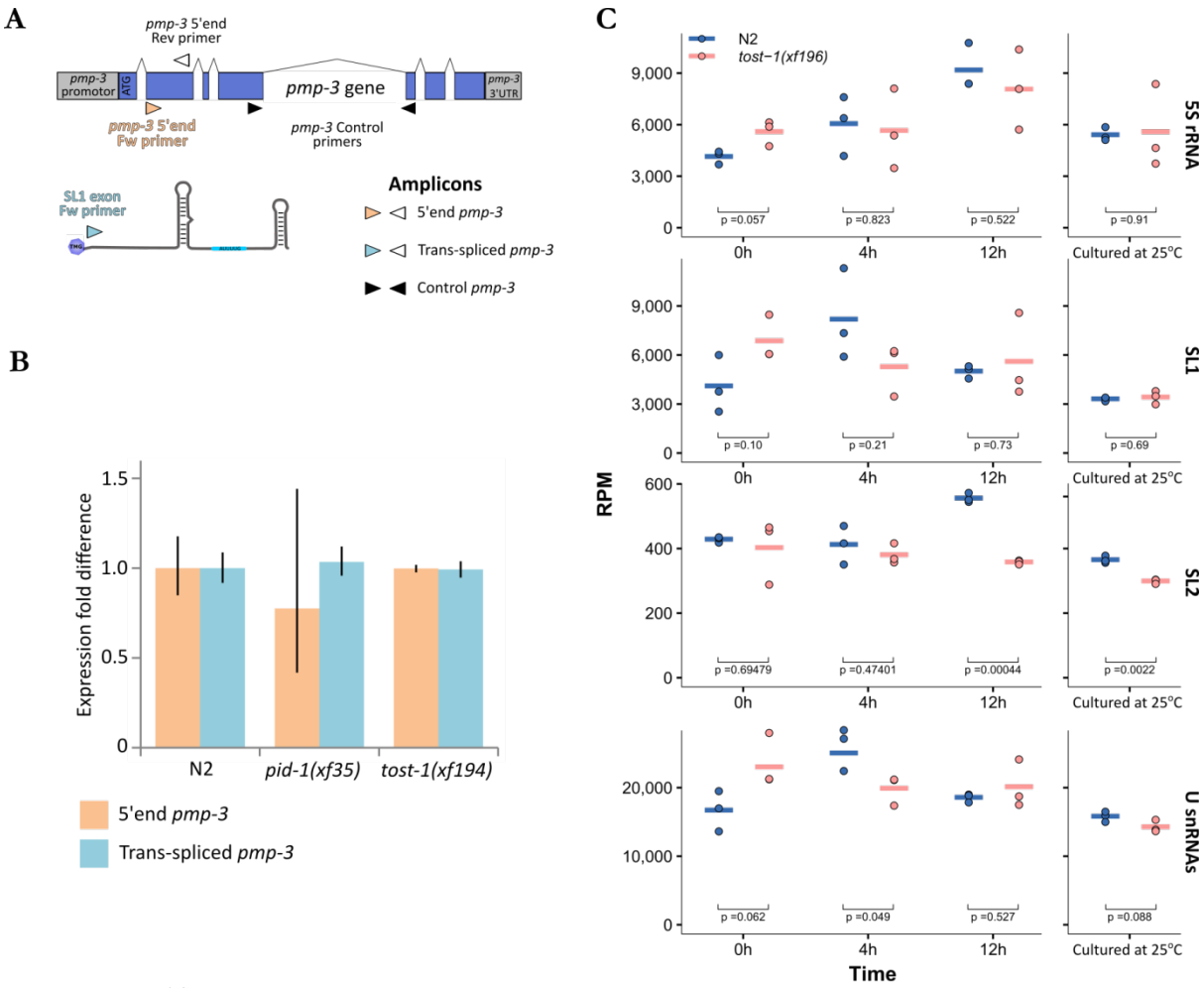

Figure R17 - PETISCO does not affect trans-splicing.
(A) Schematic representation of RT-qPCR amplicons to detect trans-splicing efficiency in *pmp-3* mRNA (B) Quantitative RT-PCR of (A) wild type (N2), *pid-1(xf35)* and *tost-1(xf194)* gravid adults. Values are obtained from experimental triplicates and normalized to pmp-3 control amplicon levels.
(C) Global levels of SL1, SL2 and U snRNAs and 5S rRNA of wild type (N2) and *tost-1(xf196)* temperature shifted samples. Samples consist of triplicate synchronized gravid adult populations. x-axis represents the time of exposure to 25ºC fully cultured in those conditions (see text for more information). Values are in reads per million mapped reads (RPM). Each dot represents a sequenced replicate and horizontal bar represents the total mean. Significance tested with Student's t-test. N2 WT in blue and *tost-1(xf196)* in pink. The individual small RNA populations and statistical p-values are indicated in the figure.

3.15. Trans-splicing defects are not the leading cause for developmental arrest in PETISCO mutants

SL1 RNA is present in 80% of the trans-spliced transcripts of *C. elegans* (Allen et al., 2011). Thus we queried if the SL1 homeostasis defects were the cause behind the Mel phenotype. We tested this hypothesis by analysing the variation of SL1 and SL2 RNA levels in our smRNAseq experiment with the temperature sensitive *tost-1(xf196)* allele.

3.15.1. SL1 and SL2 RNA homeostasis defects are not the cause of the Mel phenotype

As embryos from this strain become irreversibly Mel within 4h of exposure to 25ºC, we expect defects directly linked to this phenotype to occur within the same timeframe. Interestingly, we find no significant difference in SL1 or SL2 RNA levels in *tost-1(xf196)* and N2 at any point of

exposure to stringent conditions (Fig. R17C). This result indicates that the embryonic arrest in these mutants is unlikely to be the result of perturbations in SL1 or SL2 RNA levels.

3.15.2. PETISCO is not required for trans-splicing

PETISCO interacts with SL1 RNA, thus we queried if this complex has an active role in the process of trans-splicing. To address this we designed a RT-qPCR assay to determine the relative amount of trans-spliced transcripts in animal populations. In this assay we measure the ratio of trans-spliced transcripts using three primer pairs: a first pair targeting the SL1 sequence and the 5' end of the target gene; a second pair targeting the 5' end of the same target gene, but contained in its coding sequence; and finally a third pair targeting an amplicon distant from the transcript 5' end as an internal control for the other two pairs (Fig. R17A). *pmp-3* is a trans-spliced transcript (Gu et al., 2012) regularly used as a control due to its highly stable expression (Hoogewijs et al., 2008). For these two characteristics we chose *pmp*-3 to perform our trans-splicing assay.

We performed RT-qPCR in triplicates, using the samples of wild-type, *pid-1(xf35)* and *tost-1(xf194)* animals previously used for smRNAseq (Fig. R10). We found little difference across the genetic backgrounds used (Fig. R17B). This result shows that trans-splicing of *pmp-3* RNA is unaffected by either *pid-1(xf35)* or *tost-1(xf194)* mutations and that PETISCO does not play an active role in trans-splicing.

3.16. PETISCO binds core histone mRNAs and is required for their stability.

Splice leaders and 21U RNA precursors both bind PETISCO, but neither of these RNA species explain the deleterious effects caused by the absence of this complex. Hence, we further probed our data for other interacting RNA species that may be the leading cause for the Mel phenotype.

3.16.1. PETISCO interaction with core histone mRNAs is not mediated by a 5'end Cap

We performed differential gene expression analysis in our RIPseq datasets against the corresponding inputs to find specific transcripts interacting with PETISCO. Surprisingly, in both PID-3 and IFE-3, but not mock RIPs, we consistently find enrichments for the mRNA of core histone genes (Fig. R18A). In both cases we can find mRNA transcripts corresponding to H2A, H2B, H3 and H4 coding genes, although IFE-3 shows milder enrichments than PID-3. RppH treatment led only to minor changes in histone mRNA enrichments, implying that these interactions are not mediated by a 5' end cap (Fig. R18A).

3.16.2. PETISCO and tost-1 mutations destabilize core histone mRNAs

We further investigated the connection between PETISCO and histone mRNAs within our mutant smRNAseq datasets. Interestingly, differential gene expression analysis of these data

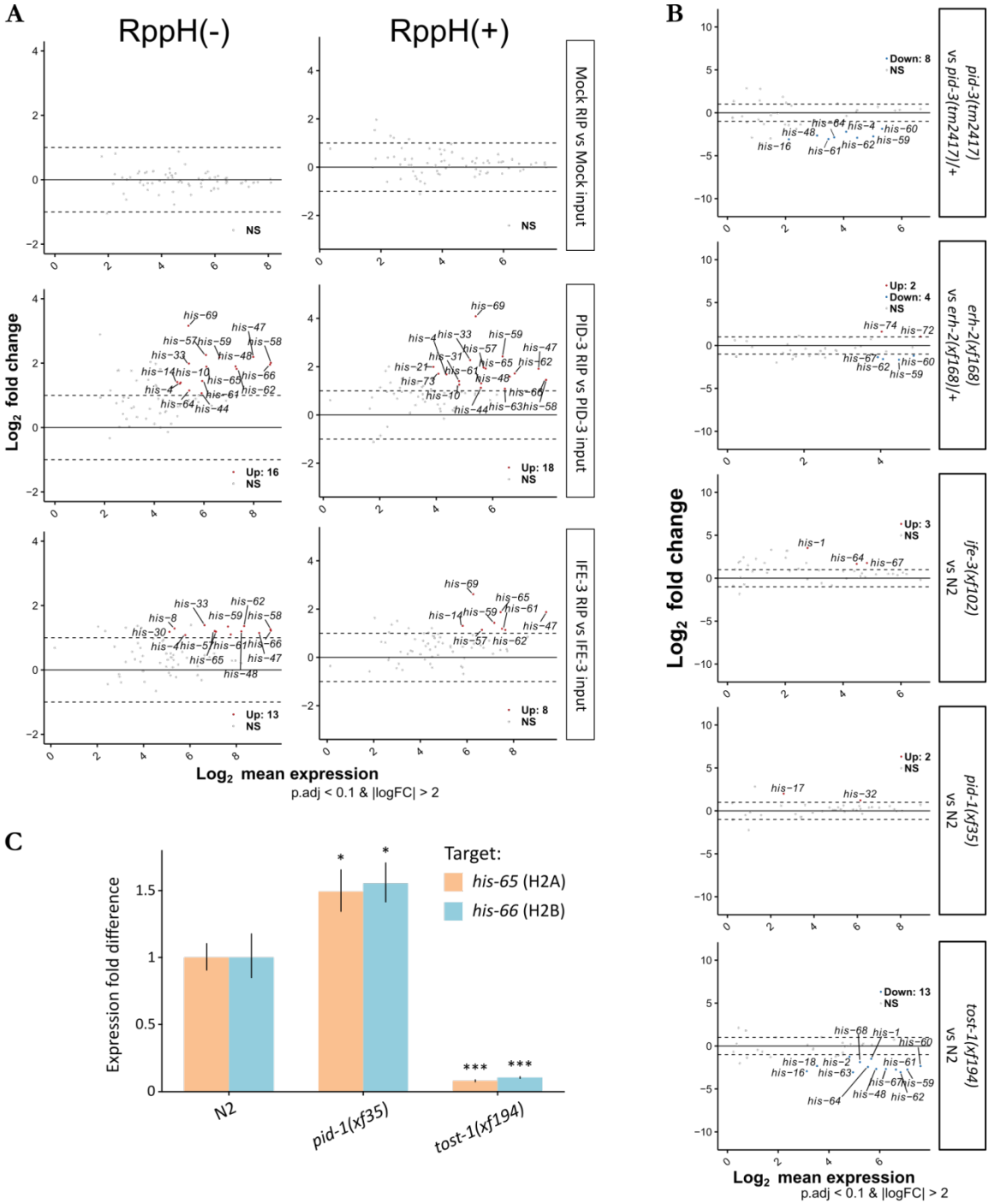

Figure R18 - PETISCO is required for core histone mRNA stability.
Pairwise differential gene expression of histone mRNAs in: (A) Mock (N2), IFE-3 and PID-3 RIPs vs paired inputs in non-gravid adult worms. Right column displays RppH treated and left column non-treated samples. RIP bait is indicated in the figure; (B) *pid-1(xf35)* and *tost-1(xf194)* gravid adult worms and *pid-3(tm2417), erh-2(xf168)* and *ife-3(xf102)* non-gravid adults vs respective controls. In (A) and (B) Points represent the average of replicates. Red and blue dots indicate upregulated and downregulated transcripts respectively. (C) Quantitative RT-PCR of *his-66*(H2A) and *his-66*(H2B) expression in wild type (N2), *pid-1(xf35)* and *tost-1(xf194)* gravid adults. Values are obtained from experimental triplicates and normalized to *pmp-3* expression. Significance tested with Student's t-test in mutant vs N2. *p-value<0.05; ***p-value<0.001.

showed a substantial decrease of histone mapping transcripts in both *pid-3(tm2417)* and *tost-1(xf194)* and a mild decrease for *erh-2(xf168)* (Fig. R18B). In contrast, we find no changes in *pid-1(xf35),* or *ife-3(xf102)* populations (Fig. R18B). It is noteworthy that these results are extracted from small RNA libraries, meaning that these sequence reads are likely derived from the

degradation of mRNAs. While these results are indicative, one has to be cautious when extracting further conclusions of these datasets in respect to mRNA. Hence, we queried the validity of these results using RT-qPCR to measure the levels of histone mRNA in *tost-1(xf194)*, *pid-1(xf35)* and wild-type animals.

We quantified the levels of total histone *his-65* and *his-66* mRNA (coding H2A and H2B, respectively) in the gravid adult populations previously used for smRNAseq. Strikingly, we find that *tost-1(xf194)* individuals express only 0,08-fold of *his-65* and 0,1-fold of *his-66* of wild-type mRNA levels (Fig. R18C). On the other hand, we find *pid-1(xf35)* to have a 1,5-fold increase in the expression of both of these mRNAs (Fig. R18C). This may be the result of a higher availability of PETISCO for TOST-1 in the absence of PID-1, further reflecting the dichotomy between these two proteins. The astonishing decrease in histone mRNA levels in *tost-1(xf194)* may account for the Mel phenotype of PETISCO mutants, as developing embryos are in strong demand of histones to sustain their mitotic rates (Pettitt et al., 2002).

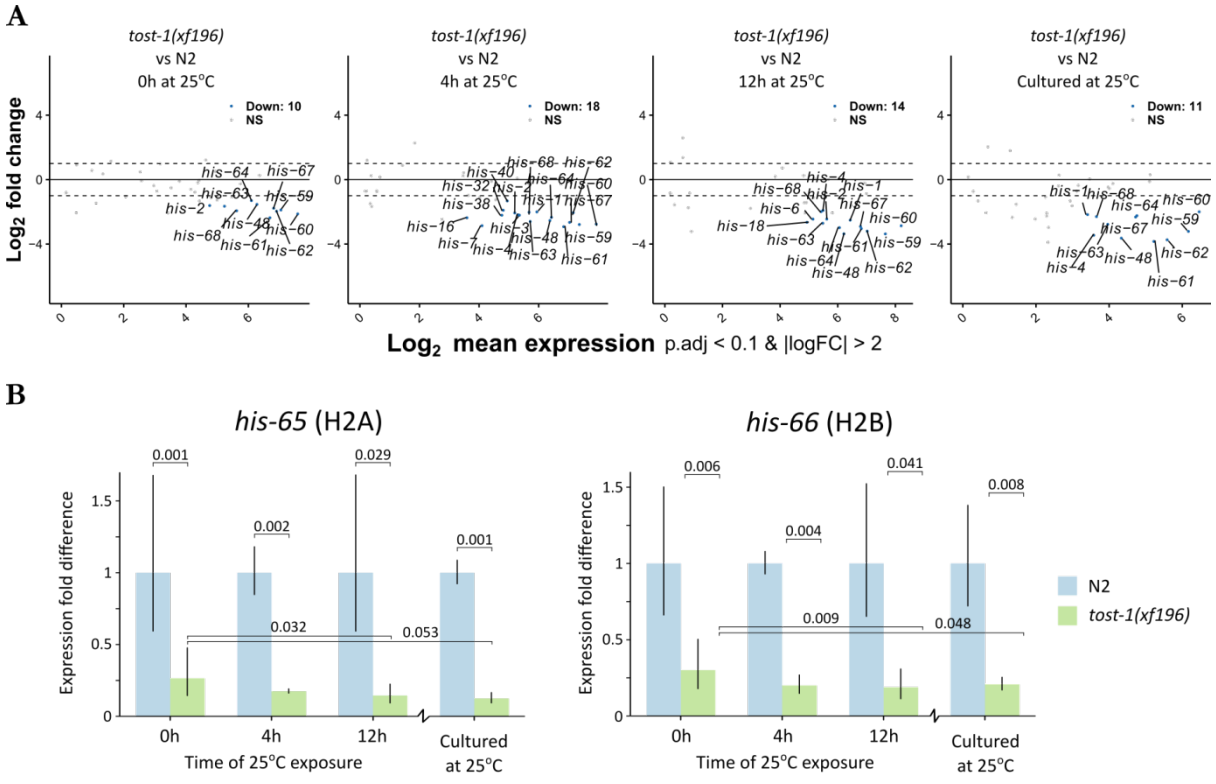

Figure R19 - Histone mRNAs are depleted in *tost-1(xf196)*.
(A) Pairwise differential gene expression of histone mRNAs in *tost-1(xf196)* vs. wildtype (N2) control. Samples consist of synchronized gravid adults exposed to stringent temperatures (25°C) for 0h, 4h or 12h. Otherwise synchronized L1 larvae were cultured at 25°C and collected as gravid adults. Red and blue dots indicate upregulated and downregulated transcripts respectively. Each dot represents the replicate average.
(B) Quantitative RT-PCR of *his-66*(H2A) and *his-66*(H2B) expression in samples of (A). Values are obtained from experimental triplicates and normalized to *pmp-3* expression. Significance tested with Student's t-test and p-values are indicated in graph.

3.16.3. Core histone mRNA destabilization coincides with embryonic arrest

In order to further confirm our findings we searched our *tost-1(xf196)* temperature shift smRNAseq experiment for the same histone mRNA defects. Interestingly, under all conditions *tost-1(xf196)* animals show a decreased core histone mRNA expression compared to WT control (Fig. R19A). Still, we found that the more time animals were exposed to stringent temperatures, the more aggravated the registered effect is (Fig. R19A and B). We further quantified this histone depletion in each of the different time points by RT-qPCR targeting *his-65* and *his-66* mRNA levels.

Confirming our observations, in each condition the total levels of histone *his-65* and *his-66* mRNA are significantly decreased in *tost-1(xf196)* vs wild-type individuals (Fig. R19B), likely a consequence of the hypomorphic nature of this allele. *tost-1(xf196)* individuals cultured under permissive conditions express only 0,26-fold and 0,3-fold of wild type *his-65* and *his-66* mRNA levels, respectively (Fig. R19B). Comparing within the *tost-1(xf196)* population we find that *his-65* and *his-66* mRNA levels are worsened by prolonged exposure to stringent temperatures. At 4h of exposure we find a mild but non-statistically significant decrease of *his-65* and *his-66* mRNA levels in *tost-1(xf196)* compared to non-exposed control. These differences become more prominent after 12h at 25°C, where *his-65* and *his-66* mRNA levels significantly decreased to 0,6 and 0,5-fold, respectively, of non-exposed animals (Fig. R19B).

Populations fully cultured at 25°C express only 0,4-fold of the *his-65* and *his-66* mRNA of corresponding permissive controls (Fig. R19B). On the other hand we find no significant changes in *his-65* and *his-66* mRNA levels within the wild-type populations in any of these conditions, eliminating the possibility that these observations are due to temperature effects. Conceivably, the hypomorphic nature of *tost-1(xf196)* is due to its ability to maintain histone mRNAs just above a threshold of embryonic lethality, which is no longer sufficient under stringent culture conditions.

Altogether, these data show that PETISCO interacts with core histone mRNAs and is required for their stability. It also suggests that PID-1 and TOST-1 compete for the available PETISCO and that histone mRNA depletion may be the source for the embryonic arrest in these mutants.

3.17. Depletion of PETISCO causes widespread chromosome segregation defects in embryos

Nucleosomes are an integral part of genome organization and cell division in eukaryotic cells. In *C. elegans* it has been shown that depletion of histone mRNA can lead to embryonic arrest and mitotic defects, such as formation of chromatin bridges (Avgousti et al., 2012; Pettitt et al., 2002). These phenotypes have also been recorded in *tofu*-6 mutants (Minasaki and Streit, 2007). Thus, we investigated if PETISCO depletion led to similar defects.

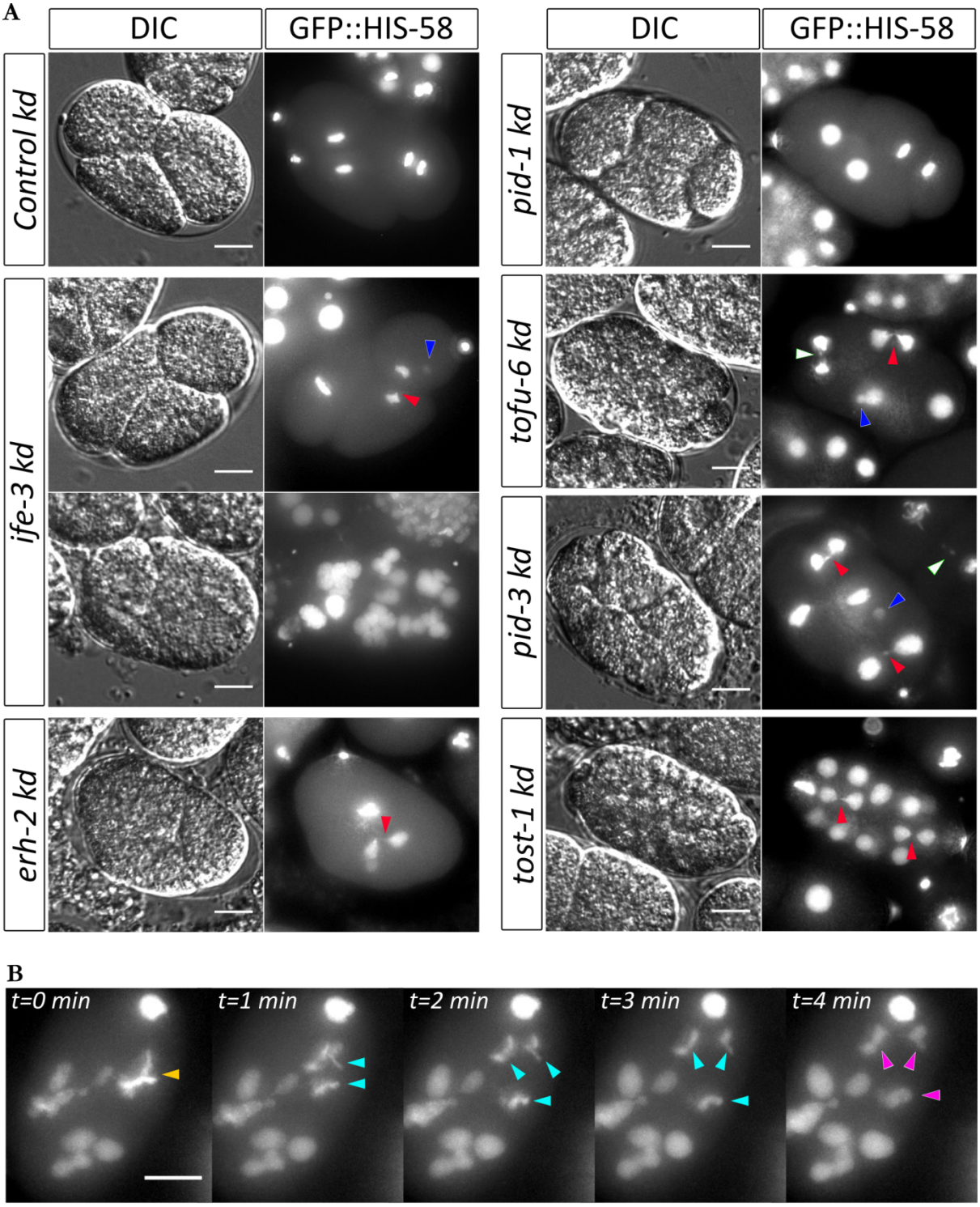

Figure R20 - Chromosome segregation defects of PETISCO depletion.
Wide field fluorescent microscopy of live embryos depleted for PETISCO genes via feeding RNAi. (A) Representative images of the found chromosome segregation defects in PETISCO depleted embryos. RNAi targets are indicated in panels. (B) Time-lapse of chromosome segregation in *ife-3* knock down embryos. Arrow colour code: red - chromatin bridge; white - lagging chromosome; dark blue - micronuclei; yellow - metaphase plate; light blue - anaphase chromosome; lilac - chromosome decondensation/nuclear envelope formation. Scale bar represents 10 μm.

We investigated early mitotic divisions of embryos using a strain carrying a transgene expressing a fluorescently tagged H2B, GFP::HIS-58. In order to avoid possible regulatory consequences of disrupting PETISCO on GFP::HIS-58, this transgene is regulated by a different

promoter and 3' UTR from those of *his-58*. We tested the effects of PETISCO depletion in dividing embryos by feeding RNAi targeting each of its subunits. Embryos of this strain were hatched in pre-seeded RNAi media and their progeny imaged. Strikingly, we find the depletion of PETISCO subunits to have a widespread impact in chromosome segregation.

While we found mitotic defects on neither control nor *pid-1* knock down plates, depletion of *tofu-6* led to the formation of chromatin bridges during cell division, as previously described (Fig. R20A, red arrow). These bridges were also commonly observed upon knock down of *pid-3, erh-2, ife-3* and *tost-1* (Fig. R20A, red arrow). Moreover, we observed the occurrence of micronuclei and lagging chromosomes in both *tofu-6* and *pid-3* depletion (Fig. R20A, dark blue and white arrows). These defects are known to cause chromosome instability (Kodama et al., 2002) which may lead to embryonic arrest.

3.17.1. *ife-3* causes spindle and cytokinesis defects

Although we found these mitotic defects for each PETISCO subunit, we observed defects unique to *ife-3* depletion. In these embryos, we found micronuclei and multinuclear cells (Fig. R20A). In order to understand how these multinuclear cells are formed we performed live imaging of *ife-3* knock down embryos. Interestingly, we found that these embryos formed apparently standard metaphase plates (Fig. R20B, t=0 min, yellow arrow) but as cells progressed into anaphase, chromosomes spread along three rather than two poles (Fig. R20B, light blue arrows), a sign of a multipolar mitotic spindle. These cells were further able to form nuclear envelopes in each of these poles, but no longer underwent cytokinesis (Fig. R20B, t=4 min, lilac arrow). These additional effects of IFE-3 depletion may be associated with its additional functions that lead to the Mog phenotype in adult individuals.

We conclude from these data that PETISCO is essential for chromosome segregation in early embryos and its maternal ablation leads to embryonic arrest. Possibly, PETISCO depletion leads the lack of sufficient core histones available for proper mitosis, leading to genomic collapse and embryonic arrest.

3.18. Evolutionary analysis of PETISCO

Nematodes are a richly diverse *phylum* known to have rapid genetic evolution (Aboobaker and Blaxter, 2010; Kasimatis and Phillips, 2018). 21U RNA conservation reflects this diversity, as many nematodes seemingly have lost this pathway (Beltran et al., 2019; Sarkies et al., 2015). We now know of the connection between PETISCO and 21U RNAs in *C. elegans* and probed the evolution of this complex in other nematode species. We determined the conservation of PETISCO resorting to reciprocal BLASTp analysis of each subunit to the whole proteome of species representing the multiple clades of *Nematoda* (Fig. R21). As useful comparisons we

included the conservation patterns of two essential elements of the 21U RNA pathway PRG-1 and PRDE-1 (Batista et al., 2008; Das et al., 2008; Weick et al., 2014), and the ERH-2 paralogue ERH-1.

We found that all PETISCO subunits are conserved within the *Elegans* and *Drosophilae* supergroups, where they possibly function like their *C. elegans* counterparts (Fig. R21). Perhaps unsurprisingly, the analysis further reveals IFE-3 and ERH-1 as proteins whose ancestry goes beyond the *Nematoda phylum* (Fig. R21). These belong to protein superfamilies known to have essential functions in other organisms as well. PETISCO conservation differences arise when inspecting clades beyond the *Caenorhabditis* genus.

PID-3, TOST-1 and TOFU-6 orthologues were likely present in the last common ancestor of nematode clades III, IV and V, though we note that TOFU-6 may have been lost in clade IV nematodes. The occurrence of these proteins together in the evolutionary timeline may be an indication of an early partnership between them. Such partnership may be related to the essential function of PETISCO in *C. elegans* and species lacking the 21U RNA pathway may still require PETISCO.

PID-1 orthologues are only found in clade V species containing PRG-1. We find the

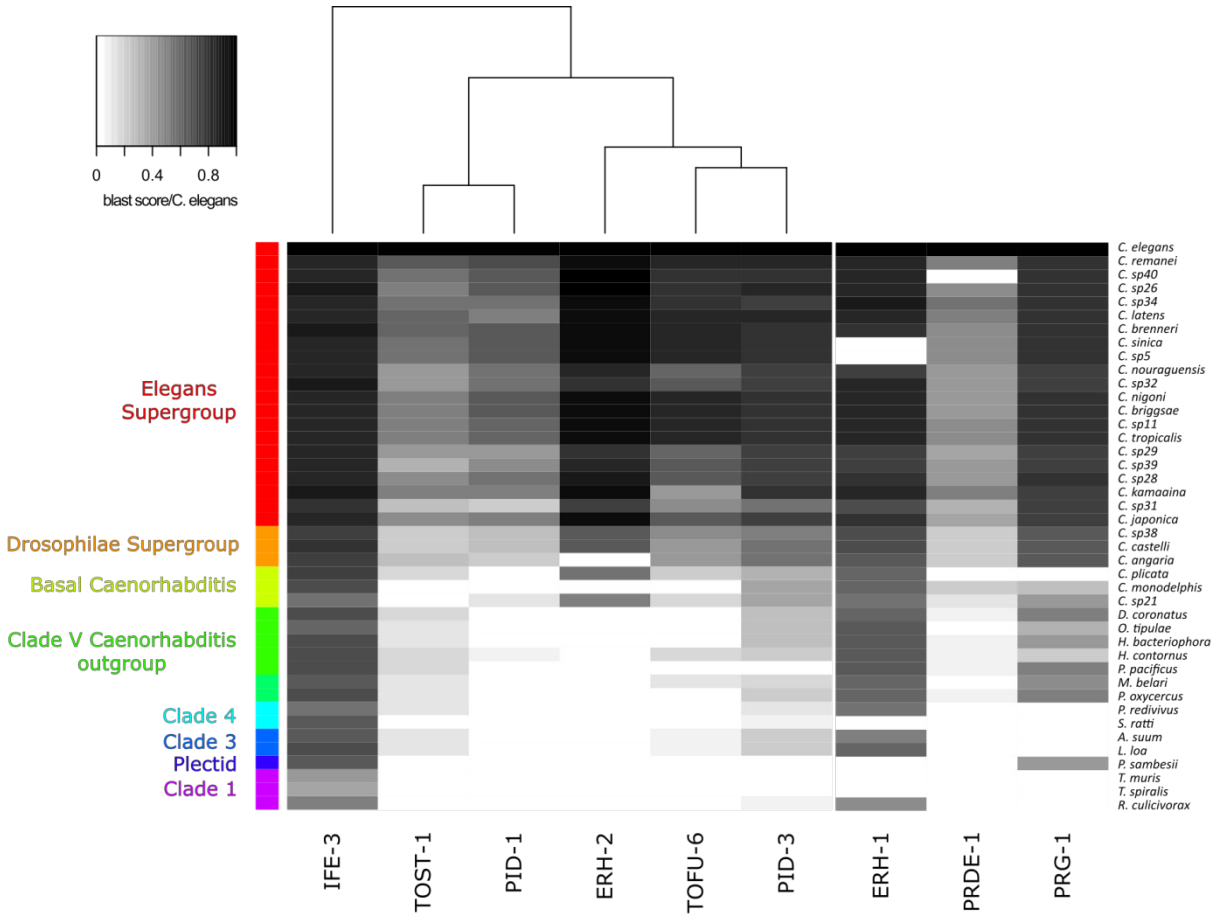

Figure R21 - PETISCO conservation in nematodes.
Heatmap representing the conservation of PETISCO, PID-1 and TOST-1 orthologues in nematodes according to BLASTp score vs *C. elegans*. ERH-1, PRDE-1 and PRG-1 are included for comparison.

same pattern for PRDE-1, a protein that adapts the SNAPc complex for 21U RNA expression (Kasper et al., 2014; Weng et al., 2019). Possibly this is an indicator that PID-1, like PRDE-1, may be a specialization of available molecular machinery into the 21U RNA pathway.

4. Discussion

In this work we describe a protein complex, PETISCO, which is required for both 21U RNA biogenesis and for embryonic viability. Additionally, we find these functions are driven by two distinct proteins, PID-1 and TOST-1, the absence of which exclusively impact 21U RNA biogenesis and embryonic viability, respectively. In this section we present our interpretation and hypotheses for the function and evolution of PETISCO in *Caenorhabditis elegans*.

4.1. PETISCO is an RNA binding complex composed of different modular units

In our IP-LFQP experiments we found PETISCO components to be difficult to separate *ex vivo*. Most of these proteins remain intact interactors independently of RNase treatment or increased stringency conditions (Fig. R6). Together with our subcellular localization assays (Fig. R4), these experiments demonstrate the presence of a robust protein complex at the adult germline and early embryos. The subunits of PETISCO, as expected for a protein complex, have similar mutant phenotypes and expression patterns. Nonetheless, a detailed analysis of each of the PETISCO proteins allows us to define different modular parts for this complex.

Ablation of any of the elements of PETISCO leads to Mel (Maternal effect lethal) phenotypes (Table R1), the exception being removal of IFE-3. The removal of the latter can also lead to a Mog (Masculinization of the germline) phenotype or multinucleated embryos (Fig. R20). IFE-3 also is the component we found to have the largest spectrum of interacting partners, which besides the PETISCO subunits include Gemins and, possibly, the complete SMN complex (Fig. R5E). *ife-3* mutants also display the most moderate effect in the 21U RNA population from each of the PETISCO components. Altogether, these data indicate a broader function for this protein relative to the other PETISCO proteins. Thus, we consider IFE-3 to be its own module within this complex.

The remaining PETISCO elements also show characteristics of a modular identity. Depletion of ERH-2 leads to a steep decrease in the total 21U RNAs levels, mirroring the effects of PID-3 depleted individuals (Fig. R106). On the other hand, mutations in *erh-2* lead to an accumulation of SL1 RNA, which remains unchanged in *pid-3* mutant animals (Fig. R15). Also, core histone mRNAs appear to be less affected in *erh-2* mutants, while they fall significantly in *pid-3* mutants (Fig. R18). Furthermore, ERH-2 mediates the interaction between PETISCO and TOST-1 or PID-1, key determinants of its end function. Based on these results we consider ERH-2 as a second module of this complex.

TOFU-6 ablation was not studied in this work. Nevertheless, publicly available datasets for *tofu-6* knock down, show us this protein is required for 21U RNA biogenesis (Goh et al., 2014). We also probed these datasets for accumulation of SL1 RNA and, as in *pid-3* mutants,

failed to find significant changes (data not shown). In their study, Goh et al., 2014 use a different experimental setup from the one we presented here, including different library preparation procedures and absence of biological replicates. These differences do not allow us to make direct comparisons between our datasets and those of Goh et al., 2014. Nonetheless, according to the current data available, *tofu-6* and *pid-3* depletion resemble each other in both 21U RNA absence and maintenance of SL1 RNA levels. Other studies further show interdependency and similarities between these two proteins. Zeng et al., 2019, show that TOFU-6:GFP is unable to form P-granules in *pid-3(tm2417),* while the same is not true neither for *pid-1* mutants nor in worms fed RNAi against *tost-1* or *erh-2*. Interestingly, in the latter conditions both TOFU-6 and PID-3 form larger perinuclear granules. On the grounds of the similarities between these two subunits, we consider them two halves of a third modular unit of PETISCO.

4.2. PID-3 and TOFU-6 form an RNA stabilising core module of PETISCO

Functional similarities between TOFU-6 and PID-3 lead us to consider them part of the same PETISCO modular unit. These two proteins also share biochemical characteristics; Particularly, they both contain an RRM domain, a domain family with a high potential for RNA binding (Maris et al., 2005). Presumably, these may be required for the RNA interactions of PETISCO. Empirically, we found that the function of these domains cannot be restricted to RNA binding. The RRMs of PID-3 and TOFU-6 are sufficient to mediate interaction between these two proteins (Fig. R7C). Nevertheless, a protein-protein binding function and an RNA binding activity are not mutually exclusive features of this domain family, as combined RNA and protein binding activities have previously been described (Deo et al., 1999; Handa et al., 1999).

Interactions between PETISCO proteins are resistant to RNase A/T1 treatment (Fig. R6). This result is a strong indicator that these interactions are RNA-independent. Yet, this experiment does not exclude the possibility that the RNA mediating these interactions is secluded and thus, protected from RNases. In this scenario PID-3 and TOFU-6 interact in an RRM-RNA-RRM manner. The fact the RRM-RRM interaction can be recapitulated in a Y2H set up further suggests this interaction is RNA independent, although one cannot exclude a *S. cerevisiae* RNA, such as an snRNA with similarities to SL1 RNA, to assume RNA bridging functions between TOFU-6 and PID-3 RRMs. Experiments with purified recombinant proteins will be needed to conclusively address this issue.

4.2.1. Potential binding partners of the TOFU-6 tudor domain

TOFU-6 also contains a Tudor domain for which we were unable to determine an interacting partner (Fig. R7). Tudor domains are known to interact with symmetrically dimethylated arginines (sDMA) (Huang et al., 2011; Nishida et al., 2009; Roovers et al., 2018), a

modification notoriously abundant in nuage and perinuclear RNA granules (Gao and Arkov, 2013), to which PETISCO localizes (Fig. R4B). Conceivably, this domain may mediate a protein-protein interaction directing PETISCO to P granules. PRG-1, the 21U RNA Piwi argonaute of *C. elegans*, may be a good candidate for this interaction. Piwi argonautes are known to have N-terminal arginines undergo symmetric dimethylation (sDMA) (Huang et al., 2011; Kirino et al., 2009; Reuter et al., 2009) and PRG-1 contains a typical Piwi N-terminal arginine cluster with sDMA motifs. We find PRG-1 to be mildly enriched in our IFE-3 IP-LFQP assay (Fig. R5A), which could be a reflection of such an interaction. Other candidate interactors of the TOFU-6 tudor domain are Sm proteins, which form heptameric rings and bind to snRNAs. SL1 RNA, like other snRNAs, is loaded with the Sm protein ring at its Sm site (Thomas et al., 1988), a process dependent on sDMAs (Friesen et al., 2001; Gonsalvez et al., 2007). The interaction between PETISCO and SL1 RNA may thus be stabilized by a TOFU-6-Sm protein interaction. On the other hand, we should be able to detect the existence of direct binding between PRG-1 or Sm proteins and TOFU-6 in our TOFU-6 IP-LFQP data, and we find no evidence of these interactions (Fig. R5). Notably, we did not detect any sDMA-enriched peptides in this experiment (data not shown) and we cannot exclude other interactions with this tudor domain.

Human tudor protein TDRD3 has diverged from its typical tudor function and has been shown to recognize an asymmetric methylated arginine of RNA Pol II (Sikorsky et al., 2012). Also, TDRD2/PAPI recognition of PIWI-like 1 has been shown to be sDMA-independent (Zhang et al., 2017). Similarly, the TOFU-6 tudor domain may have evolved to recognize a non-modified protein or a PTM either than sDMA. Additionally, interactions with TOFU-6, beyond the ones here mentioned, may be transient and difficult to detect with IP-LFQP. Proximity labelling proteomic methods, such as TurboID and BioID (Branon et al., 2018; Roux et al., 2012), may potentially circumvent these limitations and identify other TOFU-6 interactors.

Surprisingly, although we know PID-1 interacts with PETISCO, we only found low enrichments for this protein in this particular assay (Fig. R5C). Thus, the lack of, or relatively low detection of PID-1 may be attributed to our sampling of both PETISCO pools, in which PID-1:PETISCO and TOST-1:PETISCO dilute the signal of one another. In this case, a slight higher amount of TOST-1:PETISCO in our TOFU-6 IP-LFQP, may have driven PID-1 to the edge of the detection threshold of this assay.

4.2.2. PID-3 harbours a domain analogous to the Argonaute-typical MID domain

Apart from the aforementioned RRM, PID-3 also harbours a domain that likely resembles a MID domain. MID domains are to our current knowledge only found in AGO proteins, where they mediate the interaction between these proteins and the 5' end of their small RNA co-factor (Boland et al., 2010; Elkayam et al., 2012; Frank et al., 2010; Schirle and MacRae,

2012). MID domains can also occasionally confer a nucleotide binding preference (Cora et al., 2014; Frank et al., 2010), though the PID-3 MID does not present an obvious selectivity based on structure prediction (I. Macrae, personal communication). Although similar, this domain differs enough so that homology with argonaute MID domains cannot be detected by homology detection tools such as BLAST (Fig. R2C). The predicted protein fold structure of the PID-3 MID domain, as judged by HHpred an I-TASSER, highly resembles the hAGO2 MID (Fig. R2B). On the other hand, the amino acid sequence of the former differs from the hAGO2 MID domain to the extent that we cannot determine a common ancestry between these two domains (Fig. R2C). This implies that, rather than a homologue, the PID-3 MID domain is an analogous structure that may perform a specialized function within PETISCO. Hence, one could consider that this domain requires its own nomenclature, such as MID AnalogouS domain, or MIDAS. MID domains adopt a Rossmann-like fold (Boland et al., 2010), the latter creates a cavity with a high potential to bind a nucleotide ring (Rossmann et al., 1974). This fold is fairly common, and is present in up to 22% of proteins with known structures (Medvedev et al., 2018). Potentially, MIDAS has evolved from another protein containing a Rossmann-like fold that adopted a MID-like function. Moreover, MIDAS is an essential part of PID-3, as it is evidenced by *pid-3(xf151)*. This allele contains an in-frame deletion of MIDAS, leaving the remaining PID-3 potentially intact. We cannot exclude that this deletion leads to an unstable PID-3, but like *pid-3(tm2417), pid-3(xf151)* individuals are fully penetrant Mel (Table R1), which may indicate an inability to mediate interactions with proteins or RNA.

The exact function of PID-3 in PETISCO remains unclear, although this protein can potentially bind and stabilize the 5' PO_4 of the 21U precursor RNA intermediates via MIDAS. We did not succeed at detecting such RNA species in our RIPseq experiments. This could be due to a quick transit of these transcripts through PETISCO, making it difficult to detect them via RIPseq. A technique that may address this problem is individual-nucleotide resolution UV crosslinking and immunoprecipitation (iCLIP), by which RNA is covalently linked to the protein that it directly interacts with. iCLIP will not only determine which RNA species PID-3 interacts with, but also the exact interacting nucleotide. If MIDAS interacts with the 5' nucleotide it would be a good indication of a hAGO2 MID-like function. Alternatively, recombinant protein assays of MIDAS and RNAs with different 5'ends may also strengthen this view.

4.2.3. PID-3 binds and stabilizes RNA

Our studies show PID-3 to be an interactor of both core histone mRNAs and SL1 RNA (Fig. R15, R16 and R18). While we do not find a binding bias for the former, RIPseq enrichments are clearly focused on the 3' end of the latter transcript (Fig. R17). Interestingly, the only common feature we find between all these transcripts is the presence of a hairpin secondary

structure at their 3' end. Possibly, PID-3 is responsible for the binding of these RNAs via this secondary structure. The stabilization of SL1 RNA in all TOST-1:PETISCO mutants tested, except for *pid-3,* may be a reflection of this same interaction. When depleting the remaining PETISCO subunits we may prevent the downstream action of PETISCO, stabilising the SL1 RNA in its interaction with PID-3. Supporting this hypothesis is the fact that we find the same 3' end fragment of SL1 RNA to be interacting with PID-3 (Fig. R16) and to be upregulated in the other PETISCO mutants (Fig. R15B).

Interestingly, PID-3 also interacts and is vital for the presence of core histone mRNAs. Together, these are strong signals that PID-3 is a direct regulator of these transcripts but our experimental approach does not allow us to determine direct interactions of PID-3 with RNA molecules. Once again, performing iCLIP on PID-3 will help us understand if these interactions are direct and give us single nucleotide resolution, confirming a possible preference of PID-3 to RNA hairpin binding. The *in vitro* reconstitution of this core of PETISCO may also allow us to test these RNA interactions and possible preference in binding sites.

4.3. ERH-2 serves as a mediator module connecting PETISCO to function

4.3.1. 'Enhancer of Rudimentary' proteins are ancient multitasking factors

Enhancer of rudimentary proteins are a highly conserved protein family in eukaryotes, the ancestry of which can be traced back to at least the last common ancestor of animals, fungi and plants (Arai et al., 2005). This protein family is known to adopt a unique protein fold, consisting of an antiparallel β–sheet and three α-helices in a $\beta_1\beta_2\alpha_1\alpha_2\beta_3\beta_4\alpha_3$ secondary structure disposition (Arai et al., 2005; Wan et al., 2005). The β-sheet is rigid and mediates ERH dimerization, while the α-helices are prone to some flexibility, although how this influences interactions with other proteins is unclear (Arai et al., 2005; Jin et al., 2007).

Despite the available structures, the molecular function of Enhancer of Rudimentary family proteins remains elusive in most organisms. The first ERH protein was identified in a forward mutagenesis screen in *Drosophila*, *e(r)*, where the authors were searching for genes affecting the rudimentary wing phenotype (Gelsthorpe et al., 1997), thus its name. The *Drosophila* rudimentary locus encodes a pyrimidine synthesis machinery and consequently the authors suggest *e(r)* to play a role in pyrimidine biosynthesis. In humans *erh1* has been reported to interact with the Sm protein D3, a subunit of the Sm ring required for snRNA processing (Weng et al., 2012). In the same study, the authors report that the depletion of hERH1 leads to missplicing of the mitotic motor protein CENP-E, ATR and other cell cycle-related transcripts. Although the

reason for this specificity remains unclear, these events lead to chromosome segregation defects. hERH1 has also been reported to be a transcription repressor by interacting with the RNA Pol II C-terminal domain (CTD) phosphatase, FCP1 (Amente et al., 2005). In frog, XERH1 interacts with the Dimerization Cofactor of HNF1 (DCoH) and is suggested to act as a repressor of HNF1 homeobox transcription factors (Pogge von Strandmann et al., 2001), but like hERH1, its exact molecular function remains undefined.

The function of enhancer of rudimentary orthologues is clearest in *S. pombe*. There, Erh1 forms a complex with Mmi1 (Sugiyama et al., 2016; Xie et al., 2019), a protein that recognizes and determines the degradation of meiotic transcripts during the *S. pombe* vegetative life cycle (Chen et al., 2011; Sugiyama et al., 2016; Yamashita et al., 2013). Together these proteins form the EMC, an interesting complex that has two distinct functions depending on its interacting partner. On the one hand it interacts with MTREC, a complex required for meiotic transcript degradation and heterochromatin island assembly, on the other hand, it is a close partner of the CCR4-NOT complex. EMC-CCR4-NOT is required for ribosomal DNA (rDNA) heterochromatin assembly (Sugiyama et al., 2016). Interestingly, the RNAi machinery is required for heterochromatin establishment and maintenance at these sites in *S. pombe* (reviewed in Martienssen and Moazed, 2015).

4.3.2. ERH protein family characteristics are ideal for PETISCO task mediation

We find that PETISCO interacts with snRNAs and PETISCO mutants display clear chromosome segregation defects, resembling the functions of hERH1 (Amente et al., 2005; Weng et al., 2012). Conversely, *erh-2* is required for the RNAi-like 21U RNA pathway that induces the silencing of transcripts. ERH-2 also mediates the interaction between PETISCO and TOST-1 and PID-1, resembling more the functions of its *S. pombe* counterpart (Sugiyama et al., 2016). Perhaps, all these phenotypes are related, though we could not identify human or yeast orthologues of most PETISCO subunits (data not shown). Furthermore, human and yeast Erh homologues are strictly nuclear proteins (Fujimura et al., 2012; Krzyzanowski et al., 2012) and the elegans ERH-2 protein is, at least predominantly, cytoplasmic (Fig. R4). The diversity of functions associated with Erh proteins favours a model by which this protein family is not composed of catalytic proteins, but rather proteins with a high potential for modulating and bridging protein interactions. In the formation of EMC in *S. pombe,* Erh1 dimerizes via its β-sheet, forging a pseudo-β-barrel (Xie et al., 2019). The surface created by the homodimer pseudo-β-barrel allows for the docking of a Mmi1 protein molecule in each end of the barrel, creating a heterodimer with 2:2 stoichiometry. The homo-dimerization surface is a conserved, and key characteristic of Erh family proteins (Arai et al., 2005; Wan et al., 2005; Xie et al., 2019). This

surface has the ability to recognize intrinsically disordered regions (IDRs)(Xie et al., 2019), common and crucial elements of RNA-binding proteins (Protter et al., 2018). Given that this plasticity is complemented with additional binding pockets available in Erh proteins (Jin et al., 2007; Xie et al., 2019), perhaps it is not surprising that these proteins have assumed different functions in RNA biology during evolution. Protein families, with high interaction potential, being co-opted for different functions has been reported before (Almeida et al., 2018; Dönertas et al., 2013; Yoshimura et al., 2018). In fact, we find ERH-2 to have the largest spectrum of interactions in our Y2H assay (Fig. R7). ERH-2 has the potential to bind itself, the RRM of PID-3 and TOST-1 or PID-1 and according to our IP-LFQP data it at least maintains the connection between PID-3 and PID-1 or TOST-1 simultaneously (Fig. R5). We were further able to show that there is an amino acid motif responsible for the binding of ERH-2 by TOST-1 and PID-1 (Fig. R7 and R12). Key for this interaction is an arginine, the substitution of which leads to interaction loss (Fig. R7D) and even PID-1 destabilization (de Albuquerque et al., 2014). The interaction motif between Erh1 and Mmi1 contains no arginines (Xie et al., 2019) and possibly the interaction between PID-1/TOST-1 and ERH-2 is maintained by a different segment of the protein. Still, if ERH-2 dimerizes in PETISCO and creates the same pseudo-β-barrel as in *S. pombe* Erh1, it has the potential to allow the dimerization of PETISCO. In this scenario each side of the barrel surface could interact with a PID-3 RRM, creating a PETISCO dimer (Fig. D1). In fact, we demonstrate that the native PETISCO molecular weight is larger than the sum of its known subunits (Fig. R5H). PID-1 or TOST-1 would thus interact with another binding site of ERH-2. The latter interaction would direct the complex into two different functions, in resemblance to the two functions acquired by the *S. pombe* Erh1 (Sugiyama et al., 2016).

Our study did not address the binding pockets and dimerization of ERH-2 *in vivo*. A dimerization would explain the observed size of PETISCO in our chromatography assays (Fig. R5H). The Y2H assay did give us direct binding information of each subunit, but it does not allow us to determine the exact architecture or stoichiometry of PETISCO. This problem would likely be solved by reconstitution of this complex *in vitro*. These experiments would not only allow for RNA binding tests but potentially determine the existence of other subunits missed in our IP-LFQP essays by comparison to our size exclusion chromatography results. Cryo-EM of PETISCO or NMR assays of ERH-2 with its interacting partners would give insightful information on the flexibility of ERH-2 and its interactions. These assays would contribute to the overall modest knowledge of the community on Erh proteins and likely show how the binding and alternation of PID-1/TOST-1 is achieved.

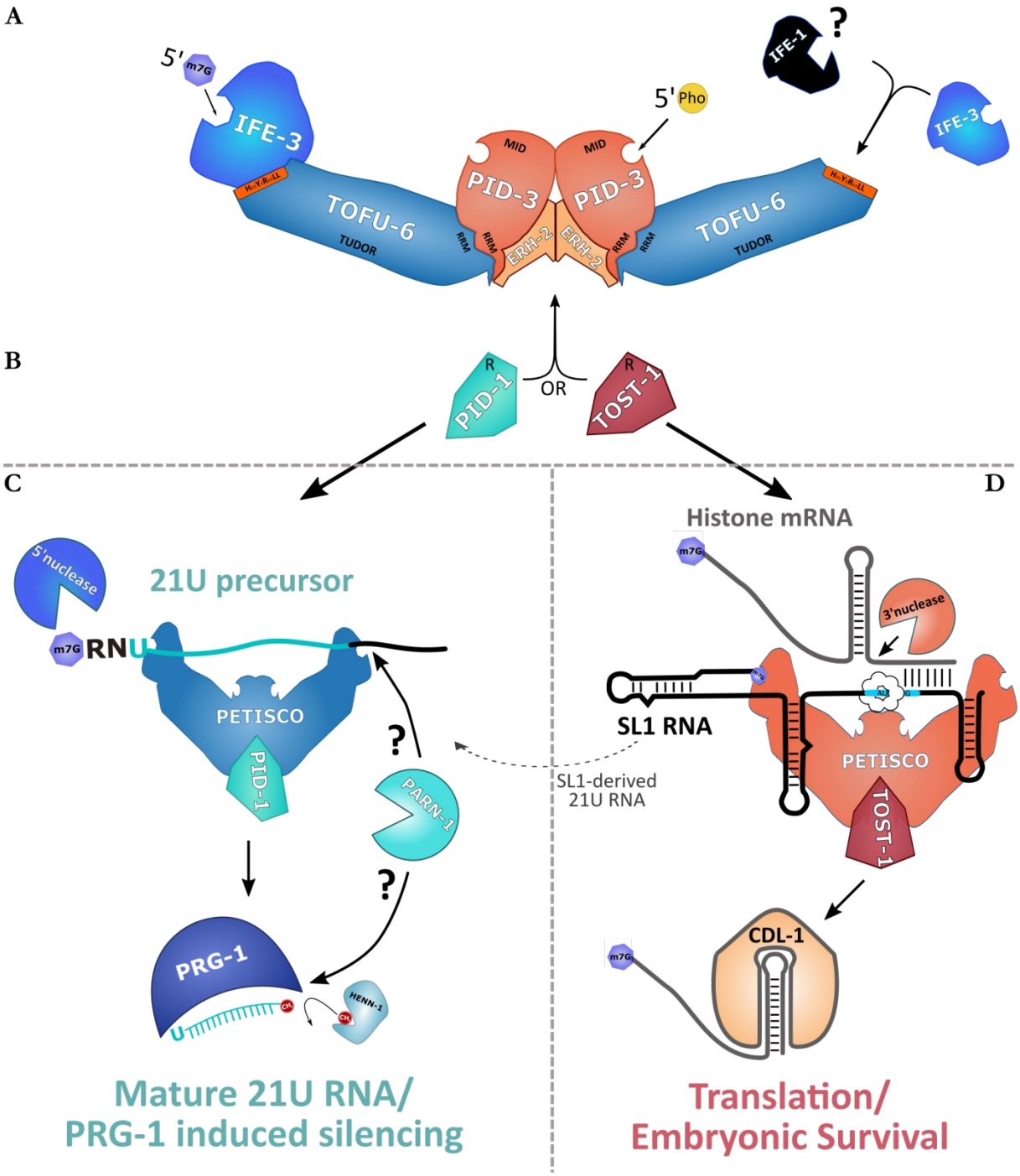

Figure D1 – Schematic representation of PETISCO function.

A schematic of the proposed PETISCO function, displaying its dual function: one in 21U biogenesis and another needed for embryonic survival.

(A) PETISCO architecture as a dimer connected by ERH-2. The m7G capped transcripts interact with the IFE-3, possibly there is redundancy with IFE-1. The presence of a 5'P binding activity in PID-3 may reflect 5' end processing of transcripts bound by PETISCO.

(B) ERH-2 serves as an anchor for PID-1 or TOST-1 driving PETISCO function either towards 21U RNA biogenesis or its embryonic survival function.

(C) PID-1:PETISCO serves as a platform for 21U RNA precursor processing, possibly both at the 5' and 3' end. The nuclease responsible former has not yet been identified. PARN-1 may cleave 21U precursor on or after PETISCO. 21U RNAs are methylated by HENN-1 after PARN-1 activity.

(D) TOST-1:PETISCO interacting with SL1 snRNA. Leakage of SL1 interacting PETISCO into the PID-1:PETISCO explains SL1-derived 21U RNA. The SL1 RNA may be used to process core histone mRNAs in substitution of mammalian U7 snRNA. Processed histone mRNAs can proceed to translation and ensure embryonic survival

4.4. IFE-3, a 5' end Cap binding module of PETISCO

4.4.1. C. elegans eIF4E orthologues are divergent and specialized

IFE-3 is one of the five *C. elegans* orthologues of eukaryotic initiation factor 4E (eIF4E), a protein superfamily whose elements typically function as translation initiator factors, by binding 5' Cap ends of mRNAs (reviewed in Mamane et al., 2004 and Piserà et al., 2018). During translation, eIF4E is part of the eIF4F protein complex which bridges the two ends of an mRNA, a process required for polysome assembly and translation. In our model organism, eIF4E orthologues have diverged and acquired additional functions (Amiri et al., 2001; Mangio et al., 2015; Syntichaki et al., 2007). One clear separation between the various *C. elegans* eIF4E proteins is their expression pattern: in adults IFE-2 and IFE-4 are soma specific, while IFE-1, IFE-3 and IFE-5 are restricted to the germline (Amiri et al., 2001).

Little is known of the exact function of the five eIF4E proteins of *C. elegans*. IFE-4 selectively binds m7G 5'Caps (Keiper et al., 2000), it is expressed in neurons, muscles and the vulva and moderate phenotypes in these tissues have been reported for *ife-4* mutants (Dinkova et al., 2005). In fact, this orthologue is more closely related to the mammalian 4EHP (Keiper et al., 2000), a protein that negatively regulates translation by competing with eIF4F (Morita et al., 2012), and is also reported to facilitate translation inhibition by miRISC (Chapat et al., 2017). IFE-2 is the only somatic eIF4E that can bind both TMG and m7G Caps and is essential for viability (Jankowska-Anyszka et al., 1998; Keiper et al., 2000). In the soma, its depletion leads to longer life spans, which is due to its regulatory function in protein translation and stress response (Rieckher et al., 2018; Syntichaki et al., 2007). Like IFE-2, IFE-1 and IFE-5 are non-selective TMG/m7G binders (Keiper et al., 2000; Miyoshi et al., 2002; Stachelska et al., 2002), with the difference that these are germline specific orthologues (Amiri et al., 2001). No functions have yet been attributed to IFE-5, and *ife-5* has been previously screened as a candidate for 21U RNA biogenesis, with no effects, and all null alleles were viable (de Albuquerque et al., 2014 - data not shown). Conversely, IFE-1 is reported to have key functions in the germline: *ife-1* individuals are sub-fertile and have decreased oocyte production and defective spermatogenesis (Amiri et al., 2001; Henderson et al., 2009). These defects are likely derived from the fact that IFE-1 is required for the translation of maternal mRNAs (Henderson et al., 2009), such as *glp-1*and *oma-1*, which are essential in these processes (Detwiler et al., 2001; Hansen et al., 2004). The worm orthologue of eIF4G, IFG-1, is also required for proper oocyte maturation (Contreras et al., 2008), evidencing a partnership between these two proteins in translation.

IFE-3 is the closest orthologue to the human eIF4E (Joshi et al., 2005) and like the latter, it selectively binds m7G 5' end Caps (Keiper et al., 2000; Miyoshi et al., 2002). In *C. elegans*, m7G is a mark of non-trans-spliced transcripts, which constitute only up to 30% of the transcripts in

this organism (Allen et al., 2011; Gu et al., 2012). From our IP-LFQP experiments we find no evidence for the association of IFE-3 with ribosomes, besides the observation of very low enrichments for two ribosomal proteins large subunits (Fig. R5). If the main function of this protein would be equivalent to that of its human homologue, one would expect to find it in the context of the translation complex eIF4F, i.e., interacting with eIF4G, named IFG-1 in *C. elegans*. Although this interaction has previously been reported *in vitro* (Peter et al., 2015), we do not find IFG-1 to interact with IFE-3 (Fig. R5A and E). Instead, we find IFE-3 associated with different protein complexes with no apparent connection besides the fact that these are RNA processing complexes: PETISCO, the SMN complex and the repressive Dhh1 complex, or CGH-1 complex in *C. elegans* (Boag et al., 2008). These complexes represent a non-PETISCO-associated pool of IFE-3, as evidenced by the fact that we do not find these complexes enriched in the IP-LFQP assays of the remaining PETISCO elements (Fig. R5E). Next, I will discuss the different possible functions of IFE-3 and how these may crossover.

4.4.2. IFE-3 is a partner of a translation-repressive machinery

The CGH-1 complex is an RNP complex that, besides CGH-1, includes CAR-1, ATX-2, PAB-1, CEY-2 to -4, and is predicted to include orthologues of enhancer of decapping EDC-3, eIF4E and its binder eIF4E-T (IFET-1 in *C. elegans*)(Boag et al., 2008). With exception of ATX-2 and PAB-1, we find all of these components moderately enriched in our IFE-3 IP-LQMS assay (Fig. R5A). Of note, the *C. elegans* eIF4E and eIF4E-T orthologues in this complex have not yet been determined, although IFE-3 and IFET-1 are natural candidates. Traditionally, this complex is associated with translation repression by mRNA decapping (Coller and Parker, 2005). In this process, decapping by DCAP-1 and -2 enables degradation by the 5'-to-3' exonuclease XRN-1 and thus translational repression (Parker and Sheth, 2007). As IFE-3 interacts with the 5' Capped 21U precursor RNAs, the IFE-3:IFET-1 interaction may be the result of IFE-3 recruiting decapping machinery for the maturation of the 5' end of the 21U precursor. In this scenario, DCAP-1/-2 and XRN-1 may be the enzymes that process the m7G-RNU$(N)_{20}$ into the mature PO_4-U$(N)_{20}$ (Fig. D2). Indeed, in the course of our work, we found RNAi depletion of these enzymes to lead to a moderate increase in fluorescence of the 21U RNA sensor (data not shown). However, we found no evidence for an interaction between the direct decapping machinery and IFE-3 (Fig. R5A) and the increase in expression may be an indirect effect caused by stabilization of the 21U sensor transcript. A better and more direct approach would be performing smRNAseq or northern blot for 21U precursors in *dcap-1/2* or *xrn-1* mutant strains. Quantitative, or length changes of precursors in these strains would be a good indicator of a role for this machinery in 21U RNA biogenesis.

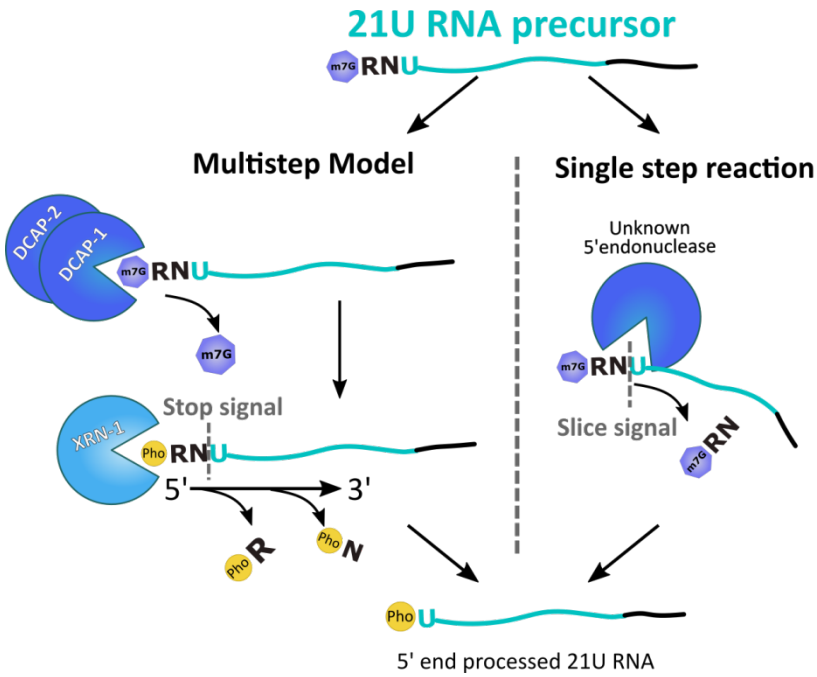

Figure D2 – Schematic representation of 21U RNA precursor 5'end processing models. In a multistep model the 5'end of the 21U RNA precursor first has to be decapped, e.g. by DCAP-1/2 complex; Followed by a 5' to 3' exonuclease, e.g. XRN-1, that is signaled to stop before the 5'end U of the 21U RNA. In a single step reaction an unknown endonuclease slices the precursor just before the 5'end U. A mixed model of decapping followed by an endonuclease is also possible. Either signal may be due to PETISCO interaction

Opposite to a role in 21U precursor decapping is the fact that the CGH-1 complex acts as a transcript destabilizer in association with PATR-1 (Sheth and Parker, 2003), which in *C. elegans* only happens in the soma (Boag et al., 2008). In the germline, this complex stabilizes and binds maternal mRNAs, such as *oma-1* and *pos-1*, and inhibits their translation until they are required at oogenesis and embryogenesis (Boag et al., 2008). This process includes the master germline regulator GLD-1 (Scheckel et al., 2012), which we also find enriched in our IFE-3 IP-LFQP data (Fig. R5A). IFE-3 together with IFET-1 may thus act as translation inhibitor within this storage RNP (Fig. D3). In support of this hypothesis, *ife-3* RNAi phenotypes in embryos differ from the depletion of the remaining PETISCO subunits. In all of these we find defects in chromosome segregation, but *ife-3* is the only factor that reveals spindle defects (Fig. R20B). Interestingly, IFET-1 was previously known as spindle orientation defective 2 (SPN-2) for causing spindle defects (Li et al., 2009). Also, *car-1* RNAi has previously been shown to cause cytokinesis defects and the formation of multi-nuclei cells (Audhya et al., 2005; Boag et al., 2005; Squirrell et al., 2006), phenocopying *ife-3* RNAi (Fig. R20A). The subcellular localization of CGH-1 complex subunits also coincides with that of IFE-3, which would be expected in case of a partnership. These proteins localize to P-granules from the distal germline to embryos (Sengupta et al., 2013), and IFE-3 is the only PETISCO subunit to show this expression pattern (Fig. R4). Conceivably, IFE-3 may have a dual role on facilitating maternal inheritance of 21U RNA precursor and mRNA RNPs.

In order to probe if *ife-3* plays a role in these germline processes, *in situ* hybridization of germline syncytia could lead to quick preliminary results. In such an experiment one would probe known CGH-1 regulated transcripts, such as *pos-1* (Boag et al., 2008), and if these would be

erratically distributed, or destabilized, this would be a good indicator of a role of IFE-3 in transcript storage. iCLIP of IFE-3 combined with ribosome profiling (Ingolia et al., 2009) may also give us an anticorrelation between translation and IFE-3 binding, which would also support this hypothesis. Looking at 21U RNA levels CGH-1 complex mutants may also show a connection between these pathways.

4.4.3. IFE-3 binds Splice Leader 1 snRNA and is part of the SMN complex

The SMN complex is responsible for processing snRNAs into snRNPs, including the snRNA uploading of the Sm ring, that allows further downstream functions (reviewed in Battle et al., 2006). Nematodes have widespread TMG 5' end Caps on translated transcripts (Allen et al., 2011), whereas TMG caps in other animals act as a nuclear import signal (Fischer and Lührmann, 1990; Plessel et al., 1994). Thus, nematodes cannot singularly depend on this signal for nuclear shuttling of snRNAs. Moreover, in comparison with other animals, this *phylum* is missing orthologues for key subunits of this complex, including Gemin4, Gemin5 and UNRIP (Kroiss et al., 2008). These facts signal that these organisms are likely to have developed a snRNA processing mechanism that differs from other animals.

In our studies, we found the Sm ring protein SNR-4 and all the nematode subunits of the SMN complex, with the exception of Gemin7, enriched in IFE-3 IPs (Fig. R5A and E), supporting a function for this protein in the SMN complex. To our knowledge, eIF4E orthologues have not been reported to be part of this complex. Still, 5' end cap binding has been reported as one of the key functions of Gemin5 (Bradrick and Gromeier, 2009; Tang et al., 2016b; Xu et al., 2016), which is absent in nematodes. Gemin5 is credited with the recognition of snRNA precursors and their delivery to the SMN complex (Battle et al., 2006b), conceivably, IFE-3 could have taken over this task in the nematode. Like Gemin5, IFE-3 can interact with the SMN complex and recognize m7G, a marker of non-trans-spliced transcripts such as snRNAs. Indeed, we find IFE-3 interacting with capped SL1 snRNAs (Fig. R15D), although the current setup of our experiments do not allow us to determine the maturation state of these RNAs. We do not observe other snRNAs to be significantly enriched in our RIPseq assay (Fig. R15D). Still, the absence of other snRNAs in our observations may be related with the fact that, in contrast to other snRNAs, splice leader RNAs are consumed upon activity and continuously produced (Philippe et al., 2017), i.e. the proportion of SMN complex associated with splice leader RNAs might be considerably higher than any other snRNA. In support of the Gemin5-like IFE-3 function is the fact that the only studied Gemin mutant of *C. elegans* has a striking resemblance to the *ife-3* phenotype. MEL-46 (Gemin3) depletion leads to the alternation between the Mog and Mel phenotypes in these animals (Minasaki et al., 2009). Furthermore, four out of the five registered *mog* genes are splicing factors (Puoti and Kimble, 1999, 2000; Zanetti et al., 2011). The

Mog phenotype is due to changes in the splicing of sex determination machinery transcripts, which hinders the normal spermatogenesis to oogenesis transition in the *C. elegans* adult (Zanetti et al., 2011). Altogether, these point towards a function of IFE-3 in snRNA homeostasis and consequently splicing homeostasis.

While indicative, the mentioned parallels between snRNA and IFE-3 biology do not settle the matter, and further experimental evidence needs to be acquired to confirm this hypothesis (Fig. D3). Performing iCLIP with IFE-3 should further elucidate which RNA species this protein interacts with, including low abundance and transient ones. Total RNAseq of *ife-3* mutants could also reveal the extent by which snRNA levels are affected and what mis-spliced events, if any, occur due to the ablation of this protein. Such data would strongly support that IFE-3 is part of the processing machinery of snRNAs.

4.4.4. IFE-3 is a key component for 21U RNA biogenesis

In our IFE-3 RIPseq experiments we only find enriched transcripts upon treatment of the RNA with RppH, which removes the 5'end Cap and allows for the cloning and sequencing of capped RNA species (Almeida et al., 2019c; Neri et al., 2017). This is a strong indicator that IFE-3 is binding the m7G Caps of 21U RNA precursors and splice leader RNAs. Interestingly, we do not find the same enrichments for 21U RNA precursors in our PID-3 RIP (Fig. R14B). Nonetheless, we demonstrate that both these proteins are required for the presence of 21U RNAs, albeit *ife-3* mutation has a lesser effect than *pid-3* mutation (Fig. R10). This may indicate that the presence of a 21U RNA precursor within PETISCO is transient and that the majority of the registered 21U precursors interacting IFE-3 are outside of the PETISCO context. In our IFE-3 IP-LFQP analysis we clearly sample multiple pools of this protein, including three different complexes and possibly free IFE-3.

In 21U RNA biogenesis the selectivity of IFE-3 for m7G could help this protein to select

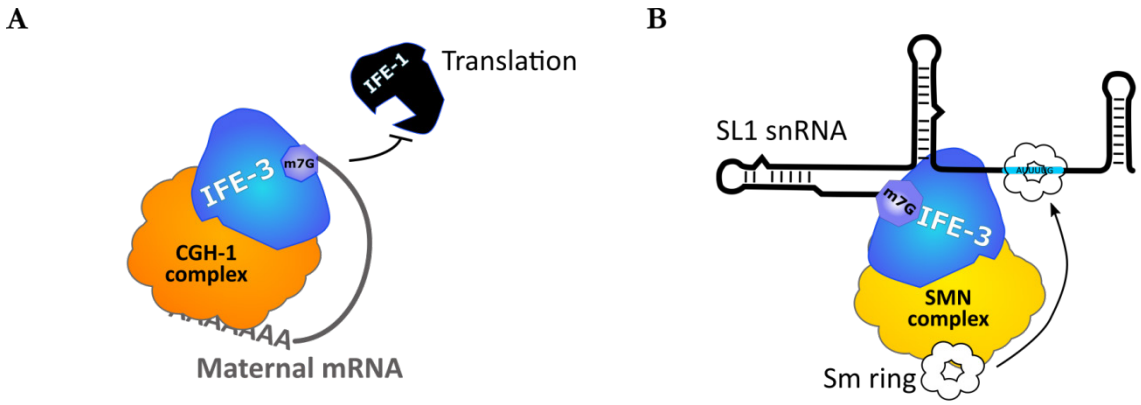

Figure D3 – Schematic representation of non-PETISCO functions of IFE-3.
(A) IFE-3 interacts with the translation inhibiting CGH-1 complex. The CGH-1 complex binds maternally provided mRNAs, e.g. *oma-1* and inhibits their translation until they reach the embryos. IFE-3 may compete with other eIF4E proteins, e.g. IFE-1.
(B) IFE-3 interacts with the SMN complex. The SMN complex ads the Sm ring to snRNAs. IFE-3 may recognize the 5' end cap of snRNAs, e.g. SL1 RNA, and act as Gemin5, which is missing in *C. elegans*.

21U precursors of both type I and II, while distinguishing them from the TMG Capped mRNA. IFE-3 could bring them to PETISCO where their 5'end is processed. This could explain the occurrence of SL1 RNA-derived type II 21U RNAs (Gu et al., 2012), whose origin would be the cross talk between the snRNA and 21U RNA functions of IFE-3. Furthermore, such a mode of action would also account for the origin of the remaining type II 21U RNAs, suggested to be early termination transcripts (Gu et al., 2012). IFE-3 would seize these m7G-carrying transcripts and channel them to 21U RNA biogenesis, significantly increasing the 21U RNA repertoire of this species. Assuming these premises, IFE-3 would be in the epicentre of 21U RNA biogenesis and perhaps one would expect the same downregulation severity in total 21U RNA levels as in other PETISCO mutants. Conceivably, in *ife-3* individuals the processing can still occur but its efficiency may be significantly impaired, as knock down of *ife-3* in 21U sensor individuals would suggest (Fig. R8). Also, there may be redundancy between IFE-3 and other eIF4E orthologues in the germline. We do observe IFE-1 enriched in both our TOFU-6 and IFE-3 IP-LFQP assays (Fig. R5A and C), which as a non-selective TMG/m7G binder may act as IFE-3 replacement in PETISCO in *ife-3* mutants. The fact that we find IFE-1 while using IFE-3 as bait in IP-LFQP can be explained in a dimer model of PETISCO (Fig. D1), where each arm could bind a different eIF4E orthologue.

Redundancy between IFE-3 and IFE-1 may be addressed by the sequencing of double mutants for these factors. The 21U RNA population in such a strain should resemble that of *pid-3* individuals. The ability of IFE-1 to bind PETISCO may be related to the eIF4E binding motif of TOFU-6 (Fig. R2A) and ablating this motif may have an equivalent effect to an *ife-1/3* double mutant. Including IFE-1 in the studies of *in vitro* reconstitution of PETISCO will also address the ability for this protein to take over the task of IFE-3. The latter's exact task, beyond binding the 5'Cap of 21U precursors, remains unanswered as it can be either bringing 21U RNA molecules to PETISCO or may be a 21U precursor receiver. An interesting experiment, although challenging, would be to change the cap nature of 21U RNAs, in a semi-synthetic system. In this setup one would change the affinity of IFE-3 from m7G to TMG by point mutations, in a similar manner to previous studies (Miyoshi et al., 2002). This would be followed by injection of TMG cap 21U RNA precursors in the worm gonad and observation of the silencing state of a reporter, e.g. the 21U sensor. Constructing such system would allow the user to create artificial 21U RNAs and test the requirements for a functional 21U RNA precursor and its transition into maturity.

4.5. PID-1 and TOST-1 define the functions of PETISCO

4.5.1. PID-1 and TOST-1 compete for available PETISCO

In the present study we find PETISCO to be required for two functions: 21U RNA biogenesis and embryonic survival. Our current model is that these functions are guided by two proteins: PID-1 and TOST-1 (Fig. D1).

Taken the experimental evidence we have obtained, it is unlikely that these two proteins coexist in PETISCO. TOST-1 is a strong interactor of all PETISCO subunits, as pointed by our IP-LFQP assays (Fig. R5). Yet, we find no TOST-1 enrichment when using PID-1as bait (Fig. R1). A similar observation is made by Zeng et al., 2019 in the reverse experiment. In their studies the authors search for interactors using mass spectrometry and TOST-1 as bait. Interestingly, they do not report finding PID-1, but they do identify the same PETISCO subunits we have. However, the technique differs from the one used here: they performed experiments on worm extracts of non-synchronized populations and subsequent proteomic analysis was non-quantitative. We cannot exclude that these differences may account for PID-1 absence from TOST-1 IPs. Nevertheless, Zeng et al., 2019 do find PID-1 in IPs of other PETISCO subunits, making it rather unlikely that technical differences lead to absence of PID-1 from their TOST-1 IPs.

The phenotypic profiles of *tost-1* and *pid-1* depletion also serve as indicators of alternate functions of these proteins. RNAi depletion of *pid-1* leads to the upregulation of the 21U sensor while the depletion of *tost-1* leads to a downregulation of the same transgene (Fig. R8). We do not find a significant increase in the expression of 21U RNAs in *tost-1* that would explain such behaviour (Fig. R10). Perhaps a small increase in 21U RNAs may lead to a much larger response from the 22G RNA pathway and consequently a more efficient silencing. We excluded the hypothesis by which the *tost-1* effect on the 21U sensor, which codes for a histone (Ashe et al., 2012), is related to the effects of *tost-1* on histone mRNA stability (Fig. R18 and R19). The transgene used in the imaging of chromosome segregation is identical in all features to the 21U sensor (Fig. R8A), with the exception that it is not recognized by 21U RNAs. Thus, an effect on the 21U sensor by means of the histone coding sequence should lead to a decrease of expression of the other transgene, which we do not observe.

The opposite phenotypes registered by the depletion of *pid-1* and *tost-1* by RNAi are confirmed in the genetic mutants for these genes. *tost-1* mutant animals have normal levels of 21U RNAs (Fig. R10 and R13), have an upregulated levels of SL1 RNA (Fig. R15A to C), a drastic downregulation of core histone mRNAs (Fig. R18 and R19) and are Mel (Table R1 and Fig. R11B). On the other hand, *pid-1* mutant animals have a complete loss of 21U RNA (Fig.

R10), upregulation of core histone mRNAs (Fig. R18), but none of the remaining phenotypes. The sum of these phenotypes corresponds to the phenotypes of PETISCO mutants. The opposing effects of *tost-1* and *pid-1* in both the 21U sensor assay and the core histone mRNA regulation indicate that these two proteins may compete with each other for PETISCO binding. Indeed, we find a small amount (<1%) of *pid-1* individuals to be Mog (Fig. R11A), which may correspond to an indirect effect, such as the sequestering of IFE-3 by larger amounts of available PETISCO.

A competition model (Fig. D1) would require that these two proteins target PETISCO in a similar manner, which is indeed the case. PID-1 and TOST-1 interact with ERH-2 via a common amino acid motif To note, this motif may be common to other enhancer of rudimentary homologue interactors, as mutating it also disrupts the interaction of TOST-1 with ERH-1 (Fig. R7D). The assignment of PETISCO to PID-1 or TOST-1 may be regulated by protein abundance, their affinity to ERH-2 or simple spatial separation, although these are not necessarily mutually exclusive. In favour of the latter, we do not observe PID-1 expression in late oogenesis or embryogenesis, while the PETISCO subunits are expressed in all these stages (Fig. R4D). Moreover, in our 21U sensor/*tost-1* RNAi experiments we occasionally observed the PID-1::mCherry transgene expression in later stages of oogenesis (data not shown). While a more systematic study should be performed, this may indicate a stabilization of PID-1 in the absence of TOST-1.

Altogether, our current knowledge points to two distinct PETISCO complexes with different functions: TOST-1:PETISCO and PID-1:PETISCO. The reconstruction of these two PETISCO complexes *in vitro* may elucidate and confirm the exclusivity and how PID-1 and TOST-1 interact with PETISCO. Unfortunately, we were not able to create a functional TOST-1 transgene or antibody, but further attempts should be encouraged. Such tools would allow us to probe TOST-1 specific interacting partners, which may elucidate its exact function, as well as for the spatial separation in relation to PID-1.

4.5.2. The framework of PID-1 and TOST-1 in PETISCO

The exact function of PID-1 or TOST-1 in PETISCO remains an open question. PID-1 and TOST-1 affect the stability of different RNA species in the context of PETISCO and seemingly direct this complex into different tasks. Are these factors active carriers of these RNA species into PETISCO or are they adaptor proteins that guide this complex following RNA binding? *pid-1* individuals show decreased total amounts of 21U RNA precursors (Fig. R14A). This could be the consequence of these RNA species not being brought to the complex, suggesting a role of PID-1 in precursor acquisition by PETISCO. Alternatively, in absence of PID-1, PETISCO may no longer be able to absorb the flow of 21U RNA precursors resulting in

their degradation. We find no putative RNA binding motifs in either PID-1 or TOST-1 that would be compatible with RNA transportation to PETISCO. Nonetheless, the unstructured regions of these proteins may still bind specific RNA species. iCLIP and *in vitro* experiments using both TOST-1 and PID-1 should be able to determine the selectivity and ability of these proteins to interact with RNA. These two proteins could also ensure different PETISCO subcellular localizations to correctly accept different RNA molecules. Thus, assays on the subcellular localization of PETISCO in *pid-1* and *tost-1* mutants may help define their function.

If PID-1 and TOST-1 are determining the functional output of PETISCO, one would expect that these are either the catalysts of the designated function or serve as adaptors to that purpose. Their small size (19KDa and 16.2KDa) and the lack of detectable domains deems unlikely that these proteins act alone. As adaptors, we would expect to find their specific partners in our IP-LFQP assays. Besides the presence of TOST-1, we do not find any systematic differences between IP-LFQP of PID-1 and the PETISCO subunits (Fig. R1 and R5). Possibly, these are transient interactions that are not easily detectable in this assay. In an effort to enrich our assay for TOST-1:PETISCO we performed PID-3 IP-LFQP in a *pid-1* mutant background, but observed no differences (Fig. R5G). In this context, performing IP-LFQP on PETISCO with the addition of a chemical crosslinking step may resolve this issue. Furthermore, the latter assay may even bring forward other components of the 21U RNA pathway, such as the unknown nuclease that catalyses the 5' end maturation of 21U RNA precursors.

4.6. PETISCO is a central player in multiple pathways

4.6.1. PETISCO as a 21U RNA precursor processing platform

Under the current understanding of the sequence of events of 21U RNA precursor processing, these small RNAs should be expressed, processed at their 5'end and finally at their 3'end (Izumi and Tomari, 2014) before loading into PRG-1 (Fig. I5). In this study, not only do we find that this complex contains the potential requirements for a 21U RNA precursor processing platform, but also, that all elements of PETISCO are required for the presence of 21U RNAs.

PETISCO contains a 5' Cap binding protein in IFE-3 and a potential 5'PO_4 binding protein in PID-3, due to its MIDAS domain. We find precursors to be bound to IFE-3, but we find no enrichments in PID-3 RIPseq assays (Fig. R15D). If PETISCO is a platform for the 5' end maturation of 21U RNA precursors, one could expect to find 21U RNA precursor intermediates in either our RIPseq or mutant smRNAseq data, which is not the case. These intermediaries could be 21U RNAs with extended 3' end in PID-3 RIPs, similar to *parn-1* mutants (Tang et al., 2016a), which would allow us to place 5' end processing upstream of the 3' end

processing. Other intermediates we would hope to find include: uncapped full length 21U RNA precursors, which would indicate that these RNAs are uncapped before the removal of the two nucleotides upstream of the consensus uracil (Fig. D2). As we do not detect either of these intermediates, or found a putative nuclease for 5' end processing, two models remain that can explain 21U precursor processing. The first would be a multistep model, where decapping is followed by recruitment of a 5' to 3' exonuclease that removes two nucleotides of the small RNA. Such a nuclease would have to be restrained from hydrolysing the whole RNA. PETISCO could act in such model by recruiting the required machinery and protecting the 21U RNA from extended degradation via binding of the exposed 5'PO_4 to the PID-3 MIDAS domain (Fig. D1 and D2). The RRM domains of TOFU-6 and PID-3, beyond being involved in protein-protein interactions, could bind and protect the small RNA as well. A second model would be an endonuclease that slices the 21U RNA precursor 5'end, removing both the two upstream nucleotides and the m7G at once. Under this model, PETISCO may be required to determine the cut site of this nuclease and stabilize the intermediate until it is channelled to the next step of the pathway (Fig. D1 and D2). Nonetheless, further experiments are required to distinguish between the proposed modes of PETISCO action. The previously mentioned iCLIP experiments may reveal the presence of the mentioned RNA intermediates, allowing for a step wise view on 21U RNA precursor processing. *In vitro* reconstruction of PETISCO may also elucidate the processing steps, but this approach is limited by the discovery of the enzymatic machinery responsible for this process.

Notably, in our IFE-3 RIPseq experiments, we find multiple precursors with intact 5'ends but deprived of the 3'end tail (data not shown). This observation may be due to non-specific nuclease activity during our experimental procedure. On the other hand, it might indicate that the 3'end processing by PARN-1 is not necessarily downstream of the 5'end processing. Conceivably, both processing steps could occur simultaneously in PETISCO, which would then stabilize the mature RNA until its 2' O-methylation and/or PRG-1 loading. Thus, 21U RNA precursor 3' end processing should be included in future 21U RNA:PETISCO studies.

4.6.2. PETISCO is a protein complex essential for embryogenesis

The subject of our studies was the 21U RNA pathway, thus it was surprising to find that PETISCO is required for embryonic survival. As no other 21U RNA factors have been reported to be essential, this is a strong indication for the requirement of this complex beyond 21U RNAs. While we were unable to pinpoint the exact mechanism by which PETISCO ablation leads to embryonic lethality, we have uncovered possible sources for this effect.

The embryonic arrest of TOST-1/PETISCO mutants is likely due to the widespread occurrence on chromatin bridges during mitosis (Fig. R20A). These are generally associated with

aneuploidy, and DNA damage (reviewed in Fernández-Casañas et al., 2018). Minasaki et al., 2009 previously described the cell cycle delays and embryonic arrest in *tofu-6* mutants. The authors find that these delays are due to an extension of the interphase and that it can be rescued by simultaneous depletion of the ATM and ATR checkpoints. They further suggest that this protein is required for the completion of DNA synthesis and its depletion leads to DNA damage and thus embryonic arrest. Minasaki et al., 2009 describe the same embryonic defects we observe in the remaining PETISCO mutants (Fig. R20A), although in the present study, we did not perform measurements of mitotic timing. Nonetheless, we do not find PETISCO to be associated with any of the known DNA synthesis or checkpoint machinery (Fig. R5). Possibly, these effects are secondary to the defects in SL1 homeostasis caused by PETISCO ablation (Fig. R15B and C). Still, we find the Mel defect to be irreversible just after two hours of *tost-1* inactivity (Fig. R13A). The determined timeframe implies immediate effects, which we do not observe in splice leader RNAs (Fig. R17C) and thus their accumulation may be a secondary effect. I propose that the mitotic defects are associated with the severe depletion of core histone mRNAs we find in PETISCO/*tost-1* mutants (Fig. R18 and R19). These effects occur within the expected timeframe (Fig. R19) and would be sufficient to explain the lethality associated with PETISCO mutants.

In *C. elegans,* histone depletion leads to defects in chromosome condensation and embryonic arrest (Pettitt et al., 2002). Chromatin condensation defects can lead to the formation of chromatin bridges (Kodama et al., 2002), the activation of Chk1 kinase and a delayed mitosis (Fasulo et al., 2012), while embryonic arrest is believed to be reached when the available histones are consumed and unable to sustain genome integrity (Pettitt et al., 2002). These phenotypes are consistent with our observations with PETISCO mutants (Fig. R20A), which possibly have a depletion of histone proteins. This hypothesis is supported by the steep decrease in histone mRNA (Fig. R18 an R19), together with the histone-depletion-like phenotypes in these mutants. We did not address this issue directly and further experiments need to be carried out in this direction. Performing immunostainings or western blots probing for core histones, in both PETISCO mutant adults and embryos, should be sufficient evidence of this depletion. If confirmed, one could go a step further and attempt to rescue the embryonic arrest by the ectopic expression of histones in a PETISCO-independent manner.

4.6.3. A blueprint for PETISCO essential function

Out work leaves a significant question unanswered: What is the source of the core histone mRNA depletion? The fact that PETISCO is required for the biogenesis of small RNA molecules and proper chromosome segregation, leads us to consider the CSR-1 pathway. This essential small RNA pathway is required both for proper chromosome segregation and core histone mRNA stability (Avgousti et al., 2012; Claycomb et al., 2009; Gerson-Gurwitz et al.,

2016). However, Gerson-Gurwitz et al., 2016 have shown that these may be indirect effects of the miss-expression of maternal transcripts, including multiple cell cycle-related machinery genes. Moreover, *csr-1* mutants are embryonic lethal with the few escapers (<5%) becoming fully sterile, which differs from our PETISCO/*tost-1* mutant observations. On the other hand, PETISCO phenotypes highly resemble those of *cdl-1* depletion, the stem-loop binding protein (SLBP) orthologue of *C. elegans*. *cdl-1* individuals have destabilized core histone mRNAs and are depleted of histone proteins (Avgousti et al., 2012; Pettitt et al., 2002). Like PETISCO mutants, these also have widespread occurrence of chromatin bridges in early cell divisions and embryos arrest before the 100 cell stage (Kodama et al., 2002; Pettitt et al., 2002). Overall, *cdl-1* and PETISCO depletion show a striking resemblance, yet, so does *csr-1*. Interestingly, the *cdl-1* embryonic defect is marked by the loss of CENP-A (HCP-3 in nematodes) along the chromosome, while *csr-1* is not (Claycomb et al., 2009; Gerson-Gurwitz et al., 2016; van Wolfswinkel et al., 2009). Hence, CENP-A immunostaining in PETISCO mutants may be a good approach to understand which of these two pathways may be under the influence of PETISCO.

It is noteworthy that both CSR-1 and CDL-1 are reported to play a role in core histone mRNA 3'end processing (Avgousti et al., 2012; Marzluff et al., 2008), a process yet to be understood in nematodes. In most animals, the 3'end of core histones is processed in nuclear histone bodies, the process requires multiple factors including U7 snRNA, LSm10/11 and ZFP100 (Marzluff et al., 2008). None of the mentioned factors has known *C. elegans* orthologues, indicating that these animals have a different mechanism for this process (López and Samuelsson, 2008). Furthermore, nematodes lack coilin, the common marker for histone bodies and Cajal bodies, and it is unknown if these animals contain either of these structures (Lee et al., 2012b; Machyna et al., 2015; Nizami et al., 2010). PETISCO interacts with snRNAs, core histone mRNAs and is required for the stability of the latter. Also, effects of the depletion of PETISCO are similar to the depletion of histone mRNA 3'end processing machinery. Together these parallels raise the hypothesis that PETISCO may have taken over core histone mRNA processing in *C. elegans*. In this hypothetical scenario, PETISCO may use SL1 snRNAs in substitution of the absent U7 snRNA and possibly use the same nucleases to process both 21U RNA and the core histone mRNA 3' end (Fig. D1). Thus, it would be interesting to look into histone transcripts in mutants for the 5' and 3' end 21U RNA precursor processing nucleases, once they are identified. Changed ratios or levels of unprocessed histone mRNAs may also serve as an indicator that PETISCO is acting on these transcripts at this step.

Another scenario compatible with our observations would be that PETISCO acts on RNA Pol II transcription termination of 3'end stem-loop-carrying transcripts. These transcripts include snRNAs and core histone mRNAs (Kasper et al., 2014; Keall et al., 2007; Weng et al.,

2019). 21U RNA precursor expression shares machinery and is believed to have evolved from snRNA expression (Beltran et al., 2019; Weng et al., 2019). Perhaps, it may also share termination machinery explaining the overlap between these pathways.

A major limitation to the two above hypotheses is the fact that we did not manage to prove a presence of PETISCO in the nucleus, where these mechanisms are expected to take place. It is noteworthy that in embryos, we found that PETISCO transgenes seem to flux into the nucleus at the onset of cell division (data not shown). However, when testing other unrelated transgenes, such as PGL-1::GFP, we saw the same nuclear shift and thus, we could not confidently exclude that this observation was an artefact. Zeng et al., 2019 also find PETISCO transgenes to localize in the nucleus in both embryos and the adult germline and suggest that this complex shuttles between the nucleus and the cytoplasm. Possibly, the bulk of PETISCO is cytoplasmic and we may require an enrichment procedure to observe its nuclear fraction. A possible approach would be the use of leptomycin B, which blocks exportin 1 and its nuclear export functions (Kudo et al., 1999). These experiments would be important to clear the discrepancy between our two experimental set ups. Furthermore confirming a nuclear fraction of PETISCO would open the door to a direct role of this complex on transcription.

4.7. The Evolution of PETISCO

PETISCO only arose in its current architecture within the clade V of nematodes, while the ancestral elements of this complex are already present in the last common ancestor between Clade III and V (Fig. R21). Interestingly, a proto-PETISCO seems to have surged relatively quickly in the evolutionary timeline as TOST-1, PID-3 and TOFU-6 orthologues are all strictly present since the branching of Clade III nematodes (Fig. R21). While ERH-2 only arose in the *Caenorhabditis* genus it is possible that its paralogue ERH-1 would have been originally performing its task in the complex. Remarkably, ERH-1 does have the ability to interact with TOST-1 (Fig. R7D), possibly a vestigial property signalling an ancestor link between them. As TOST-1, PID-3 and TOFU-6 seemingly appear together, it may be an indication of strong positive selection for a PETISCO-like machinery. Possibly, a TOST-1:PETISCO with ERH-1 mediation may already be performing its essential task in earlier ancestors, however the exact moment in which PETISCO took over an essential task in the evolutionary timeline is not known. Conceivably, within these clades one will find PETISCO versions that are not essential or are in transition of becoming so. In that line it would be interesting to expand PETISCO studies to other nematodes in order to pinpoint its primordial function and strengthen our knowledge of its function in *C. elegans*.

21U RNAs are only present in Plectids and Clade V of nematodes (Beltran et al., 2019; Sarkies et al., 2015). Thus, it is unlikely that PETISCO was already performing 21U RNA

processing functions in the common ancestors of clade III, IV and V. PRDE-1 is only present on the latter clade (Fig. R21), coinciding with the prevalence of 21U RNAs (Beltran et al., 2019; Sarkies et al., 2015). In these nematodes, and likely for the first time, the 21U RNA pathway and PETISCO coexist. Conceivably, PETISCO may play a role in 21U RNA precursor processing outside of the *Caenorhabditis* genus, where PID-1 is not present (Fig. R21). In this scenario, TOST-1:PETISCO would be acting on 21U RNA processing in these species. Testing if PETISCO or TOST-1 are required for 21U RNA biogenesis in these species may answer if the two functions of this complex were early or progressively tangled and give insight into stepwise evolution of this complex. Interestingly, ERH-2 and PID-1 emerge in the *Caenorhabditis* genus simultaneously (Fig. R21), pointing towards an adaptation of PETISCO to the 21U RNA pathway. Perchance, ERH proteins could have been the central gear for such adaptation. Each ERH homologue is capable of interacting with TOST-1, but ERH-1 and ERH-2 are unable to dimerize with each other (Fig. R7A). At the evolutionary moment that this constraint occurred, the evolution of the ERH paralogues became no longer constrained by each other. The liberation of ERH-2 may have allowed for a specialization of PETISCO, which was then no longer constrained by ERH-1 function. The emergence of PID-1 may have further consolidated the function of PETISCO in 21U RNA biogenesis. Such evolutionary steps would be facilitated by the previously mentioned modular character of PETISCO. The exchange of modules between PETISCO tasks allows parallel evolutionary changes and loosens the selection pressure of one function over another.

The co-option of existing machinery into small RNA pathways is an emergent theme in the field. In *Drosophila*, recruitment of specialized orthologues of nuclear export factor 1 are required for transposon transcriptional gene silencing and to export heterochromatic transcripts for piRNA production (Batki et al., 2019; ElMaghraby et al., 2019; Fabry et al., 2019). In nematodes, the snRNA expression protein SNAPc has been co-opted for 21U RNA expression (Beltran et al., 2019; Kasper et al., 2014; Weng et al., 2019). The latter is recruited to the 21U RNA genes by PRDE-1 (Weick et al., 2014; Weng et al., 2019), in similarity to the recruitment of PETISCO by PID-1. Once again the *C. elegans* 21U RNA pathway seems to have refurbished the already available machinery for its own purposes, rather than creating it *de novo*. PETISCO co-option is thus another example of what appears to be a common feature of small RNA pathways as well as the striking molecular plasticity of a cell.

5. Concluding Remarks

The Piwi pathway is a fast evolving, highly adaptive small RNA pathway, whose evolutionary pace mirrors that of the foreign sequences it counteracts. Perhaps due to this fast space, animals rely on a myriad of mechanisms by which these small RNAs are generated. PETISCO is a novel protein complex that stands on the crossroads between 21U RNAs and key house keeping pathways such as histone mRNA and snRNAs.

While we bring to light some of the mechanisms of this complex, a long path remains ahead to fully understand it. How does PETISCO precisely support 21U RNA biogenesis? How does PETISCO sustain histone mRNA levels? How are splice leader RNAs connected to this? These are just some of the open questions that are left from this work. Further studies on PETISCO may not only enlighten the scientific community about the mechanisms of these pathways, but also the mechanism by which they have intertwined and contribute to the general understanding of evolutionary mechanisms.

In our studies we set out to further understand 21U RNA biogenesis. Serendipity brought forward an exciting example of how ancient proteins, like Erh and eIF4E, can be co-opted and used in novel manners. This work is thus another reflection of the evolutionary fluidity of small RNA pathways.

6. References

Aboobaker, A., and Blaxter, M. (2010). The nematode story: Hox gene loss and rapid evolution. Adv. Exp. Med. Biol. *689*, 101–110.

Akkouche, A., Mugat, B., Barckmann, B., Varela-Chavez, C., Li, B., Raffel, R., Pélisson, A., and Chambeyron, S. (2017). Piwi Is Required during Drosophila Embryogenesis to License Dual-Strand piRNA Clusters for Transposon Repression in Adult Ovaries. Mol. Cell *66*, 411-419.e4.

de Albuquerque, B.F.M., Luteijn, M.J., Cordeiro Rodrigues, R.J., van Bergeijk, P., Waaijers, S., Kaaij, L.J.T., Klein, H., Boxem, M., and Ketting, R.F. (2014). PID-1 is a novel factor that operates during 21U-RNA biogenesis in Caenorhabditis elegans. Genes Dev. *28*, 683–688.

Allen, M.A., Hillier, L.W., Waterston, R.H., and Blumenthal, T. (2011). A global analysis of C. elegans trans-splicing. Genome Res. *21*, 255–264.

Almeida, M.V., Dietz, S., Redl, S., Karaulanov, E., Hildebrandt, A., Renz, C., Ulrich, H.D., König, J., Butter, F., and Ketting, R.F. (2018). GTSF-1 is required for formation of a functional RNA-dependent RNA Polymerase complex in Caenorhabditis elegans. EMBO J. *37*, e99325.

Almeida, M.V., de Jesus Domingues, A.M., and Ketting, R.F. (2019a). Maternal and zygotic gene regulatory effects of endogenous RNAi pathways. PLOS Genet. *15*, e1007784.

Almeida, M.V., Andrade-Navarro, M.A., Ketting, R.F., Almeida, M.V., Andrade-Navarro, M.A., and Ketting, R.F. (2019b). Function and Evolution of Nematode RNAi Pathways. Non-Coding RNA *5*, 8.

Almeida, M.V., de Jesus Domingues, A.M., Lukas, H., Mendez-Lago, M., and Ketting, R.F. (2019c). RppH can faithfully replace TAP to allow cloning of 5'-triphosphate carrying small RNAs. MethodsX *6*, 265–272.

Altschul, S. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

Altun, Z.F., and Hall, D.H. (2009). Introduction. WormAtlas doi:10.3908/wormatlas.1.1.

Alva, V., Nam, S.-Z., Söding, J., Lupas, A.N., I., S., S., C., S., C., M., F., A., G., and L.J., G. (2016). The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. Nucleic Acids Res. *44*, W410–W415.

Alvarez-Saavedra, E., and Horvitz, H.R. (2010). Many Families of C. elegans MicroRNAs Are Not Essential for Development or Viability. Curr. Biol. *20*, 367–373.

Ambros, V., and Ruvkun, G. (2018). Recent Molecular Genetic Explorations of Caenorhabditis elegans MicroRNAs. Genetics *209*, 651–673.

Amente, S., Napolitano, G., Licciardo, P., Monti, M., Pucci, P., Lania, L., and Majello, B. (2005). Identification of proteins interacting with the RNAPII FCP1 phosphatase: FCP1 forms a complex with arginine methyltransferase PRMT5 and it is a substrate for PRMT5-mediated methylation. FEBS Lett. *579*, 683–689.

Ameres, S.L., Horwich, M.D., Hung, J.-H., Xu, J., Ghildiyal, M., Weng, Z., and Zamore, P.D. (2010). Target RNA-Directed Trimming and Tailing of Small Silencing RNAs. Science (80-.). *328*, 1534–1539.

Amiri, A., Keiper, B.D., Kawasaki, I., Fan, Y., Kohara, Y., Rhoads, R.E., and Strome, S. (2001). An isoform of eIF4E is a component of germ granules and is required for spermatogenesis in C. elegans. Development *128*, 3899–3912.

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

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

Andrei, M.A., Ingelfinger, D., Heintzmann, R., Achsel, T., Rivera-Pomar, R., and Lührmann, R. (2005). A role for eIF4E and eIF4E-transporter in targeting mRNPs to mammalian processing bodies. RNA *11*, 717–727.

Arai, R., Kukimoto-Niino, M., Uda-Tochio, H., Morita, S., Uchikubo-Kamo, T., Akasaka, R., Etou, Y., Hayashizaki, Y., Kigawa, T., Terada, T., et al. (2005). Crystal structure of an enhancer of rudimentary homolog (ERH) at 2.1 Å resolution. Protein Sci. *14*, 1888–1893.

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

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

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

Audhya, A., Hyndman, F., McLeod, I.X., Maddox, A.S., Yates, J.R., Desai, A., and Oegema, K. (2005). A complex containing the Sm protein CAR-1 and the RNA helicase CGH-1 is required for embryonic cytokinesis in Caenorhabditis elegans. J. Cell Biol. *171*, 267–279.

Avgousti, D.C., Palani, S., Sherman, Y., and Grishok, A. (2012). CSR-1 RNAi pathway positively regulates histone expression in C. elegans. EMBO J. *31*, 3821–3832.

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

Banani, S.F., Lee, H.O., Hyman, A.A., and Rosen, M.K. (2017). Biomolecular condensates: organizers of cellular biochemistry. Nat. Rev. Mol. Cell Biol. *18*, 285–298.

Barnett, D.W., Garrison, E.K., Quinlan, A.R., Stromberg, M.P., and Marth, G.T. (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics *27*, 1691–1692.

Bartel, D.P. (2018). Metazoan MicroRNAs. Cell *173*, 20–51.

Batista, P.J., Ruby, J.G., Claycomb, J.M., Chiang, R., Fahlgren, N., Kasschau, K.D., Chaves, D. a, Gu, W., Vasale, J.J., Duan, S., et al. (2008). PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans. Mol. Cell *31*, 67–78.

Batki, J., Schnabl, J., Wang, J., Handler, D., Andreev, V.I., Stieger, C.E., Novatchkova, M., Lampersberger, L., Kauneckaite, K., Mechtler, K., et al. (2019). The nascent RNA binding

complex SFiNX licenses piRNA-guided heterochromatin formation. BioRxiv 609693.

Battle, D.J., Kasim, M., Yong, J., Lotti, F., Lau, C.-K., Mouaikel, J., Zhang, Z., Han, K., Wan, L., and Dreyfuss, G. (2006a). The SMN complex: an assembly machine for RNPs. Cold Spring Harb. Symp. Quant. Biol. *71*, 313–320.

Battle, D.J., Lau, C.-K., Wan, L., Deng, H., Lotti, F., and Dreyfuss, G. (2006b). The Gemin5 Protein of the SMN Complex Identifies snRNAs. Mol. Cell *23*, 273–279.

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

Beanan, M.J., and Strome, S. (1992). Characterization of a germ-line proliferation mutation in C. elegans. Development *116*, 755–766.

Beltran, T., Barroso, C., Birkle, T.Y., Stevens, L., Schwartz, H.T., Sternberg, P.W., Fradin, H., Gunsalus, K., Piano, F., Sharma, G., et al. (2019). Comparative Epigenomics Reveals that RNA Polymerase II Pausing and Chromatin Domain Organization Control Nematode piRNA Biogenesis. Dev. Cell *48*, 793-810.e6.

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

Billi, A.C. (2014). Endogenous RNAi pathways in C. elegans. WormBook 1–49.

Billi, A.C., Alessi, A.F., Khivansara, V., Han, T., Freeberg, M., Mitani, S., and Kim, J.K. (2012). The Caenorhabditis elegans HEN1 Ortholog, HENN-1, Methylates and Stabilizes Select Subclasses of Germline Small RNAs. PLoS Genet. *8*, e1002617.

Billi, A.C., Freeberg, M.A., Day, A.M., Chun, S.Y., Khivansara, V., and Kim, J.K. (2013). A conserved upstream motif orchestrates autonomous, germline-enriched expression of Caenorhabditis elegans piRNAs. PLoS Genet. *9*, e1003392.

Bluhm, A., Casas-Vila, N., Scheibe, M., and Butter, F. (2016). Reader interactome of epigenetic histone marks in birds. Proteomics *16*, 427–436.

Blumenfeld, A.L., and Jose, A.M. (2016). Reproducible features of small RNAs in *C. elegans* reveal NU RNAs and provide insights into 22G RNAs and 26G RNAs. RNA *22*, 184–192.

Blumenthal, T. (2004). Operons in eukaryotes. Briefings Funct. Genomics Proteomics *3*, 199–211.

Blumenthal, T. (2005). Trans-splicing and operons. WormBook.

Boag, P.R., Nakamura, A., and Blackwell, T.K. (2005). A conserved RNA-protein complex component involved in physiological germline apoptosis regulation in C. elegans. Development *132*, 4975–4986.

Boag, P.R., Atalay, A., Robida, S., Reinke, V., and Blackwell, T.K. (2008). Protection of specific maternal messenger RNAs by the P body protein CGH-1 (Dhh1/RCK) during *Caenorhabditis elegans* oogenesis. J. Cell Biol. *182*, 543–557.

Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M., and Benning, C. (1998). AGO1 defines a novel locus of Arabidopsis controlling leaf development. EMBO J. *17*, 170–180.

Boland, A., Tritschler, F., Heimstädt, S., Izaurralde, E., and Weichenrieder, O. (2010). Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein. EMBO Rep.

11, 522–527.

Bontems, F., Stein, A., Marlow, F., Lyautey, J., Gupta, T., Mullins, M.C., and Dosch, R. (2009). Bucky Ball Organizes Germ Plasm Assembly in Zebrafish. Curr. Biol. *19*, 414–422.

Boxem, M., Maliga, Z., Klitgord, N., Li, N., Lemmens, I., Mana, M., de Lichtervelde, L., Mul, J.D., van de Peut, D., Devos, M., et al. (2008). A Protein Domain-Based Interactome Network for C. elegans Early Embryogenesis. Cell *134*, 534–545.

Bradrick, S.S., and Gromeier, M. (2009). Identification of Gemin5 as a Novel 7-Methylguanosine Cap-Binding Protein. PLoS One *4*, e7030.

Brancati, G., and Großhans, H. (2018). An interplay of miRNA abundance and target site architecture determines miRNA activity and specificity. Nucleic Acids Res. *46*, 3259–3269.

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

Braukmann, F., Jordan, D., and Miska, E. (2017). Artificial and natural RNA interactions between bacteria and *C. elegans*. RNA Biol. *14*, 415–420.

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

Brenner, S. (1974). THE GENETICS OF CAENORHABDITIS ELEGANS. Genetics *77*.

Brenner, J.L., Jasiewicz, K.L., Fahley, A.F., Kemp, B.J., and Abbott, A.L. (2010). Loss of Individual MicroRNAs Causes Mutant Phenotypes in Sensitized Genetic Backgrounds in C. elegans. Curr. Biol. *20*, 1321–1325.

Brenner, S., Jacob, F., and Meselson, M. (1961). An unstable intermediate carrying information from genes to ribosomes for protein synthesis. Nature *190*, 576–581.

Buchan, J.R. (2014). mRNP granules. RNA Biol. *11*, 1019–1030.

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

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

Carmell, M.A., Girard, A., van de Kant, H.J.G., Bourc'his, D., Bestor, T.H., de Rooij, D.G., and Hannon, G.J. (2007). MIWI2 Is Essential for Spermatogenesis and Repression of Transposons in the Mouse Male Germline. Dev. Cell *12*, 503–514.

Carthew, R.W., and Sontheimer, E.J. (2009). Origins and Mechanisms of miRNAs and siRNAs. Cell *136*, 642–655.

Cecere, G., Zheng, G.X.Y., Mansisidor, A.R., Klymko, K.E., and Grishok, A. (2012). Promoters recognized by forkhead proteins exist for individual 21U-RNAs. Mol. Cell *47*, 734–745.

Chapat, C., Jafarnejad, S.M., Matta-Camacho, E., Hesketh, G.G., Gelbart, I.A., Attig, J., Gkogkas, C.G., Alain, T., Stern-Ginossar, N., Fabian, M.R., et al. (2017). Cap-binding protein 4EHP effects translation silencing by microRNAs. Proc. Natl. Acad. Sci. U. S. A. *114*, 5425–5430.

Cheeks, R.J., Canman, J.C., Gabriel, W.N., Meyer, N., Strome, S., and Goldstein, B. (2004). C. elegans PAR Proteins Function by Mobilizing and Stabilizing Asymmetrically Localized Protein Complexes. Curr. Biol. *14*, 851–862.

Chen, H.-M., Futcher, B., and Leatherwood, J. (2011). The Fission Yeast RNA Binding Protein Mmi1 Regulates Meiotic Genes by Controlling Intron Specific Splicing and Polyadenylation Coupled RNA Turnover. PLoS One *6*, e26804.

Chung, W.-J., Okamura, K., Martin, R., and Lai, E.C. (2008). Endogenous RNA Interference Provides a Somatic Defense against Drosophila Transposons. Curr. Biol. *18*, 795–802.

Claycomb, J.M., Batista, P.J., Pang, K.M., Gu, W., Vasale, J.J., van Wolfswinkel, J.C., Chaves, D.A., Shirayama, M., Mitani, S., Ketting, R.F., et al. (2009). The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. Cell *139*, 123–134.

Coller, J., and Parker, R. (2005). General Translational Repression by Activators of mRNA Decapping. Cell *122*, 875–886.

Collins, J., Saari, B., and Anderson, P. (1987). Activation of a transposable element in the germ line but not the soma of Caenorhabditis elegans. Nature *328*, 726–728.

Conine, C.C., Batista, P.J., Gu, W., Claycomb, J.M., Chaves, D.A., Shirayama, M., and Mello, C.C. (2010). Argonautes ALG-3 and ALG-4 are required for spermatogenesis-specific 26G-RNAs and thermotolerant sperm in Caenorhabditis elegans. Proc. Natl. Acad. Sci. U. S. A. *107*, 3588–3593.

Conine, C.C.C., Moresco, J.J.J., Gu, W., Shirayama, M., Conte, D., Yates, J.R.R., and Mello, C.C.C. (2013). Argonautes promote male fertility and provide a paternal memory of germline gene expression in C. elegans. Cell *155*, 1532–1544.

Conrad, R., Thomas, J., Spieth, J., and Blumenthal, T. (1991). Insertion of part of an intron into the 5' untranslated region of a Caenorhabditis elegans gene converts it into a trans-spliced gene. Mol. Cell. Biol. *11*, 1921–1926.

Conrad, R., Liou, R.F., and Blumenthal, T. (1993). Conversion of a trans-spliced C. elegans gene into a conventional gene by introduction of a splice donor site. EMBO J. *12*, 1249–1255.

Conrad, R., Lea, K., and Blumenthal, T. (1995). SL1 trans-splicing specified by AU-rich synthetic RNA inserted at the 5' end of Caenorhabditis elegans pre-mRNA. RNA *1*, 164–170.

Contreras, V., Richardson, M.A., Hao, E., and Keiper, B.D. (2008). Depletion of the cap-associated isoform of translation factor eIF4G induces germline apoptosis in C. elegans. Cell Death Differ. *15*, 1232–1242.

Cora, E., Pandey, R.R., Xiol, J., Taylor, J., Sachidanandam, R., McCarthy, A.A., and Pillai, R.S. (2014). The MID-PIWI module of Piwi proteins specifies nucleotide- and strand-biases of piRNAs. RNA *20*, 773–781.

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

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.

Crick, F.H. (1958). On protein synthesis. Symp. Soc. Exp. Biol. *12*, 138–163.

Czech, B., and Hannon, G.J. (2016). One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. Trends Biochem. Sci. *41*, 324–337.

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

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

Danks, G.B., Raasholm, M., Campsteijn, C., Long, A.M., Manak, J.R., Lenhard, B., and Thompson, E.M. (2015). Trans-Splicing and Operons in Metazoans: Translational Control in Maternally Regulated Development and Recovery from Growth Arrest. Mol. Biol. Evol. *32*, 585–599.

Das, P.P., Bagijn, M.P., Goldstein, L.D., Woolford, J.R., Lehrbach, N.J., Sapetschnig, A., Buhecha, H.R., Gilchrist, M.J., Howe, K.L., Stark, R., et al. (2008). Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the Caenorhabditis elegans germline. Mol. Cell *31*, 79–90.

de Albuquerque, B.F.M., Placentino, M., Ketting, R.F., Alcazar, R.M., Lin, R., Fire, A.Z., Ameres, S.L., Horwich, M.D., Hung, J.H., Xu, J., et al. (2015). Maternal piRNAs Are Essential for Germline Development following De Novo Establishment of Endo-siRNAs in Caenorhabditis elegans. Dev. Cell *34*, 448–456.

Decker, C.J., and Parker, R. (2012). P-Bodies and Stress Granules: Possible Roles in the Control of Translation and mRNA Degradation. Cold Spring Harb. Perspect. Biol. *4*, a012286–a012286.

Deng, W., and Lin, H. (2002). Miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. Dev. Cell *2*, 819–830.

Deo, R.C., Bonanno, J.B., Sonenberg, N., and Burley, S.K. (1999). Recognition of polyadenylate RNA by the poly(A)-binding protein. Cell *98*, 835–845.

Detwiler, M.R., Reuben, M., Li, X., Rogers, E., and Lin, R. (2001). Two zinc finger proteins, OMA-1 and OMA-2, are redundantly required for oocyte maturation in C. elegans. Dev. Cell *1*, 187–199.

Dinkova, T.D., Keiper, B.D., Korneeva, N.L., Aamodt, E.J., and Rhoads, R.E. (2005). Translation of a small subset of Caenorhabditis elegans mRNAs is dependent on a specific eukaryotic translation initiation factor 4E isoform. Mol. Cell. Biol. *25*, 100–113.

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

Douris, V., Telford, M.J., and Averof, M. (2010). Evidence for Multiple Independent Origins of trans-Splicing in Metazoa. Mol. Biol. Evol. *27*, 684–693.

Dreyfuss, G., Philipson, L., and Mattaj, I.W. (1988). Ribonucleoprotein particles in cellular processes. J. Cell Biol. *106*, 1419–1425.

Drinnenberg, I.A., Weinberg, D.E., Xie, K.T., Mower, J.P., Wolfe, K.H., Fink, G.R., and Bartel, D.P. (2009). RNAi in Budding Yeast. Science (80-.). *326*, 544–550.

Duchaine, T.F., Wohlschlegel, J.A., Kennedy, S., Bei, Y., Conte, D., Pang, K., Brownell, D.R.,

Harding, S., Mitani, S., Ruvkun, G., et al. (2006). Functional proteomics reveals the biochemical niche of C. elegans DCR-1 in multiple small-RNA-mediated pathways. Cell *124*, 343–354.

Duerr, J.S. (2013). Antibody Staining in C. Elegans Using "Freeze-Cracking" J. Vis. Exp.

Edgar, R.C. (2004a). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

Edgar, R.C. (2004b). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5*, 113.

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

ElMaghraby, M.F., Andersen, P.R., Pühringer, F., Meixner, K., Lendl, T., Tirian, L., and Brennecke, J. (2019). A heterochromatin-specific RNA export pathway facilitates piRNA production. BioRxiv 596171.

Eulalio, A., Huntzinger, E., Nishihara, T., Rehwinkel, J., Fauser, M., and Izaurralde, E. (2008). Deadenylation is a widespread effect of miRNA regulation. RNA *15*, 21–32.

Fabry, M.H., Ciabrelli, F., Munafò, M., Eastwood, E.L., Kneuss, E., Falciatori, I., Falconio, F.A., Hannon, G.J., and Czech, B. (2019). piRNA-guided co-transcriptional silencing coopts nuclear export factors. BioRxiv 611343.

Fasulo, B., Koyama, C., Yu, K.R., Homola, E.M., Hsieh, T.S., Campbell, S.D., and Sullivan, W. (2012). Chk1 and Wee1 kinases coordinate DNA replication, chromosome condensation, and anaphase entry. Mol. Biol. Cell *23*, 1047–1057.

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*, 259–263.

Fernández-Casañas, M., Chan, K.-L., Fernández-Casañas, M., and Chan, K.-L. (2018). The Unresolved Problem of DNA Bridging. Genes (Basel). *9*, 623.

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.

Fischer, U., and Lührmann, R. (1990). An essential signaling role for the m3G cap in the transport of U1 snRNP to the nucleus. Science *249*, 786–790.

Fischer, S.E.J., Montgomery, T.A., Zhang, C., Fahlgren, N., Breen, P.C., Hwang, A., Sullivan, C.M., Carrington, J.C., and Ruvkun, G. (2011). The ERI-6/7 Helicase Acts at the First Stage of an siRNA Amplification Pathway That Targets Recent Gene Duplications. PLoS Genet. *7*, e1002369.

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

Friedland, A.E., Tzur, Y.B., Esvelt, K.M., Colaiácovo, M.P., Church, G.M., and Calarco, J.A. (2013). Heritable genome editing in C. elegans via a CRISPR-Cas9 system. Nat. Methods *10*, 741–743.

Friesen, W.J., Massenet, S., Paushkin, S., Wyce, A., and Dreyfuss, G. (2001). SMN, the Product of the Spinal Muscular Atrophy Gene, Binds Preferentially to Dimethylarginine-Containing Protein Targets. Mol. Cell *7*, 1111–1117.

Frøkjær-Jensen, C., Davis, M.W., Sarov, M., Taylor, J., Flibotte, S., LaBella, M., Pozniakovsky, A., Moerman, D.G., and Jorgensen, E.M. (2014). Random and targeted transgene insertion in Caenorhabditis elegans using a modified Mos1 transposon. Nat. Methods *11*, 529–534.

Frøkjær-Jensen, C., Jain, N., Hansen, L., Davis, M.W., Li, Y., Zhao, D., Rebora, K., Millet, J.R.M., Liu, X., Kim, S.K., et al. (2016). An Abundant Class of Non-coding DNA Can Prevent Stochastic Gene Silencing in the C. elegans Germline. Cell *166*, 343–357.

Fujimura, A., Kishimoto, H., Yanagisawa, J., and Kimura, K. (2012). Enhancer of rudimentary homolog (ERH) plays an essential role in the progression of mitosis by promoting mitotic chromosome alignment. Biochem. Biophys. Res. Commun. *423*, 588–592.

Gao, M., and Arkov, A.L. (2013). Next generation organelles: structure and role of germ granules in the germline. Mol. Reprod. Dev. *80*, 610–623.

Gelsthorpe, M., Pulumati, M., McCallum, C., Dang-Vu, K., and Tsubota, S.I. (1997). The putative cell cycle gene, enhancer of rudimentary, encodes a highly conserved protein found in plants and animals. Gene *186*, 189–195.

Gent, J.I., Schvarzstein, M., Villeneuve, A.M., Gu, S.G., Jantsch, V., Fire, A.Z., and Baudrimont, A. (2009). A Caenorhabditis elegans RNA-directed RNA polymerase in sperm development and endogenous RNA interference. Genetics *183*, 1297–1314.

Gent, J.I., Lamm, A.T., Pavelec, D.M., Maniar, J.M., Parameswaran, P., Tao, L., Kennedy, S., and Fire, A.Z. (2010). Distinct phases of siRNA synthesis in an endogenous RNAi pathway in C. elegans soma. Mol. Cell *37*, 679–689.

Gerson-Gurwitz, A., Wang, S., Sathe, S., Green, R., Yeo, G.W., Oegema, K., and Desai, A. (2016). A Small RNA-Catalytic Argonaute Pathway Tunes Germline Transcript Levels to Ensure Embryonic Divisions. Cell *165*, 396–409.

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

Gilbert, W., and Müller-Hill, B. (1966). Isolation of the lac repressor. Proc. Natl. Acad. Sci. U. S. A. *56*, 1891–1898.

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*, 199–202.

Goh, W.-S.S., Seah, J.W.E., Harrison, E.J., Chen, C., Hammell, C.M., and Hannon, G.J. (2014). A genome-wide RNAi screen identifies factors required for distinct stages of C. elegans piRNA biogenesis. Genes Dev. *28*, 797–807.

Gonsalvez, G.B., Tian, L., Ospina, J.K., Boisvert, F.-M., Lamond, A.I., and Matera, A.G. (2007). Two distinct arginine methyltransferases are required for biogenesis of Sm-class ribonucleoproteins. J. Cell Biol. *178*, 733–740.

Gosselin, P., Martineau, Y., Morales, J., Czjzek, M., Glippa, V., Gauffeny, I., Morin, E., Le Corguillé, G., Pyronnet, S., Cormier, P., et al. (2013). Tracking a refined eIF4E-binding motif reveals Angel1 as a new partner of eIF4E. Nucleic Acids Res. *41*, 7783–7792.

Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B.J., Chiang, H.R., King, N., Degnan, B.M., Rokhsar, D.S., and Bartel, D.P. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature *455*, 1193–1197.

Gu, W., Shirayama, M., Conte, D., Vasale, J., Batista, P.J., Claycomb, J.M., Moresco, J.J., Youngman, E.M., Keys, J., Stoltz, M.J., et al. (2009). Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the C. elegans germline. Mol. Cell *36*, 231–244.

Gu, W., Lee, H.-C., Chaves, D., Youngman, E.M., Pazour, G.J., Conte, D., and Mello, C.C. (2012). CapSeq and CIP-TAP Identify Pol II Start Sites and Reveal Capped Small RNAs as C. elegans piRNA Precursors. Cell *151*, 1488–1500.

Guang, S., Bochner, A.F., Pavelec, D.M., Burkhart, K.B., Harding, S., Lachowiec, J., and Kennedy, S. (2008). An Argonaute transports siRNAs from the cytoplasm to the nucleus. Science *321*, 537–541.

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell *35*, 849–857.

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 (80-.). *315*, 1587–1590.

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

Guzzardo, P.M., Muerdter, F., and Hannon, G.J. (2013). The piRNA pathway in flies: highlights and future directions. Curr. Opin. Genet. Dev. *23*, 44–52.

Ha, M., and Kim, V.N. (2014). Regulation of microRNA biogenesis. Nat. Rev. Mol. Cell Biol. *15*, 509–524.

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 (80-.). *348*, 817–821.

Han, T., Manoharan, A.P., Harkins, T.T., Bouffard, P., Fitzpatrick, C., Chu, D.S., Thierry-Mieg, D., Thierry-Mieg, J., and Kim, J.K. (2009). 26G endo-siRNAs regulate spermatogenic and zygotic gene expression in Caenorhabditis elegans. Proc. Natl. Acad. Sci. U. S. A. *106*, 18674–18679.

Hanazawa, M., Yonetani, M., and Sugimoto, A. (2011). PGL proteins self associate and bind RNPs to mediate germ granule assembly in C. elegans. J. Cell Biol. *192*, 929–937.

Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y., and Yokoyama, S. (1999). Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. Nature *398*, 579–585.

Hannon, G.J., Maroney, P.A., and Nilsen, T.W. (1991). U small nuclear ribonucleoprotein requirements for nematode cis- and trans-splicing in vitro. J. Biol. Chem. *266*, 22792–22795.

Hansen, D., Wilson-Berry, L., Dang, T., and Schedl, T. (2004). Control of the proliferation versus meiotic development decision in the C. elegans germline through regulation of GLD-1 protein accumulation. Development *131*, 93–104.

Hastings, K.E.M. (2005). SL trans-splicing: easy come or easy go? Trends Genet. *21*, 240–247.

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

20, 814–817.

Henderson, M.A., Cronland, E., Dunkelbarger, S., Contreras, V., Strome, S., and Keiper, B.D. (2009). A germline-specific isoform of eIF4E (IFE-1) is required for efficient translation of stored mRNAs and maturation of both oocytes and sperm. J. Cell Sci. *122*, 1529–1539.

Hendrickson, D.G., Hogan, D.J., McCullough, H.L., Myers, J.W., Herschlag, D., Ferrell, J.E., and Brown, P.O. (2009). Concordant Regulation of Translation and mRNA Abundance for Hundreds of Targets of a Human microRNA. PLoS Biol. *7*, e1000238.

Hoagland, M.B., Stephenson, M.L., Scott, J.F., Hecht, L.I., and Zamecnik, P.C. (1958). A soluble ribonucleic acid intermediate in protein synthesis. J. Biol. Chem. *231*, 241–257.

Hoogewijs, D., Houthoofd, K., Matthijssens, F., Vandesompele, J., and Vanfleteren, J.R. (2008). Selection and validation of a set of reliable reference genes for quantitative sod gene expression analysis in C. elegans. BMC Mol. Biol. *9*, 9.

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

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

Huang, X.Y., and Hirsh, D. (1989). A second trans-spliced RNA leader sequence in the nematode Caenorhabditis elegans. Proc. Natl. Acad. Sci. *86*, 8640–8644.

Huang, H.-Y., Houwing, S., Kaaij, L.J.T., Meppelink, A., Redl, S., Gauci, S., Vos, H., Draper, B.W., Moens, C.B., Burgering, B.M., et al. (2011). Tdrd1 acts as a molecular scaffold for Piwi proteins and piRNA targets in zebrafish. EMBO J. *30*, 3298–3308.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. Science (80-.). *324*, 218–223.

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

Ishidate, T., Ozturk, A.R., Durning, D.J., Sharma, R., Shen, E., Chen, H., Seth, M., Shirayama, M., and Mello, C.C. (2018). ZNFX-1 Functions within Perinuclear Nuage to Balance Epigenetic Signals. Mol. Cell *70*, 639-649.e6.

Izumi, N., and Tomari, Y. (2014). Diversity of the piRNA pathway for nonself silencing: worm-specific piRNA biogenesis factors. Genes Dev. *28*, 665–671.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. *3*, 318–356.

James, P., Halladay, J., and Craig, E.A. (1996). Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. Genetics *144*, 1425–1436.

Jankowska-Anyszka, M., Lamphear, B.J., Aamodt, E.J., Harrington, T., Darzynkiewicz, E., Stolarski, R., and Rhoads, R.E. (1998). Multiple isoforms of eukaryotic protein synthesis initiation factor 4E in Caenorhabditis elegans can distinguish between mono- and trimethylated mRNA cap structures. J. Biol. Chem. *273*, 10538–10542.

Jin, T., Guo, F., Serebriiskii, I.G., Howard, A., and Zhang, Y.-Z. (2007). A 1.55 Å resolution X-ray crystal structure of HEF2/ERH and insights into its transcriptional and cell-cycle interaction networks. Proteins Struct. Funct. Bioinforma. *68*, 427–437.

Joshi, B., Lee, K., Maeder, D.L., and Jagus, R. (2005). Phylogenetic analysis of eIF4E-family members. BMC Evol. Biol. *5*, 48.

Juang, B.-T., Gu, C., Starnes, L., Palladino, F., Goga, A., Kennedy, S., and L'Etoile, N.D. (2013). Endogenous nuclear RNAi mediates behavioral adaptation to odor. Cell *154*, 1010–1022.

Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., et al. (2003). Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. Nature *421*, 231–237.

Kamminga, L.M., Luteijn, M.J., den Broeder, M.J., Redl, S., Kaaij, L.J.T., Roovers, E.F., Ladurner, P., Berezikov, E., and Ketting, R.F. (2010). Hen1 is required for oocyte development and piRNA stability in zebrafish. EMBO J. *29*, 3688–3700.

Kamminga, L.M., van Wolfswinkel, J.C., Luteijn, M.J., Kaaij, L.J.T., Bagijn, M.P., Sapetschnig, A., Miska, E.A., Berezikov, E., and Ketting, R.F. (2012). Differential impact of the HEN1 homolog HENN-1 on 21U and 26G RNAs in the germline of Caenorhabditis elegans. PLoS Genet. *8*, e1002702.

Kappei, D., Butter, F., Benda, C., Scheibe, M., Draškovič, I., Stevense, M., Novo, C.L., Basquin, C., Araki, M., Araki, K., et al. (2013). HOT1 is a mammalian direct telomere repeat-binding protein contributing to telomerase recruitment. EMBO J. *32*, 1681–1701.

Kasimatis, K.R., and Phillips, P.C. (2018). Rapid Gene Family Evolution of a Nematode Sperm Protein Despite Sequence Hyper-conservation. G3 (Bethesda). *8*, 353–362.

Kasper, D.M., Wang, G., Gardner, K.E., Johnstone, T.G., and Reinke, V. (2014). The C. elegans SNAPc Component SNPC-4 Coats piRNA Domains and Is Globally Required for piRNA Abundance. Dev. Cell *31*, 145–158.

Kato, H., Goto, D.B., Martienssen, R.A., Urano, T., Furukawa, K., and Murakami, Y. (2005). RNA polymerase II is required for RNAi-dependent heterochromatin assembly. Science *309*, 467–469.

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

Keall, R., Whitelaw, S., Pettitt, J., and Müller, B. (2007). Histone gene expression and histone mRNA 3' end structure in Caenorhabditis elegans. BMC Mol. Biol. *8*, 51.

Keiper, B.D., Lamphear, B.J., Deshpande, A.M., Jankowska-Anyszka, M., Aamodt, E.J., Blumenthal, T., and Rhoads, R.E. (2000). Functional characterization of five eIF4E isoforms in Caenorhabditis elegans. J. Biol. Chem. *275*, 10590–10596.

Ketting, R.F. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes Dev. *15*, 2654–2659.

Ketting, R.F., Haverkamp, T.H., van Luenen, H.G., and Plasterk, R.H. (1999). Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. Cell *99*, 133–141.

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

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*, 511-519.e6.

Kirino, Y., and Mourelatos, Z. (2007a). Mouse Piwi-interacting RNAs are 2'-O-methylated at their 3' termini. Nat. Struct. Mol. Biol. *14*, 347–348.

Kirino, Y., and Mourelatos, Z. (2007b). The mouse homolog of HEN1 is a potential methylase for Piwi-interacting RNAs. RNA *13*, 1397–1401.

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

Kiuchi, T., Koga, H., Kawamoto, M., Shoji, K., Sakai, H., Arai, Y., Ishihara, G., Kawaoka, S., Sugano, S., Shimada, T., et al. (2014). A single female-specific piRNA is the primary determiner of sex in the silkworm. Nature *509*, 633–636.

Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Khurana, J.S., Zhang, F., Schultz, N., Koppetsch, B.S., Nowosielska, A., et al. (2009). The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. Cell *138*, 1137–1149.

Klosin, A., Casas, E., Hidalgo-Carcedo, C., Vavouri, T., and Lehner, B. (2017). Transgenerational transmission of environmental information in *C. elegans*. Science (80-.). *356*, 320–323.

Kodama, Y., Rothman, J.H., Sugimoto, A., and Yamamoto, M. (2002). The stem-loop binding protein CDL-1 is required for chromosome condensation, progression of cell death and morphogenesis in Caenorhabditis elegans. Development *129*, 187–196.

Krause, M., and Hirsh, D. (1987). A trans-spliced leader sequence on actin mRNA in C. elegans. Cell *49*, 753–761.

Kroiss, M., Schultz, J., Wiesner, J., Chari, A., Sickmann, A., and Fischer, U. (2008). Evolution of an RNP assembly system: a minimal SMN complex facilitates formation of UsnRNPs in Drosophila melanogaster. Proc. Natl. Acad. Sci. U. S. A. *105*, 10045–10050.

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

Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., and Cech, T.R. (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. Cell *31*, 147–157.

Krzyzanowski, M.K., Kozlowska, E., and Kozlowski, P. (2012). Identification and Functional Analysis of the erh1+ Gene Encoding Enhancer of Rudimentary Homolog from the Fission Yeast Schizosaccharomyces pombe. PLoS One *7*, e49059.

Kudo, N., Matsumori, N., Taoka, H., Fujiwara, D., Schreiner, E.P., Wolff, B., Yoshida, M., and Horinouchi, S. (1999). Leptomycin B inactivates CRM1/exportin 1 by covalent modification at a cysteine residue in the central conserved region. Proc. Natl. Acad. Sci. U. S. A. *96*, 9112–9117.

Kuramochi-Miyagawa, S., Kimura, T., Ijiri, T.W., Isobe, T., Asada, N., Fujita, Y., Ikawa, M., Iwai, N., Okabe, M., Deng, W., et al. (2004). Mili, a mammalian member of piwi family gene, is

essential for spermatogenesis. Development *127*, 503–514.

Kwak, P.B., and Tomari, Y. (2012). The N domain of Argonaute drives duplex unwinding during RISC assembly. Nat. Struct. Mol. Biol. *19*, 145–151.

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

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

Lee, J.T., Strauss, W.M., Dausman, J.A., and Jaenisch, R. (1996). A 450 kb transgene displays properties of the mammalian X-inactivation center. Cell *86*, 83–94.

Lee, L.-W., Lee, C.-C., Huang, C.-R., and Lo, S.J. (2012b). The Nucleolus of *Caenorhabditis elegans*. J. Biomed. Biotechnol. *2012*, 1–11.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843–854.

Li, W., DeBella, L.R., Guven-Ozkan, T., Lin, R., and Rose, L.S. (2009). An eIF4E-binding protein regulates katanin protein levels in C. elegans embryos. J. Cell Biol. *187*, 33–42.

Li, X.Z., Roy, C.K., Dong, X., Bolcun-Filas, E., Wang, J., Han, B.W., Xu, J., Moore, M.J., Schimenti, J.C., Weng, Z., et al. (2013). An Ancient Transcription Factor Initiates the Burst of piRNA Production during Early Meiosis in Mouse Testes. Mol. Cell *50*, 67–81.

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

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

Lisitskaya, L., Aravin, A.A., and Kulbachinskiy, A. (2018). DNA interference and beyond: structure and functions of prokaryotic Argonaute proteins. Nat. Commun. *9*, 5165.

Liu, J., Carmell, M.A., Rivas, F. V, Marsden, C.G., Thomson, J.M., Song, J.-J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. (2004). Argonaute2 Is the Catalytic Engine of Mammalian RNAi. Science (80-.). *305*, 1437–1441.

López, M.D., and Samuelsson, T. (2008). Early evolution of histone mRNA 3' end processing. RNA *14*, 1–10.

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

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

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

Machyna, M., Neugebauer, K.M., and Staněk, D. (2015). Coilin: The first 25 years. RNA Biol. *12*, 590–596.

MacRae, I.J., Ma, E., Zhou, M., Robinson, C. V., and Doudna, J.A. (2008). In vitro reconstitution

of the human RISC-loading complex. Proc. Natl. Acad. Sci. *105*, 512–517.

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

Mamane, Y., Petroulakis, E., Rong, L., Yoshida, K., Ler, L.W., and Sonenberg, N. (2004). eIF4E – from translation to transformation. Oncogene *23*, 3172–3179.

Mangio, R.S., Votra, S., and Pruyne, D. (2015). The canonical eIF4E isoform of C. elegans regulates growth, embryogenesis, and germline sex-determination. Biol. Open *4*, 843–851.

Maniar, J.M., and Fire, A.Z. (2011). EGO-1, a C. elegans RdRP, modulates gene expression via production of mRNA-templated short antisense RNAs. Curr. Biol. *21*, 449–459.

Maris, C., Dominguez, C., and Allain, F.H.-T. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. FEBS J. *272*, 2118–2131.

Maroney, P.A., Hannon, G.J., Denker, J.A., and Nilsen, T.W. (1990). The nematode spliced leader RNA participates in trans-splicing as an Sm snRNP. EMBO J. *9*, 3667–3673.

Marré, J., Traver, E.C., and Jose, A.M. (2016). Extracellular RNA is transported from one generation to the next in *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. *113*, 12496–12501.

Martienssen, R., and Moazed, D. (2015). RNAi and Heterochromatin Assembly. Cold Spring Harb. Perspect. Biol. *7*, a019323.

Marzluff, W.F., Wagner, E.J., and Duronio, R.J. (2008). Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat. Rev. Genet. *9*, 843–854.

Medvedev, K.E., Kinch, L.N., and Grishin, N. V. (2018). Functional and evolutionary analysis of viral proteins containing a Rossmann-like fold. Protein Sci. *27*, 1450–1463.

Miesen, P., Girardi, E., and van Rij, R.P. (2015). Distinct sets of PIWI proteins produce arbovirus and transposon-derived piRNAs in *Aedes aegypti* mosquito cells. Nucleic Acids Res. *43*, 6545–6556.

Minasaki, R., and Streit, A. (2007). MEL-47, a novel protein required for early cell divisions in the nematode Caenorhabditis elegans. Mol. Genet. Genomics *277*, 315–328.

Minasaki, R., Puoti, A., and Streit, A. (2009). The DEAD-box protein MEL-46 is required in the germ line of the nematode Caenorhabditis elegans. BMC Dev. Biol. *9*, 35.

Miska, E.A., Alvarez-Saavedra, E., Abbott, A.L., Lau, N.C., Hellman, A.B., McGonagle, S.M., Bartel, D.P., Ambros, V.R., and Horvitz, H.R. (2007). Most Caenorhabditis elegans microRNAs Are Individually Not Essential for Development or Viability. PLoS Genet. *3*, e215.

Miyoshi, H., Dwyer, D.S., Keiper, B.D., Jankowska-Anyszka, M., Darzynkiewicz, E., and Rhoads, R.E. (2002). Discrimination between mono- and trimethylated cap structures by two isoforms of Caenorhabditis elegans eIF4E. EMBO J. *21*, 4680–4690.

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

Mohn, F., Handler, D., and Brennecke, J. (2015). piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. Science (80-.). *348*, 812–817.

Montgomery, T.A., Rim, Y.-S., Zhang, C., Dowen, R.H., Phillips, C.M., Fischer, S.E.J., and Ruvkun, G. (2012). PIWI Associated siRNAs and piRNAs Specifically Require the Caenorhabditis elegans HEN1 Ortholog henn-1. PLoS Genet. *8*, e1002616.

Moore, R.S., Kaletsky, R., and Murphy, C.T. (2019). Piwi/PRG-1 Argonaute and TGF-β Mediate Transgenerational Learned Pathogenic Avoidance. Cell *177*, 1827-1841.e12.

Morazzani, E.M., Wiley, M.R., Murreddu, M.G., Adelman, Z.N., and Myles, K.M. (2012). Production of Virus-Derived Ping-Pong-Dependent piRNA-like Small RNAs in the Mosquito Soma. PLoS Pathog. *8*, e1002470.

Morimoto, M., Boerkoel, C., Morimoto, M., and Boerkoel, C.F. (2013). The Role of Nuclear Bodies in Gene Expression and Disease. Biology (Basel). *2*, 976–1033.

Morita, M., Ler, L.W., Fabian, M.R., Siddiqui, N., Mullin, M., Henderson, V.C., Alain, T., Fonseca, B.D., Karashchuk, G., Bennett, C.F., et al. (2012). A Novel 4EHP-GIGYF2 Translational Repressor Complex Is Essential for Mammalian Development. Mol. Cell. Biol. *32*, 3585–3593.

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

Naito, Y., Hino, K., Bono, H., and Ui-Tei, K. (2015). CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. Bioinformatics *31*, 1120–1123.

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

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., and Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. Nature *543*, 72–77.

Newman, M.A., Ji, F., Fischer, S.E.J., Anselmo, A., Sadreyev, R.I., and Ruvkun, G. (2018). The surveillance of pre-mRNA splicing is an early step in C. elegans RNAi of endogenous genes. Genes Dev. *32*, 670–681.

Nilsen, T.W. (1993). Trans-Splicing of Nematode Premessenger RNA. Annu. Rev. Microbiol. *47*, 413–440.

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

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

Nizami, Z., Deryusheva, S., and Gall, J.G. (2010). The Cajal body and histone locus body. Cold Spring Harb. Perspect. Biol. *2*, a000653.

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

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

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

Pak, J., and Fire, A. (2007). Distinct populations of primary and secondary effectors during RNAi in C. elegans. Science *315*, 241–244.

Pak, J., Maniar, J.M., Mello, C.C., and Fire, A. (2012). Protection from feed-forward amplification in an amplified RNAi mechanism. Cell *151*, 885–899.

Palade, G.E. (1955). A small particulate component of the cytoplasm. J. Biophys. Biochem. Cytol. *1*, 59–68.

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

Parker, R., and Sheth, U. (2007). P bodies and the control of mRNA translation and degradation. Mol. Cell *25*, 635–646.

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

Parker, J.S., Roe, S.M., and Barford, D. (2004). Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. EMBO J. *23*, 4727–4737.

Pavelec, D.M., Lachowiec, J., Duchaine, T.F., Smith, H.E., and Kennedy, S. (2009). Requirement for the ERI/DICER Complex in Endogenous RNA Interference and Sperm Development in *Caenorhabditis elegans*. Genetics *183*, 1283–1295.

Peter, D., Weber, R., Köne, C., Chung, M.-Y., Ebertsch, L., Truffault, V., Weichenrieder, O., Igreja, C., and Izaurralde, E. (2015). Mextli proteins use both canonical bipartite and novel tripartite binding modes to form eIF4E complexes that display differential sensitivity to 4E-BP regulation. Genes Dev. *29*, 1835–1849.

Pettitt, J., Crombie, C., Schümperli, D., and Müller, B. (2002). The Caenorhabditis elegans histone hairpin-binding protein is required for core histone gene expression and is essential for embryonic and postembryonic cell division. J. Cell Sci. *115*, 857–866.

Philippe, L., Pandarakalam, G.C., Fasimoye, R., Harrison, N., Connolly, B., Pettitt, J., and Müller, B. (2017). An in vivo genetic screen for genes involved in spliced leader trans-splicing indicates a crucial role for continuous de novo spliced leader RNP assembly. Nucleic Acids Res. *45*, 8474–8483.

Phillips, C.M., Montgomery, T.A., Breen, P.C., and Ruvkun, G. (2012). MUT-16 promotes formation of perinuclear mutator foci required for RNA silencing in the C. elegans germline. Genes Dev. *26*, 1433–1444.

Phillips, C.M., Montgomery, B.E., Breen, P.C., Roovers, E.F., Rim, Y.-S., Ohsumi, T.K., Newman, M.A., van Wolfswinkel, J.C., Ketting, R.F., Ruvkun, G., et al. (2014). MUT-14 and SMUT-1 DEAD box RNA helicases have overlapping roles in germline RNAi and endogenous siRNA formation. Curr. Biol. *24*, 839–844.

Phillips, C.M., Brown, K.C., Montgomery, B.E., Ruvkun, G., and Montgomery, T.A. (2015).

piRNAs and piRNA-Dependent siRNAs Protect Conserved and Essential C. elegans Genes from Misrouting into the RNAi Pathway. Dev. Cell *34*, 457–465.

Piserà, A., Campo, A., and Campo, S. (2018). Structure and functions of the translation initiation factor eIF4E and its role in cancer development and treatment. J. Genet. Genomics *45*, 13–24.

Pitt, J.N., Schisa, J.A., and Priess, J.R. (2000). P Granules in the Germ Cells of Caenorhabditis elegans Adults Are Associated with Clusters of Nuclear Pores and Contain RNA. Dev. Biol. *219*, 315–333.

Plessel, G., Fischer, U., and Lührmann, R. (1994). m3G cap hypermethylation of U1 small nuclear ribonucleoprotein (snRNP) in vitro: evidence that the U1 small nuclear RNA-(guanosine-N2)-methyltransferase is a non-snRNP cytoplasmic protein that requires a binding site on the Sm core domain. Mol. Cell. Biol. *14*, 4160–4172.

Pogge von Strandmann, E., Senkel, S., and Ryffel, G.U. (2001). ERH (enhancer of rudimentary homologue), a conserved factor identical between frog and human, is a transcriptional repressor. Biol. Chem. *382*, 1379–1385.

Posner, R., Toker, I.A., Antonova, O., Star, E., Anava, S., Azmon, E., Hendricks, M., Bracha, S., Gingold, H., and Rechavi, O. (2019). Neuronal Small RNAs Control Behavior Transgenerationally. Cell *177*, 1814-1826.e15.

Protter, D.S.W., Rao, B.S., Van Treeck, B., Lin, Y., Mizoue, L., Rosen, M.K., and Parker, R. (2018). Intrinsically Disordered Regions Can Contribute Promiscuous Interactions to RNP Granule Assembly. Cell Rep. *22*, 1401–1412.

Puoti, A., and Kimble, J. (1999). The Caenorhabditis elegans sex determination gene mog-1 encodes a member of the DEAH-Box protein family. Mol. Cell. Biol. *19*, 2189–2197.

Puoti, A., and Kimble, J. (2000). The hermaphrodite sperm/oocyte switch requires the Caenorhabditis elegans homologs of PRP2 and PRP22. Proc. Natl. Acad. Sci. *97*, 3276–3281.

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*, 693–707.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160-5.

Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat. Protoc. *2*, 1896–1906.

Rashid, U.J., Paterok, D., Koglin, A., Gohlke, H., Piehler, J., and Chen, J.C.-H. (2007). Structure of Aquifex aeolicus argonaute highlights conformational flexibility of the PAZ domain as a potential regulator of RNA-induced silencing complex function. J. Biol. Chem. *282*, 13824–13832.

Raz, E. (2003). Primordial germ-cell development: the zebrafish perspective. Nat. Rev. Genet. *4*, 690–700.

Rechavi, O., Houri-Ze'evi, L., Anava, S., Goh, W.S.S., Kerk, S.Y., Hannon, G.J., and Hobert, O. (2014). Starvation-Induced Transgenerational Inheritance of Small RNAs in C. elegans. Cell *158*, 277–287.

Reuter, M., Chuma, S., Tanaka, T., Franz, T., Stark, A., and Pillai, R.S. (2009). Loss of the Mili-interacting Tudor domain–containing protein-1 activates transposons and alters the Mili-associated small RNA profile. Nat. Struct. Mol. Biol. *16*, 639–646.

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*, 264–267.

Rieckher, M., Markaki, M., Princz, A., Schumacher, B., and Tavernarakis, N. (2018). Maintenance of Proteostasis by P Body-Mediated Regulation of eIF4E Availability during Aging in Caenorhabditis elegans. Cell Rep. *25*, 199-211.e6.

Rivas, F. V, Tolia, N.H., Song, J.-J., Aragon, J.P., Liu, J., Hannon, G.J., and Joshua-Tor, L. (2005). Purified Argonaute2 and an siRNA form recombinant human RISC. Nat. Struct. Mol. Biol. *12*, 340–349.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. *42*, W320–W324.

Rojas-Ríos, P., and Simonelig, M. (2018). piRNAs and PIWI proteins: regulators of gene expression in development and stem cells. Development *145*, dev161786.

Roovers, E.F., Rosenkranz, D., Mahdipour, M., Han, C.-T., He, N., Chuva de Sousa Lopes, S.M., van der Westerlaken, L.A.J., Zischler, H., Butter, F., Roelen, B.A.J., et al. (2015). Piwi Proteins and piRNAs in Mammalian Oocytes and Early Embryos. Cell Rep. *10*, 2069–2082.

Roovers, E.F., Kaaij, L.J.T., Redl, S., Bronkhorst, A.W., Wiebrands, K., de Jesus Domingues, A.M., Huang, H.-Y., Han, C.-T., Riemer, S., Dosch, R., et al. (2018). Tdrd6a Regulates the Aggregation of Buc into Functional Subcellular Compartments that Drive Germ Cell Specification. Dev. Cell *46*, 285-301.e9.

Rossmann, M.G., Moras, D., and Olsen, K.W. (1974). Chemical and biological evolution of nucleotide-binding protein. Nature *250*, 194–199.

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

Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., and Bartel, D.P. (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. Cell *127*, 1193–1207.

Sarkies, P., Selkirk, M.E., Jones, J.T., Blok, V., Boothby, T., Goldstein, B., Hanelt, B., Ardila-Garcia, A., Fast, N.M., Schiffer, P.M., et al. (2015). Ancient and Novel Small RNA Pathways Compensate for the Loss of piRNAs in Multiple Independent Nematode Lineages. PLOS Biol. *13*, e1002061.

Scheckel, C., Gaidatzis, D., Wright, J.E., and Ciosk, R. (2012). Genome-Wide Analysis of GLD-1–Mediated mRNA Regulation Suggests a Role in mRNA Storage. PLoS Genet. *8*, e1002742.

Schirle, N.T., and MacRae, I.J. (2012). The crystal structure of human Argonaute2. Science *336*, 1037–1040.

Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative CT method. Nat. Protoc. *3*, 1101–1108.

Schnettler, E., Donald, C.L., Human, S., Watson, M., Siu, R.W.C., McFarlane, M., Fazakerley, J.K., Kohl, A., and Fragkoudis, R. (2013). Knockdown of piRNA pathway proteins results in enhanced Semliki Forest virus production in mosquito cells. J. Gen. Virol. *94*, 1680–1689.

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

Schweinsberg, P.J., and Grant, B.D. (2013). C. elegans gene transformation by microparticle bombardment. WormBook 1–10.

Sengupta, M.S., Low, W.Y., Patterson, J.R., Kim, H.-M., Traven, A., Beilharz, T.H., Colaiacovo, M.P., Schisa, J.A., and Boag, P.R. (2013). ifet-1 is a broad-scale translational repressor required for normal P granule formation in C. elegans. J. Cell Sci. *126*, 850–859.

Seth, M., Shirayama, M., Gu, W., Ishidate, T., Conte, D., and Mello, C.C. (2013). The C. elegans CSR-1 argonaute pathway counteracts epigenetic silencing to promote germline gene expression. Dev. Cell *27*, 656–663.

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 Rep. *22*, 2254–2264.

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

Shabalina, S.A., and Koonin, E. V. (2008). Origins and evolution of eukaryotic RNA interference. Trends Ecol. Evol. *23*, 578–587.

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*, 937-951.e18.

Sheth, U., and Parker, R. (2003). Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. Science *300*, 805–808.

Sheth, U., Pitt, J., Dennis, S., and Priess, J.R. (2010). Perinuclear P granules are the principal sites of mRNA export in adult C. elegans germ cells. Development *137*, 1305–1314.

Shevchenko, A., Tomas, H., Havli, J., Olsen, J. V, and Mann, M. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat. Protoc. *1*, 2856–2860.

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

Shirayama, M., Stanney, W., Gu, W., Seth, M., and Mello, C.C.C. (2014). The Vasa Homolog RDE-12 engages target mRNA and multiple argonaute proteins to promote RNAi in C. elegans. Curr. Biol. *24*, 845–851.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539.

Sijen, T., Steiner, F.A., Thijssen, K.L., and Plasterk, R.H.A. (2007). Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. Science *315*, 244–247.

Sikorsky, T., Hobor, F., Krizanova, E., Pasulka, J., Kubicek, K., and Stefl, R. (2012). Recognition

of asymmetrically dimethylated arginine by TDRD3. Nucleic Acids Res. *40*, 11748–11755.

Simon, B., Kirkpatrick, J.P., Eckhardt, S., Reuter, M., Rocha, E.A., Andrade-Navarro, M.A., Sehr, P., Pillai, R.S., and Carlomagno, T. (2011). Recognition of 2'-O-Methylated 3'-End of piRNA by the PAZ Domain of a Piwi Protein. Structure *19*, 172–180.

Simon, M., Sarkies, P., Ikegami, K., Doebley, A.-L., Goldstein, L.D., Mitchell, J., Sakaguchi, A., Miska, E.A., and Ahmed, S. (2014). Reduced insulin/IGF-1 signaling restores germ cell immortality to caenorhabditis elegans Piwi mutants. Cell Rep. *7*, 762–773.

Simonelig, M. (2014). piRNAs, master regulators of gene expression. Cell Res. *24*, 779–780.

Simonis, N., Rual, J.-F., Carvunis, A.-R., Tasan, M., Lemmens, I., Hirozane-Kishikawa, T., Hao, T., Sahalie, J.M., Venkatesan, K., Gebreab, F., et al. (2009). Empirically controlled mapping of the Caenorhabditis elegans protein-protein interactome network. Nat. Methods *6*, 47–54.

Siomi, M.C., Sato, K., Pezic, D., and Aravin, A.A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. Nat. Rev. Mol. Cell Biol. *12*, 246–258.

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

Song, J.-J., Smith, S.K., Hannon, G.J., and Joshua-Tor, L. (2004). Crystal Structure of Argonaute and Its Implications for RISC Slicer Activity. Science (80-.). *305*, 1434–1437.

Squirrell, J.M., Eggers, Z.T., Luedke, N., Saari, B., Grimson, A., Lyons, G.E., Anderson, P., and White, J.G. (2006). CAR-1, a Protein That Localizes with the mRNA Decapping Component DCAP-1, Is Required for Cytokinesis and ER Organization in *Caenorhabditis elegans* Embryos. Mol. Biol. Cell *17*, 336–344.

Stachelska, A., Wieczorek, Z., Ruszczyńska, K., Stolarski, R., Pietrzak, M., Lamphear, B.J., Rhoads, R.E., Darzynkiewicz, E., and Jankowska-Anyszka, M. (2002). Interaction of three Caenorhabditis elegans isoforms of translation initiation factor eIF4E with mono- and trimethylated mRNA 5' cap analogues. Acta Biochim. Pol. *49*, 671–682.

Steiner, F. a, Okihara, K.L., Hoogstrate, S.W., Sijen, T., and Ketting, R.F. (2009). RDE-1 slicer activity is required only for passenger-strand cleavage during RNAi in Caenorhabditis elegans. Nat. Struct. Mol. Biol. *16*, 207–211.

Strome, S., and Wood, W.B. (1982). Immunofluorescence visualization of germ-line-specific cytoplasmic granules in embryos, larvae, and adults of Caenorhabditis elegans. Proc. Natl. Acad. Sci. *79*, 1558–1562.

Sugiyama, T., Thillainadesan, G., Chalamcharla, V.R., Meng, Z., Balachandran, V., Dhakshnamoorthy, J., Zhou, M., and Grewal, S.I.S. (2016). Enhancer of Rudimentary Cooperates with Conserved RNA-Processing Factors to Promote Meiotic mRNA Decay and Facultative Heterochromatin Assembly. Mol. Cell *61*, 747–759.

Sulston, J.E., and Horvitz, H.R. (1977). Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. Dev. Biol. *56*, 110–156.

Sulston, J.E., Schierenberg, E., White, J.G., and Thomson, J.N. (1983). The embryonic cell lineage of the nematode Caenorhabditis elegans. Dev. Biol. *100*, 64–119.

Swarts, D.C., Makarova, K., Wang, Y., Nakanishi, K., Ketting, R.F., Koonin, E. V, Patel, D.J.,

and van der Oost, J. (2014a). The evolutionary journey of Argonaute proteins. Nat. Struct. Mol. Biol. *21*, 743–753.

Swarts, D.C., Jore, M.M., Westra, E.R., Zhu, Y., Janssen, J.H., Snijders, A.P., Wang, Y., Patel, D.J., Berenguer, J., Brouns, S.J.J., et al. (2014b). DNA-guided DNA interference by a prokaryotic Argonaute. Nature *507*, 258–261.

Swarts, D.C., Hegge, J.W., Hinojo, I., Shiimori, M., Ellis, M.A., Dumrongkulraksa, J., Terns, R.M., Terns, M.P., and van der Oost, J. (2015). Argonaute of the archaeon Pyrococcus furiosus is a DNA-guided nuclease that targets cognate DNA. Nucleic Acids Res. *43*, 5120–5129.

Syntichaki, P., Troulinaki, K., and Tavernarakis, N. (2007). eIF4E function in somatic cells modulates ageing in Caenorhabditis elegans. Nature *445*, 922–926.

Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C.C. (1999). The rde-1 Gene, RNA Interference, and Transposon Silencing in C. elegans. Cell *99*, 123–132.

Tabara, H., Yigit, E., Siomi, H., and Mello, C.C. (2002). The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DExH-box helicase to direct RNAi in C. elegans. Cell *109*, 861–871.

Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M., et al. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature *453*, 534–538.

Tang, W., Tu, S., Lee, H.-C., Weng, Z., Mello, C.C., Giraldez, A.J., Tomari, Y., Martinez-Perez, E., Miska, E.A., Kryvych, N., et al. (2016a). The RNase PARN-1 Trims piRNA 3' Ends to Promote Transcriptome Surveillance in C. elegans. Cell *164*, 974–984.

Tang, W., Seth, M., Tu, S., Shen, E.-Z., Li, Q., Shirayama, M., Weng, Z., and Mello, C.C. (2018). A Sex Chromosome piRNA Promotes Robust Dosage Compensation and Sex Determination in C. elegans. Dev. Cell *44*, 762-770.e3.

Tang, X., Bharath, S.R., Piao, S., Tan, V.Q., Bowler, M.W., and Song, H. (2016b). Structural basis for specific recognition of pre-snRNA by Gemin5. Cell Res. *26*, 1353–1356.

Thivierge, C., Makil, N., Flamand, M., Vasale, J.J., Mello, C.C., Wohlschlegel, J., Conte, D., and Duchaine, T.F. (2012). Tudor domain ERI-5 tethers an RNA-dependent RNA polymerase to DCR-1 to potentiate endo-RNAi. Nat. Struct. Mol. Biol. *19*, 90–97.

Thomas, J., Lea, K., Zucker-Aprison, E., and Blumenthal, T. (1990). The spliceosomal snRNAs of *Caenorhabditis elegans*. Nucleic Acids Res. *18*, 2633–2642.

Thomas, J.D., Conrad, R.C., and Blumenthal, T. (1988). The C. elegans Trans-spliced leader RNA is bound to Sm and has a trimethylguanosine cap. Cell *54*, 533–539.

Timmons, L., and Fire, A. (1998). Specific interference by ingested dsRNA. Nature *395*, 854–854.

Tolia, N.H., and Joshua-Tor, L. (2007). Slicer and the argonautes. Nat. Chem. Biol. *3*, 36–43.

Updike, D.L., and Strome, S. (2009). A genomewide RNAi screen for genes that affect the stability, distribution and function of P granules in Caenorhabditis elegans. Genetics *183*, 1397–1419.

Vasale, J.J., Gu, W., Thivierge, C., Batista, P.J., Claycomb, J.M., Youngman, E.M., Duchaine, T.F., Mello, C.C., and Conte, D. (2010). Sequential rounds of RNA-dependent RNA

transcription drive endogenous small-RNA biogenesis in the ERGO-1/Argonaute pathway. Proc. Natl. Acad. Sci. U. S. A. *107*, 3582–3587.

Vastenhouw, N.L., Fischer, S.E.J., Robert, V.J.P., Thijssen, K.L., Fraser, A.G., Kamath, R.S., Ahringer, J., and Plasterk, R.H.A. (2003). A genome-wide screen identifies 27 genes involved in transposon silencing in C. elegans. Curr. Biol. *13*, 1311–1316.

Verdel, A., and Moazed, D. (2005). RNAi-directed assembly of heterochromatin in fission yeast. FEBS Lett. *579*, 5872–5878.

Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I.S., and Moazed, D. (2004). RNAi-Mediated Targeting of Heterochromatin by the RITS Complex. Science (80-.). *303*, 672–676.

Volpe, T.A. (2002). Regulation of Heterochromatic Silencing and Histone H3 Lysine-9 Methylation by RNAi. Science (80-.). *297*, 1833–1837.

Voronina, E. (2013). The diverse functions of germline P-granules in *Caenorhabditis elegans*. Mol. Reprod. Dev. *80*, 624–631.

Wan, C., Tempel, W., Liu, Z., Wang, B., and Rose, R. (2005). Structure of the conserved transcriptional repressor enhancer of rudimentary homolog. Biochemistry *44*, 5017–5023.

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

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

Wang, H., Ma, Z., Niu, K., Xiao, Y., Wu, X., Pan, C., Zhao, Y., Wang, K., Zhang, Y., and Liu, N. (2016). Antagonistic roles of Nibbler and Hen1 in modulating piRNA 3' ends in Drosophila. Development *143*, 530–539.

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

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 Dev. *20*, 1732–1743.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., et al. (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature *453*, 539–543.

Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature *171*, 737–738.

Wedeles, C.J., Wu, M.Z., and Claycomb, J.M. (2013). Protection of germline gene expression by the C. elegans Argonaute CSR-1. Dev. Cell *27*, 664–671.

Weick, E.-M., Sarkies, P., Silva, N., Chen, R. a, Moss, S.M.M., Cording, A.C., Ahringer, J., Martinez-Perez, E., and Miska, E. a (2014). PRDE-1 is a nuclear factor essential for the biogenesis of Ruby motif-dependent piRNAs in C. elegans. Genes Dev. *28*, 783–796.

Welker, N.C., Pavelec, D.M., Nix, D.A., Duchaine, T.F., Kennedy, S., and Bass, B.L. (2010). Dicer's helicase domain is required for accumulation of some, but not all, C. elegans endogenous siRNAs. RNA *16*, 893–903.

Weng, C., Kosalka, J., Berkyurek, A.C., Stempor, P., Feng, X., Mao, H., Zeng, C., Li, W.-J., Yan, Y.-H., Dong, M.-Q., et al. (2019). The USTC co-opts an ancient machinery to drive piRNA transcription in C. elegans. Genes Dev. *33*, 90–102.

Weng, M.-T., Lee, J.-H., Wei, S.-C., Li, Q., Shahamatdar, S., Hsu, D., Schetter, A.J., Swatkoski, S., Mannan, P., Garfield, S., et al. (2012). Evolutionarily conserved protein ERH controls CENP-E mRNA splicing and is required for the survival of KRAS mutant cancer cells. Proc. Natl. Acad. Sci. U. S. A. *109*, E3659-67.

Wojcik, E., Murphy, A.M., Fares, H., Dang-Vu, K., and Tsubota, S.I. (1994). Enhancer of rudimentaryp1, e(r)p1, a highly conserved enhancer of the rudimentary gene. Genetics *138*, 1163–1170.

van Wolfswinkel, J.C. (2014). Piwi and Potency: PIWI Proteins in Animal Stem Cells and Regeneration. Integr. Comp. Biol. *54*, 700–713.

van Wolfswinkel, J.C., Claycomb, J.M., Batista, P.J., Mello, C.C., Berezikov, E., and Ketting, R.F. (2009). CDE-1 affects chromosome segregation through uridylation of CSR-1-bound siRNAs. Cell *139*, 135–148.

Xie, G., Vo, T. V., Thillainadesan, G., Holla, S., Zhang, B., Jiang, Y., Lv, M., Xu, Z., Wang, C., Balachandran, V., et al. (2019). A conserved dimer interface connects ERH and YTH family proteins to promote gene silencing. Nat. Commun. 2019 101 *10*, 251.

Xu, C., Ishikawa, H., Izumikawa, K., Li, L., He, H., Nobe, Y., Yamauchi, Y., Shahjee, H.M., Wu, X.-H., Yu, Y., et al. (2016). Structural insights into Gemin5-guided selection of pre-snRNAs for snRNP assembly. Genes Dev. *30*, 2376–2390.

Xu, F., Feng, X., Chen, X., Weng, C., Yan, Q., Xu, T., Hong, M., and Guang, S. (2018). A Cytoplasmic Argonaute Protein Promotes the Inheritance of RNAi. Cell Rep. *23*, 2482–2494.

Yamashiro, H., and Siomi, M.C. (2018). PIWI-Interacting RNA in *Drosophila*: Biogenesis, Transposon Regulation, and Beyond. Chem. Rev. *118*, 4404–4421.

Yamashita, A., Takayama, T., Iwata, R., and Yamamoto, M. (2013). A novel factor Iss10 regulates Mmi1-mediated selective elimination of meiotic transcripts. Nucleic Acids Res. *41*, 9680–9687.

Yan, K.S., Yan, S., Farooq, A., Han, A., Zeng, L., and Zhou, M.-M. (2003). Structure and conserved RNA binding of the PAZ domain. Nature *426*, 469–474.

Yang, H., Zhang, Y., Vallandingham, J., Li, H., Florens, L., Mak, H.Y., and Mak, H.Y. (2012). The RDE-10/RDE-11 complex triggers RNAi-induced mRNA degradation by association with target mRNA in C. elegans. Genes Dev. *26*, 846–856.

Yang, H., Vallandingham, J., Shiu, P., Li, H., Hunter, C.P., and Mak, H.Y. (2014). The DEAD Box Helicase RDE-12 Promotes Amplification of RNAi in Cytoplasmic Foci in C. elegans. Curr. Biol. *24*, 832–838.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. Nat. Methods *12*, 7–8.

Yang, Y.-F., Zhang, X., Ma, X., Zhao, T., Sun, Q., Huan, Q., Wu, S., Du, Z., and Qian, W.

(2017). Trans-splicing enhances translational efficiency in C. elegans. Genome Res. *27*, 1525–1535.

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

Yoshimura, T., Watanabe, T., Kuramochi-Miyagawa, S., Takemoto, N., Shiromoto, Y., Kudo, A., Kanai-Azuma, M., Tashiro, F., Miyazaki, S., Katanaya, A., et al. (2018). Mouse GTSF1 is an essential factor for secondary piRNA biogenesis. EMBO Rep. *19*, e42054.

Yuan, Y.-R., Pei, Y., Ma, J.-B., Kuryavyi, V., Zhadina, M., Meister, G., Chen, H.-Y., Dauter, Z., Tuschl, T., and Patel, D.J. (2005). Crystal Structure of A. aeolicus Argonaute, a Site-Specific DNA-Guided Endoribonuclease, Provides Insights into RISC-Mediated mRNA Cleavage. Mol. Cell *19*, 405–419.

Zanetti, S., Meola, M., Bochud, A., and Puoti, A. (2011). Role of the C. elegans U2 snRNP protein MOG-2 in sex determination, meiosis, and splice site selection. Dev. Biol. *354*, 232–241.

Zaslaver, A., Baugh, L.R., and Sternberg, P.W. (2011). Metazoan Operons Accelerate Recovery from Growth-Arrested States. Cell *145*, 981–992.

Zeng, C., Weng, C., Wang, X., Yan, Y.-H., Li, W.-J., Xu, D., Hong, M., Liao, S., Dong, M.-Q., Feng, X., et al. (2019). Functional Proteomics Identifies a PICS Complex Required for piRNA Maturation and Chromosome Segregation. Cell Rep. *27*, 3561-3572.e3.

Zhang, C., Montgomery, T. a, Fischer, S.E.J., Garcia, S.M.D. a, Riedel, C.G., Fahlgren, N., Sullivan, C.M., Carrington, J.C., and Ruvkun, G. (2012a). The Caenorhabditis elegans RDE-10/RDE-11 complex regulates RNAi by promoting secondary siRNA amplification. Curr. Biol. *22*, 881–890.

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 (80-.). *359*, 587–592.

Zhang, F., Wang, J., Xu, J., Zhang, Z., Koppetsch, B.S., Schultz, N., Vreven, T., Meignin, C., Davis, I., Zamore, P.D., et al. (2012b). UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. Cell *151*, 871–884.

Zhang, H., Liu, K., Izumi, N., Huang, H., Ding, D., Ni, Z., Sidhu, S.S., Chen, C., Tomari, Y., and Min, J. (2017). Structural basis for arginine methylation-independent recognition of PIWIL1 by TDRD2. Proc. Natl. Acad. Sci. U. S. A. *114*, 12483–12488.

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

Zhao, H., Sun, Z., Wang, J., Huang, H., Kocher, J.-P., and Wang, L. (2014). CrossMap: a versatile tool for coordinate conversion between genome assemblies. Bioinformatics *30*, 1006–1007.

Zhao, S., Gou, L.-T.T., Zhang, M., Zu, L.-D.D., Hua, M.-M.M., Hua, Y., Shi, H.-J.J., Li, Y., Li, J., Li, D., et al. (2013). PiRNA-Triggered MIWI Ubiquitination and Removal by APC/C in Late Spermatogenesis. Dev. Cell *24*, 13–25.

Zhou, X., Xu, F., Mao, H., Ji, J., Yin, M., Feng, X., and Guang, S. (2014). Nuclear RNAi contributes to the silencing of off-target genes and repetitive sequences in Caenorhabditis

elegans. Genetics *197*, 121–132.

Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A.N., and Alva, V. (2018). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. J. Mol. Biol. *430*, 2237–2243.

Acknowledgements

Curriculum Vitae

Personal Data

Name: Ricardo José Cordeiro Machado Rodrigues

Nationality: Portuguese

Place and Date of Birth: Lisbon, 7th March 1988

Address: Raupelsweg 15, 55118 Mainz, Germany

Education

2010 - 2013	**Master of Science**	**Utrecht University, the Netherlands –** Master of Sciences in Biomedical Sciences - Cancer Genomics and Developmental Biology Graduated *Cum Laude*. Final GPA - 4.0 (on a scale of 0.0 to 4.0)
2006 - 2010	**Licenciatura** (Portuguese equivalent to BSc)	**University of Lisbon, Portugal –** Licenciatura in Biology, Branch of Molecular Biology and Genetics. Final Grade - 16 (on a scale of 1 to 20)

Research Experience

December 2012/ July 2019	**PhD Candidate** Institute for Molecular Biology, Mainz, Germany	**Project Title:** Searching and characterizing novel 21U RNA biogenesis factors in *Caenorhabditis elegans*. **Supervisor:** Prof. Dr. René F. Ketting
October 2011/ October 2012	**Minor Research project** (within MSc. program) - Cell Biology, Faculty of Sciences, Utrecht University	**Project Title:** LIS1/NudE1/NudEL1 and the Axon Initial Segment **Supervisor:** Prof. Dr. Casper Hoogenraad
September 2010/ September 2011	**Major Research project** (within MSc. program) **-** Hubrecht Institute for Developmental Biology and Stem Cell Research, Utrecht	**Project Title:** Hen1 Localization in Nuage. **Supervisor:** Prof. Dr. René F. Ketting

Scientific Publications

Cordeiro Rodrigues RJ, *et al.* (2019) "PETISCO is a novel protein complex required for 21U RNA biogenesis and embryonic viability". *Genes & Development*

Kuijpers M, *et al.* (2016) "Dynein Regulator NDEL1 Controls Polarized Cargo Transport at the Axon Initial Segment". *Neuron*

de Albuquerque BF, *et al.* (2014) "PID-1 is a novel factor that operates during 21U-RNA biogenesis in Caenorhabditis elegans". *Genes & Development*