Insights into the function and evolution of nematode RNAi pathways

Dissertation

Zur Erlangung des Grades

Doktor der Naturwissenschaften

Am Fachbereich Biologie

Der Johannes Gutenberg-Universität Mainz

Miguel Duarte Dias de Vasconcelos Almeida

Geb. am 03.11.1989 in Barreiro, Portugal

Mainz, 2019

Dekan:

1. Berichterstatter:
2. Berichterstatter:

Tag der mündlichen Prüfung:

[redacted]

Table of Contents

Summary

Argonaute proteins bind small RNAs (sRNAs) and together regulate gene expression across all domains of life. Multiple sRNA pathways are involved in controlling selfish genetic elements, like transposable elements (TEs). The nematode *Caenorhabditis elegans* expresses multiple classes of sRNAs, including 21U-RNAs, which are considered the Piwi-interacting RNAs (piRNAs) of *C. elegans*, as well as 26G- and 22G-RNAs, which are primary and secondary endogenous small interfering RNAs (endo-siRNAs), respectively.

GTSF1 proteins are evolutionarily conserved factors that are required for normal fertility and TE silencing in the animal germline, by physically interacting with Piwi Argonautes. Given the lack of conserved factors acting in the 21U-RNA/piRNA pathway, we wanted to dissect the role of the single GTSF-1 ortholog in *C. elegans*. To do this, we first created *gtsf-1* mutants using CRISPR/Cas9 technology. We found that *gtsf-1* mutants display fertility defects similar to its orthologs in mouse and flies. Surprisingly, we found that GTSF-1 is not required for TE silencing nor for 21U-RNA biogenesis or function. Instead, we have shown that GTSF-1 is required for the biogenesis of endo-siRNAs by interacting, via its CHHC zinc fingers, with the RNA-dependent RNA Polymerase RRF-3. Importantly, GTSF-1 is required for the assembly of the protein complex that assists RRF-3 in the biogenesis of endo-siRNAs. This work suggests that a common denominator of GTSF1 function may be to promote the assembly of multi-subunit effector complexes, in the context of sRNA pathways.

Upon the genetic dissection of GTSF-1 function, we uncovered a remarkable maternal effect in the transmission of its mutant phenotypes. We used high-throughput sequencing and genetics to characterize the sRNA and target mRNA dynamics subjacent to this maternal effect. We found that primary endo-siRNAs are maternally deposited to initiate secondary endo-siRNA production, which, in turn, will exert target gene silencing throughout development. Furthermore, we explored additional aspects of gene regulation by endo-siRNAs. We have shown that two redundant paralog Argonaute proteins, termed ALG-3 and ALG-4, which interact with primary endo-siRNAs, fine-tune their own expression in a negative feedback loop. Moreover, we identified several determinants of the regulatory outcome of ALG-3/4 targets.

The last facet of my PhD work concerns the characterization of GTSF-1 homologs in additional nematode species. We were driven to do so because of the apparent functional diversity of GTSF-1 proteins. To this end, we identified and mutated *gtsf-1* homologs in *C. briggsae* and *Pristionchus pacificus*. We found that *gtsf-1* is required for normal fertility in these two species, but is required for endo-siRNA biogenesis only in *P. pacificus*. Therefore, Cbr-GTSF-1 may not be required for the endo-siRNA pathway, in line with the robust evolutionary plasticity of GTSF-1 proteins.

Zusammenfassung

Argonautenproteine binden kleine RNS-Sequenzen (small RNAs - sRNAs) und regulieren, gemeinsam mit diesen sRNAs, die Genexpression in allen Domänen des Lebens. Mehrere sRNA-Wege sind an der Kontrolle von eigennützigen genetischen Elementen, wie den transponierbaren Elementen (TEs), beteiligt. Der Nematode *Caenorhabditis elegans* exprimiert mehrere Klassen von sRNAs, welche die 21U-RNAs sowie die 26G- und 22G-RNAs einschließen. 21U-RNAs gelten als Piwi-interagierende RNAs (piRNAs) von *C. elegans*. Die 26G- und 22G-RNAs sind primäre beziehungsweise sekundäre endogene kleine interferierende RNAs (Endo-siRNAs).

GTSF1-Proteine sind evolutionär konservierte Faktoren, die für die normale Fertilität und das TE-Silencing in der Keimbahn des Tieres erforderlich sind indem sie mit den Piwi-Argonautenproteinen interagieren. Angesichts des Mangels an konservierten Faktoren, die im 21U-RNA / piRNA-Weg wirken, wollten wir die Rolle des einzigen GTSF-1-Orthologen in *C. elegans* untersuchen. Zu diesem Zweck wurden zunächst *gtsf-1*-Mutanten mithilfe der CRISPR/Cas9-Technologie erstellt. Wir fanden heraus, dass *gtsf-1*-Mutanten Fertilitätsdefekte aufweisen, die den der Orthologen von Mäusen und Fliegen ähneln. Überraschenderweise stellten wir fest, dass GTSF-1 weder für das TE-Silencing noch für die 21U-RNA-Biogenese oder -Funktion benötigt wird. Stattdessen haben wir gezeigt, dass GTSF-1 für die Biogenese von Endo-siRNAs erforderlich ist, indem es über seine CHHC-Zinkfinger mit der RNA-abhängigen RNA-Polymerase RRF-3 interagiert. Zudem ist GTSF-1 an dem Aufbau des Proteinkomplexes beteiligt, der RRF-3 bei der Biogenese von Endo-siRNAs unterstützt. Diese Doktorarbeit deutet darauf hin, dass ein gemeinsamer Nenner der GTSF1-Funktion darin besteht, den Aufbau von großen Effektor-Komplexen im Zusammenhang mit sRNA-Wegen zu fördern.

Bei der genetischen Untersuchung der GTSF-1-Funktion haben wir einen bemerkenswerten maternalen Effekt bei der Übertragung seiner Mutanten-Phänotypen entdeckt. Wir verwendeten Hochdurchsatz-Sequenzierung und genetische Ansätze, um die Dynamik der sRNA und der Ziel-mRNA zu charakterisieren, die diesem maternalen Effekt zugrunde liegen. Wir fanden heraus, dass primäre Endo-siRNAs maternal hinterlegt werden, um die Produktion der sekundären Endo-siRNA zu initiieren. Diese wiederum bewirken während der gesamten Entwicklung die Stilllegung des Zielgens. Darüber hinaus untersuchten wir weitere Aspekte der Genregulation durch Endo-siRNAs. Wir haben gezeigt, dass zwei redundante paraloge Argonautenproteine, ALG-3 und ALG-4, die mit primären Endo-siRNAs interagieren, ihre eigene Expression in einer negativen Rückkopplungsschleife kontrollieren. Darüber hinaus haben wir mehrere Faktoren zur Regulierung von ALG-3/4-Zielgenen identifiziert.

Aufgrund der funktionellen Vielfalt der GTSF-1-Proteine, behandelt der letzte Teil dieser Doktorarbeit die Charakterisierung von GTSF-1-Homologen in weiteren Nematodenarten. Zu diesem Zweck haben wir *gtsf-1*-Homologe in *C. briggsae* und *Pristionchus pacificus* identifiziert und mutiert. Wir haben festgestellt, dass *gtsf-1* für die normale Fertilität in diesen beiden Spezies erforderlich ist, für die Endo-siRNA-Biogenese allerdings nur in *P. pacificus*. Angesichts der robusten evolutionären Plastizität von GTSF-1-Proteinen, ist Cbr-GTSF-1 möglicherweise nicht erforderlich für den Endo-siRNA-Weg.

List of Figures

List of Tables

List of Abbreviations and Gene Names

AGO	Argonaute
ALG	Argonaute-like gene
Cbr-GTSF-1	*C. briggsae* GTSF-1
Cel-GTSF-1	*C. elegans* GTSF-1
Cin-GTSF-1	*C. inopinata* GTSF-1
CLRC	Clr4-Rik1-Cul4 complex
CRISPR	Clustered regularly interspaced short palindromic repeats
CSR	Chromosome segregation and RNAi defective
DCL	Dicer-like
DCPM	Depth of coverage per base per million reads
DCR	Dicer related
DIC	Differential interference contrast microscopy
Dmel-Gtsf1	*Drosophila melanogaster* Gtsf1
DNMT	DNA methyltransferase
DRH	Dicer related helicase
dsRNA	Double-stranded RNA
EGO	Enhancer of glp-one
GFP	Enhanced green fluorescent protein
GTSF-1/Gtsf1	Gametocyte-specific factor 1
Endo-siRNA	Endogenous siRNA
EuAgo	Eukaryotic Argonaute
Eri	Enhanced RNAi
Exo-RNAi	Exogenous RNAi
FDR	False discovery rate
GTSF	Gametocyte-specific factor
HEN1	HUA ENHANCER 1
HENN	HEN-1 of nematode
HMT	Histone methyltransferase
HP1	Heterochromatin protein 1
HPL	HP1 like
HRDE	Heritable RNAi deficient
H3K9me2 and me3	Histone H3 lysine 9 di- or trimethylation
IP	Immunoprecipitation

KRAB	Kruppel-associated box
LECA	Last eukaryotic common ancestor
lncRNA	Long non-coding RNA
MID	Middle
miRISC	Micro RNA RISC
miRNA	Micro RNA
Mmus-GTSF1	*Mus musculus* GTSF1
mRNA	Messenger RNA
MUT	Mutator
NRDE	Nuclear RNAi defective
PARN	Poly(A)-specific ribonuclease
PATCs	Periodic A_n/T_n clusters
PAZ	Piwi Argonaute and Zwille
PCR	Polymerase chain reaction
PID	piRNA-induced silencing defective
PIR	Phosphatase interacting with RNA/RNP
piRNA	Piwi-interacting RNA
Piwi	P-element induced Wimpy testis
Ppa-PIR-1::GTSF-1	*Pristionchus pacificus* PIR-1::GTSF-1
PRDE	piRNA-dependent silencing defective
PRG	Piwi related gene
ProAgo	Prokaryotic Ago
PTGS	Post-transcriptional gene silencing
RDE	RNAi defective
RDRC	RNA-dependent RNA Polymerase complex
RdRP	RNA-dependent RNA Polymerase
RIP	RNA immunoprecipitation
RISC	RNA-induced silencing complex
RITS	RNA-induced transcriptional silencing complex
RNAe	RNA-induced epigenetic silencing
RNAi	RNA interference
RNA PolII	RNA Polymerase II
RPKM	Reads per kilobase million
RPM	Reads per million
RRF	RNA-dependent RNA Polymerase family

RT-PCR	Reverse transcription PCR
RT-qPCR	Reverse transcription quantitative PCR
sgRNA	Single guide RNA
siRNA	Small interfering RNA
SNPC	Small nuclear RNA activating complex
snRNA	Small nuclear RNA
sRNA	Small RNA
Stc	siRNA to chromatin
TE	Transposable element
TGS	Transcriptional gene silencing
TOFU	Twenty one u-RNA biogenesis fouled up
USTC	Upstream sequence transcription complex
WAGO	Worm Argonaute

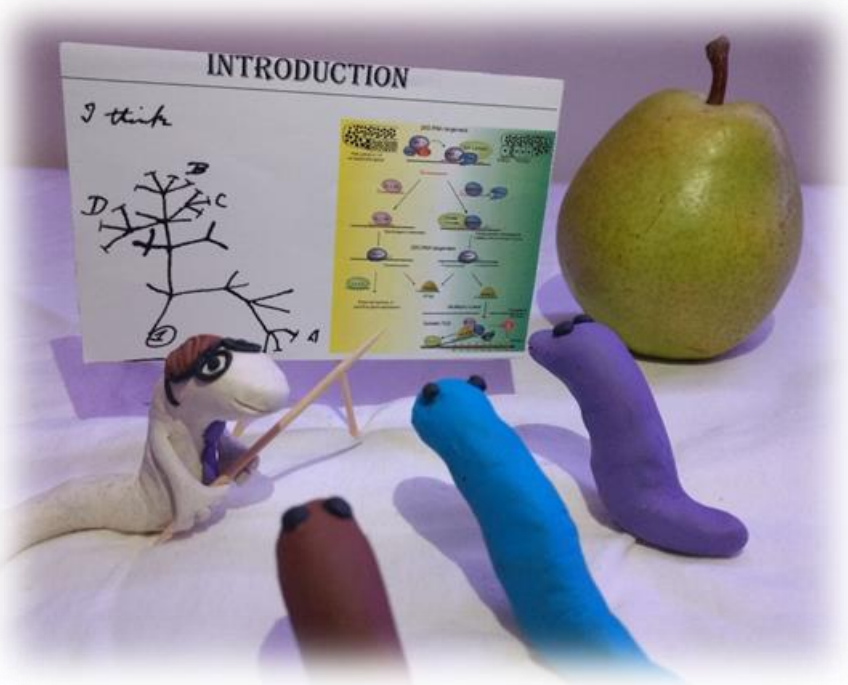
Chapter I

Introduction

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

Almeida, M.V., Andrade-Navarro M., and Ketting R.F. (2019). Function and evolution of nematode RNAi pathways. Noncoding RNA *5*, 1-24.

Miguel Andrade-Navarro constructed the phylogenetic tree shown in **Figure I.3C**.



Managing genetic conflicts

The emergence of life and genetic conflict

It is believed that an RNA world was the first, or one of the first stages of life on Earth (Pressman et al., 2015). The main arguments supporting this view lie in the intrinsic self-replicating capability of several RNA molecules, as well as in the presumable existence, in a primordial Earth, of all chemical building blocks required to produce RNA. One can thus envision the first proto-cell consisting of a self-replicating RNA molecule encased in a lipid bilayer. At that moment, natural selection came into work: cells divided, and from then on cells that could divide faster and better than others would have an advantage and grow in numbers, while less fit cells would divide less and eventually perish (Dawkins, 1976). A parasitic nucleic acid that could hitchhike on a successfully diving cell, would also increase its frequency in the population of cells. It is perhaps appealing to think that the birth of genetic conflict happened in parallel to, or shortly after, the emergence of life in an RNA world. And to this day, life and genetic conflict go hand in hand.

Mechanisms of self versus non-self recognition

At any given point, organisms are engaged in a relentless fight for survival. If a threat comes in the form of a pathogen, organisms fight back using innate or adaptive immune systems. A key feature of immune systems is the ability to recognize the self from the non-self in order to protect the self, while ultimately neutralizing the non-self (Janeway, 1992).

At the nucleic acid level, these non-self invaders take the form of transposable elements (TEs) and viruses. These are considered selfish genetic parasites, since they propagate at the expense of the host, often resulting in a decrease in host fitness (Ågren and Clark, 2018). As a result, these selfish elements can be highly detrimental to the host and thus represent a source of genetic conflict. A great diversity of pathways has evolved to hamper the propagation of such elements. For example, in prokaryotes, restriction-modification systems and a myriad of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) pathways control invasions by bacteriophages and plasmids (Koonin et al., 2017). In vertebrates, RIG-1-type I interferon (IFN) pathways are important antiviral surveillance platforms (Schlee and Hartmann, 2016). In tetrapods, Krüppel-associated box (KRAB) zinc-finger proteins are a major class of DNA-binding transcription factors that specifically bind and repress TEs (Yang et al., 2017). Collectively, small RNA (sRNA)-based RNA interference (RNAi)-like pathways are heavily studied platforms of gene regulation. These pathways exist in all clades of life and constitute an important defensive line against non-self genetic elements (Borges and Martienssen, 2015; Holoch and Moazed, 2015; Ketting, 2011; Ozata et al., 2018; Swarts et al., 2014).

It has been hypothesized that protection against invading selfish genetic elements actually stimulated the emergence of increasingly more complex modes of gene regulation in eukaryotes (Madhani, 2013). Indeed, it is appealing to think of chromatin condensation as a denier of DNA access, of the nucleus as a physical gate and of RNA modifications as self "identity cards". In this view, these, amongst many other steps in the normal life of a gene, constitute incredibly complex and often redundant processes, which simultaneously assure bona fide self gene expression and work as hurdles for selfish genetic elements. All such hurdles may have evolved as a result of a constant arms race between non-self genetic elements and their hosts, consistent with the "Red Queen" hypothesis (Madhani, 2013; Pearson, 2001). A simplified, modern interpretation of the "Red Queen" hypothesis predicts that, for example, a positive adaptation of a predator species will favor the evolution of a counteradaptation in the prey species, or vice-versa. As a consequence of their vanguard position in combating non-self sequences, RNAi-like pathways are evolving very fast (Obbard et al., 2009; Palmer et al., 2018; Simkin et al., 2013).

A brief history of RNAi

The "Central Dogma" of Molecular Biology proposed a linear flow of information from DNA, to RNA and finally to protein (Crick, 1970). By this time, RNA was considered as a fairly inert intermediate in the information flow. The road to RNAi was paved by an enormous body of work, which ultimately changed the fundamental way we look at RNA molecules. In its origin are studies that increasingly explored the special properties of RNA, for instance the remarkable discoveries that RNAs have enzymatic activity and regulatory regions (Caughman et al., 1988; Kruger et al., 1982; Parkin et al., 1988; Stark et al., 1978).

Several independent studies pointed at a role for RNA in mediating gene silencing, including important groundwork in plants and in the fungi *Neurospora Crassa* (Napoli et al., 1990; Romano and Macino, 1992). RNA molecules in an antisense (Fire et al., 1991; Guo and Kemphues, 1995; Izant and Weintraub, 1984; Rosenberg et al., 1985) and sense (Guo and Kemphues, 1995) orientation, relative to coding RNA, were shown to be able to induce gene silencing independently, arguing against a mechanism merely mediated by base-pairing of complementary sequences. Fire and Mello, using the nematode *Caenorhabditis elegans* as a model system, were the first to show that double-stranded RNA (dsRNA) was in fact the molecule triggering silencing, and coined the term RNAi (Fire et al., 1998). It turned out that the contradicting results observed by many others before were not due to sense or antisense RNA per se, but due to residual contamination with dsRNA. These new insights, obtained using *C. elegans*, preluded the swift discovery of RNAi in plants, fungi and other animals, highlighting

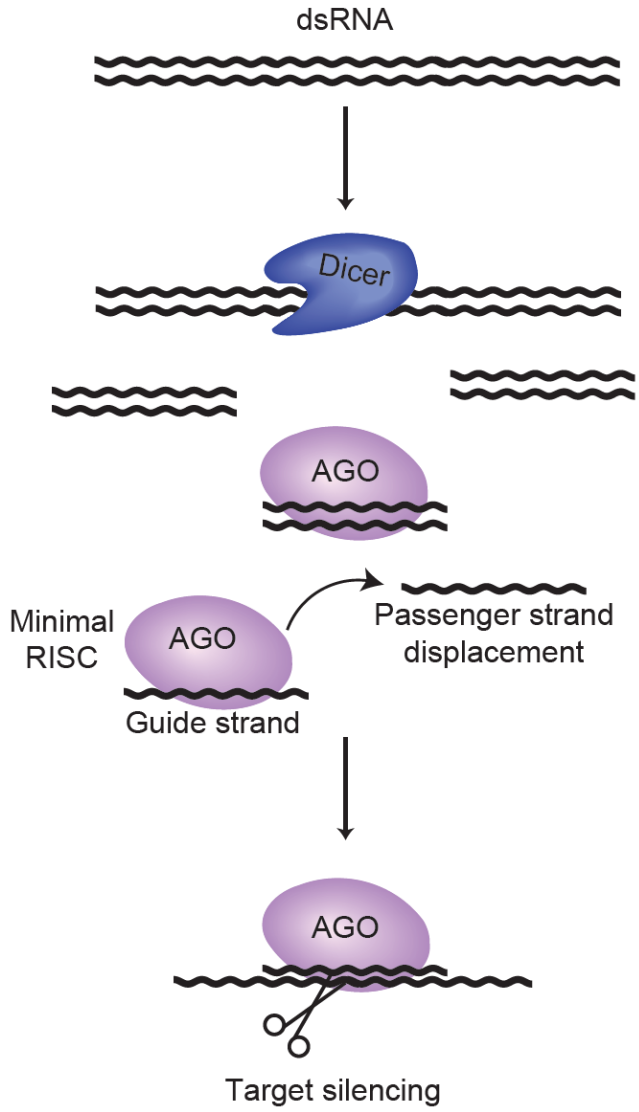


Figure I.1. **Stereotypical RNAi pathway**. dsRNA, either from endogenous or exogenous sources, is processed by Dicer enzymes into shorter duplexes that are bound by AGO proteins. The AGO protein utilizes its catalytic activity to displace the passenger strand. The subsequent complex of an AGO protein plus the retained guide strand comprise a minimal RISC. The sRNA guides the AGO protein to RNA targets with sequence complementarity. The AGO protein cleaves the target, or recruits other factors, ultimately resulting in gene silencing. AGO, Argonaute; dsRNA, double-stranded RNA; RISC, RNA-induced silencing complex.

its conserved nature (Elbashir et al., 2001; Kennerdell and Carthew, 1998; Volpe et al., 2002; Wargelius et al., 1999; Waterhouse et al., 1998).

Seminal groundwork established a stereotypical core RNAi mechanism (**Figure I.1**). Long dsRNA is processed by Dicer enzymes into shorter 21-23 nucleotide duplexes (Bernstein et al., 2001; Zamore et al., 2000), which associate with an Argonaute (AGO) protein (Hammond et al., 2001; Tabara et al., 1999). The AGO protein cleaves the duplex and remains associated with one of the strands, the so-called guide strand, and releases the other strand, termed the passenger strand (Martinez et al., 2002; Matranga et al., 2005; Miyoshi et al., 2005; Rand et al., 2005; Steiner et al., 2009). The guide strand directs its partner AGO protein to its RNA targets by base complementarity. Once a target is recognized, the AGO may cleave the target and/or recruit other inhibitory factors, thus resulting in target silencing (Ender and Meister, 2010; Hutvagner and Simard, 2008; Liu et al., 2004; Rand et al., 2004; Rivas et al., 2005; Zamore et al., 2000). The 21-23 nucleotide sRNA and the AGO protein cofactors comprise the minimal core unit required for RNAi, also termed as RNA-induced silencing complex (RISC) (Caudy and Hannon, 2004; Liu et al., 2004; Rand et al., 2004; Rivas et al., 2005). In spite of illustrating key principles, a prototypic RNAi pathway as that shown in **Figure I.1** is not universally representative. In fact, many sRNA pathways have diverged and specialized in multiple ways.

Function and evolution of core RNAi factors

It was proposed that eukaryotic RNAi factors used phage and prokaryotic protein modules that subsequently radiated greatly (Shabalina and Koonin, 2008). The last eukaryotic common ancestor (LECA) was believed to possess a minimal RNAi pathway comprised of an RNA-dependent RNA Polymerase (RdRP), which produced dsRNA from a single-stranded RNA target, a Dicer enzyme to process dsRNA intermediates into smaller duplexes, and two AGO proteins to execute gene silencing, one of the Ago-clade and another of the P-element induced Wimpy testis (Piwi) clade (Cerutti and Casas-Mollano, 2006; Shabalina and Koonin, 2008). This RNAi protein complement present in the LECA is thought to have been an ancestral eukaryotic defense platform against non-self nucleic acids like viruses and TEs (Obbard et al., 2009; Shabalina and Koonin, 2008). Even though eukaryotic RNAi-like pathways have expanded and diversified profusely, existing pathways maintain roles in TE and viral silencing. Next, the players that take center stage in RNAi-like pathways will be introduced: sRNAs, AGO proteins, Dicer enzymes and RdRPs.

sRNA classes and biogenesis mechanisms

At the core of RNAi-like pathways lie sRNAs that are single-stranded and shorter than 35 ribonucleotides. A plethora of sRNA classes have been described and it is plausible that many more are yet to be uncovered. The most prominent and best characterized classes are Piwi-interacting RNAs (piRNAs), micro RNAs (miRNAs), and small interfering RNAs (siRNAs). piRNA precursors are single-stranded, thus rendering piRNA biogenesis independent of Dicer enzymes (Huang et al., 2017; Ozata et al., 2018). Conversely, the biogenesis of miRNAs and siRNAs involves the cleavage of dsRNA into shorter duplexes by Dicer proteins (Ender and Meister, 2010; Hutvagner and Simard, 2008; Ketting, 2011). dsRNAs that are used by Dicer enzymes can originate from exogenous or endogenous sources (**Figure I.2i-2vi**). Endogenous dsRNA can arise, for example, from structured *loci* that fold back by intramolecular base-pairing (**Figure I.2iii**); from *loci* undergoing convergent or bidirectional transcription (**Figure I.2iv**); and by base-pairing *in trans* of homologous RNA molecules, like a gene-pseudogene pair (**Figure I.2v**) (Ghildiyal and Zamore, 2009; Ketting, 2011). In addition, some eukaryotes encode RdRPs that can synthesize RNA antisense to a pre-existing template, resulting in a dsRNA substrate for Dicer (**Figure I.2vi**) (Ghildiyal and Zamore, 2009; Ketting, 2011).

Under an evolutionary standpoint, it is harder to reconcile the ability of several organisms to acquire and respond to exogenous dsRNA from their environment, the same ability that makes RNAi an invaluable tool to manipulate endogenous gene expression. It has been proposed that this ability may provide an RNA-based communication system between organisms, of the same or distinct species (Braukmann et al., 2017; Sarkies and Miska, 2013). In this view, the acquisition of exogenous dsRNA

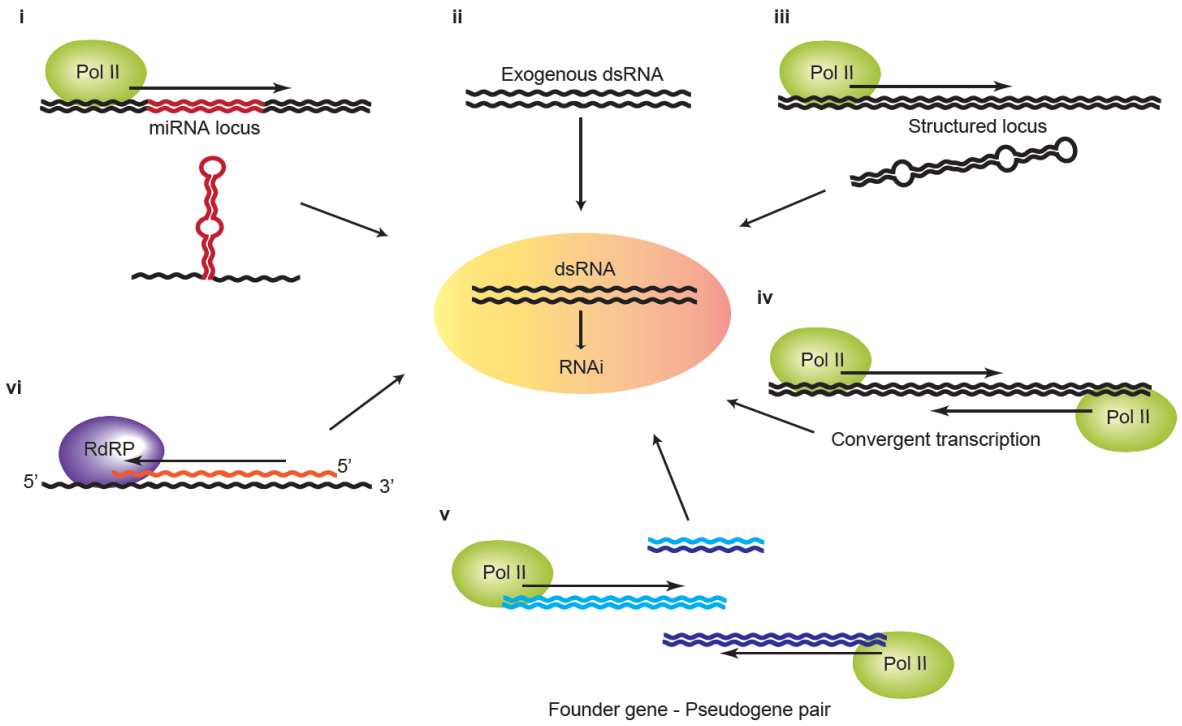


Figure I.2. **Sources of dsRNA**. Schematics depicting possible sources of dsRNA that can initiate RNAi. (i) miRNA transcripts form stem-loop structures that are processed into Dicer dsRNA substrates. (ii) exogenous dsRNA can, in some organisms, be absorbed from the environment and processed by Dicer enzymes. (iii) *Loci* that form secondary structures can be Dicer substrates. (iv) Convergent transcription generates dsRNAs. (v) Transcripts from homologous sequences, like founder genes and related pseudogenes, may anneal thereby originating a substrate for Dicer enzymes. (vi) RdRPs are specialized Polymerases that produce complementary RNA from a template RNA molecule. RdRPs have two distinct modes of biogenesis, processive and non-processive, see Figure I.4. dsRNA, double-strand RNA; Pol II, RNA Polymerase II; RdRP, RNA-dependent RNA Polymerase; RNAi, RNA interference.

would allow organisms to quickly integrate signals from their surrounding environment, creating potentially adaptive gene expression changes.

Argonaute proteins

AGO proteins are, without a doubt, the most important cofactors of sRNAs. The resemblance between the *Argonauta argo* octopus and the phenotype of *Arabidopsis thaliana* AGO1 mutants, gave the AGO family its name. AGO protein homologs have been identified in all realms of life, from Bacteria and Archaea to humans, supporting an ancient origin and functional importance (Olina et al., 2018; Swarts et al., 2014). Eukaryotic AGOs (EuAGOs) have been characterized in multiple plants, fungi and animals, where they have been shown to be involved in a myriad of processes. However, the first functional studies on prokaryotic AGOs (ProAGOs) have only emerged more recently (Olina et al., 2018; Swarts et al., 2014).

EuAGOs display four characteristic, conserved domains (**Figure I.3A**): N-terminal; Piwi-Argonaute-Zwille (PAZ); middle (MID); and PIWI (Olina et al., 2018; Swarts et al., 2014). Many structural and biochemical studies have elucidated the function of these domains. The N-terminal domain is

required for target RNA cleavage and to unwind RNA duplexes bound to AGOs (Kwak and Tomari, 2012). The PAZ domain provides a binding pocket for the 3' end of the sRNA (Lingel et al., 2003; Song et al., 2004; Yan et al., 2003), while the MID domain provides a binding pocket for the 5' phosphate of the sRNA (Frank et al., 2010). The PIWI domain has an RNase H fold, which comprises the catalytic center of the AGO protein (Parker et al., 2004; Rivas et al., 2005; Song et al., 2004). A conserved tetrad of Asp-Glu-Asp-His/Asp (DEDH/D) amino acid residues in the RNase H fold is required for target cleavage (Rivas et al., 2005; Song et al., 2004). Lack of this amino acid combination renders AGO proteins catalytically inactive. Lastly, there are two linker domains: L1, between the N-terminal and the PAZ domains, and L2, between the PAZ and MID domains (Olina et al., 2018; Swarts et al., 2014). Overall, AGOs display a bilobed architecture, with one lobe including the N, L1 and PAZ domains, and another lobe with the MID and PIWI domains (**Figure I.3B**). The L2 linker connects both lobes (Olina et al., 2018; Swarts et al., 2014).

EuAGOs comprise a broad family of proteins that can be subdivided into three clades according to sequence homology (**Figure I.3C**): the Ago clade, the Piwi clade and the worm-specific (Wago) clade (Olina et al., 2018). Ago clade AGOs are ubiquitously expressed and interact with miRNAs or siRNAs. Piwi clade AGOs interact with piRNAs, typically in the animal germline, and regulate TEs. Wago clade AGOs are present in nematodes, bind to secondary siRNAs termed 22G-RNAs, and regulate a variety of processes, in some cases redundantly. Most eukaryotic genomes encode a varying number of AGO proteins of both the Ago and Piwi clades (Olina et al., 2018). EuAGOs were lost in rare cases, like in the budding yeast *Saccharomyces cerevisiae*. Of note, there are major expansions of AGO proteins in plants and nematodes, exemplified by 19 and more than 25 AGOs in rice and in *Caenorhabditis nematodes*, respectively.

EuAGO proteins can localize both in the nucleus and in the cytoplasm, orchestrating transcriptional gene silencing (TGS) or post-transcriptional gene silencing (PTGS), respectively (Ketting, 2011; Olina et al., 2018). When in the cytoplasm, AGOs can be commonly found in RNA- and protein-rich membrane-less granules, often in the perinuclear region, that are considered RNA processing factories (Ketting, 2011). An extensive description of these granules, as well as the dynamic interplay between distinct types of granules, goes beyond the scope of this thesis. Therefore, from now on RNAi reactions will be described as nuclear or, for simplicity, cytoplasmic.

The little that is known about ProAGOs suggests roles in foreign nucleic acid recognition and silencing, much like EuAGOs (Olina et al., 2018; Swarts et al., 2014). These observations are in line with eukaryotic RNAi evolving from an ancient prokaryotic system to defend against foreign genetic elements (Shabalina and Koonin, 2008; Swarts et al., 2014). The current knowledge on ProAGOs illuminates several differences to their eukaryotic counterparts (Swarts et al., 2014). A crucial difference is that ProAGOs can bind both RNA and DNA guides and are believed to target mostly DNA.

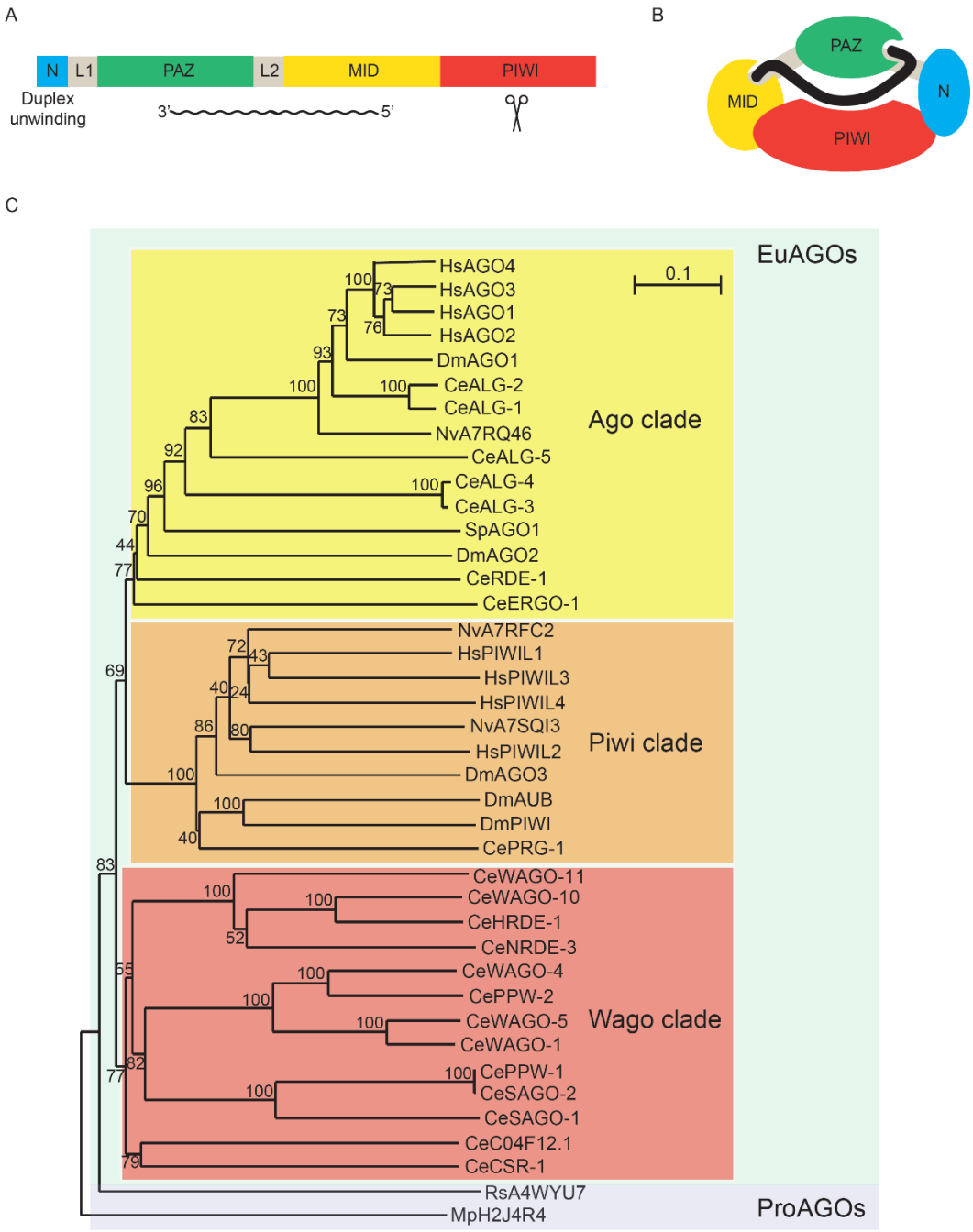


Figure I.3. **Structure and phylogeny of AGO proteins**. (A) Schematics depicting the domain organization of AGO proteins. The PAZ domain binds the 3' end of the sRNA, whereas the MID domain binds to the 5' end. The PIWI domain is the catalytic center of the protein, represented by scissors. (B) 3D AGO architecture. sRNA is represented in black. (C) Phylogenetic tree depicting the three main eukaryotic AGO clades. To provide a broad phylogenetic representation across eukaryotes, the AGO proteins of *C. elegans* were aligned with those of human, fruit fly, fission yeast and starlet sea anemone. For tractability, we chose not to include plant AGOs. Two prokaryotic AGOs were used to root the tree. A multiple sequence alignment of AGO sequences was constructed using MUSCLE (Edgar, 2004). The phylogenetic tree was constructed with ClustalW (Larkin et al., 2007), excluding positions with gaps and represented with NJPlot (Perrière and Gouy, 1996). Numbers indicate bootstrapping values. AGOs are represented by a two letter prefix indicating the species, followed by the AGO name or UniProt ID. Species legend: Ce, *Caenorhabditis elegans*; Dm, *Drosophila melanogaster*; Hs, *Homo sapiens*; Mp, *Marinitoga piezophila*; Nv, *Nematostella vectensis*; Rs, *Rhodobacter sphaeroides*; Sp, *Schizosaccharomyces pombe*.

Also, ProAGOs show striking structural diversity. ProAGOs share all the known domains present in EuAGOs, but often display extensive rearrangements. There are long ProAGOs containing all the domains, as well as shorter ProAGOs with only MID and PIWI domains (Swarts et al., 2014). Interestingly, most short ProAGO genes are co-expressed in the same operon with nucleases, helicases or genes with domains analogous to the PAZ domain. Thus, it is likely that these proteins assist or complement short ProAGOs in accomplishing gene silencing (Swarts et al., 2014).

Dicer proteins

Dicer enzymes are endoribonucleases that associate with, and cleave dsRNAs into shorter dsRNAs that can be loaded onto an AGO protein. The catalytic core of Dicer proteins consists of two RNase III domains and a PAZ domain (MacRae et al., 2006). The PAZ domain accommodates the 3' end of the dsRNA, while each of the RNase III domains cleaves one strand of the duplex (Zhang et al., 2004). Dicer proteins act as molecular rulers, i.e. specific Dicers display specific cleavage product lengths (Ha and Kim, 2014; MacRae et al., 2006; Zhang et al., 2004). The length of the dsRNA products is defined by the distance between the RNase III and PAZ domains. Dicer enzymes may also contain other domains that further promote RNA-binding and provide interaction surfaces with other cofactors (Mukherjee et al., 2013). Many Dicers have, for example, a DEAD-box helicase domain, believed to be required for the movement of Dicer along dsRNAs.

Although its separate protein modules can be found in Bacteria and Archaea, Dicer enzymes are absent in prokaryotes (Shabalina and Koonin, 2008). It is therefore believed that Dicers had an early eukaryotic origin by recombination of preexisting domains. Dicer enzymes underwent moderate expansions in some eukaryotic groups, and were lost in other groups that completely lack functional RNAi pathways (Mukherjee et al., 2013). Phylogenetic analysis indicates that Dicers have independently diversified in animal and plant lineages (Cerutti and Casas-Mollano, 2006). Plant genomes encode four distinct Dicer-like genes that show functional specification in miRNA or antiviral siRNA pathways (Borges and Martienssen, 2015; Henderson et al., 2006). Likewise, insects have two Dicers with distinct functions: while Dicer1 is involved in miRNA processing, Dicer2 is dedicated to antiviral RNAi (Lee et al., 2004). Dicer duplication occurred early in metazoans, with subsequent loss of Dicer2 in lineages bearing other antiviral mechanisms (Mukherjee et al., 2013). However, the single Dicer gene of vertebrates and nematodes can code for multiple isoforms with diverging or complementing functions (Flemr et al., 2013; Mukherjee et al., 2013; Sawh and Duchaine, 2013).

RNA-dependent RNA Polymerases

Of the core RNAi factors present in the LECA, RdRPs are the least studied. RdRPs catalyze primer-independent synthesis of RNA, antisense to an RNA template (Aoki et al., 2007; Makeyev and Bamford, 2002; Pak and Fire, 2007; Sijen et al., 2007). RdRP activity thus generates dsRNA intermediates that are Dicer substrates, ultimately producing sRNAs (**Figure I.4A**). Alternatively, nematode RdRPs can produce sRNAs in a non-processive manner, thereby generating sRNAs without the participation of Dicer enzymes (Sarkies et al., 2015) (**Figure I.4B**). It is likely that extant RdRPs, occurring in viruses and eukaryotes, originated from bacteriophage DNA-dependent RNA Polymerase proteins (Iyer et al., 2003; Salgado et al., 2006). Eukaryotic RdRPs are found in plants, fungi, nematodes and some Arthropod groups (Borges and Martienssen, 2015; Holoch and Moazed, 2015; Lewis et al., 2017; Sarkies et al., 2015). Organisms that lack RdRPs, like most Arthropod groups and vertebrates, use multiple other mechanisms for generating dsRNA. Supporting this claim, arthropods lacking RdRPs have increased antisense transcription of TEs, which can presumably base-pair with homolog sense transcripts leading to the production of sRNAs (Lewis et al., 2017).

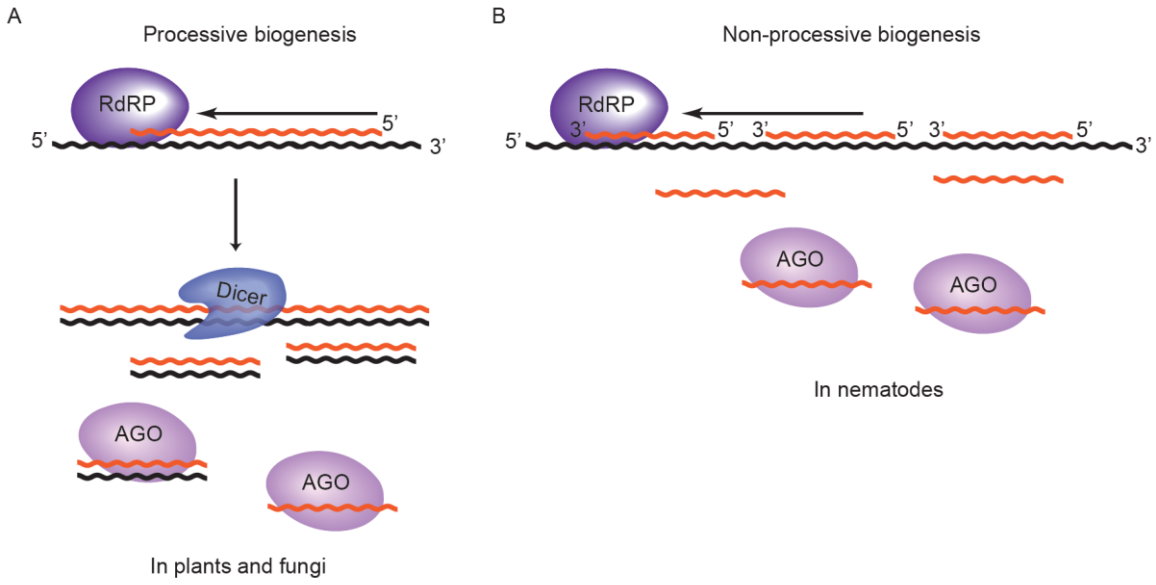


Figure I.4. **Two distinct modes of RdRP biogenesis**. (A) RdRPs can synthesize a long RNA molecule from an RNA template in a processive manner. This dsRNA intermediate is a substrate for Dicer enzymes and leads to the production of sRNAs. Plant and fungi RdRPs display such activity. (B) In nematodes, RdRPs can directly synthesize sRNAs in a non-processive biogenesis mode that bypasses the requirement for Dicer activity. RdRP, RNA-dependent RNA Polymerase.

General aspects of RNAi-like pathways

miRNA pathways

miRNAs are a prolific class of small RNAs, broadly present in animals, plants, protists, and viruses (Bartel, 2018; Ha and Kim, 2014). Absence of individual miRNAs, or of multiple miRNAs

simultaneously, may lead to a broad range of defects, illustrating their contribution to normal development and physiology. Despite such concrete phenotypes associated with some miRNAs, loss-of-function of many miRNAs does not lead to any discernible phenotypes. Their lack of obvious phenotypes in the face of strong conservation seems somewhat incoherent. This is believed to stem from a broad function of miRNAs in fine-tuning gene regulatory networks on the post-transcriptional level (Bartel, 2018). In fact, the function of particular miRNAs may only become apparent under certain stress conditions.

miRNA genes are transcribed by RNA Polymerase II (RNA Pol II) from dedicated *loci*, typically within introns of coding or non-coding transcription units (**Figure I.2i**). Primary miRNA transcripts contain a stem-loop structure, which serves as substrate for Dicer (Bartel, 2018; Ha and Kim, 2014). Dicer cleavage generates a 21-25 ribonucleotide long dsRNA duplex that is loaded onto an AGO protein of the Ago clade. A mature miRNA is created after the AGO ejects the passenger strand. The retained guide strand is selected by its thermodynamic stability (Martinez et al., 2002; Matranga et al., 2005; Miyoshi et al., 2005; Rand et al., 2005; Steiner et al., 2009). miRNA-RISCs generally target the 3' untranslated region (UTR) of mRNAs. The so-called seed region, comprising positions 2 to 7 of the miRNA, is most important for target recognition. Upon target recognition, miRNA-RISCs repress the translation of target mRNAs, often by eliciting their deadenylation and degradation. However, the exact mechanism of translation repression by miRNA-RISCs remains a matter of great controversy (Bartel, 2018; Ha and Kim, 2014).

Endogenous siRNA pathways

Endogenous siRNA (endo-siRNA) pathways comprise a very broad class of sRNA pathways across the eukaryotic kingdoms (Borges and Martienssen, 2015; Holoch and Moazed, 2015; Kim et al., 2009). Below, I will provide an overview on the current understanding of the more intensely studied endo-siRNA pathways.

The fission yeast *Schizosaccharomyces pombe* has served as a paradigm for RNAi studies. Within the nucleus of *S. pombe*, siRNAs direct heterochromatin formation on pericentromeric repeat sequences (Holoch and Moazed, 2015). A minimal genomically encoded RNAi complement is at the heart of this process: one AGO, one Dicer gene, and one RdRP, termed *ago1*$^{+}$, *dcr1*$^{+}$, and *rdp1*$^{+}$, respectively. In brief, an RNA-dependent RNA Polymerase complex (RDRC) containing Rdp1, creates dsRNA from long transcripts originating from the pericentromeric region (Motamedi et al., 2004). Next, Dcr1 processes the dsRNA into siRNAs, which, in turn, associate with Ago1 within the RNA-induced transcriptional silencing complex (RITS) (Bühler et al., 2006; Verdel et al., 2004). The RITS then targets newly transcribed pericentromeric transcripts and recruits the Clr4-Rik1-Cul4 (CLRC) complex, which deposits H3K9 di- and tri-methyl (H3K9me2 and H3K9me3) in pericentromeric regions (Bühler et al.,

2006; Horn et al., 2005; Jia et al., 2005; Nakayama et al., 2001). Clr4 is critical in this complex because it constitutes the single genomically encoded H3K9 methyltransferase in fission yeast. Another key element in the RNAi pathway of *S. pombe* is Stc1, a small protein with a LIM domain, which consists of two tandem zinc-fingers that can mediate protein-protein interactions (Kadrmas and Beckerle, 2004). Stc1 can interact with Clr4 and Ago1, thereby bridging the transcript-bound RITS with the CLRC (Bayne et al., 2010; He et al., 2013).

Another prolific endo-siRNA field studies the incredibly complex sRNA pathways of plants (Axtell, 2013; Borges and Martienssen, 2015). The genome of *Arabidopsis thaliana* encodes four Dicer-like (DCL) proteins (Gasciolli et al., 2005; Henderson et al., 2006), six RdRPs (Willmann et al., 2011), ten AGOs, and two plant-specific RNA Pol II-related RNA polymerases, termed PolIV and PolV (Borges and Martienssen, 2015). Several distinct subpopulations of sRNAs have been identified and can be distinguished by: 1) the origin of their precursor dsRNA; 2) their cofactors, namely the RdRP and DCL involved in their processing, and the AGO with which the mature sRNA will associate with; 3) the length of the mature sRNA; and 4) the ultimate target regulatory outcome, be it PTGS or TGS (Axtell, 2013; Borges and Martienssen, 2015).

Endo-siRNAs can be produced from hybridized sense and antisense transcripts, structured *loci*, and as a product of RdRP activity (Axtell, 2013). TE transcripts are abundantly triggering endo-siRNA biogenesis and their regulation can switch from PTGS to TGS in ways that are still not clear. This switch occurs when there is a change in the sRNAs being produced: from Pol II-driven to PolIV- and PolV-driven (Borges and Martienssen, 2015; Matzke and Mosher, 2014). RNA-directed DNA methylation (RdDM) at genomic target *loci* is a hallmark of the switch to TGS (Matzke and Mosher, 2014). Of note, the TE sequences that undergo TGS are enriched in pericentromeric regions, thus defining a parallel between RdDM and RNAi-directed heterochromatin formation in *S. pombe*, as described above. Plant sRNA pathways are critical for genome defense, by controlling TEs and multiple other less-than-optimal transcripts. Overall, these sRNA pathways have been implicated in several processes essential for plant development, such as gametogenesis and embryogenesis (Borges and Martienssen, 2015).

Gene regulation by self-reinforcing positive feedback loops is a recurrent theme of RNAi in plants and *S. pombe*. sRNA biogenesis stimulates the formation of heterochromatin by H3K9me3 deposition in *S. pombe* and DNA methylation in plants (**Figure I.5**). In turn, sRNA biogenesis from these heterochromatic regions is further stimulated through interactions between heterochromatin-associated factors and the sRNA biogenesis machinery (Borges and Martienssen, 2015; Holoch and Moazed, 2015). These positive feedback mechanisms allow faithful maintenance of heterochromatin at pericentromeric regions and TEs.

In the fruit fly *Drosophila melanogaster*, endo-siRNAs are produced from dsRNA intermediates originating from convergent transcription and structured *loci* (**Figure I.2iii-2iv**) (Chung et al., 2008;

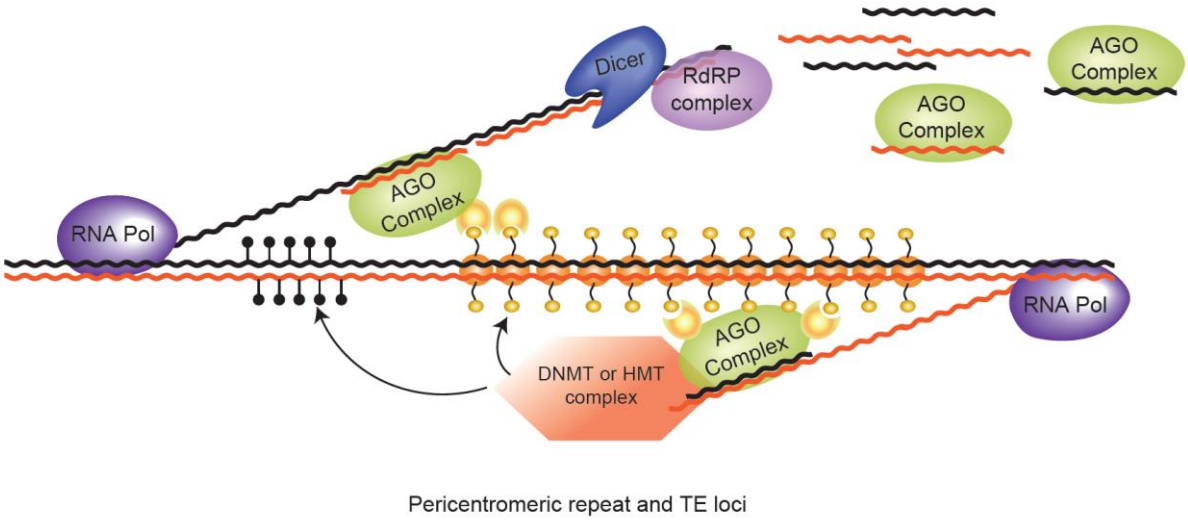


Figure I.5. **Endo-siRNA-driven positive feedback loops**. Depiction of the stereotypical nuclear reactions occurring in the endo-siRNA pathways of plants and fission yeast. Dicer enzymes and RdRP complexes synthesize sRNAs from template transcripts. The sRNAs subsequently interact with AGO proteins that recruit DNMT and/or HMT complexes that methylate DNA and deposit chemical modifications to histone tails, respectively. These modifications to DNA and histones promote heterochromatin formation that will consequently stimulate the production of more sRNAs. In plants, such RNA-directed DNA methylation mechanism is employed to silence a variety of targets, including TEs. In fission yeast, sRNAs drive the assembly of heterochromatin at pericentromeric repeats. AGO, Argonaute; DNMT, DNA methyltransferase; HMT, histone methyltransferase; RdRP, RNA-dependent RNA Polymerase; RNA Pol, RNA Polymerase.

Czech et al., 2008; Ghildiyal et al., 2008; Kawamura et al., 2008; Okamura et al., 2008). Dicer2 processes these dsRNAs that will subsequently associate with Ago2. The resulting Ago2-RISCs target protein-coding genes and TEs in the soma and in the germline. TE targeting is thought to provide a somatic silencing platform for TEs in the absence of the piRNA pathway. siRNAs originating from testes-expressed hairpin RNAs are required for normal spermatogenesis and male fertility (Wen et al., 2015). Interestingly, these hairpin-derived siRNAs are de-escalating genetic conflicts in sperm, by controlling selfish genetic elements engaged in meiotic drive systems (Lin et al., 2018).

Endo-siRNAs have been also described in mammals, both in mouse oocytes (Tam et al., 2008; Watanabe et al., 2006, 2008) and in human cells, for example in HeLa and human osteosarcoma 143B TK^- (Yang and Kazazian Jr., 2006). Mammalian endo-siRNAs are profusely expressed and better characterized in mouse oocytes. Much like in *Drosophila*, mammalian endo-siRNAs depend on Dicer and Ago2 for their biogenesis, and are generated from repetitive sequences, especially TEs, structured *loci*, or *loci* transcribed in a convergent manner (Flemr et al., 2013; Tam et al., 2008; Watanabe et al., 2008). Interestingly, pseudogenes are a strong source of oocyte endo-siRNAs, potentially as a result of dsRNA intermediates of hybridized pseudogene and founder gene transcripts (**Figure I.2**) (Tam et al., 2008; Watanabe et al., 2008). This is functionally relevant since these endo-siRNAs are capable of silencing the founder protein-coding gene, thereby implicating pseudogenes in the regulation of gene expression.

Antiviral RNAi

As mentioned above, the ancestral eukaryotic RNAi may have evolved as a defense mechanism against viruses and TEs (Obbard et al., 2009; Shabalina and Koonin, 2008). The genomes of dsRNA viruses and replication intermediates of single-stranded RNA viruses are sources of dsRNA molecules amenable to Dicer processing. These viral dsRNAs are subsequently processed in a manner similar to endo-siRNAs. Therefore, the mature viral sRNA shares an identical length and structure with endo-siRNAs, including 5' monophosphates (Ding, 2010; Obbard et al., 2009). AGO proteins loaded with viral sRNAs can then target the viral genome and elicit its degradation. RNAi is a major defense platform against viruses in plants and invertebrates. In plants and nematodes, host RdRPs often produce viral dsRNA intermediates (Ashe et al., 2013; Ding, 2010; Obbard et al., 2009). Contrary to plants and invertebrates, vertebrates have adaptive immune systems consisting of type I IFN, which recognize viral dsRNA and initiate an antiviral response (Schlee and Hartmann, 2016). The physiological relevance of type I IFN in vertebrates seems to render RNAi redundant. Indeed, there is a great deal of controversy around the physiological importance of mammalian antiviral RNAi (Schlee and Hartmann, 2016; Svoboda, 2014).

Another prominent endogenous sRNA class does not depend on Dicer enzymes for its biogenesis and serves to control selfish genetic elements in the germline – these are the piRNAs. Despite their germline-associated "canonical" functions in TE control, piRNAs silence viruses in the soma of invertebrate species (Lewis et al., 2017; Ozata et al., 2018). In the next section, we will delve further into piRNA biology.

General aspects of piRNA biogenesis and function

In animals, Piwi clade AGOs and piRNAs are co-expressed and interact predominantly in the germline (Aravin et al., 2006; Batista et al., 2008; Brennecke et al., 2007; Carmell et al., 2007; Cox et al., 1998, 2000; Das et al., 2008; Girard et al., 2006; Grivna et al., 2006; Gunawardane et al., 2007; Houwing et al., 2007; Lau et al., 2006; Saito et al., 2006; Vagin et al., 2006; Wang and Reinke, 2008). Their "canonical" function is to protect the germline from the detrimental effects of TEs (Huang et al., 2017; Luteijn and Ketting, 2013; Ozata et al., 2018). The germline can be considered as an "immortal" tissue, since it continuously gives rise to the next generation. Hence, piRNAs have the important function of preventing uncontrolled TE mobilization from wreaking havoc in the genomes of following generations. Their importance can be attested by the sterility or broad range of fertility defects displayed by piRNA-defective mutants, such as Piwi mutants (Aravin et al., 2007; Batista et al., 2008; Carmell et al., 2007; Cox et al., 1998; Das et al., 2008; Deng and Lin, 2002; Harris and Macdonald, 2001; Kuramochi-Miyagawa et al., 2008; Li et al., 2009; Schüpbach and Wieschaus, 1991; Wilson et al., 1996).

piRNA biogenesis and function are intensively studied in fruit flies, silkworm, mouse, and the nematode *C. elegans*. Although there are several organism-specific nuances and exceptions, several universal principles of piRNA biogenesis and function can be extracted. Of note, some of the principles presented below do not necessarily apply to the nematode pathway considered analogous to the piRNA pathway, which will be introduced later (see "the 21U-RNA pathway" section).

Long piRNA precursor transcripts are typically transcribed by RNA Pol II from large genomic regions known as piRNA clusters (**Figure I.6**) (Aravin et al., 2006; Brennecke et al., 2007; Girard et al., 2006; Houwing et al., 2007; Huang et al., 2017; Vagin et al., 2006). These clusters harbor degenerate TE copies, relics of old TE invasions (Brennecke et al., 2007; Ozata et al., 2018). piRNA clusters are considered TE traps because once a TE jumps in a cluster at random, complementary piRNAs are produced and this will lead to the silencing of other homolog TE copies, even *in trans*.

After export of piRNA precursors to the cytoplasm, piRNA production starts by cleavage of piRNA precursors by Piwi AGOs and specialized endo- and exonucleases (**Figure I.6**) (Gainetdinov et al., 2018; Huang et al., 2017; Ozata et al., 2018). piRNA maturation is completed when HEN1 enzymes 2'-O-methylate the 3' end of the piRNA (Billi et al., 2012; Horwich et al., 2007; Kamminga et al., 2010, 2012; Lim et al., 2015; Montgomery et al., 2012; Saito et al., 2007). This modification is thought to provide stability to sRNAs. Mature piRNAs are consensually 23-30 nucleotides in length and some have a 5' bias for uridine (Huang et al., 2017; Ozata et al., 2018). One of the amplification mechanisms is the so-called "ping-pong" amplification cycle, and it typically involves two relaying Piwi AGOs (Brennecke et al., 2007; Gunawardane et al., 2007; Huang et al., 2017; Ozata et al., 2018). The catalytic activity of one Piwi generates a piRNA, which is accepted by another Piwi, and this event is repeated in a loop. This positive feedback loop allows for robust amplification of the piRNA pool and faithful silencing. The "ping-pong" amplification cycle seems to be an evolutionarily conserved mechanism in Piwi/piRNA pathways (Gainetdinov et al., 2018).

Piwi-RISCs silence their targets both by PTGS and TGS mechanisms (**Figure I.6**) (Huang et al., 2017; Ozata et al., 2018; Sato and Siomi, 2018). PTGS is mainly dependent on target cleavage by Piwi AGOs, whereas TGS involves at least one Piwi AGO that is shuttled into the nucleus to target nascent RNAs. Nuclear Piwi AGOs are not sufficient for TGS. Interactions with other factors, such as histone methyltransferases, are required to establish repressive chromatin at target *loci* (Sato and Siomi, 2018; Sienski et al., 2015; Yu et al., 2015).

Piwi/piRNA pathways function as an adaptive immune system against genetic parasites. Several features of adaptive immune systems are shared by Piwi/piRNA pathways, like the ability to recognize the threat, initiate a response, amplify the response, and keep a memory of the response for further encounters (Ozata et al., 2018). Memory of past encounters is embedded in piRNA clusters and thus

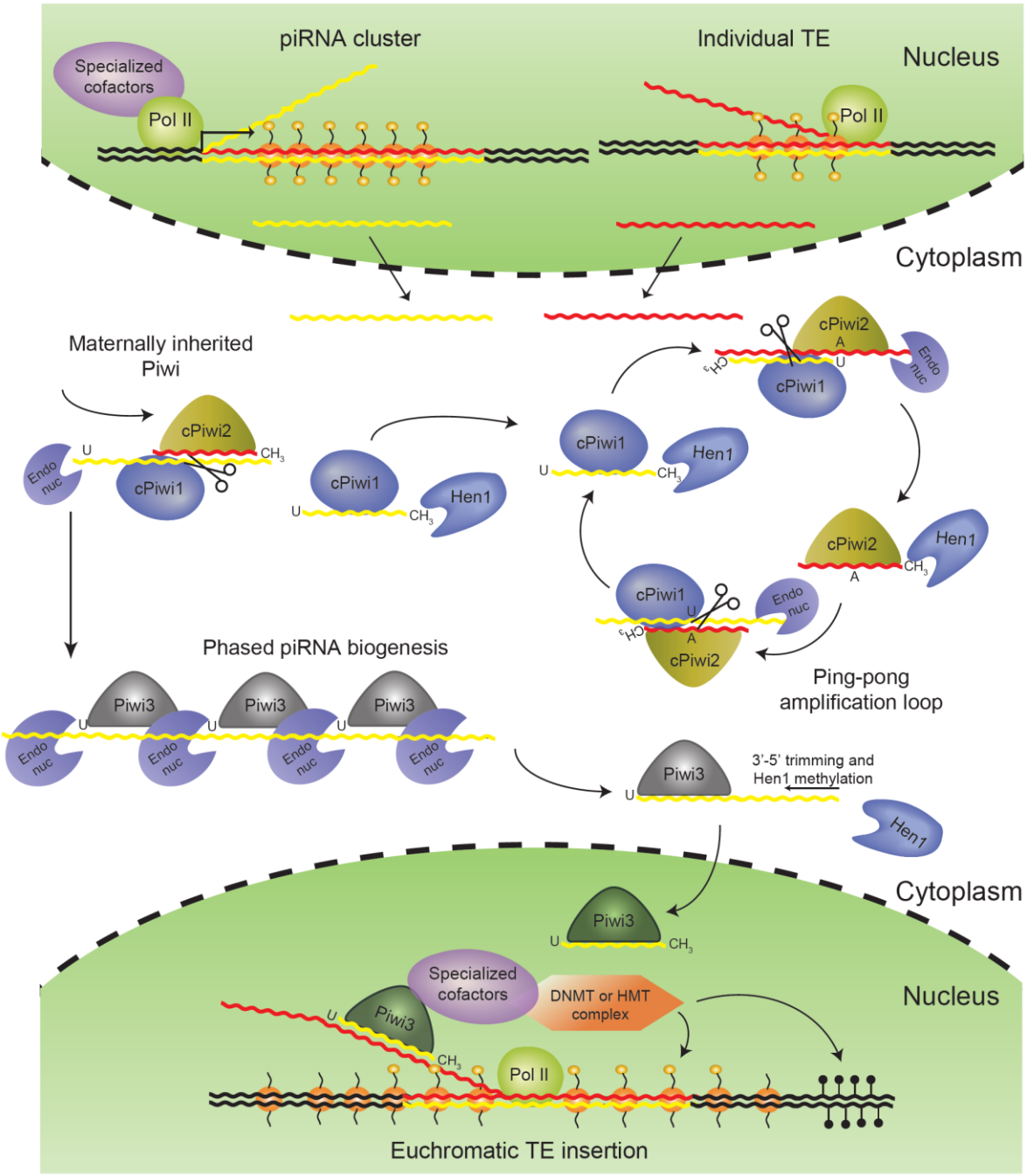


Figure I.6. **Conserved aspects of piRNA pathways**. Long piRNA precursor transcripts are produced by RNA Pol II in the nucleus (upper nucleus). Both sense and antisense transcripts can be produced by piRNA clusters and individual TE copies. After export into the cytoplasm, precursors are processed into mature piRNAs by two distinct biogenesis mechanisms. In early development, maternally provided Piwi-RISCs may initiate piRNA biogenesis. piRNA maturation is considered complete when Hen1 enzymes add a methyl group to the 3' end of the piRNA. The first biogenesis mechanism is the so-called ping-pong amplification loop that encompasses two cytoplasmic Piwi AGOs reciprocally engaged in production of sense and antisense piRNAs. These Piwi AGOs engage RNA targets with sequence complementarity in the cytoplasm, leading to post-transcriptional gene silencing. The second biogenesis mechanism produces piRNAs in a phased manner, by the combined action of a Piwi protein and an endonuclease. In some cases, the Piwi AGO loaded with phased piRNAs can be shuttled into the nucleus, where it elicits transcriptional gene silencing of TEs (lower nucleus). Silencing is executed with the assistance of specialized cofactors, as well as DNMT and HMT complexes, which methylate DNA and chemically modify histone tails, respectively. cPiwi, cytoplasmic Piwi; DNMT, DNA methyltransferase; Endo nuc, endonuclease; HMT, histone methyltransferase; Pol II, RNA Polymerase II; TE, transposable element.

transmitted to the next generation (Brennecke et al., 2007; Ozata et al., 2018). In addition, Piwi-RISCs may be directly inherited by the progeny in order to jump-start piRNA biogenesis in the next generation.

Several lines of evidence suggest that Piwi/piRNA pathways and TEs are engaged in an evolutionary arms race, consistent with the "Red Queen" hypothesis (Palmer et al., 2018; Parhad et al., 2017; Simkin et al., 2013; Vermaak et al., 2005). Widely adopted to explain host-parasite interactions, this theory may be applicable in the nucleic acid world to Piwi/piRNA pathways, as hosts, and TEs, as genetic parasites. In this light, genetic changes beneficial to TEs are counteracted by genetic changes in piRNA pathway factors that eliminate or attenuate the TE advantage. In fact, the *Drosophila* genus seems to be rich in examples supporting an arms race between hosts and TEs (Parhad et al., 2017; Simkin et al., 2013; Vermaak et al., 2005). Also, many factors involved in Piwi/piRNA pathways in diverse organisms are not evolutionarily conserved, in agreement with fast evolution of these pathways, potentially in response to TEs.

Non-canonical piRNA functions

Other than regulating TE expression in the germline and in somatic gonadal tissues, Piwi-RISCs have been implicated in multiple other processes (Ozata et al., 2018; Rojas-Ríos and Simonelig, 2018). For example, regulation of protein-coding genes by piRNAs has far-reaching implications in sex determination. In fact, piRNAs are involved in the regulation of specific targets involved in this process in the silkworm *bombyx mori* and in *C. elegans* (Kiuchi et al., 2014; Tang et al., 2018). Future studies should determine if piRNA regulation is more broadly integrated in animal gene regulatory networks orchestrating sex determination.

A recurring functional theme of Piwi proteins seems to be the maintenance of the stem cell state (Rojas-Ríos and Simonelig, 2018; van Wolfswinkel, 2014). Beyond germline expression, Piwi AGOs are expressed in adult somatic stem cells in basal animal lineages that have the ability to regenerate parts of their body, or their entire body (Rojas-Ríos and Simonelig, 2018; van Wolfswinkel, 2014). Furthermore, Piwi depletion in cnidarian, ascidian and planarian species precludes regeneration (Rojas-Ríos and Simonelig, 2018; van Wolfswinkel, 2014). It is currently not clear what determines Piwi requirement for stem cell maintenance and regeneration. In this light, Piwi proteins may have a critical role in maintaining genome stability in proliferating cells, for example by inhibiting genome instability secondary to TE mobility.

Another recurring non-canonical piRNA functional theme is neuronal activity. piRNAs have been implicated in memory formation and persistence in *Aplysia* (Rajasethupathy et al., 2012), and in axon regeneration in *C. elegans* (Kim et al., 2018). Furthermore, Piwi AGO expression has been detected, although typically in very low levels, in the brains of fruit flies, mice and macaque (Lee et al.,

2011; Perrat et al., 2013; Yan et al., 2011). However, a more concrete characterization of piRNAs co-precipitating with Piwi AGOs in these tissues is required in order to elucidate the existence of functional Piwi-RISCs and regulated target RNAs.

In sum, a growing body of work, some circumstantial, implies the involvement of Piwi AGOs and piRNAs in several physiologically relevant processes, including sex determination, regeneration and neuronal activity. However, future studies should clarify the underlying mechanisms.

Clinical relevance of RNAi research

Studying sRNA pathways, piRNA pathways in particular, is of high clinical relevance. Piwi mutants display gametogenesis defects and infertility in a variety of animals (see the "general aspects of piRNA biogenesis and function" section) and humans may not be an exception. There are several lines of evidence that indicate a possible correlation of piRNA pathway defects with spermatogenesis defects and infertility in human males. Polymorphisms in HIWI2 and HIWI3 are correlated with oligozoospermia risk in a Chinese population (Gu et al., 2010), and reduced expression of several piRNA pathway factors was associated with impaired spermatogenesis in human infertility syndromes (Heyn et al., 2012). Furthermore, a recent study has described HIWI mutations in infertile patients with azoospermia (Gou et al., 2017). Recreation of these mutations in a mouse model revealed defects in HIWI turnover that impede efficient histone-to-protamine exchange, ultimately leading to spermatogenesis defects (Gou et al., 2017).

Deregulation of Piwi AGO expression in humans has also been correlated with testicular tumors (Ferreira et al., 2014; Qiao et al., 2002). In fact, a vast number of studies implicated Piwi proteins or specific piRNAs in the most diverse cancer types. While there may be an association of ectopic expression of germline genes in cancerous tissues (Janic et al., 2010), it is hard to establish this as a cause for, or a consequence of cancer. Moreover, most studies in cancer cells typically do not address co-expression and interaction of Piwi proteins and piRNAs. More concrete and adequately designed experiments should be employed in cancer cell lines and tissues in order to shed some light on the influence of Piwi proteins and piRNAs in cancer biology.

Other than its relevance for fertility studies and, to a lesser extent, to cancer studies, studying RNAi-like pathways is relevant in itself. Increasing the mechanistic understanding of these pathways may pave the ground for future drug discovery. Development of RNAi-based therapeutics has been a productive approach, attested by many ongoing clinical trials (Bobbin and Rossi, 2016). In fact, the first RNAi-based therapeutic was very recently approved by the United States Food and Drug Administration (Mathias, 2018). This new drug was developed to treat a rare, fatal condition: polyneuropathy arising in hereditary transthyretin-mediated amyloidosis. Approval of more RNAi-

based therapeutics may follow soon. Fundamental research on RNAi pathways contributed and will continue to contribute to the development of this new class of therapeutics.

Tiny nematodes, enormous Biology

When Sydney Brenner adopted *C. elegans* as a model organism, little did he know the impact that this tiny nematode would have in biological research. In retrospect, the choice was obvious for multiple reasons. In the lab, these small nematodes (adults can reach up to 1 mm in length) are easily grown in Agar plates supplemented with bacteria as food, conventionally *Escherichia coli* (Brenner, 1974; Corsi et al., 2015). *C. elegans* are androdioecious, meaning their populations include hermaphrodites and males, the latter occurring spontaneously at a frequency of <0.2% (Corsi et al., 2015). Hermaphrodites have two X chromosomes, whereas males have only one X, commonly referred to as XO. These animals are transparent, have a short life cycle of approximately 3 days and a large brood size of approximately 300 progeny, when grown at 20 ^{o}C. All these features make *C. elegans* an excellent model organism for genetics and cell biology studies (Corsi et al., 2015). Vouching for its importance, the number of scientific breakthroughs achieved using *C. elegans* is enormous (Corsi et al., 2015). In the hall of fame, amongst many others, there is the discovery of RNAi, by Fire, Mello, and colleagues (Fire et al., 1998); the introduction of GFP as a biological marker, by Chalfie and colleagues (Chalfie et al., 1994); as well as the first determination of a complete cell lineage of an animal throughout development, meticulous work by Sulston and Horvitz (Sulston and Horvitz, 1977). The latter research ultimately led to the discovery of apoptosis (Ellis and Horvitz, 1986; Hedgecock et al., 1983). Moreover, *C. elegans* was the first multicellular organism having its genome sequenced in 1998 (The C. elegans Sequencing Consortium, 1998). The genome of *C. elegans* has 100 Mb split into six chromosomes, historically called Linkage Groups (LG): I-V autosomes, plus the X chromosome (Brenner, 1974; The *C. elegans* Sequencing Consortium, 1998). The latest genome annotation indicates that the genome of *C. elegans* has a total of 20213 protein-coding genes, and 26697 pseudogenes plus non-coding RNAs (WormBase release WS267, November 2018, at wormbase.org). Next, we will dive into the astonishingly complex world of tiny non-coding RNAs existing within this tiny animal.

The *C. elegans* RNAi-like pathways

C. elegans can initiate RNAi upon acquisition of dsRNA from its surroundings. The fact that exogenous, or environmental dsRNA can be absorbed so easily from the environment reflects the existence of specialized machinery to internalize it (Braukmann et al., 2017). Several factors help in bringing the dsRNA from the gut lumen into the intestinal epithelium. Interestingly, dsRNA internalization was recently visualized *in vivo*, including the deposition of dsRNA into oocytes and

embryos (Marré et al., 2016). Another key feature of RNAi in *C. elegans* is that it is systemic, i.e. dsRNA spreads throughout the pseudocoelom, potentially initiating an RNAi response in most somatic tissues and germline (Braukmann et al., 2017). After dsRNA enters the cytoplasm of a cell it is bound by the dsRNA-binding protein RDE-4, which in turn interacts with DCR-1 and the AGO protein RDE-1 (**Figure I.7**) (Parker et al., 2006; Tabara et al., 2002). DCR-1 chops the dsRNA molecule into shorter duplexes that associate with RDE-1 (Ketting et al., 2001; Tabara et al., 1999). RDE-1 then removes the passenger strand using its RNase H fold, but does not cleave the target mRNA (Steiner et al., 2009). Silencing reactions downstream of RDE-1 will be described below (see section "22G-RNA pathways: a nexus of gene regulation").

In addition, *C. elegans* has a complex endogenous sRNA landscape (Ambros et al., 2003; Ruby et al., 2006). *C. elegans* expresses several classes of endogenous sRNAs other than miRNAs: 21U-, 26G- and 22G-RNAs. These distinct sRNA populations can be distinguished by size, 5' nucleotide bias and 5'-end phosphorylation: 21U-RNAs are 21 nucleotides long and have a 5' bias for uridine monophosphate; 26G-RNAs are consensually 26 nucleotides long and have a 5' bias for guanosine monophosphate; and 22G-RNAs are 22 nucleotides in length and have a 5' bias for guanosine triphosphate (Ruby et al., 2006). 21U-RNAs are likely produced by RNA Pol II (Cecere et al., 2012), while 26G- and 22G-RNAs are products of RdRP activity (Conine et al., 2010; Gent et al., 2009, 2010; Han et al., 2009; Pak and Fire, 2007; Sijen et al., 2007; Vasale et al., 2010). The RdRPs of *C. elegans* synthesize sRNAs from template target RNAs in a non-primed and non-processive manner (**Figure I.4B**) (Aoki et al., 2007). Therefore, all RdRP products have an antisense orientation in respect to the gene feature they map to. The genome of *C. elegans* encodes four RdRP genes: *ego-1* and *rrf-1*, necessary for 22G-RNA biogenesis; *rrf-3*, required for 26G-RNA biogenesis; and *rrf-2* of unknown function. 21U- and 26G-RNAs represent primary sRNA species, which dictate downstream production of secondary 22G-RNAs (Bagijn et al., 2012; Conine et al., 2010; Das et al., 2008; Gent et al., 2010; Lee et al., 2012; Vasale et al., 2010). Furthermore, secondary 22G-RNA production is also triggered in response to exogenous dsRNA (Pak and Fire, 2007; Sijen et al., 2007). Therefore, several primary sRNA inputs result in 22G-RNA production, largely by a common set of factors (Duchaine et al., 2006; Gu et al., 2009; Phillips et al., 2012, 2014; Yigit et al., 2006). Since all sRNA pathways converge on this 22G-RNA nexus, many sRNA pathways are in theory competing for shared factors required for 22G-RNA amplification and silencing (Duchaine et al., 2006; Lee et al., 2006). Distinct subpopulations of 22G-RNAs can be defined by their AGO cofactors and their targets (Gu et al., 2009; Yigit et al., 2006).

All the described sRNA species associate with at least one of the 27 AGOs that are encoded in the *C. elegans* genome (Yigit et al., 2006). These AGOs can have overlapping roles, making it challenging to genetically and biochemically dissect these very intricate pathways. The biogenesis of 21U-, 26G- and 22G-RNA species of *C. elegans*, their cofactors and effector functions will be described below. The

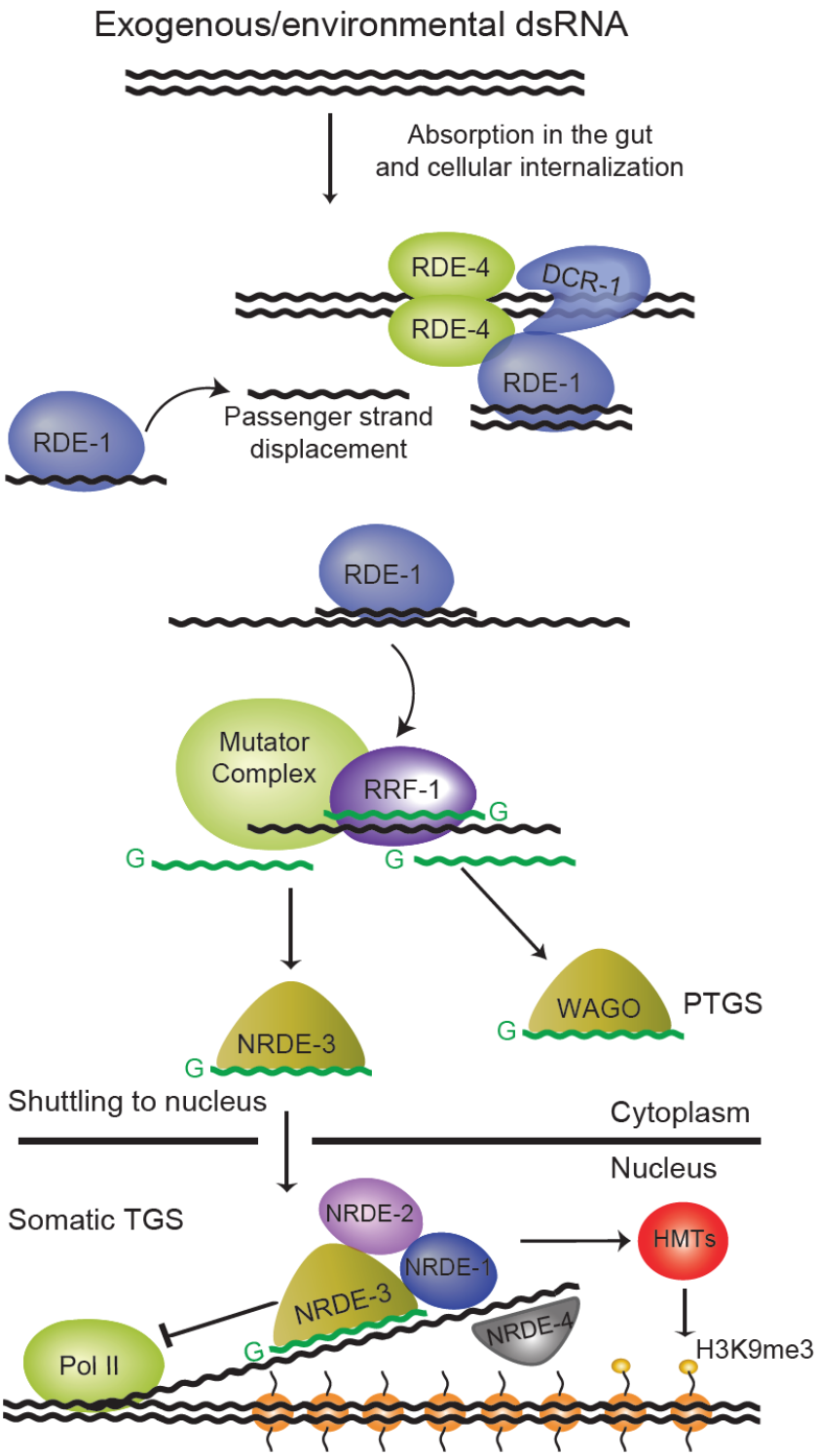


Figure I.7. **Exogenous RNAi pathway in *C. elegans***. After cellular internalization, dsRNA is bound by a homodimer of RDE-4 and processed by DCR-1 into shorter duplexes that are in turn bound by the AGO RDE-1. This AGO uses its catalytic activity to remove the passenger strand, but not to cleave complementary target RNAs. Instead, the RdRP RRF-1 is recruited to target transcripts to produce 22G-RNAs, with the assistance of accessory Mutator proteins. These interact with cytoplasmic and nuclear WAGOs, resulting in PTGS or TGS, respectively. The somatic nuclear AGO NRDE-3 is engaged in exogenous RNAi. Upon 22G-RNA binding, NRDE-3 is shuttled into the nucleus, where it elicits TGS, in a concerted action with other NRDE factors and HMTs, by inhibiting Pol II elongation and promoting the deposition of inhibitory chromatin marks. dsRNA, double-stranded RNA; HMTs, histone methyltransferases; Pol II, RNA Polymerase II; PTGS, post-transcriptional gene silencing; TGS, transcriptional gene silencing.

biogenesis and function of miRNAs in worms have been reviewed elsewhere (Ambros and Ruvkun, 2018) and will not be discussed here.

The 26G-RNA pathway

26G-RNAs are produced by the so-called ERI complex in the germline (**Figure I.8**). The ERI complex consists of several factors, including the RdRP RRF-3 and DCR-1 (Duchaine et al., 2006; Thivierge et al., 2012). The sequence of biochemical events underlying 26G-RNA biogenesis still has to be determined. Also, it is not known how ERI complex factors are recruited to specific target transcripts. A current model of 26G-RNA biogenesis implies RNA synthesis by RRF-3 antisense to a template target RNA, creating a dsRNA intermediate (Blumenfeld and Jose, 2016). A blunt end is subsequently formed by an exonuclease, presumably ERI-1. The blunt end creates a substrate for DCR-1 cleavage and selectively stimulates the production of a 26 nucleotide long cleavage product (Welker et al., 2011). In this model, RRF-3 would then synthesize another 26G-RNA from the next cytosine available 3' of the DCR-1 cleavage site, thereby initiating another cycle of biogenesis.

ERI complex mutants (e.g. *rrf-3* and multiple Eri genes) display a broad range of defects (Duchaine et al., 2006; Gent et al., 2009; Pavelec et al., 2009; Simmer et al., 2002; Thivierge et al., 2012; Welker et al., 2010): sperm-derived fertility defects including reduced brood size at 20 ^{o}C and sterility at 25 ^{o}C, high incidence of males (Him), and enhanced RNAi (Eri). The latter phenotype is characterized by hypersensitivity to exogenous RNAi (exo-RNAi) and it is believed to be a reflection of competition between exogenous and endogenous RNAi pathways for shared factors (Duchaine et al., 2006; Gu et al., 2009; Lee et al., 2006; Yigit et al., 2006).

Two distinct subpopulations of 26G-RNAs are produced in two spatially and temporally distinct stages (**Figure I.8**). One subpopulation of 26G-RNAs binds to the AGOs ALG-3 and ALG-4 (henceforth referred to as ALG-3/4) in the spermatogenic germline in L4 hermaphrodites and in male worms (Conine et al., 2010, 2013; Han et al., 2009). The other subpopulation of 26G-RNAs associates with ERGO-1 in the oogenic germline and in embryos (Gent et al., 2010; Han et al., 2009; Pavelec et al., 2009; Vasale et al., 2010). Both can trigger downstream production of secondary 22G-RNAs that exert effector functions. However, it remains to be established whether ALG-3/4 and ERGO-1 have the potential to directly regulate targets. In fact, the catalytic center of ALG-3/4 and ERGO-1 may be compatible with cleavage (Yigit et al., 2006).

ALG-3/4 branch 26G-RNAs

ALG-3/4 are two apparently redundant AGO paralog genes with high sequence similarity, which are expressed specifically during spermatogenesis, but are absent from mature sperm. Their loss

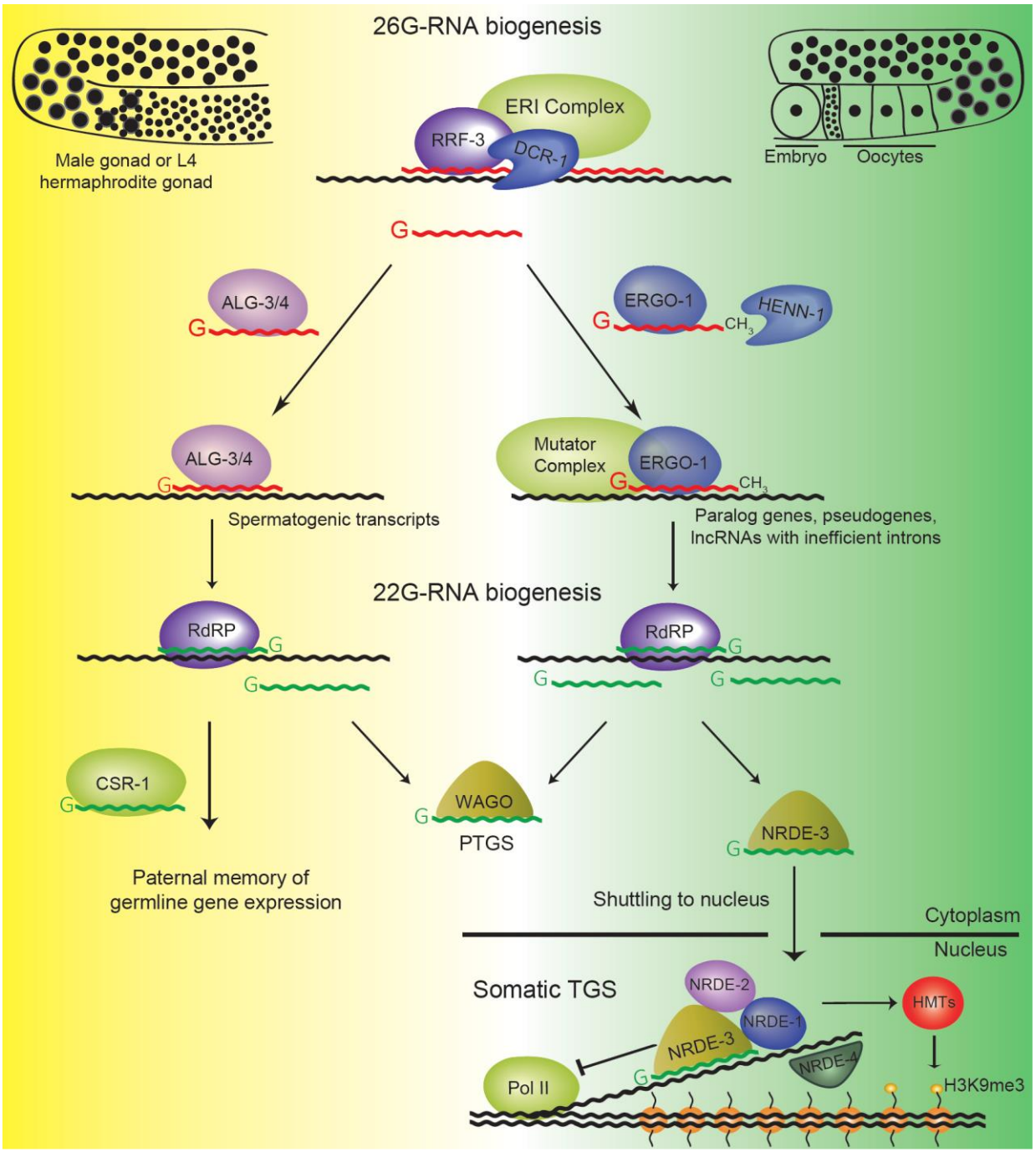


Figure I.8. **Biogenesis and function of 26G-RNA pathways**. Two distinct subpopulations of 26G-RNAs are produced by the ERI complex, which includes the RdRP RRF-3 and DCR-1. One subpopulation of 26G-RNAs is produced in the male germline and in the spermatogenic gonad in L4 hermaphrodites (on the left). These 26G-RNAs associate with ALG-3/4 and target spermatogenic transcripts. This is thought to trigger the biogenesis of 22G-RNAs that interact with CSR-1, providing a transgenerational memory of paternal gene expression. In addition, such ALG-3/4-triggered 22G-RNAs may associate with other WAGOs and negatively regulate target gene expression. Another subpopulation of 26G-RNAs is expressed in the oogenic gonad and in embryos (on the right). These associate with ERGO-1 and are 2'-O-methylated by HENN-1 at their 3' end. ERGO-1-bound 26G-RNAs target a diverse set of genes, which tend to be short, with few introns and with weak splicing signals. ERGO-1 triggers 22G-RNA biogenesis, which in turn can promote PTGS through their association with cytoplasmic WAGOs, as well as TGS through association with the nuclear WAGO protein NRDE-3. NRDE-3, with the assistance of additional NRDE factors and HMTs, perpetuates silencing of 26G-RNA targets in the soma throughout animal development. Of note, NRDE-3 acts in a manner analogous to HRDE-1 in the germline (see Figure I.9A). HMTs, Histone methyltransferases; PTGS, post-transcriptional gene silencing; RdRP, RNA-dependent RNA Polymerase; TGS, transcriptional gene silencing.

of function phenocopies the abovementioned Him and sperm-derived fertility defects (Conine et al., 2010; Gent et al., 2009; Pavelec et al., 2009). The fertility defects have been attributed to the inability of sperm to activate, specifically by impeding the formation of its characteristic pseudopods that allow for fertilization (Conine et al., 2010, 2013). ALG-3/4-bound 26G-RNAs target spermatogenesis-specific genes, such as major sperm protein genes, kinases and phosphatases required for spermatogenesis (Conine et al., 2010, 2013). The mechanisms of ALG-3/4 target regulation are largely unknown.

ERGO-1 branch 26G-RNAs

The enhanced RNAi (Eri) phenotype of Eri mutants stems from defects in 26G-RNAs that bind the AGO protein ERGO-1 (Han et al., 2009; Pavelec et al., 2009; Yigit et al., 2006; Zhuang and Hunter, 2011). ERGO-1 is expressed in the first two larval stages, in the adult oogenic germline and in embryos (Billi et al., 2012; Vasale et al., 2010). ERGO-1 binds a subpopulation of 26G-RNAs, which is distinct from ALG-3/4-bound 26G-RNAs (**Figure I.8**). Unlike ALG-3/4-bound 26G-RNAs, ERGO-1 binds 26G-RNAs that are 3' 2'-O-methylated by the conserved HEN1 methyltransferase ortholog HENN-1 (Billi et al., 2012; Kamminga et al., 2012; Montgomery et al., 2012; Vasale et al., 2010). This methylation of the ribose ring is required for stability of ERGO-1-class 26G-RNAs, and influences the accumulation of downstream 22G-RNA populations (Billi et al., 2012; Kamminga et al., 2012; Montgomery et al., 2012). ERGO-1-RISCs target non-conserved, non-germline specific, repeat-enriched sequences like pseudogenes, long non-coding RNAs (lncRNAs), and recently duplicated genes, triggering secondary 22G-RNA production in these targets (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 21U-RNA pathway

21U-RNAs were found to interact with the Piwi clade AGO PRG-1 in the germline and were therefore considered the piRNAs of *C. elegans* (Bagijn et al., 2012; Batista et al., 2008; Cox et al., 1998; Das et al., 2008; Wang and Reinke, 2008). *prg-2* is another Piwi gene in *C. elegans*, but it is currently annotated as a pseudogene (WormBase release WS267, November 2018, at wormbase.org). *prg-1* mutants have a variety of germline defects but are nevertheless viable (Batista et al., 2008; Cox et al., 1998; Das et al., 2008). However, when *prg-1* mutant animal populations are subjected to consecutive bottlenecks, animals show a progressive deterioration of germline health throughout generations, the so-called mortal germline phenotype (Simon et al., 2014).

21U-RNAs can be subdivided into two types: type-I and type-II (Gu et al., 2012; Ruby et al., 2006; Weick et al., 2014). Type-I 21U-RNA precursors are likely transcribed by RNA Pol II from discrete autonomous genomic *loci* with a distinctive 5' upstream conserved motif (Billi et al., 2013; Cecere et

al., 2012; Ruby et al., 2006; Weick et al., 2014). The precursors are approximately 26 nucleotides long, with a conventional 5' cap but are not polyadenylated (Gu et al., 2012). Although type-I 21U-RNA *loci* are present throughout the genome, their distribution seems to be highly concentrated in two ~3 Mb clusters on chromosome IV (Ruby et al., 2006). There are specialized factors dedicated to 21U-RNA precursor transcription. For example, an additional complex termed the upstream sequence transcription complex (USTC) recognizes the upstream motif and promotes transcription of 21U-RNA precursors (**Figure I.9A**) (Weng et al., 2018). PRDE-1, TOFU-4/5 and SNPC-4 are part of the USTC and all these factors are individually required for 21U-RNA biogenesis (Goh et al., 2014; Kasper et al., 2014). Curiously, SNPC-4 is also interacting with and regulating the expression of small nuclear RNA (snRNA) genes (Kasper et al., 2014). It was recently shown that the upstream motif seems to be derived from promoter elements that drive snRNA biogenesis (Beltran et al., 2018; Weng et al., 2018), consistent with the co-option of SNPC-4 in type-I 21U-RNA biogenesis. Altogether, the participation of these specialized factors, together with an amenable chromatin environment (Beltran et al., 2018), provide a framework of how RNA Pol II can synthesize these very short, atypical transcripts.

Type-II 21U-RNAs are derived from capped sRNAs that are transcribed by RNA Pol II bidirectionally from transcription start-sites (Gu et al., 2012). When YRNT motifs (where Y is a pyrimidine, R is a purine, N is any nucleotide, and T encodes the 5' U of the 21U-RNA) are present near the transcription start site, type-II 21U-RNA precursors are processed into mature 21U-RNAs. These results, together with the abovementioned new observations, indicate that both type-I and -II 21U-RNA *loci* lack the cis sequences and the chromatin environment that is typically associated with canonical RNA Pol II elongation, and this may be a reason why these *loci* generate such short transcripts (Beltran et al., 2018; Gu et al., 2012). It is unknown whether type-II 21U-RNAs have any biological function or are just a by-product of RNA Pol II pausing at promoters.

Following transcription, the precursors are exported out of the nucleus but it is not clear how this is achieved. PID-1 is a factor required for 21U-RNA biogenesis that has predicted nuclear localization and export signals (de Albuquerque et al., 2014). Although the functionality of these localization signals was not experimentally verified, a working model implies that PID-1 may be cycling between the cytoplasm and the nucleus, possibly exporting a 21U-RNA precursor. Next, the 5' cap is removed and the 5' and 3' ends of the precursor 21U-RNA are trimmed. PETISCO, a recently identified protein complex containing PID-1, is required for 21U-RNA biogenesis, and has the potential to bind the 5'-cap and 5'-phosphate of RNA (Rodrigues et al., 2018; Zeng et al., 2018) (**Figure I.9A**). Future biochemical studies should determine whether PETISCO is involved in processing the 5' ends of 21U-RNA precursors. PARN-1, a conserved 3'-to-5'exonuclease, trims the 3' end of 21U-RNA precursors (Tang et al., 2016). Finally, the conserved RNA methyltransferase HENN-1 2'-O-methylates the 3' ends

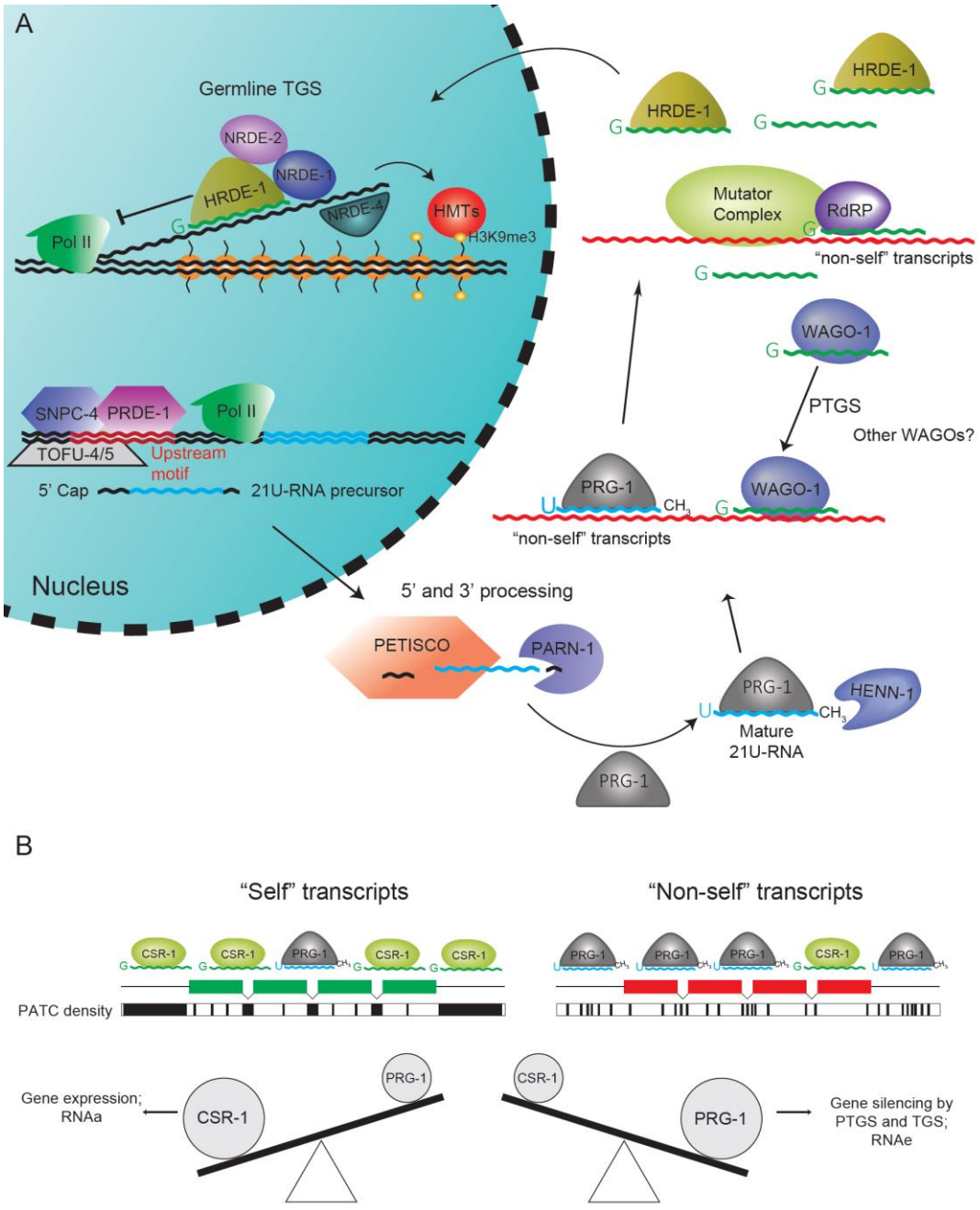


Figure I.9. **Biogenesis and function of the PRG-1/21U-RNA pathway and its antagonistic signals**. (A) In the *C. elegans* germline, the USTC complex, with SNPC-4, PRDE-1 and TOFU-4/5, recognize the upstream motif and promote 21U-RNA transcription by RNA Pol II. 21U-RNAs are exported to the cytoplasm for further processing. 5' processing may be undertaken by the PETISCO complex, whereas PARN trims the 3' ends of 21U-RNAs. 21U-RNAs subsequently associate with PRG-1 and are 2'-O-methylated at the 3' end by HENN-1. PRG-1 scans the germline transcriptome and stimulates 22G-RNA production by RdRPs and mutator factors from foreign transcripts. Then, 22G-RNAs bind to cytoplasmic WAGOs, leading to PTGS of targets or to HRDE-1, which promotes TGS. Accessory NRDE factors are required for HMT recruitment and deposition of repressive chromatin marks. Moreover, NRDE factors may directly impair the activity of RNA Pol II. (B) CSR-1 and PATCs inhibit repressive PRG-1 signals. The regulatory outcome of a germline-expressed gene will depend on the balance between associated CSR-1/22G-RNAs and PRG-1/21U-RNAs. Germline genes are strongly recognized by CSR-1 and have PATCs throughout their sequence, thereby inhibiting PRG-1 targeting. In contrast, non-self sequences tend to lack the PATC signature and will not be strongly targeted by CSR-1. In this case, stronger PRG-1 targeting will flag these transcripts for degradation. HMT, Histone methyltransferases; Pol II, RNA Polymerase II; PTGS, post-transcriptional gene silencing; RdRP, RNA-dependent RNA Polymerase; RNAa, RNA-induced epigenetic gene activation; RNAe, RNA-induced epigenetic silencing; TGS, transcriptional gene silencing.

of 21U-RNAs, likely after loading onto PRG-1 (Billi et al., 2012; Kamminga et al., 2012; Montgomery et al., 2012). In absence of HENN-1, 21U-RNAs become destabilized.

21U-RNAs loaded into PRG-1 will subsequently guide PRG-1 to RNA targets with relaxed sequence complementarity (Bagijn et al., 2012; Lee et al., 2012). Binding of PRG-1-RISCs to target RNAs tolerates mismatches in most positions, except for positions 2-8 of the 21U-RNA, similarly to miRNA targeting (Shen et al., 2018; Zhang et al., 2018). Such target binding results in a broad potential target range, which implies that these PRG-1-RISCs have the potential to scan a large sequence-space, potentially the whole germline transcriptome, for non-self transcripts. It follows that there have to be additional pathways or signals inhibiting PRG-1-RISCs from eliciting gene silencing of bona fide self transcripts. Indeed, there is a complex interplay of several factors and pathways contributing to the ultimate outcome of expressing or silencing a given transcript (see section below "CSR-1 pathway and periodic A_n/T_n clusters inhibit PRG-1-mediated silencing").

22G-RNA pathways: a nexus of gene regulation

RNAi pathways in *C. elegans* typically involve a primary sRNA that triggers the biogenesis of abundant secondary 22G-RNAs. As a result, 22G-RNAs can be subdivided into several subpopulations distinguished by the original primary trigger, interacting AGOs and sets of targets.

Several factors have been implicated in secondary 22G-RNA biogenesis downstream of RDE-1 (Sijen et al., 2007), 21U- (Ashe et al., 2012; Bagijn et al., 2012; Luteijn et al., 2012; Shirayama et al., 2012) and 26G-RNAs (Gent et al., 2010; Montgomery et al., 2012; Phillips et al., 2012, 2014; Vasale et al., 2010). The RdRPs RRF-1 and EGO-1 are redundantly required for 21U-RNA-driven 22G-RNA biogenesis (Bagijn et al., 2012). In contrast, only RRF-1 was implicated in RDE-1- and 26G-RNA-driven 22G-RNA biogenesis (Gent et al., 2010; Sijen et al., 2007; Vasale et al., 2010). Other factors assumed to be collectively part of a large accessory complex termed Mutator complex, are required for the production of these three 22G-RNA subpopulations (Bagijn et al., 2012; Ketting et al., 1999; Phillips et al., 2012, 2014; Sijen et al., 2007) (**Figure I.9A**). Subsequently, 22G-RNAs can associate with cytoplasmic AGOs, like WAGO-1, which provoke PTGS of their targets (Gu et al., 2009; Yigit et al., 2006). Alternatively, 22G-RNAs can bind to nuclear AGOs, like HRDE-1 and NRDE-3, which lead to TGS of their targets (Buckley et al., 2012; Guang et al., 2008). TGS is achieved with the assistance of other NRDE nuclear factors that may affect RNA Pol II elongation, and recruit histone methyltransferases that deposit H3K9me3 and H3K27me3 (Ashe et al., 2012; Buckley et al., 2012; Mao et al., 2015; Shirayama et al., 2012) (**Figure I.9A**).

RDE-1, 21U- and 26G-RNA targets are believed to be silenced both by PTGS and TGS (Ashe et al., 2012; Conine et al., 2010; Duchaine et al., 2006; Gu et al., 2009; Guang et al., 2008; Luteijn et al., 2012; Shirayama et al., 2012; Vasale et al., 2010; Yigit et al., 2006; Zhou et al., 2014) (**Figure I.7-8, 9A**).

RDE-1 and ERGO-1 couple to the somatic nuclear AGO NRDE-3 (Guang et al., 2008; Montgomery et al., 2012), whereas PRG-1 drives loading of HRDE-1 with 22G-RNAs in the germline (Ashe et al., 2012; Luteijn et al., 2012; Shirayama et al., 2012). Silencing by both of these nuclear AGOs depends on NRDE factors and histone methyltransferases (Ashe et al., 2012; Buckley et al., 2012; Burkhart et al., 2011; Guang et al., 2010; Luteijn et al., 2012; Mao et al., 2015; Shirayama et al., 2012). Overall, endogenous RNAi-like pathways have both cytoplasmic and nuclear elements, which regulate targets by a combination of PTGS and TGS mechanisms. Loss of HRDE-1 leads to a dramatic mortal germline, highlighting the importance of establishing and maintaining gene silencing (Buckley et al., 2012). Conversely, *nrde-3* mutants do not have a distinctive phenotype.

In some instances, silencing of transgenes and endogenous genes targeted by the 21U-RNA pathway can become stable, independent of the initial PRG-1 trigger, and perpetuated transgenerationally, most likely by a combination of 22G-RNA and histone tail modification signals (Ashe et al., 2012; Luteijn et al., 2012; Sapetschnig et al., 2015; Shirayama et al., 2012). This form of stable silencing is termed RNA-induced epigenetic silencing (RNAe). Importantly, HRDE-1, NRDE factors, the HP1 homolog HPL-2 and histone methyltransferases seem to be central to RNAe establishment and maintenance.

CSR-1 pathway and periodic A_n/T_n clusters inhibit PRG-1-mediated silencing

CSR-1 is a very enigmatic AGO. Amongst the 27 AGOs of *C. elegans*, CSR-1 is the only WAGO protein required for viability (Claycomb et al., 2009; Yigit et al., 2006). Moreover, CSR-1 is expressed in the germline and seems to promote gene expression, rather than gene silencing (Claycomb et al., 2009; Conine et al., 2013; Seth et al., 2013, 2018; Wedeles et al., 2013). In fact, this pathway was proposed to provide a memory of self gene expression in the germline (Claycomb et al., 2009; Conine et al., 2013; Seth et al., 2013, 2018; Wedeles et al., 2013). CSR-1 does so by associating with an abundant subpopulation of 22G-RNAs, made by the RdRP EGO-1, which targets germline-expressed genes. Mutator genes are dispensable for CSR-1-associated 22G-RNAs (Phillips et al., 2014).

PRG-1-RISCs have been shown to perform a transcriptome-wide surveillance of germline transcripts (Bagijn et al., 2012; Lee et al., 2012; Shen et al., 2018; Zhang et al., 2018), while CSR-1-RISCs target germline-expressed genes (Seth et al., 2013, 2018; Wedeles et al., 2013). The current model is that either silencing or licensing of a targeted transcript will depend on the outcome of a balance between PRG-1 and CSR-1 targeting (**Figure I.9B**). If PRG-1 targeting prevails, a transcript is flagged as non-self and targeted for silencing. If CSR-1 targeting is stronger, the transcript is recognized as self and is expressed.

Besides the CSR-1 pathway, another prominent signal has been shown to counteract the germline-wide surveillance activity of PRG-1, and to positively affect gene expression: periodic A_n/T_n-clusters (PATCs). PATCs are short AT-rich sequences embedded in introns and UTRs of genes that are correlated with germline expression (Frøkjær-Jensen et al., 2016). Single-copy germline-expressed transgenes with foreign sequences, like gfp, tend to be silenced in the germline by PRG-1 (Ashe et al., 2012; Lee et al., 2012; Luteijn et al., 2012; Shirayama et al., 2012). Equivalent transgenes with added PATCs to introns and UTRs can bypass silencing by PRG-1 (Frøkjær-Jensen et al., 2016; Zhang et al., 2018) (**Figure I.9B**). Of note, PATCs do not seem to be a universal feature of genomes. The genomic signature of PATCs seems to be largely absent in eukaryotes outside the nematode *Caenorhabditis* group. In sum, PATCs and CSR-1 provide signals that antagonize PRG-1 activity in *Caenorhabditis* animals, thereby protecting germline genes from erroneous silencing.

Gtsf1 genes: small but powerful regulators of TEs

The core sRNA cofactors presumably present in the LECA, i.e. AGOs, RdRPs and Dicer enzymes, are deeply conserved in eukaryotes (Cerutti and Casas-Mollano, 2006; Shabalina and Koonin, 2008). In addition, multiple other factors involved in piRNA pathways show high conservation in animals (Ozata et al., 2018). However, plenty of factors have less conserved roles in piRNA pathways. The existence of such lineage-specific cofactors is consistent with the rapid evolution of piRNA pathways, as the result of an arms race against TEs (Obbard et al., 2009; Palmer et al., 2018; Parhad et al., 2017; Simkin et al., 2013; Vermaak et al., 2005). Furthermore, it was recently shown that proteins involved in piRNA-dependent TGS have stronger signatures of positive selection, and therefore may be evolving faster (Palmer et al., 2018). Factors involved in transcription of piRNA precursors seem to be similarly fast evolving (Ozata et al., 2018). For example, Deadlock, Cuttoff, Moonshiner and Rhino are Drosophilid-specific factors required for the transcription of dual-strand piRNA clusters (Andersen et al., 2017; Chen et al., 2016; Mohn et al., 2014).

In 2013, three independent studies in *D. melanogaster* reported the results of genome-wide RNAi screens for factors involved in the piRNA pathway and in TE control (Czech et al., 2013; Handler et al., 2013; Muerdter et al., 2013). These screens identified *asterix* as a key factor involved in TGS of TEs. The name *asterix* originated “because of its small size, yet powerful role in transposon silencing” (Muerdter et al., 2013). Orthologs of *asterix* had already been described in mouse, and were previously named *Gametocyte-specific factor 1* (*Gtsf1*) (Yoshimura et al., 2007, 2009). The latter name was adopted in later publications for simplicity (Dönertas et al., 2013; Ohtani et al., 2013). Throughout this dissertation, this gene family will be addressed as *Gtsf1*.

Gtsf1 genes encode small proteins of approximately 20 kDa, which are expressed in the animal germline. GTSF1 proteins have two novel CHHC zinc fingers (Andreeva and Tidow, 2008) at their N-

terminus, followed by an unstructured C-terminal tail with acidic amino acid residues (**Figure I.10A**). Many animals encode multiple paralogs of GTSF1, typically termed GTSF1-Like (GTSF1L) and GTSF2. Depletion of *Drosophila melanogaster* Gtsf1 (Dmel-Gtsf1) leads to TE derepression and loss of H3K9me3 at euchromatic TE insertions (**Figure I.10B**) (Dönertas et al., 2013; Ohtani et al., 2013). Furthermore, Dmel-Gtsf1 physically interacts with Piwi via its C-terminal tail in the nucleus. In the absence of Dmel-Gtsf1, loaded Piwi is in the nucleus but is incapable of eliciting TGS. Downstream of GTSF1, the Drosophilid-specific Panoramix/Silencio, which also interacts physically with Piwi, is required to recruit the histone methyltransferase Eggless to TE loci, allowing H3K9me3 deposition and consequent silencing (**Figure I.10B**) (Sienski et al., 2015; Yu et al., 2015).

Similarly to *Drosophila*, *Mus muculus* GTSF1 (Mmus-GTSF1) is required for TE silencing (**Figure I.10C**) (Yoshimura et al., 2009, 2018). Mmus-GTSF1 is mainly expressed in the cytoplasm, where it interacts with MILI, the Piwi AGO that drives homotypic ping-pong amplification during spermatogenesis (Yoshimura et al., 2018). Therefore, Mmus-GTSF1 is assisting a Piwi AGO in piRNA biogenesis, not silencing TEs on the transcriptional level, like its *Drosophila* ortholog. *Gtsf1* mutation in flies renders females sterile, whereas in mice mutant males become sterile. Human *GTSF1* is expressed in oocytes, in the fetal ovary, and in preimplantation embryos (Huntriss et al., 2017). Given this expression pattern and the requirement of its orthologs for normal fertility, it is possible to envision roles in fertility or early development for human *GTSF1*. Thus, understanding the function of *Gtsf1* genes has the potential of being clinically relevant.

Aim of the thesis research

Given the general lack of conservation of factors involved in 21U-RNA biogenesis and function, we were surprised to identify one single conserved GTSF1 homolog in *C. elegans* (Cel-GTSF-1). The initial aim of this thesis work was to characterize the function and molecular mechanism of Cel-GTSF-1, especially in the context of sRNA pathways. Specifically, we were expecting a conserved role in the nuclear RNAi pathway downstream of 21U-RNAs, analogous to the role of Dmel-GTSF-1 in TGS of TEs. However, as will be shown in the following chapter we were mistaken. This fact opened an alternative line of research that allowed us to address several aspects regarding the biogenesis, function and evolution of sRNA pathways.

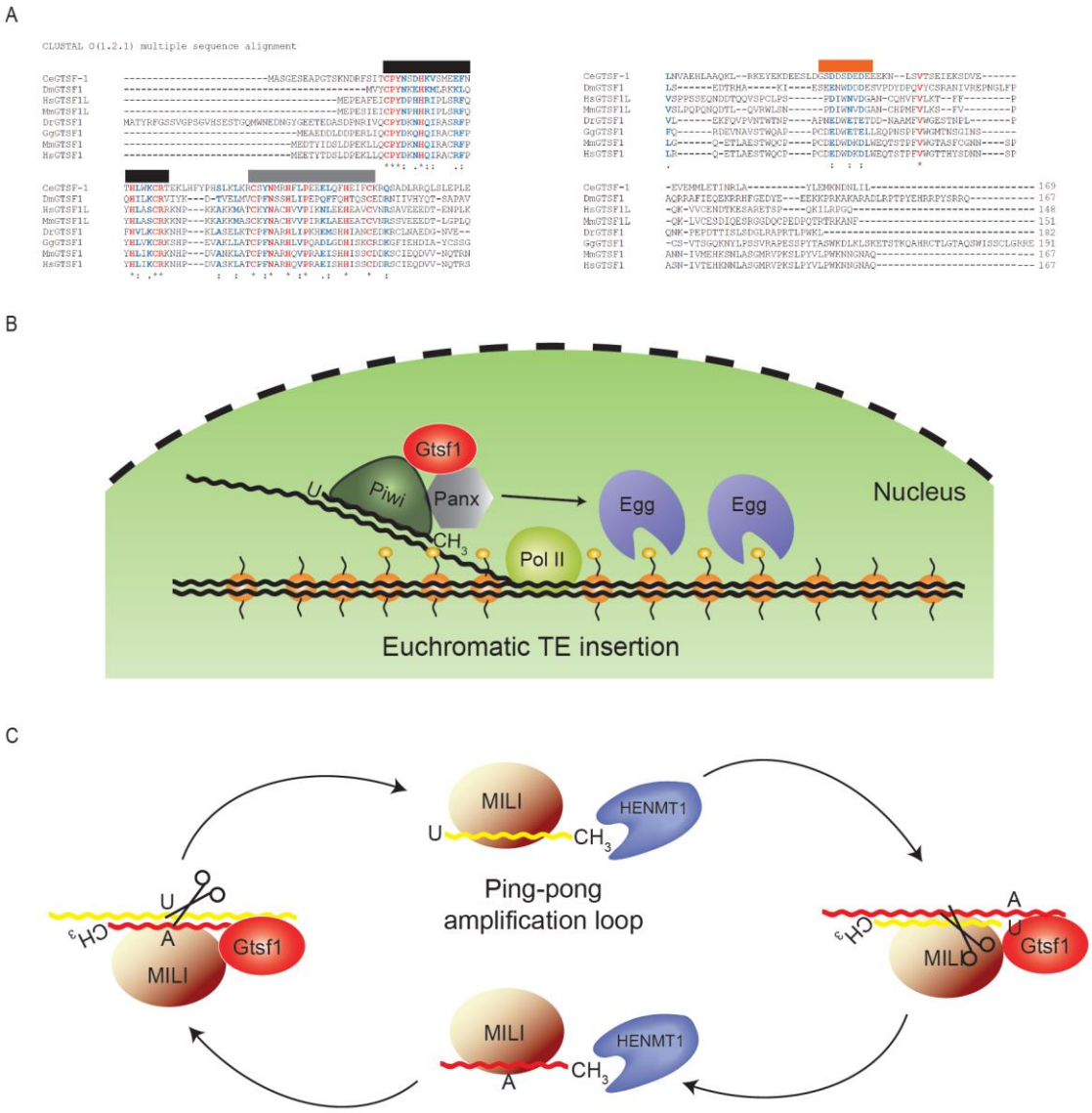


Figure I.10. **Conserved GTSF1 proteins act in sRNA pathways**. (A) Multiple sequence alignment, using ClustalO, of GTSF-1 and GTSF-1-Like proteins. An asterisk and red coloring highlight fully conserved residues. The cysteines and histidines of the CHHC Zinc fingers are fully conserved. The first and second CHHC zinc fingers are highlighted with black and grey horizontal bars, respectively. A colon indicates strong conservation of the properties of the residue, while a period indicates weakly conservation of properties. Both cases are also highlighted in blue. Of note, there is a conserved acidic region on the C-terminal tail of GTSF-1, highlighted by a horizontal orange bar. Also, *C. elegans* GTSF-1 has an extended acidic region with more Glutamic and Aspartic acid residues. Ce, *Caenorhabditis elegans*. Dm, *Drosophila melanogaster*. Mm, *Mus musculus*. Dr, *Danio rerio*. Gg, *Gallus gallus*. Hs, *Homo sapiens*. (B) In *Drosophila melanogaster*, Gtsf1 physically interacts with Piwi. Gtsf1 is required for Piwi-dependent TGS of euchromatic TE insertions. Panx acts downstream of Gtsf1 to recruit histone methyltransferases, as for example Egg, which deposit inhibitory histone modifications. Egg, Eggless; Panx, Panoramix; Pol II, RNA Polymerase II. (C) During mouse spermatogenesis, GTSF1 is required for piRNA biogenesis in MILI homotypic ping-pong amplification. Accordingly, mouse GTSF1 locates mainly to the cytoplasm.

References

Ågren, J.A., and Clark, A.G. (2018). Selfish genetic elements. PLOS Genet. *14*, e1007700.

de Albuquerque, B.F.M., Luteijn, M.J., Rodrigues, R.J.C., Bergeijk, P. van, 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.

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

Ambros, V., Lee, R.C., Lavanway, A., Williams, P.T., and Jewell, D. (2003). MicroRNAs and Other Tiny Endogenous RNAs in C. elegans. Curr. Biol. *13*, 807–818.

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

Andreeva, A., and Tidow, H. (2008). A novel CHHC Zn-finger domain found in spliceosomal proteins and tRNA modifying enzymes. Bioinformatics *24*, 2277–2280.

Aoki, K., Moriguchi, H., Yoshioka, T., Okawa, K., and Tabara, H. (2007). In vitro analyses of the production and activity of secondary small interfering RNAs in C. elegans. EMBO J. *26*, 5007–5019.

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

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.

Ashe, A., Bélicard, T., Pen, J.L., Sarkies, P., Frézal, L., Lehrbach, N.J., Félix, M.-A., and Miska, E.A. (2013). A deletion polymorphism in the Caenorhabditis elegans RIG-I homolog disables viral RNA dicing and antiviral immunity. ELife *2*, e00994.

Axtell, M.J. (2013). Classification and Comparison of Small RNAs from Plants. Annu. Rev. Plant Biol. *64*, 137–159.

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.

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.

Bayne, E.H., White, S.A., Kagansky, A., Bijos, D.A., Sanchez-Pulido, L., Hoe, K.-L., Kim, D.-U., Park, H.-O., Ponting, C.P., Rappsilber, J., et al. (2010). Stc1: A Critical Link between RNAi and Chromatin Modification Required for Heterochromatin Integrity. Cell *140*, 666–677.

Beltran, T., Barruso, C., Birkle, T., Stevens, L., Schwartz, H.T., Sternberg, P., Fradin, H., Gunsalus, K., Piano, F., Martinez-Perez, E., et al. (2018). Evolutionary analysis implicates RNA polymerase II pausing and chromatin structure in nematode piRNA biogenesis. BioRxiv 281360.

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., 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.

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.

Bobbin, M.L., and Rossi, J.J. (2016). RNA Interference (RNAi)-Based Therapeutics: Delivering on the Promise? Annu. Rev. Pharmacol. Toxicol. *56*, 103–122.

Borges, F., and Martienssen, R.A. (2015). The expanding world of small RNAs in plants. Nat. Rev. Mol. Cell Biol. *16*, 727–741.

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*, 71–94.

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.

Burkhart, K.B., Guang, S., Buckley, B.A., Wong, L., Bochner, A.F., and Kennedy, S. (2011). A Pre-mRNA–Associating Factor Links Endogenous siRNAs to Chromatin Regulation. PLOS Genet. *7*, e1002249.

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.

Caudy, A.A., and Hannon, G.J. (2004). Induction and Biochemical Purification of RNA-Induced Silencing Complex From Drosophila S2 Cells. In RNA Interference, Editing, and Modification: Methods and Protocols, J.M. Gott, ed. (Totowa, NJ: Humana Press), pp. 59–72.

Caughman, S.W., Hentze, M.W., Rouault, T.A., Harford, J.B., and Klausner, R.D. (1988). The iron-responsive element is the single element responsible for iron-dependent translational regulation of ferritin biosynthesis. Evidence for function as the binding site for a translational repressor. J. Biol. Chem. *263*, 19048–19052.

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.

Cerutti, H., and Casas-Mollano, J.A. (2006). On the origin and functions of RNA-mediated silencing: from protists to man. Curr. Genet. *50*, 81–99.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. Science *263*, 802–805.

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

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.

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

Corsi, A.K., Wightman, B., and Chalfie, M. (2015). A Transparent window into biology: A primer on Caenorhabditis elegans. WormBook 1–31.

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

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

Crick, F. (1970). Central Dogma of Molecular Biology. Nature *227*, 561–563.

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

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.

Dawkins, R. (1976). The Selfish Gene (Oxford: Oxford University Press).

Deng, W., and Lin, H. (2002). miwi, a Murine Homolog of piwi, Encodes a Cytoplasmic Protein Essential for Spermatogenesis. Dev. Cell *2*, 819–830.

Ding, S.-W. (2010). RNA-based antiviral immunity. Nat. Rev. Immunol. *10*, 632–644.

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.

Duchaine, T.F., Wohlschlegel, J.A., Kennedy, S., Bei, Y., Conte Jr., 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.

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

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

Ellis, H.M., and Horvitz, H.R. (1986). Genetic control of programmed cell death in the nematode C. elegans. Cell *44*, 817–829.

Ender, C., and Meister, G. (2010). Argonaute proteins at a glance. J Cell Sci *123*, 1819–1823.

Ferreira, H.J., Heyn, H., Muro, X.G. del, Vidal, A., Larriba, S., Muñoz, C., Villanueva, A., and Esteller, M. (2014). Epigenetic loss of the PIWI/piRNA machinery in human testicular tumorigenesis. Epigenetics *9*, 113–118.

Fire, A., Albertson, D., Harrison, S.W., and Moerman, D.G. (1991). Production of antisense RNA leads to effective and specific inhibition of gene expression in C. elegans muscle. Development *113*, 503–514.

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, 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*.

Flemr, M., Malik, R., Franke, V., Nejepinska, J., Sedlacek, R., Vlahovicek, K., and Svoboda, P. (2013). A Retrotransposon-Driven Dicer Isoform Directs Endogenous Small Interfering RNA Production in Mouse Oocytes. Cell *155*, 807–816.

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.

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.

Gainetdinov, I., Colpan, C., Arif, A., Cecchini, K., and Zamore, P.D. (2018). A Single Mechanism of Biogenesis, Initiated and Directed by PIWI Proteins, Explains piRNA Production in Most Animals. Mol. Cell *71*, 775-790.e5.

Gasciolli, V., Mallory, A.C., Bartel, D.P., and Vaucheret, H. (2005). Partially Redundant Functions of Arabidopsis DICER-like Enzymes and a Role for DCL4 in Producing trans-Acting siRNAs. Curr. Biol. *15*, 1494–1500.

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.

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

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.

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.

Gou, L.-T., Kang, J.-Y., Dai, P., Wang, X., Li, F., Zhao, S., Zhang, M., Hua, M.-M., Lu, Y., Zhu, Y., et al. (2017). Ubiquitination-Deficient Mutations in Human Piwi Cause Male Infertility by Impairing Histone-to-Protamine Exchange during Spermiogenesis. Cell *169*, 1090-1104.e13.

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

Gu, A., Ji, G., Shi, X., Long, Y., Xia, Y., Song, L., Wang, S., and Wang, X. (2010). Genetic variants in Piwi-interacting RNA pathway genes confer susceptibility to spermatogenic failure in a Chinese population. Hum. Reprod. *25*, 2955–2961.

Gu, W., Shirayama, M., Conte Jr., 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.

Guang, S., Bochner, A.F., Burkhart, K.B., Burton, N., Pavelec, D.M., and Kennedy, S. (2010). Small regulatory RNAs inhibit RNA polymerase II during the elongation phase of transcription. Nature *465*, 1097–1101.

Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A Slicer-Mediated Mechanism for Repeat-Associated siRNA 5′ End Formation in Drosophila. Science *315*, 1587–1590.

Guo, S., and Kemphues, K.J. (1995). par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. Cell *81*, 611–620.

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

Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R., and Hannon, G.J. (2001). Argonaute2, a Link Between Genetic and Biochemical Analyses of RNAi. Science *293*, 1146–1150.

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. *106*, 18674–18679.

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

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

He, C., Pillai, S.S., Taglini, F., Li, F., Ruan, K., Zhang, J., Wu, J., Shi, Y., and Bayne, E.H. (2013). Structural analysis of Stc1 provides insights into the coupling of RNAi and chromatin modification. Proc. Natl. Acad. Sci. *110*, E1879–E1888.

Hedgecock, E.M., Sulston, J.E., and Thomson, J.N. (1983). Mutations affecting programmed cell deaths in the nematode Caenorhabditis elegans. Science *220*, 1277–1279.

Henderson, I.R., Zhang, X., Lu, C., Johnson, L., Meyers, B.C., Green, P.J., and Jacobsen, S.E. (2006). Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing and DNA methylation patterning. Nat. Genet. *38*, 721–725.

Heyn, H., Ferreira, H.J., Bassas, L., Bonache, S., Sayols, S., Sandoval, J., Esteller, M., and Larriba, S. (2012). Epigenetic Disruption of the PIWI Pathway in Human Spermatogenic Disorders. PLOS ONE *7*, e47892.

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

Horn, P.J., Bastie, J.-N., and Peterson, C.L. (2005). A Rik1-associated, cullin-dependent E3 ubiquitin ligase is essential for heterochromatin formation. Genes Dev. *19*, 1705–1714.

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., Fejes Tóth, K., and Aravin, A.A. (2017). piRNA Biogenesis in Drosophila melanogaster. Trends Genet. *33*, 882–894.

Huntriss, J., Lu, J., Hemmings, K., Bayne, R., Anderson, R., Rutherford, A., Balen, A., Elder, K., and Picton, H.M. (2017). Isolation and expression of the human gametocyte-specific factor 1 gene (GTSF1) in fetal ovary, oocytes, and preimplantation embryos. J. Assist. Reprod. Genet. *34*, 23–31.

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

Iyer, L.M., Koonin, E.V., and Aravind, L. (2003). Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerases. BMC Struct. Biol. *3*, 1.

Izant, J.G., and Weintraub, H. (1984). Inhibition of thymidine kinase gene expression by anti-sense RNA: A molecular approach to genetic analysis. Cell *36*, 1007–1015.

Janeway, C.A. (1992). The immune system evolved to discriminate infectious nonself from noninfectious self. Immunol. Today *13*, 11–16.

Janic, A., Mendizabal, L., Llamazares, S., Rossell, D., and Gonzalez, C. (2010). Ectopic Expression of Germline Genes Drives Malignant Brain Tumor Growth in Drosophila. Science *330*, 1824–1827.

Jia, S., Kobayashi, R., and Grewal, S.I.S. (2005). Ubiquitin ligase component Cul4 associates with Clr4 histone methyltransferase to assemble heterochromatin. Nat. Cell Biol. *7*, 1007–1013.

Kadrmas, J.L., and Beckerle, M.C. (2004). The LIM domain: from the cytoskeleton to the nucleus. Nat. Rev. Mol. Cell Biol. *5*, 920–931.

Kamminga, L.M., Luteijn, M.J., Broeder, M.J. den, 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.

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.

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.

Kennerdell, J.R., and Carthew, R.W. (1998). Use of dsRNA-Mediated Genetic Interference to Demonstrate that frizzled and frizzled 2 Act in the Wingless Pathway. Cell *95*, 1017–1026.

Ketting, R.F. (2011). The Many Faces of RNAi. Dev. Cell *20*, 148–161.

Ketting, R.F., Haverkamp, T.H.A., van Luenen, H.G.A.M., and Plasterk, R.H.A. (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.

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

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.

Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. Nat. Rev. Mol. Cell Biol. *10*, 126–139.

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.

Koonin, E.V., Makarova, K.S., and Wolf, Y.I. (2017). Evolutionary Genomics of Defense Systems in Archaea and Bacteria. Annu. Rev. Microbiol. *71*, 233–261.

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.

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T.W., et al. (2008). DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes Dev. *22*, 908–917.

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.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947–2948.

Lau, N.C., Seto, A.G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D.P., and Kingston, R.E. (2006). Characterization of the piRNA Complex from Rat Testes. Science *313*, 363–367.

Lee, E.J., Banerjee, S., Zhou, H., Jammalamadaka, A., Arcila, M., Manjunath, B.S., and Kosik, K.S. (2011). Identification of piRNAs in the central nervous system. RNA.

Lee, H.-C., Gu, W., Shirayama, M., Youngman, E., Conte Jr., D., and Mello, C.C. (2012). C. elegans piRNAs Mediate the Genome-wide Surveillance of Germline Transcripts. Cell *150*, 78–87.

Lee, R.C., Hammell, C.M., and Ambros, V. (2006). Interacting endogenous and exogenous RNAi pathways in Caenorhabditis elegans. RNA *12*, 589–597.

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

Lewis, S.H., Quarles, K.A., Yang, Y., Tanguy, M., Frézal, L., Smith, S.A., Sharma, P.P., Cordaux, R., Gilbert, C., Giraud, I., et al. (2017). Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. Nat. Ecol. Evol. 1.

Li, C., Vagin, V.V., Lee, S., Xu, J., Ma, S., Xi, H., Seitz, H., Horwich, M.D., Syrzycka, M., Honda, B.M., et al. (2009). Collapse of Germline piRNAs in the Absence of Argonaute3 Reveals Somatic piRNAs in Flies. Cell *137*, 509–521.

Lim, S.L., Qu, Z.P., Kortschak, R.D., Lawrence, D.M., Geoghegan, J., Hempfling, A.-L., Bergmann, M., Goodnow, C.C., Ormandy, C.J., Wong, L., et al. (2015). HENMT1 and piRNA Stability Are Required for Adult Male Germ Cell Transposon Repression and to Define the Spermatogenic Program in the Mouse. PLOS Genet. *11*, e1005620.

Lin, C.-J., Hu, F., Dubruille, R., Vedanayagam, J., Wen, J., Smibert, P., Loppin, B., and Lai, E.C. (2018). The hpRNA/RNAi Pathway Is Essential to Resolve Intragenomic Conflict in the Drosophila Male Germline. Dev. Cell *46*, 316-326.e5.

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

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

Luteijn, M.J., and Ketting, R.F. (2013). PIWI-interacting RNAs: from generation to transgenerational epigenetics. Nat. Rev. Genet. *14*, 523–534.

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.

MacRae, I.J., Zhou, K., Li, F., Repic, A., Brooks, A.N., Cande, W.Z., Adams, P.D., and Doudna, J.A. (2006). Structural Basis for Double-Stranded RNA Processing by Dicer. Science *311*, 195–198.

Madhani, H.D. (2013). The Frustrated Gene: Origins of Eukaryotic Gene Expression. Cell *155*, 744–749.

Makeyev, E.V., and Bamford, D.H. (2002). Cellular RNA-Dependent RNA Polymerase Involved in Posttranscriptional Gene Silencing Has Two Distinct Activity Modes. Mol. Cell *10*, 1417–1427.

Mao, H., Zhu, C., Zong, D., Weng, C., Yang, X., Huang, H., Liu, D., Feng, X., and Guang, S. (2015). The Nrde Pathway Mediates Small-RNA-Directed Histone H3 Lysine 27 Trimethylation in Caenorhabditis elegans. Curr. Biol. *25*, 2398–2403.

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.

Martinez, J., Patkaniowska, A., Urlaub, H., Lührmann, R., and Tuschl, T. (2002). Single-Stranded Antisense siRNAs Guide Target RNA Cleavage in RNAi. Cell *110*, 563–574.

Mathias, T. (2018). Alnylam's gene silencing drug wins FDA approval. Reuters. Available at https://www.reuters.com/article/us-alnylam-pharms-fda/alnylams-gene-silencing-drug-wins-fda-approval-idUSKBN1KV22G

Matranga, C., Tomari, Y., Shin, C., Bartel, D.P., and Zamore, P.D. (2005). Passenger-Strand Cleavage Facilitates Assembly of siRNA into Ago2-Containing RNAi Enzyme Complexes. Cell *123*, 607–620.

Matzke, M.A., and Mosher, R.A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat. Rev. Genet. *15*, 394–408.

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

Mohn, F., 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.

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.

Motamedi, M.R., Verdel, A., Colmenares, S.U., Gerber, S.A., Gygi, S.P., and Moazed, D. (2004). Two RNAi Complexes, RITS and RDRC, Physically Interact and Localize to Noncoding Centromeric RNAs. Cell *119*, 789–802.

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

Mukherjee, K., Campos, H., and Kolaczkowski, B. (2013). Evolution of Animal and Plant Dicers: Early Parallel Duplications and Recurrent Adaptation of Antiviral RNA Binding in Plants. Mol. Biol. Evol. *30*, 627–641.

Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., and Grewal, S.I.S. (2001). Role of Histone H3 Lysine 9 Methylation in Epigenetic Control of Heterochromatin Assembly. Science *292*, 110–113.

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.

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.

Obbard, D.J., Gordon, K.H.J., Buck, A.H., and Jiggins, F.M. (2009). The evolution of RNAi as a defence against viruses and transposable elements. Philos. Trans. R. Soc. Lond. B Biol. Sci. *364*, 99–115.

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

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

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

Ozata, D.M., Gainetdinov, I., Zoch, A., O'Carroll, D., and Zamore, P.D. (2018). PIWI-interacting RNAs: small RNAs with big functions. Nat. Rev. Genet. 1.

Pak, J., and Fire, A. (2007). Distinct Populations of Primary and Secondary Effectors During RNAi in C. elegans. Science *315*, 241–244.

Palmer, W.H., Hadfield, J.D., and Obbard, D.J. (2018). RNA-Interference Pathways Display High Rates of Adaptive Protein Evolution in Multiple Invertebrates. Genetics *208*, 1585–1599.

Parhad, S.S., Tu, S., Weng, Z., and Theurkauf, W.E. (2017). Adaptive Evolution Leads to Cross-Species Incompatibility in the piRNA Transposon Silencing Machinery. Dev. Cell *43*, 60-70.e5.

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.

Parkin, N., Darveau, A., Nicholson, R., and Sonenberg, N. (1988). cis-acting translational effects of the 5' noncoding region of c-myc mRNA. Mol. Cell. Biol. *8*, 2875–2883.

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.

Pearson, P.N. (2001). Red Queen Hypothesis. Encycl. Life Sci. 1–4.

Perrat, P.N., DasGupta, S., Wang, J., Theurkauf, W., Weng, Z., Rosbash, M., and Waddell, S. (2013). Transposition-Driven Genomic Heterogeneity in the Drosophila Brain. Science *340*, 91–95.

Perrière, G., and Gouy, M. (1996). WWW-query: An on-line retrieval system for biological sequence banks. Biochimie *78*, 364–369.

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.

Pressman, A., Blanco, C., and Chen, I.A. (2015). The RNA World as a Model System to Study the Origin of Life. Curr. Biol. *25*, R953–R963.

Qiao, D., Zeeman, A.-M., Deng, W., Looijenga, L.H.J., and Lin, H. (2002). Molecular characterization of *hiwi*, a human member of the *piwi* gene family whose overexpression is correlated to seminomas. Oncogene *21*, 3988–3999.

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.

Rand, T.A., Ginalski, K., Grishin, N.V., and Wang, X. (2004). Biochemical identification of Argonaute 2 as the sole protein required for RNA-induced silencing complex activity. Proc. Natl. Acad. Sci. *101*, 14385–14389.

Rand, T.A., Petersen, S., Du, F., and Wang, X. (2005). Argonaute2 Cleaves the Anti-Guide Strand of siRNA during RISC Activation. Cell *123*, 621–629.

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.

Rodrigues, R.J.C., Domingues, A.M. de J., Hellmann, S., Dietz, S., Albuquerque, B. de, Renz, C., Ulrich, H.D., Butter, F., and Ketting, R. (2018). PETISCO is a novel protein complex required for 21U RNA biogenesis and embryonic viability. BioRxiv 463711.

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

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

Rosenberg, U.B., Preiss, A., Seifert, E., Jäckle, H., and Knipple, D.C. (1985). Production of phenocopies by Krüppel antisense RNA injection into Drosophila embryos. Nature *313*, 703–706.

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.

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

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

Salgado, P.S., Koivunen, M.R.L., Makeyev, E.V., Bamford, D.H., Stuart, D.I., and Grimes, J.M. (2006). The Structure of an RNAi Polymerase Links RNA Silencing and Transcription. PLOS Biol. *4*, e434.

Sapetschnig, A., Sarkies, P., Lehrbach, N.J., and Miska, E.A. (2015). Tertiary siRNAs Mediate Paramutation in C . elegans. PLOS Genet *11*, e1005078.

Sarkies, P., and Miska, E.A. (2013). Is There Social RNA? Science *341*, 467–468.

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.

Sato, K., and Siomi, M.C. (2018). Two distinct transcriptional controls triggered by nuclear Piwi-piRISCs in the Drosophila piRNA pathway. Curr. Opin. Struct. Biol. *53*, 69–76.

Sawh, A.N., and Duchaine, T.F. (2013). A Truncated Form of Dicer Tilts the Balance of RNA Interference Pathways. Cell Rep. *4*, 454–463.

Schlee, M., and Hartmann, G. (2016). Discriminating self from non-self in nucleic acid sensing. Nat. Rev. Immunol. *16*, 566–580.

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

Seth, M., Shirayama, M., Gu, W., Ishidate, T., Conte Jr., 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. *0*.

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 *0*.

Shirayama, M., Seth, M., Lee, H.-C., Gu, W., Ishidate, T., Conte Jr., D., and Mello, C.C. (2012). piRNAs Initiate an Epigenetic Memory of Nonself RNA in the C. elegans Germline. Cell *150*, 65–77.

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

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.

Simkin, A., Wong, A., Poh, Y.-P., Theurkauf, W.E., and Jensen, J.D. (2013). Recurrent and recent selective sweeps in the piRNA pathway. Evolution *67*, 1081–1090.

Simmer, F., Tijsterman, M., Parrish, S., Koushika, S.P., Nonet, M.L., Fire, A., Ahringer, J., and Plasterk, R.H.A. (2002). Loss of the Putative RNA-Directed RNA Polymerase RRF-3 Makes C. elegans Hypersensitive to RNAi. Curr. Biol. *12*, 1317–1319.

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.

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

Stark, B.C., Kole, R., Bowman, E.J., and Altman, S. (1978). Ribonuclease P: an enzyme with an essential RNA component. Proc. Natl. Acad. Sci. *75*, 3717–3721.

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.

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

Svoboda, P. (2014). Renaissance of mammalian endogenous RNAi. FEBS Lett. *588*, 2550–2556.

Swarts, D.C., Makarova, K., Wang, Y., Nakanishi, K., Ketting, R.F., Koonin, E.V., Patel, D.J., and Oost, J. van der (2014). The evolutionary journey of Argonaute proteins. Nat. Struct. Mol. Biol. *21*, 743–753.

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., and Mello, C.C. (2016). 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 *0*.

The C. elegans Sequencing Consortium (1998). Genome Sequence of the Nematode C. elegans: A Platform for Investigating Biology. Science *282*, 2012–2018.

Thivierge, C., Makil, N., Flamand, M., Vasale, J.J., Mello, C.C., Wohlschlegel, J., Jr, D.C., 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.

Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P.D. (2006). A Distinct Small RNA Pathway Silences Selfish Genetic Elements in the Germline. Science *313*, 320–324.

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. *107*, 3582–3587.

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 *303*, 672–676.

Vermaak, D., Henikoff, S., and Malik, H.S. (2005). Positive Selection Drives the Evolution of rhino, a Member of the Heterochromatin Protein 1 Family in Drosophila. PLOS Genet. *1*, e9.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I.S., and Martienssen, R.A. (2002). Regulation of Heterochromatic Silencing and Histone H3 Lysine-9 Methylation by RNAi. Science *297*, 1833–1837.

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

Wargelius, A., Ellingsen, S., and Fjose, A. (1999). Double–Stranded RNA Induces Specific Developmental Defects in Zebrafish Embryos. Biochem. Biophys. Res. Commun. *263*, 156–161.

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.

Waterhouse, P.M., Graham, M.W., and Wang, M.-B. (1998). Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. Proc. Natl. Acad. Sci. *95*, 13959–13964.

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.

Welker, N.C., Maity, T.S., Ye, X., Aruscavage, P.J., Krauchuk, A.A., Liu, Q., and Bass, B.L. (2011). Dicer's Helicase Domain Discriminates dsRNA Termini to Promote an Altered Reaction Mode. Mol. Cell *41*, 589–599.

Wen, J., Duan, H., Bejarano, F., Okamura, K., Fabian, L., Brill, J.A., Bortolamiol-Becet, D., Martin, R., Ruby, J.G., and Lai, E.C. (2015). Adaptive Regulation of Testis Gene Expression and Control of Male Fertility by the Drosophila Hairpin RNA Pathway. Mol. Cell *57*, 165–178.

Weng, C., Kosalka, A., Berkyurek, A.C., Stempor, P., Feng, X., Mao, H., Zeng, C., Li, W.-J., Yan, Y.-H., Dong, M.-Q., et al. (2018). The USTC complex co-opts an ancient machinery to drive piRNA transcription in C. elegans. BioRxiv 377390.

Willmann, M.R., Endres, M.W., Cook, R.T., and Gregory, B.D. (2011). The Functions of RNA-Dependent RNA Polymerases in Arabidopsis. Arab. Book Am. Soc. Plant Biol. *9*.

Wilson, J.E., Connell, J.E., and Macdonald, P.M. (1996). aubergine enhances oskar translation in the Drosophila ovary. Development *122*, 1631–1639.

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

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.

Yan, Z., Hu, H.Y., Jiang, X., Maierhofer, V., Neb, E., He, L., Hu, Y., Hu, H., Li, N., Chen, W., et al. (2011). Widespread expression of piRNA-like molecules in somatic tissues. Nucleic Acids Res. *39*, 6596–6607.

Yang, N., and Kazazian Jr., H.H. (2006). L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. Nat. Struct. Mol. Biol. *13*, 763–771.

Yang, P., Wang, Y., and Macfarlan, T.S. (2017). The Role of KRAB-ZFPs in Transposable Element Repression and Mammalian Evolution. Trends Genet. *33*, 871–881.

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., Miyazaki, T., Toyoda, S., Miyazaki, S., Tashiro, F., Yamato, E., and Miyazaki, J.I. (2007). Gene expression pattern of Cue110: A member of the uncharacterized UPF0224 gene family preferentially expressed in germ cells. Gene Expr. Patterns *8*, 27–35.

Yoshimura, T., Toyoda, S., Kuramochi-Miyagawa, S., Miyazaki, T., Miyazaki, S., Tashiro, F., Yamato, E., Nakano, T., and Miyazaki, J.I. (2009). Gtsf1/Cue110, a gene encoding a protein with two copies of a CHHC Zn-finger motif, is involved in spermatogenesis and retrotransposon suppression in murine testes. Dev. Biol. *335*, 216–227.

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. e42054.

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

Zamore, P.D., Tuschl, T., Sharp, P.A., and Bartel, D.P. (2000). RNAi: Double-Stranded RNA Directs the ATP-Dependent Cleavage of mRNA at 21 to 23 Nucleotide Intervals. Cell *101*, 25–33.

Zeng, C., Weng, C., Wang, X., Yan, Y.-H., Li, W.-J., Xu, D., Hong, M., Liao, S., Feng, X., Dong, M.-Q., et al. (2018). Differential phase partition of a PICS complex is required for piRNA processing and chromosome segregation in C. elegans. BioRxiv 463919.

Zhang, D., Tu, S., Stubna, M., Wu, W.-S., Huang, W.-C., Weng, Z., and Lee, H.-C. (2018). The piRNA targeting rules and the resistance to piRNA silencing in endogenous genes. Science *359*, 587–592.

Zhang, H., Kolb, F.A., Jaskiewicz, L., Westhof, E., and Filipowicz, W. (2004). Single Processing Center Models for Human Dicer and Bacterial RNase III. Cell *118*, 57–68.

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.

Zhuang, J.J., and Hunter, C.P. (2011). Tissue-specificity of Caenorhabditis elegans Enhanced RNAi Mutants. Genetics genetics.111.127209.

Chapter II

GTSF-1 is required for formation of a functional RNA-dependent RNA Polymerase complex in *C. elegans*

Miguel Vasconcelos Almeida[1], Sabrina Dietz[2], Stefan Redl[1], Emil Karaulanov[3], Andrea Hildebrandt[4], Christian Renz[5], Helle D. Ulrich[5], Julian König[4], Falk Butter[2] and René F. Ketting[1]

Institute of Molecular Biology, Ackermannweg 4, 55128 Mainz, Germany

[1]Biology of Non-coding RNA Group

[2]Quantitative Proteomics Group

[3]Bioinformatics Core Facility

[4]Genomic Views of Splicing Regulation Group

[5]Maintenance of Genome Stability Group

This chapter is a reproduction of the following scientific paper:

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

Author contributions:

Conceptualization: MVA and RFK; Investigation: MVA, SD, SR, AH, CR; Formal analysis: MVA, SD, SR, EK, AH; Visualization: MVA, SD, SR, EK, AH; Writing - original draft: MVA and RFK; Writing - review & editing: all authors contributed; Funding acquisition: HDU, JK, FB, RFK.

Abstract

AGO proteins and their associated sRNAs are evolutionarily conserved regulators of gene expression. GTSF1 proteins, characterized by two tandem CHHC zinc fingers and an unstructured, C-terminal tail, are conserved in animals and have been shown to interact with Piwi clade AGOs, thereby assisting their activity. We identified the *C. elegans Gtsf1* homolog, named it *gtsf-1* and sought out to characterize it in the context of the sRNA pathways of *C. elegans*. We report that GTSF-1 is not required for Piwi-mediated gene silencing. Instead, *gtsf-1* mutants show a striking depletion of 26G-RNAs, a class of endogenous sRNAs, fully phenocopying *rrf-3* mutants. We show, both *in vivo* and *in vitro*, that GTSF-1 interacts with RRF-3 via its CHHC zinc fingers. Furthermore, we demonstrate that GTSF-1 is required for the assembly of a larger RRF-3 and DCR-1-containing complex (ERI complex) thereby allowing for 26G-RNA generation. We propose that GTSF-1 homologs may act to drive the assembly of larger complexes that act in sRNA production and/or in imposing sRNA-mediated silencing activities.

Introduction

Endogenous small non-coding RNAs are responsible for regulating gene expression in many organisms. These sRNAs) act within the context of RNAi or RNAi-like pathways. In a variety of situations, these pathways provide an RNA-based protection against “foreign” genetic elements such as TEs and viruses (Ketting, 2011; Luteijn and Ketting, 2013).

In many RNAi-like pathways, sRNAs are generated from dsRNA precursors by Dicer, a conserved RNase III-related enzyme (Ketting, 2011). Subsequently, sRNAs associate with AGO family proteins, and guide them to target transcripts with complete or partial sequence complementarity. Upon AGO binding, transcripts are usually destabilized or translationally inhibited in the cytoplasm. However, some AGOs have nuclear localization and regulate gene expression on the transcriptional level. For instance, in *C. elegans*, NRDE-3 and HRDE-1 are nuclear AGOs that silence genes on the transcriptional level in the soma and in the germline, respectively (Buckley et al., 2012; Guang et al., 2008).

C. elegans, like plants and yeast, has RdRPs dedicated to the production of sRNAs. *C. elegans* has four RdRP genes, RRF-1/-2/-3 and EGO-1. It is believed that these RdRPs synthesize sRNA fragments in an unprimed manner (Billi et al., 2014). Two of these RdRPs, RRF-1 and EGO-1, generate sRNAs after target recognition by a primary AGO. These secondary sRNAs (22G-RNAs) contain a 5’-triphosphate group, have a bias for a 5’ guanosine and are mostly 22 nucleotides long (Billi et al., 2014). The RdRP enzyme RRF-3 is required for the biogenesis of another endogenous sRNA population, known as 26G-RNAs, which are mainly 26 nucleotides long, have a 5’ guanosine bias and a 5’-monophosphate (Conine

et al., 2010; Gent et al., 2009, 2010; Han et al., 2009; Pavelec et al., 2009; Vasale et al., 2010). The fourth RdRP gene, *rrf-2*, has no described function in RNAi-related pathways.

26G-RNAs can associate with three AGOs. During spermatogenesis, 26G-RNAs associate with the AGOs ALG-3 and ALG-4 (from here on indicated as ALG-3/4). These AGOs are required for normal fertility and mostly target spermatogenic transcripts, mediating post-transcriptional gene silencing (Conine et al., 2010, 2013; Han et al., 2009). Also, ALG-3/4 targets show a significant overlap with targets of CSR-1, an AGO protein that has been suggested to potentiate gene expression, rather than gene silencing (Conine et al., 2013). During oogenesis and embryogenesis, 26G-RNAs associate with the AGO ERGO-1 (Gent et al., 2010; Han et al., 2009; Vasale et al., 2010). In contrast to the ALG-3/4-bound 26G-RNAs, ERGO-1-bound 26G-RNAs are 2'O-methylated by HENN-1, which increases their stability (Billi et al., 2012; Kamminga et al., 2012; Montgomery et al., 2012). The main targets of ERGO-1 are recently duplicated paralogs and pseudogenes (Vasale et al., 2010). Upon target recognition, ERGO-1 triggers the production of 22G-RNAs. In turn, these 22G-RNAs direct gene-silencing and presumably associate with unknown cytoplasmic AGOs, as well as the somatic nuclear AGO protein NRDE-3 (Guang et al., 2008; Vasale et al., 2010). NRDE-3 and other NRDE factors lead to TGS of their targets, a process accompanied by H3K9 trimethylation of the target *locus* (Burkhart et al., 2011; Burton et al., 2011).

Mutants defective in the generation of 26G-RNAs, in particular those associated with ERGO-1, are hypersensitive to exo-RNAi. This enhanced RNAi (Eri) phenotype, is believed to stem from the fact that 26G-RNA pathways share common components with the exo-RNAi pathway (Duchaine et al., 2006; Gu et al., 2009; Yigit et al., 2006). Interestingly, many of the identified proteins that restrict exo-RNAi in wild-type animals form a complex: the ERI complex (Duchaine et al., 2006; Thivierge et al., 2012; Yigit et al., 2006). The ERI complex has a core module that has been proposed to consist of the RdRP RRF-3 and its close interacting partners, the DExD/H box helicase DRH-3 and the Tudor domain-containing protein ERI-5 (Duchaine et al., 2006; Thivierge et al., 2012). To become active, this core complex needs to interact with DCR-1, an interaction that requires ERI-5 (Thivierge et al., 2012). Additionally, ERI-1 and ERI-3 are accessory factors of the ERI complex that promote 26G-RNA biogenesis (Billi et al., 2014; Duchaine et al., 2006). Further mechanistic insights into ERI complex assembly and function are severely lacking.

Besides 22G- and 26G-RNAs, *C. elegans* produces 21U-RNAs (Billi et al., 2014). The 21U-RNAs interact with PRG-1, one of the *C. elegans* Piwi protein homologs, and are also known as the piRNAs of *C. elegans* (Billi et al., 2014). In many organisms, the Piwi-piRNA pathway provides protection against TEs (Luteijn and Ketting, 2013), and also in *C. elegans*, 21U-RNAs contribute to the defense against TE activity (Bagijn et al., 2012; de Albuquerque et al., 2015; Phillips et al., 2015). Interestingly, 21U-RNAs can initiate a nuclear, 22G-RNA-mediated pathway. These 22G-RNAs, bound by the nuclear AGO HRDE-

1 can affect histone modification patterns on targeted *loci*, and can establish a very stably inherited form of gene silencing (named RNAe) that no longer depends on continued exposure to 21U-RNAs (Ashe et al., 2012; Buckley et al., 2012; Lee et al., 2012; Luteijn et al., 2012; Shirayama et al., 2012).

Genome-wide screens have uncovered many factors involved in the piRNA pathway and TE silencing in *D. melanogaster* (Czech et al., 2013; Handler et al., 2013; Muerdter et al., 2013). Many of these factors are poorly conserved evolutionarily. Gtsf1, a double CHHC zinc finger protein, represents one of the few Piwi pathway components that displays clear evolutionary conservation. Dmel-Gtsf1 is required for fertility and associates directly with Piwi (Dönertas et al., 2013; Ohtani et al., 2013). Interestingly, in the absence of Dmel-Gtsf1, Piwi is still nuclear and loaded with piRNAs, but cannot silence TEs. Hence, Dmel-Gtsf-1 has been proposed to be required for the execution of Piwi-mediated silencing activities following target recognition. Also in mice, Mmus-GTSF1 is required for fertility and Mmus-GTSF1-related proteins have been shown to interact with Piwi proteins (Takemoto et al., 2016; Yoshimura et al., 2007, 2009, 2018).

The precise molecular function of GTSF1, or of its isolated domains is unknown. GTSF1 homologs have two tandem CHHC zinc finger domains and an unstructured C-terminal tail. *In silico* studies showed that CHHC zinc fingers are found in three protein families (Andreeva and Tidow, 2008): 1) U11-48K proteins, members of the alternative spliceosome, 2) TRM13 tRNA methyltransferases and 3) GTSF1-related proteins. These CHHC-domains behave as independent folding units and bind stoichiometrically to zinc (Andreeva and Tidow, 2008). The CHHC zinc finger of human U11-48K was shown to bind to the 5’ splice site of U12-dependent introns (Tidow et al., 2009), suggesting that CHHC zinc fingers bind RNA. Interestingly, the GTSF family is the only family of proteins that has two CHHC zinc fingers in tandem (Andreeva and Tidow, 2008).

Given its strong participation in Piwi-induced TE silencing in *Drosophila* and mouse, and that it is one of the few factors acting with piRNAs that displays wide conservation, we decided to characterize the function of GTSF-1 in *C. elegans*. Strikingly, we find that GTSF-1 is not involved in TE silencing and does not affect 21U-RNA production or activity in *C. elegans*. Instead, GTSF-1 associates with the RdRP RRF-3 and is required to assemble the ERI complex. We propose that GTSF1 proteins in general, may be present in smaller pre-complexes that may promote the assembly of larger protein-RNA complexes that elicit downstream enzymatic activities, such as sRNA production or the establishment of TGS.

Results

GTSF-1 is enriched in the germline but not in P-granules

T06A10.3, the downstream partner of *lsy-13* in an operon on chromosome IV, was identified by reciprocal BLAST as the *C. elegans gtsf1* homolog, and was named *gtsf-1* (**Figure II.1A**). GTSF-1, like its mouse and fly homologs, has two predicted CHHC zinc fingers (Andreeva and Tidow, 2008). The cysteine and histidine residues of the zinc fingers, as well as several acidic residues on the C-terminal region are conserved from worms and flies to mouse, zebrafish and human (**Figure I.10A**). We produced three independent *gtsf-1* deletion alleles using CRISPR-Cas9 technology (Friedland et al., 2013) (**Figures II.1A-B** and **II.S1A**). Five times outcrossed, homozygous *gtsf-1* mutants are fertile and do not show any obvious morphological defects. No GTSF-1 protein is detected in the mutants by Western blot, using an anti-GTSF-1 polyclonal antibody (**Figure II.1C**). Expression of *lsy-13*, the operon partner, does not seem to be affected in *gtsf-1(xf43)* mutants (**Figure II.S1B**).

To address the expression pattern of *gtsf-1* throughout development, we used publicly available RNA-sequencing datasets (Boeck et al., 2016). During embryonic development, larval development and adulthood, *gtsf-1* is moderately expressed (levels ranging from 0.4 to 7.2 depth of coverage per base per million reads [DCPM], **Figure II.S1C-D**). Notably, *gtsf-1* RNA levels are highest

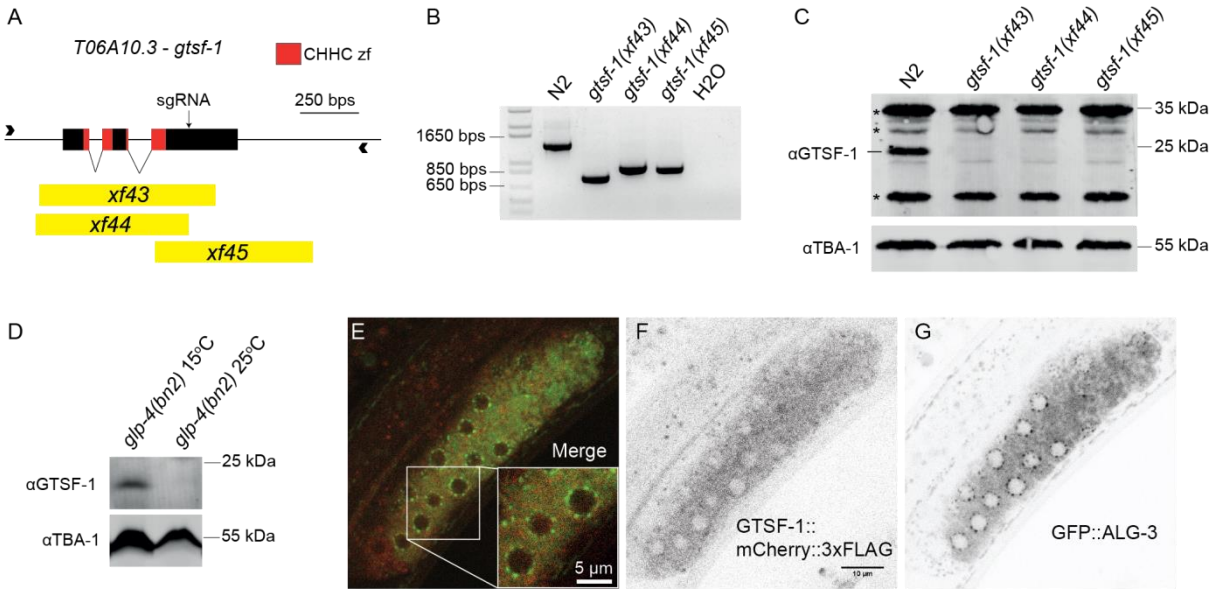


Figure II.1. *T06A10.3*, the *C. elegans* homolog of *gtsf-1* is expressed in the cytoplasm and is germline-enriched. (A) Overview of the *T06A10.3* gene in chromosome IV of *C. elegans*. The exons are represented as black boxes, the CHHC zinc finger domains are shown in red, and the black arrow corresponds to the cut site of the sgRNA used. The deletion alleles are represented in yellow. (B) PCR analysis of the deletion alleles using primers represented by arrowheads in (A). (C) Western blot analysis of mixed-stage wild-type and mutant worm extracts using a polyclonal anti-GTSF-1 antibody. TBA-1, one of the *C. elegans* alpha-tubulins was used as a loading control. (D) Western blot analysis of *glp-4(bn2)* mutant worms grown at the non-permissive temperature of 25 ºC, which precludes the development of the germline, and 15 ºC. (E-G) Representative confocal fluorescence microscopy images showing the presence of GTSF-1 and ALG-3 tagged proteins in a gonad of a L4 double transgenic worm, in the *alg-3/4; gtsf-1* triple mutant background. Scale bars correspond to 10 µm and 5 µm in the case of the inset.

during the first 300 minutes of embryonic development (2.38-7.2 DCPM), suggesting that *gtsf-1* mRNA may be maternally deposited (**Figure II.S1C**). During larval development, *gtsf-1* mRNA reaches highest levels during the L4 and young adult stage (0.89-1.2 DCPM), correlating with germline development (**Figure II.S1D**).

To address potential germline enrichment of GTSF-1, we used *glp-4(bn2)* worms, which lack a germline when grown at 25 ^{o}C. Western blot experiments on these animals (**Figure II.1D**) indicate that GTSF-1 is enriched in the germline, since we could not detect GTSF-1 in *glp-4(bn2)* worms grown at 25 ^{o}C. These data are supported by recent germline transcriptomes using dissected male and female gonads (Ortiz et al., 2014) that detected *gtsf-1* transcript in gonads irrespective of gender (**Figure II.S1E**). To address subcellular localization, we produced a *gtsf-1::mCherry::3xflag* single-copy transgene controlled by the germline-specific *gld-1* promoter (Merritt et al., 2008) and introduced it into a *gtsf-1(xf43)*; *alg-3(tm1155)*; *alg-4(ok1041)* triple mutant background, also expressing a GFP::ALG-3 fusion protein. In these animals, we observed GTSF-1::mCherry::3xFLAG protein localized throughout the germline cytoplasm in L4 stage animals. GTSF-1 does not appear to be concentrated in P-granules, marked by GFP-tagged ALG-3 (**Figure II.1E-G**), a known P-granule component (Conine et al., 2010).

These data indicate that *C. elegans* GTSF-1 is enriched in the germline cytoplasm, but mostly outside perinuclear granules.

GTSF-1 is not involved in the 21U-RNA pathway and transposon silencing in *C. elegans*

Next, we wanted to address whether *gtsf-1* is involved in TE silencing. To test this, we used a strain with the *unc-22(st136)* allele, which has the *unc-22* gene interrupted by a Tc1 transposon (Ketting et al., 1999) (**Figure II.S2A**). Animals carrying the *unc-22(st136)* allele exhibit the so-called twitcher phenotype. When a gene that participates in TE silencing, such as *mut-7* (Ketting et al., 1999), is impaired in the *unc-22(st136)* background, TEs will become mobile and phenotypical reversions to wild-type movement can be observed. All three *gtsf-1* mutant alleles were crossed into the *unc-22(st136)* background and no reversions of the twitcher phenotype were observed after culturing the strains for several generations, in ten biological replicates per allele (comprising a reversion frequency of $<10^{-5}$, **Figure II.S2B**).

To further characterize the role of *gtsf-1* in the sRNA pathways of *C. elegans*, we sequenced sRNAs from wild-type and *gtsf-1* synchronized gravid adults, in triplicates (experimental design in **Figure II.2A,** sequencing statistics in the **Supplementary Information**). To enrich for different sRNA species we employed different library preparations to each biological replicate. To increase the likelihood of cloning 22G-RNAs, which have a 5' triphosphate, we used Tobacco Acid Phosphatase

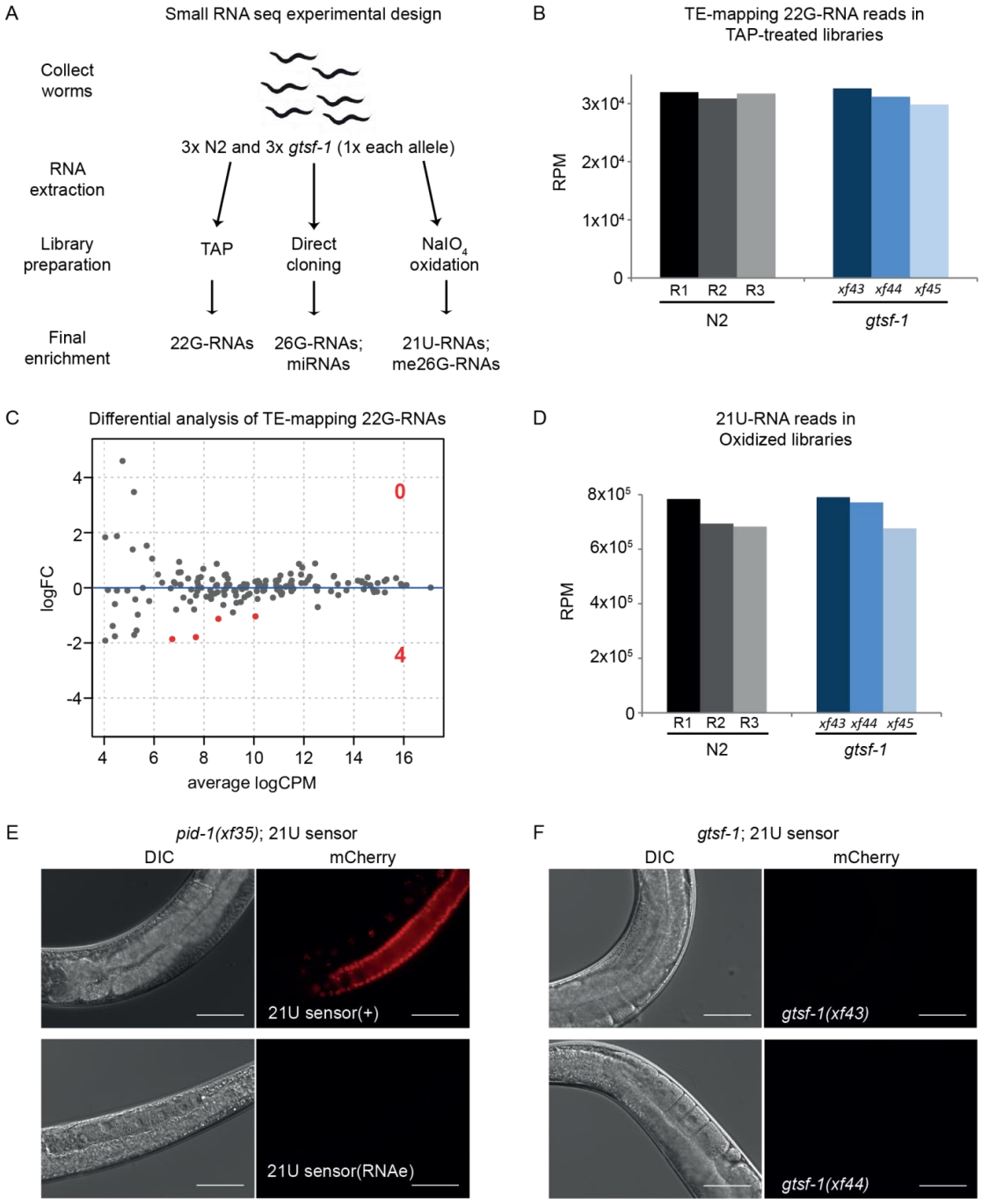


Figure II.2. *C. elegans* GTSF-1 is not involved in the 21U-RNA pathway and TE silencing. (A) Experimental design of sRNA sequencing. Wild-type and *gtsf-1* mutant gravid adult worms were collected in triplicates. For *gtsf-1*, one sample of each allele was used as a biological replicate. Libraries were subjected to a triad of treatments to enrich for different sRNA species. TAP, Tobacco Acid Pyrophosphatase. (B) Similar abundance of TE-mapping 22G-RNA reads in TAP-treated libraries in wild-type (N2) and *gtsf-1* mutants (Welch two sample t-tests *p-value* = 0.75). Normalized levels in Reads Per Million (RPM) for each biological replicate are shown. (C) Differential analysis (MA-plot) of TE-mapping 22G-RNAs in *gtsf-1* mutants vs wild-type. sRNA reads from TAP-treated libraries were used for this analysis. Only four TEs show significantly downregulated (1% FDR) sRNA levels in *gtsf-1* mutants (see **Table II.S1**). LogFC, Log2 Fold Change. logCPM, log2 Counts Per Million. (D) Similar abundance of 21U-RNA reads in oxidized libraries in wild-type (N2) and *gtsf-1* mutants (Welch two sample t-tests *p-value* = 0.62). Normalized levels in Reads Per Million (RPM) for each biological replicate are shown. (E-F) Testing the participation of *gtsf-1* in the 21U-RNA pathway. For each figure, left panels are DIC while right panels show mCherry fluorescence channel. (E) Photomicrographs of adult worms carrying a 21U-RNA reporter in the *pid-1(xf35)* background. The panels above show a strain in which the 21U-sensor is still dependent on the 21U-RNA pathway, because in the absence of PID-1, mCherry can be observed in the germline. The panels below show a strain in which reporter silencing became independent of the 21U-RNA pathway, a state known as RNAe. (F) Micrographs of 21U-sensor;*gtsf-1* worms exhibiting the sensor repressed. This images are representative of 21U-sensor;*gtsf-1* worms originating from the crosses with both strains shown in (E) (schematics of the crosses are shown in Figure EV2F-G). Scale bars represent 50 µm.

(TAP). To enrich for sRNA species with a 2'-O methyl group on their 3' end (21U-RNAs and ERGO-1-associated 26G-RNAs), we oxidized the RNA before library preparation with $NaIO_4$ (the 2'-O-methyl group protects sRNAs from oxidation). Finally, we used untreated RNA to capture a higher fraction of sRNAs carrying a 5' monophosphate, irrespective of their 3' end methylation status (ERGO-1 and ALG-3/4- bound 26G-RNAs and miRNAs). The latter type of libraries will be hereafter referred to as "directly cloned". Sequences between 18-30 nucleotides were analyzed and read counts were normalized to the total number of mapped reads in each sample, excluding structural reads (see **Supplementary Experimental Procedures**).

Consistent with the phenotypic experiments using the *unc-22(st136)* Tc1-transposition reporter, we did not observe major differences in sRNA reads mapping to TEs between wild-type and *gtsf-1* animals (**Figure II.2B-C** and **Table II.S1**). Likewise, only two miRNAs were affected in *gtsf-1* mutants (**Figure II.S2C** and **Table II.S1**).

Also, the steady-state 21U-RNA levels are not significantly affected in *gtsf-1* mutants (**Figure II.2D** and **II.S2D**). To further test participation of *gtsf-1* in the 21U-RNA pathway we performed crosses combining *gtsf-1* mutant alleles with an mCherry reporter for 21U-RNA activity (Bagijn et al., 2012; Luteijn et al., 2012) (**Figure II.S2E**). The reporter strains have a *pid-1(xf35)* mutation in the background to inform on the status of the sensor (**Figure II.2E**) (de Albuquerque et al., 2014), which can be under RNAe (insensitive to the presence of PID-1, **Figure II.2E**, lower panels) or not (**Figure II.2E**, upper panels). Loss of *gtsf-1* does not activate this reporter in either state, indicating it is not required for 21U-RNA-mediated silencing activity and RNAe (**Figure II.2E-F** and **II.S2F-G**).

Overall, these data indicate that GTSF-1 is neither involved in TE silencing, nor in the 21U-RNA/RNAe pathway in *C. elegans,* in sharp contrast with the described function of GTSF1 proteins in mouse and fly.

gtsf-1 mutants recapitulate phenotypes of 26G-RNA pathway mutants

Given that *gtsf-1* is not involved in 21U-RNA-mediated gene silencing in *C. elegans*, we looked for other phenotypes that might be indicative of a role for GTSF-1 in other endogenous sRNA pathways. We noticed that populations of *gtsf-1* mutant animals grow slower compared to wild-type. This could reflect either developmental or fertility defects. When synchronized by bleaching, *gtsf-1* animals grew synchronous with wild-type. In contrast, we noticed a striking reduction in brood size at 20 ^{o}C, and temperature-sensitive sterility at 25 ^{o}C (**Figure II.3A**). When grown at 25 ^{o}C, *gtsf-1* mutant animals mostly produced unfertilized oocytes (**Figure II.S3A-C**). Importantly, two independent germline-specific *gtsf-1::mCherry::3xflag* transgenes (including *xfIs47*, the transgene shown in **Figure II.1E-F**) completely rescue these defects (**Figures II.3A** and **II.S3C**). These data clearly demonstrate that *gtsf-1* mutants display a temperature-sensitive fertility defect.

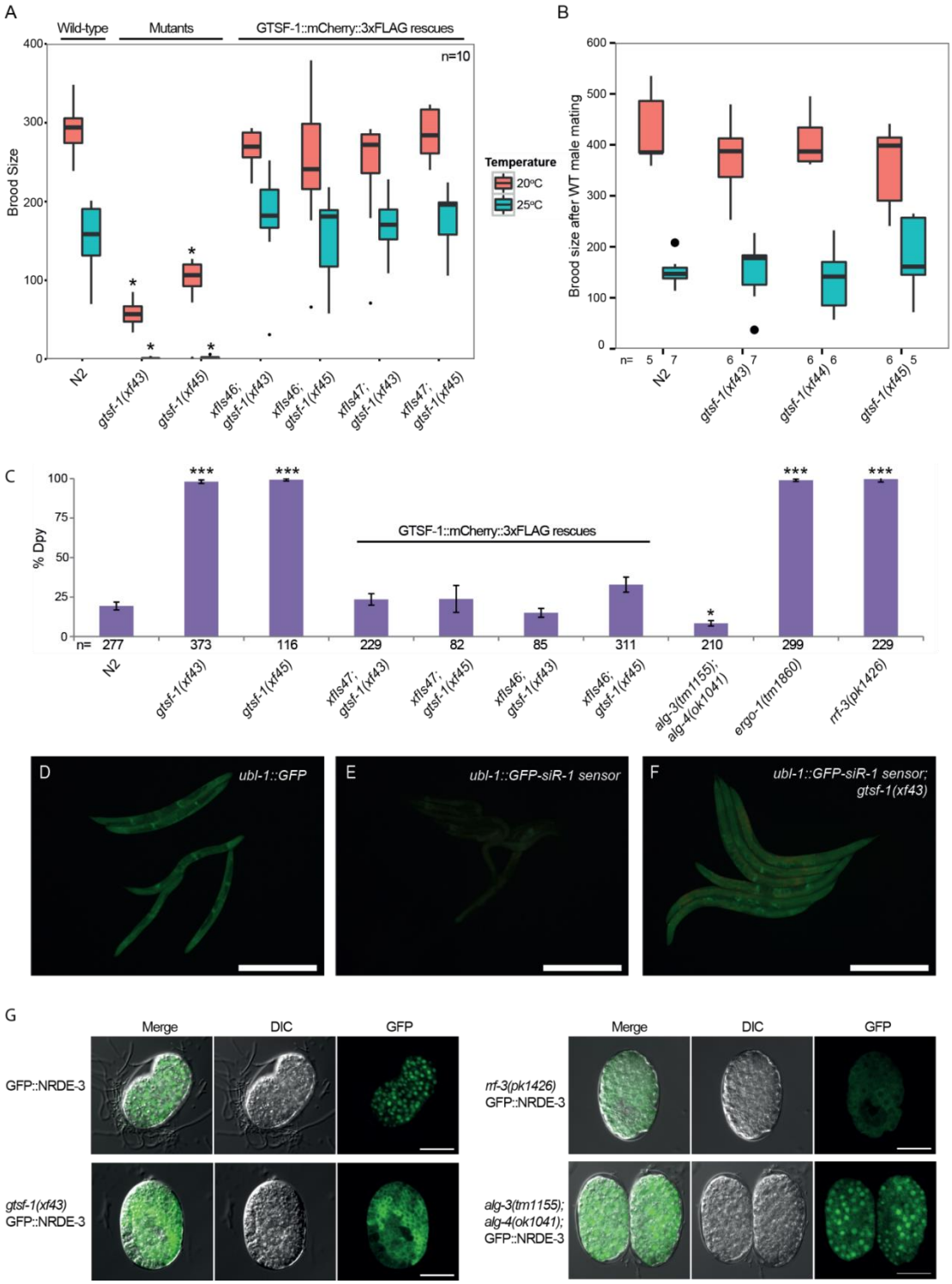


Figure II.3. *gtsf-1* animals phenocopy 26G-RNA pathway mutants. (A) Boxplot of brood size counts at 20 ºC and 25 ºC. The progenies of 10 worms were counted for each strain and each temperature. Asterisks indicate *p-value*<0.0002 as assessed by Mann-Whitney-Wilcoxon tests comparing N2 worms with the other strains. Comparisons were done for each respective temperature. (B) Hermaphrodites with the genotypes indicated on the x-axis were mated with wild-type males and the progeny was counted. n for each condition is indicated in the figure below the x-axis. Mann-Whitney-Wilcoxon tests yielded *p-values*>0.4. (C) Assaying sensitivity to somatic *dpy-13* RNAi. The rescuing transgenes shown in Figure II.3A are also assayed here. Total number of worms assayed is represented in the figure. Mann-Whitney-Wilcoxon tests were used to test if penetrance of *dpy-13* RNAi treatment was significantly different between N2 and mutant worms. Single asterisk indicates *p-value*=0.027 while triple asterisks indicate *p-values*<2.3e-05, *p-values* calculated using Mann-Whitney-Wilcoxon tests. Error bars represent the S.E.M. (D-F) GFP fluorescence images of worms carrying 22G-siR-1 sensor transgenes (see also Figure EV3G). Scale bars represent 0.5 mm. (D) Animals carrying the control transgene with no 22G-siR-1 binding site. (E) Strains carrying the 22G-siR-1 sensor. (F) GFP signal in the absence of GTSF-1. (G) Micrographs of GFP::NRDE-3 embryos in various genetic backgrounds. Scale bars represent 10 µm.

Temperature-sensitive sterility and embryonic lethality are recurring phenotypes of factors acting in endogenous sRNA pathways in *C. elegans*. For example, mutations in mutator genes, Eri genes, *rrf-3*, *drh-3* and *alg-3/4*, result in temperature-sensitive sterility at 25 °C (Billi et al., 2014; Conine et al., 2010; Duchaine et al., 2006; Gent et al., 2009; Gu et al., 2009; Han et al., 2009; Ketting et al., 1999; Pavelec et al., 2009). In some of those mutants, like *alg-3/4, eri-1* and *rrf-3*, these fertility defects can be rescued by wild-type sperm, indicative of a sperm defect (Conine et al., 2010; Gent et al., 2009; Pavelec et al., 2009). Upon crossing *gtsf-1* hermaphrodites with wild-type males, both the reduced brood size at 20 °C and the temperature-sensitive sterility at 25 °C were rescued practically to wild-type levels (**Figure II.3B**). Furthermore, we noticed that *gtsf-1* mutants have a mild high-incidence of males (him) phenotype (**Figure II.S3D**), again, similar to *alg-3/4*, many Eri and mutator mutants (Conine et al., 2010; Gent et al., 2009; Ketting et al., 1999).

One phenotype that distinguishes mutator mutants from Eri mutants is RNAi-sensitivity. Mutators are resistant to exogenous RNAi while Eri mutants are hypersensitive. *gtsf-1* mutants displayed normal sensitivity to RNAi against the germline gene *pos-1* (**Figures II.S3E, Table II.1**), but showed hypersensitivity to RNAi targeting somatic genes, as *dpy-13* (**Figures II.3C** and **II.S3F, Table II.1**), *lir-1* and *pop-1* (**Table II.1**), similarly to *rrf-3* and *ergo-1* mutant worms (Duchaine et al., 2006; Yigit et al., 2006). In contrast, *alg-3/4* double mutants did not display RNAi-hypersensitivity. Two independent, germline-specifically expressed *gtsf-1* transgenes rescued the RNAi hypersensitivity almost to wild-type levels (**Figures II.3C** and **II.S3F**). We note that this rescue of a somatic phenotype with a germline-expressed transgene likely derives from the strong maternal effect of the 26G-RNA pathway (Zhuang and Hunter, 2011). We conclude that *gtsf-1* mutants have an Eri phenotype.

Loss of ERGO-1 and RRF-3, but not ALG-3/4, derepresses a ubiquitously expressed GFP transgene that reports on the activity of a specific 22G-RNA (**Figures II.3D-E** and **II.S3G-H**), that is produced in response to ERGO-1 (Montgomery et al., 2012). GTSF-1 is also required for proper

	Germline targets	**Somatic targets**		
	pos-1[1]	*dpy-13*[2]	*lir-1*[3]	*pop-1*[4]
N2	+	+	-	-
gtsf-1(xf43)	+	+++	+++	+++
gtsf-1(xf44)	+	+++	+++	+++
gtsf-1(xf45)	+	+++	+++	+++

Table II.1. *gtsf-1* mutants have exogenous RNAi defects.

"-"Denotes no response to RNAi; "+" indicates response to RNAi; and "+++" indicates enhanced response to RNAi.
[1]Scored for Embryonic lethality;
[2]Scored for strength and penetrance of Dpy phenotype;
[3]Scored for larval arrest;
[4]Scored for bursted and protruded vulva phenotypes.

silencing of this transgene, indicating that the activity of GTSF-1 is required for ERGO-1/RRF-3-driven silencing (**Figures II.3F** and **II.S3H**). We further tested GTSF-1 participation in the ERGO-1-dependent 26G-RNA pathway more broadly, by using a GFP::NRDE-3 expressing transgene. GFP::NRDE-3 in wild-type animals displays nuclear localization, but is cytoplasmic in *ergo-1* mutants because it fails to be loaded with 22G-RNAs (Guang et al., 2008). Nuclear localization is similarly affected by *gtsf-1(xf43)* and *rrf-3(pk1426)* mutation (**Figure II.3G**). In contrast, *alg-3/4* mutations did not cause mislocalization of GFP::NRDE-3 from the nucleus.

Overall, we conclude that *gtsf-1* mutants display phenotypes of *alg-3/4* and *ergo-1* mutants. As such, loss of GTSF-1 perfectly phenocopies loss of the RdRP enzyme RRF-3, suggesting that GTSF-1 acts at a very upstream step in the 26G-RNA pathway.

26G-RNA levels are strongly reduced in *gtsf-1* mutants

Given our phenotypic analysis, we reasoned that GTSF-1 may affect 26G-RNA biogenesis. Indeed, 26G-RNA levels are severely depleted in *gtsf-1* mutants (**Figure II.4A**). This effect is observed both in the directly cloned as well as in the oxidized libraries, suggesting that both classes of 26G-RNAs, unmethylated (ALG-3/4-bound) and 2'O-methylated (ERGO-1-bound), respectively, are affected by GTSF-1 (**Figure II.4A**). The levels of 26G-RNAs derived from all gene classes are similarly reduced upon loss of GTSF-1 (**Figure II.4B**).

Next, we defined high-confidence targets (at 1% FDR) of GTSF-1-dependent sRNAs for each library treatment (**Figure II.4C-E,** lists of targets in **Table II.S1**). Targets were defined as genes that have a significant depletion of sRNAs in the mutant, in comparison with wild-type. The targets defined in the oxidized libraries (enriching for methylated 26G-RNAs) significantly overlapped with the targets of the TAP-treated libraries (enriching for 22G-RNAs, **Figure II.S4A**). These results suggest that genes that lose 2'-O-methylated 26G-RNAs also tend to lose downstream 22G-RNAs. This tendency is observed for all gene classes (**Figure II.4F**). Next, we wanted to address if there are changes in GTSF-1 target gene expression concomitantly with loss of 26G-/22G-RNAs. Indeed, in the absence of GTSF-1, its targets are upregulated as assessed by RT-qPCR (**Figure II.4G**, in levels consistent with previously published RT-qPCR data, see Duchaine et al., 2006; Pavelec et al., 2009; Vasale et al., 2010). Furthermore, our sets of GTSF-1 targets significantly overlap with a publicly available dataset from an ERGO-1 RIP (Vasale et al., 2010) (**Figure II.4H**). Consistently, genes identified in the ERGO-1 RIP are depleted of 22G-RNAs in our TAP library dataset (**Figure II.4I**). Of note, several of the GTSF-1 targets that were shown to be upregulated in **Figure II.4G** were also identified as ERGO-1 targets (Vasale et al., 2010, namely *E01G4.5*, *K02E2.6*, *W04B5.2* and *Y37E11B.2*). ERGO-1 targets include paralog genes and pseudogenes (Vasale et al., 2010). Accordingly, we did not find strongly enriched gene ontology terms for the targets defined in the oxidized and TAP-treated libraries (**Table II.S1**). Furthermore,

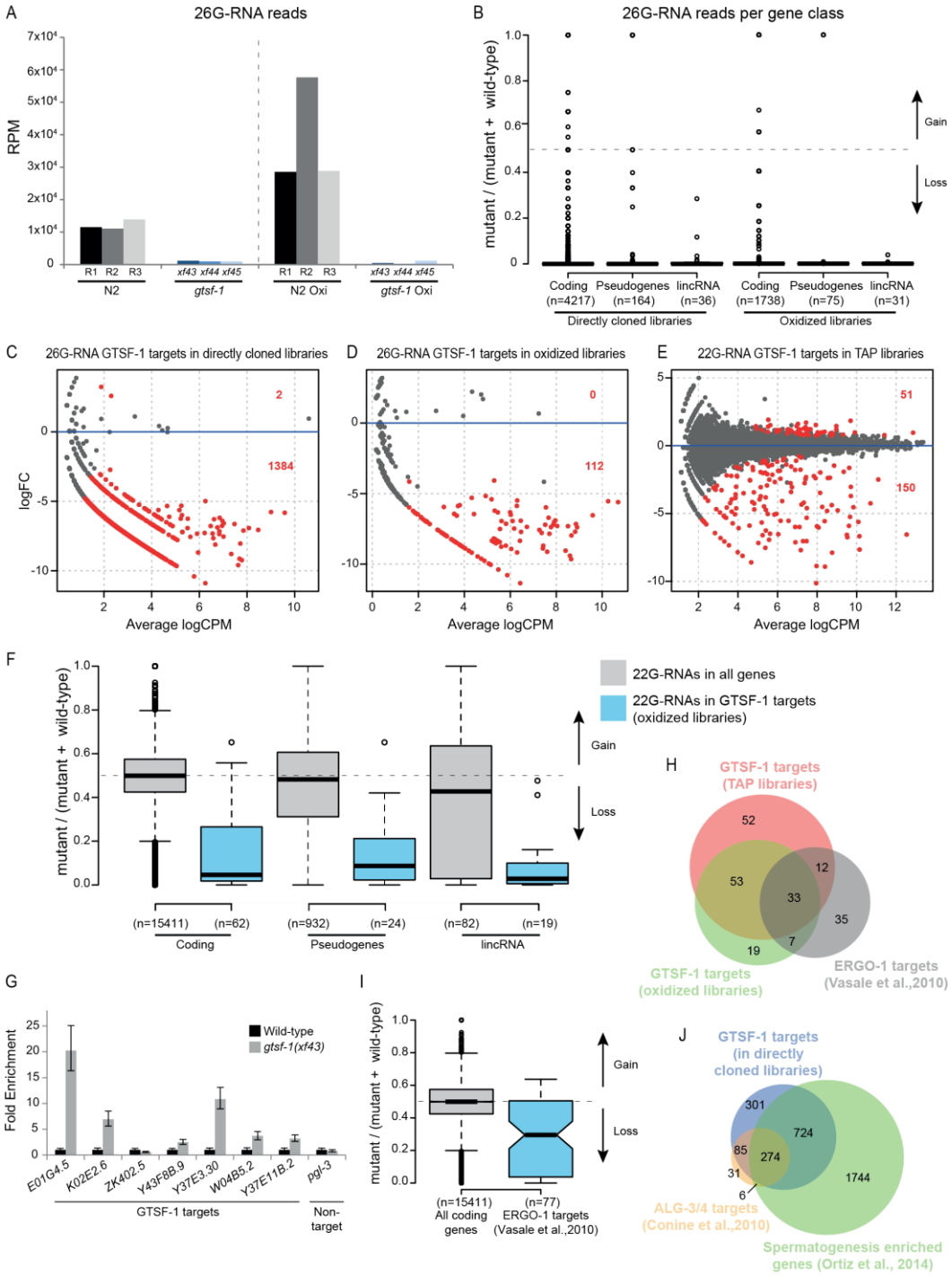


Figure II.4. 26G-RNAs are severely depleted in *gtsf-1* mutants. (A) Global levels of 26G-RNAs in wild-type and *gtsf-1* mutant worms, in RPM (Reads Per Million). Three biological replicates are shown, represented as R1-R3 for wild-type N2 worms. The dashed line separates the levels of 26G-RNAs in different library treatments: directly cloned libraries on the left, and oxidized libraries on the right. (B) Boxplot showing enrichment/depletion of normalized 26G-RNA reads per gene in *gtsf-1* mutants relative to wild-type, separated by gene class. All the genes in each class that had 26G-RNA mapped reads were used for this analysis. (C-E) Identification of GTSF-1 target genes that are significantly depleted of 26G- or 22G-RNA reads in the mutants in comparison to wild-type. Separate MA-plots are shown for the different library treatments. Statistically significant changes (1% FDR) are highlighted in red and their number is indicated. LogFC, Log2 Fold Change. LogCPM, log2 Counts Per Million. (F) Boxplot showing enrichment/depletion of 22G-RNA reads in *gtsf-1* mutant in TAP libraries, by gene class, using all genes with mapped 22G-RNAs (grey boxes), and only 22G-RNAs that map to GTSF-1 targets (blue boxes), as defined in the oxidized libraries (D). (G) RT-qPCR of seven GTSF-1 targets and a non-target (*pgl-3*). Error bars represent the standard deviation of two biological replicates. *pmp-3* was used as the normalizing gene. (H) Venn diagram showing overlap of targets of the indicated libraries with previously defined ERGO-1 targets (Vasale et al., 2010). (I) Boxplot indicating enrichment/depletion of 22G-RNA levels (from the TAP-treated libraries) in all coding genes (grey box), and in ERGO-1 targets as defined by others. We used only 77/87 ERGO-1 RIP targets from Vasale et al, 2010, since for the remaining 10 targets, we did not have mapped reads. Notches represent the 95% confidence interval for each median. (J) Venn diagram showing overlap of targets of the indicated libraries with previously defined ALG-3/4 targets (Conine et al., 2010) and with genes enriched in the spermatogenic gonad (Ortiz et al., 2014).

consistent with a role for GTSF-1 upstream of NRDE-3, we observed a significant overlap between GTSF-1-dependent sRNA targets and NRDE-3 targets (Zhou et al., 2014) (**Figure II.S4B**).

The 1384 targets defined by the directly cloned libraries (**Figure II.S4A**), extensively overlapped with ALG-3/4 targets as defined by sRNA sequencing of *alg-3/4* double mutants (Conine et al., 2010) (**Figure II.4J**). Consistent with this, functional analysis for these 1384 GTSF-1 targets shows enrichment for sperm proteins, kinases and phosphatases (**Table II.S1**). As expected for ALG-3/4 targets, these GTSF-1-dependent *loci* extensively overlapped with spermatogenesis-specific genes as defined by others (Ortiz et al., 2014) (**Figure II.4J** and **Table II.S1**).

To illustrate loss of sRNAs in *gtsf-1* mutants, exemplary genome tracks of GTSF-1 targets are shown in **Figure II.S4C-E**. Also, WormExp gene set enrichment analysis on GTSF-1 targets retrieved ERGO-1, ALG-3/4 and RRF-3 datasets, amongst many other datasets related to factors belonging to 26G- and 22G-RNA pathways (**Table II.S1**). Altogether, we conclude that both ALG-3/4-associated and ERGO-1-associated 26G-RNA populations, as well as the 22G-RNAs downstream of ERGO-1, are severely impacted by the loss of GTSF-1.

GTSF-1 interacts with RRF-3

To identify interactors of GTSF-1 we performed Immunoprecipitation (IP) followed by label-free quantitative proteomics. IPs were performed in quadruplicates, in wild-type and *gtsf-1* mutant synchronized gravid adults using an anti-GTSF-1 antibody. Additionally, using an anti-FLAG antibody, we immunoprecipitated FLAG-tagged GTSF-1 from a strain carrying a rescuing transgene (the same used in **Figures II.1E-F** and **II.3A,C**), using wild-type animals as a negative control. In both IP-mass spectrometry experiments, RRF-3 was the most enriched interactor (**Figure II.5A-B**). Notably, in the transgene pull-downs (potentially an overexpression setup, because of the use of the *gld-1* promoter) we also observed slight enrichment of other known cofactors of RRF-3 in the 26G-RNA-producing ERI complex (**Figure II.5B**, represented by black dots). These IP experiments were also performed under more stringent wash conditions (600 mM NaCl), in which case only the RRF-3 interaction was maintained (**Figure II.S5A-B**). We note that previous interactomics studies on Eri factors recovered GTSF-1 peptides, albeit with very low peptide coverage and without experiments addressing functionality (Duchaine et al., 2006; Thivierge et al., 2012). These observations support our results that GTSF-1 is associated with RRF-3 in the context of the ERI complex.

To further characterize this interaction, we produced a single-copy transgene of 3xFLAG-tagged RRF-3. This transgene rescues the Eri phenotype and the fertility defects associated with loss of RRF-3 (**Figure II.S5C-D**), indicating it recapitulates wild-type RRF-3 function. We then used this transgene to validate the GSTF-1-RRF-3 interaction via Co-IP followed by Western Blot (**Figure**

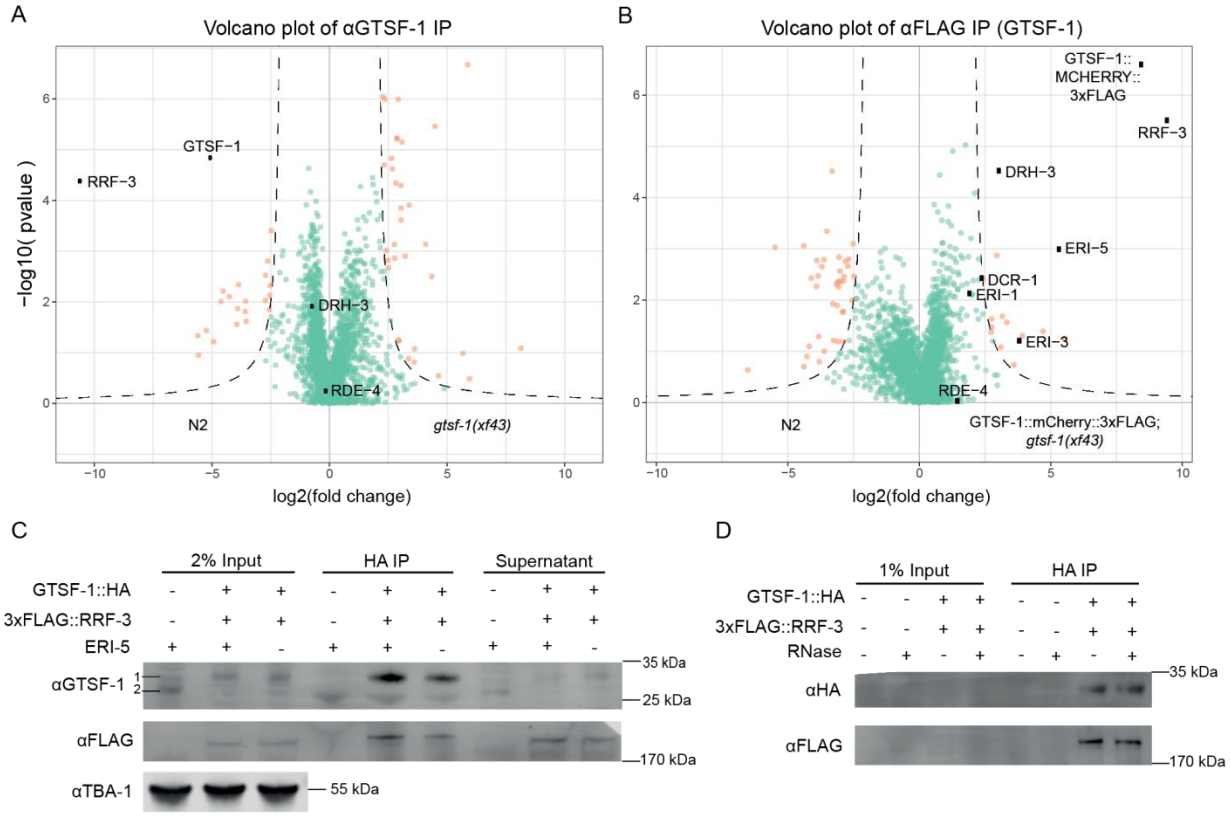


Figure II.5. GTSF-1 interacts with RRF-3 in the adult germline, independently of RNA. (A-B) Volcano plots representing label-free proteomic quantification of GTSF-1 IPs from adult worm extracts. For each strain, IPs were performed and measured in quadruplicates. Log_2 fold enrichment of individual proteins in one strain vs another is given on the x-axis. The y-axis indicates the Log_{10}-transformed probability of the observed enrichments. Proteins in the background are represented as green dots while orange dots show enriched proteins. In (A) GTSF-1 was immunoprecipitated using our polyclonal anti-GTSF-1 antibody (in wild-type and *gtsf-1* mutant worms), while in (B) an anti-FLAG antibody was used to pull-down GTSF-1::mCherry::3xFLAG (in wild-type and strains carrying the rescuing transgene). (C) To test interaction between GTSF-1 and RRF-3 in adult worms by Western blot, GTSF-1::HA was pulled-down via HA immunoprecipitation. Interaction was also tested in the presence/absence of ERI-5 by introducing an *eri-5(tm2528)* mutation in the background. Multi-channel secondary antibody detection was performed with an Odyssey CLx apparatus (see Supplementary Experimental Procedures). For the anti-GTSF-1, 1 represents GTSF-1::HA and 2 represents untagged GTSF-1. (D) Testing RNA-dependency on the interaction between GTSF-1 and RRF-3 by adding RNase. Extracts from adult worms were used. Secondary antibody detection was performed with the Odyssey CLx setup.

II.5C). This interaction is not abrogated by RNase A treatment, indicating it is RNA-independent (**Figure II.5D**).

These data clearly demonstrate that GTSF-1 interacts robustly with the RdRP enzyme RRF-3 and not with an AGO protein like its fly and mouse orthologs.

The CHHC zinc fingers of GTSF-1 mediate the interaction with RRF-3

Next, we aimed to pinpoint the determinants of the GTSF-1/RRF-3 interaction. For this, we cloned and expressed GST-fused constructs with different GTSF-1 fragments (**Figure II.6A-B**). Subsequently, we incubated these GST-fusions with embryonic extracts of a 3xFLAG::RRF-3; *gtsf-1(xf43)*; *rrf-3(pk1426)* strain and pulled-down GST. Full length (FL) GTSF-1 pulls down 3xFLAG::RRF-3 (**Figure II.6B**), corroborating the results described above. The GST fusions to the

individual zinc-fingers and the C-terminal tail did not pull-down 3xFLAG::RRF-3 over background. Interestingly, when both CHHC zinc fingers are fused to GST, 3xFLAG::RRF-3 can be efficiently retrieved (**Figure II.6B**). None of the fusion proteins interacted with DCR-1 above background. We also created GST-GTSF-1 full length proteins with mutated zinc finger residues. Specifically, we mutated the cysteines of the zinc fingers to alanines (see **Figure I.10A**). Notably, when we mutate the cysteines of individual zinc fingers, the interaction with 3xFLAG::RRF-3 is slightly disturbed (**Figure II.6C**, see Znf1- and Znf2-), but when all the four cysteines from both zinc fingers are simultaneously mutated, the interaction with 3xFLAG::RRF-3 is abrogated (**Figure II.6C**, see Znf12-). These results demonstrate that the zinc fingers of GTSF-1 are responsible for RRF-3 binding and suggest that both zinc fingers may act as a unit to mediate RRF-3 binding.

To address the *in vivo* relevance of the GTSF-1/RRF-3 interaction, we produced single-copy transgenes expressing GTSF-1 with mCherry and 3xFLAG tags, in which the CHHC cysteines in GTSF-1 were mutated to alanines (henceforth indicated as *gtsf-1*[*Znf12*-], see **Figure I.10A**). Two independent *gtsf-1(znf12-)* transgene insertions do not rescue the Eri phenotype nor the fertility defects associated with GTSF-1, thereby phenocopying *gtsf-1* mutants (**Figures II.6D** and **Figure II.S5D**). The lack of rescue is not due to poor expression of the (Znf12-) transgenes in the germline, although some degradation is observed (**Figure II.S5E**). Such partial degradation might be triggered by the disruption of the structural role that the zinc fingers have in GTSF-1. Moreover, subcellular localization of GTSF-1 is not affected by the zinc finger mutations (**Figure II.S5F**). FLAG pull-down followed by quantitative proteomics revealed that the GTSF-1(Znf12-) protein does not stably interact with RRF-3 (**Figure II.6E**).

In the literature, several examples can be found of zinc fingers mediating both protein-protein and protein-nucleic acid interactions (Gamsjaeger et al., 2007). To address if GTSF-1 is interacting with RNA, we performed *in vitro* iCLIP (Sutandy et al., 2018). We sought for a holistic approach, so we used *C. elegans* total RNA (rRNA-depleted) to test the binding of GTSF-1. Surprisingly, GTSF-1 was found not to crosslink with RNA above background levels (**Figure II.S5G**).

We conclude that GTSF-1 interacts with RRF-3 via its two tandem CHHC zinc fingers *in vitro* and *in vivo*. Since the GTSF-1/RRF-3 interaction is stable in presence of RNase (**Figure II.5D**), and GTSF-1 does not seem to interact with RNA (**Figure II.S5G**), this suggests that the two CHHC zinc fingers in GTSF-1 act strictly as a protein-protein interaction domain.

GTSF-1 is both in a precursor complex that is required for ERI complex assembly, and in the mature ERI complex

Previous studies on the ERI complex mostly focused on embryos (Duchaine et al., 2006; Thivierge et al., 2012). Next, we used embryonic extracts to probe the effect of GTSF-1 on the ERI

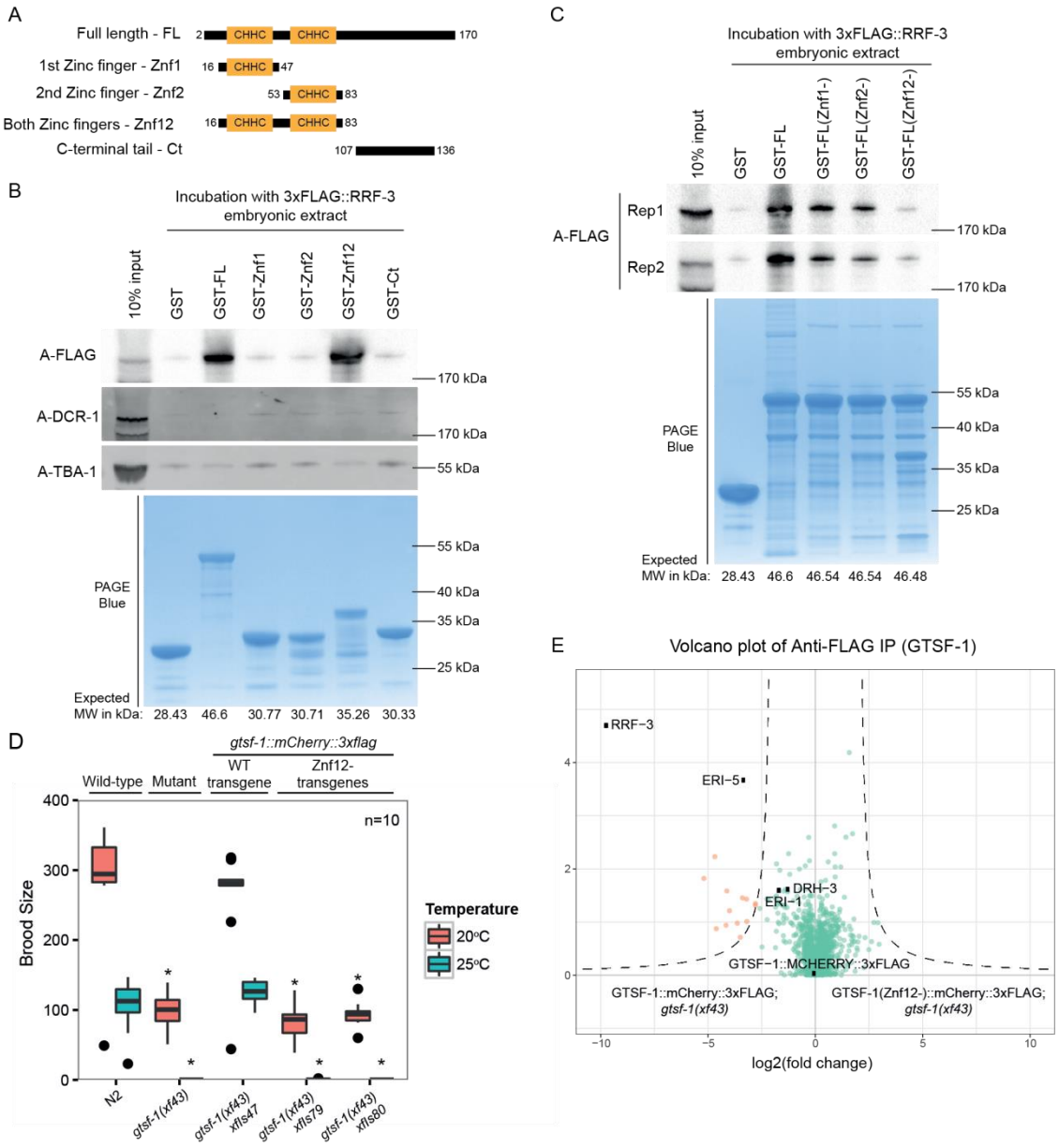


Figure II.6. The tandem CHHC zinc fingers of GTSF-1 mediate the interaction with RRF-3. (A) Schematic representation of GST-fused GTSF-1 constructs produced for this study. Amino acid residues bordering the cloned regions are indicated in the figure. (B-C) Western blot analysis of GST-GTSF-1 pull-downs. (B) 5µg of GST-GTSF-1 fusion protein (conjugated with Sepharose GSH beads, see lower PAGE Blue panel) were each incubated with approximately 1 mg of total embryonic protein extract from a 3xFLAG::RRF-3; *rrf-3(pk1426)*; *gtsf-1(xf43)* strain. 3xFLAG::RRF-3 was detected using ECL while DCR-1 and TBA-1 were detected using the Odyssey CLx apparatus. (C) 5µg of various GST-GTSF-1 fusion proteins (conjugated with Sepharose GSH beads, see lower PAGE Blue panel), with the indicated cysteine to alanine mutations in the zinc fingers (1st Zinc finger mutated – Znf1-; 2nd zinc finger mutated – Znf2-; and both zinc fingers mutated - Znf12-) were each incubated with approximately 0.5 mg of total embryonic protein extract from a 3xFLAG::RRF-3; *rrf-3(pk1426)*; *gtsf-1(xf43)* strain. 3xFLAG::RRF-3 pull-down is shown for two independent biological replicates. (D) Brood size assay at 20 °C and 25 °C. The progenies of worms of the indicated genotype are plotted. n=10 for every strain. Asterisks indicate p-value<0.0037 as assessed by Mann-Whitney Wilcoxon tests comparing wild-type worms with the other strains. (E) Volcano plots showing label-free protein quantification of GTSF-1::mCherry::3xFLAG pull-downs. Pull-downs were performed in quadruplicate with adult worm extract. Wild-type GTSF-1 fusion proteins are compared with GTSF-1 fusion proteins with zinc finger mutations (Znf12-). Proteins in the background are represented as green dots while orange dots show enriched proteins.

complex. As in the adult germline, HA-tagged GTSF-1 pulls down 3xFLAG-tagged RRF-3 in embryos, as visualized by Western blot (**Figure II.7A**). To circumvent potential overexpression (brought about by the transgene *gld-1* promoter), and to probe GTSF-1 interactions more broadly, we immunoprecipitated endogenous GTSF-1 and analyzed the precipitate with label-free quantitative mass spectrometry (**Figure II.7B**). In this experiment, we observed a strong enrichment for RRF-3 and ERI-5, while all other known components of the ERI complex are either only mildly enriched, or not enriched at all, contrasting with the previously published molecular niche of RRF-3 in embryos (Duchaine et al., 2006; Thivierge et al., 2012).

In order to test whether we can detect the ERI complex in our experimental setup, we performed IP-mass spectrometry on 3xFLAG::RRF-3 from embryo extracts. This experiment clearly identified all known ERI complex components (**Figure II.7C**, black dots). In addition to the known ERI complex components, we also found RDE-8 to strongly co-IP with RRF-3 under these conditions. RDE-8 and ERI-9, another previously identified ERI complex factor (Thivierge et al., 2012), are paralog endonucleases that have been implicated in the 26G-RNA pathway (Gent et al., 2010; Tsai et al., 2015).

Given that the 3xFLAG::RRF-3 IP results in the identification of the ERI complex in its entirety (**Figure II.7C**), while the GTSF-1 IP retrieves only RRF-3 and ERI-5, we hypothesized that GTSF-1 binds non-ERI complex-bound RRF-3. Is this non-ERI complex-bound pool of RRF-3 perhaps a precursor complex that is required for ERI complex formation? To test this, we performed a 3xFLAG::RRF-3 IP in a *gtsf-1* mutant background, and again detected RRF-3 interactors through label-free quantitative mass spectrometry. Strikingly, in absence of GTSF-1, ERI complex components no longer co-IP with RRF-3 (**Figure II.7D**), with the sole exception of ERI-5. We then tested whether ERI-5 is required for interaction between GTSF-1 and RRF-3 and found that GTSF-1::HA can still pull-down 3xFLAG::RRF-3 in *eri-5* mutants (**Figures II.5C** and **II.7A**). Interestingly, we noticed that in the absence of ERI-5, both GTSF-1 and RRF-3 are partially destabilized in embryonic extracts (**Figure II.S6A**), while 3xFLAG::RRF-3 is not destabilized in the absence of GTSF-1 (**Figure II.S6B**). These results suggest that 1) GTSF-1 is required to form mature ERI complex from a RRF-3-ERI-5 precursor complex, where ERI-5 stabilizes RRF-3; 2) that GTSF-1 does not require ERI-5 to bind to RRF-3; and 3) ERI-5 does not require GTSF-1 to bind RRF-3.

To further test the idea that GTSF-1 is required to incorporate RRF-3 into ERIC, we performed size-exclusion chromatography with 3xFLAG::RRF-3-containing embryonic extracts, followed by Western blot for GTSF-1 and FLAG. In wild-type embryos 3xFLAG::RRF-3 displays a bimodal elution pattern. The main pool elutes in a broad range between 1-4 MDa, while a smaller fraction elutes at roughly 300-400kDa (**Figure II.7E, G**). In absence of GTSF-1, 3xFLAG::RRF-3 displays a single peak at roughly 250 kDa (**Figure II.7F-G**), consistent with RRF-3 bound to ERI-5 (61.6 kDa and 18.6 kDa are the predicted molecular weights for ERI-5 isoform A and B, respectively). These data support the

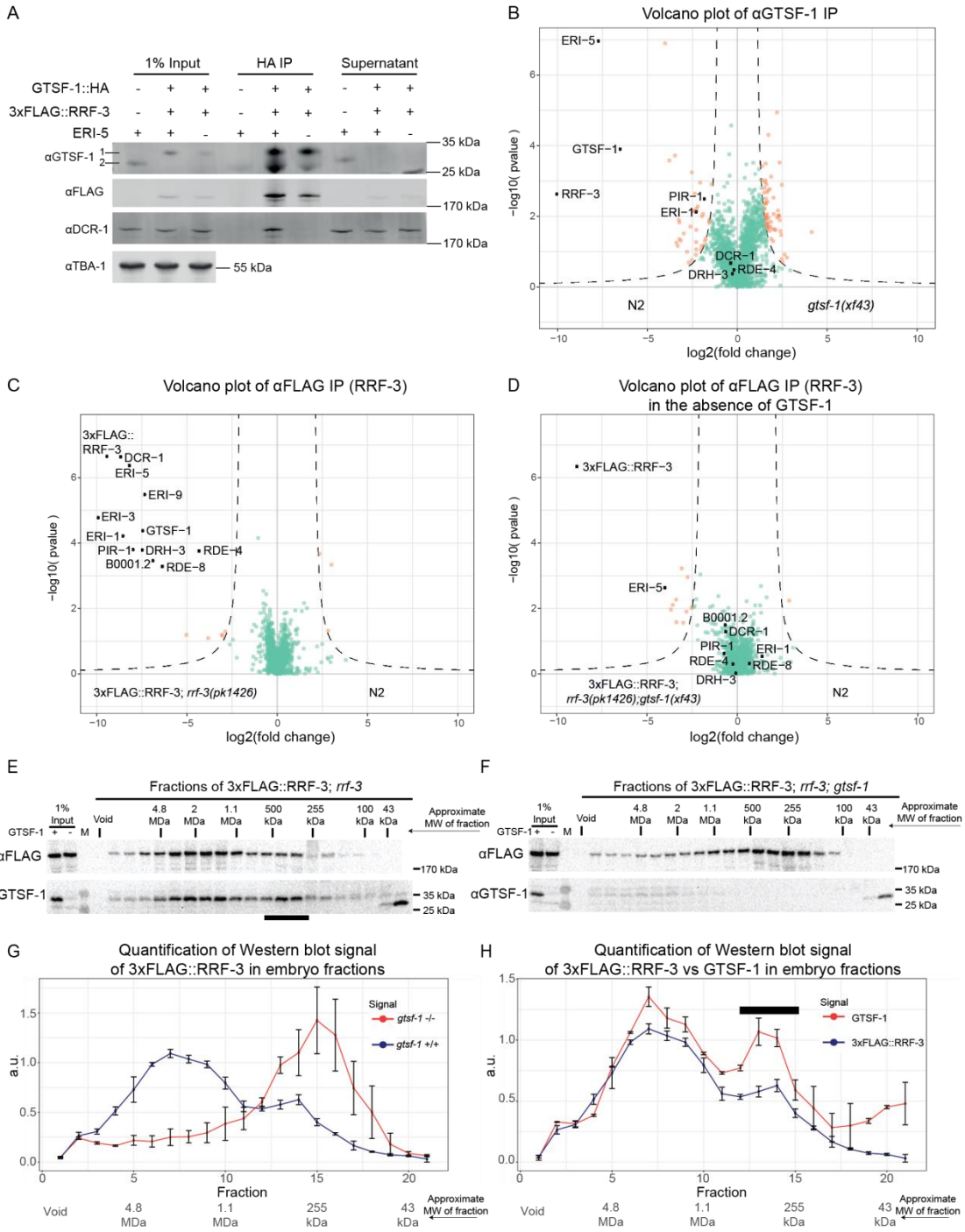


Figure II.7. GTSF-1 is required for ERI complex assembly. (A) Probing the interaction between GTSF-1 and RRF-3 by Western blot analysis, in embryonic extracts. GTSF-1::HA was pulled-down via HA immunoprecipitation. Interaction was also tested in the presence/absence of ERI-5 by introducing an *eri-5(tm2528)* mutation in the background. Multi-channel secondary antibody detection was performed with an Odyssey CLx apparatus (see Supplementary Experimental Procedures). (B) Label-free quantification of GTSF-1 IPs in embryos (comparing wild-type and *gtsf-1* mutant worms). IPs were done in quadruplicates, and a polyclonal anti-GTSF-1 antibody was used. (C-D) Volcano plots depicting quantitative proteomic analysis of RRF-3 pull-downs in the presence (C) and absence (D) of GTSF-1, in embryos. IPs were performed in quadruplicates. Proteins in the background are represented as green dots while orange dots show enriched proteins. (E-F) Size exclusion chromatography of 3xFLAG::RRF-3-containing embryo extracts. Fractions were collected and probed for GTSF-1 and 3xFLAG::RRF-3. Approximate Molecular Weight (MW) of the fractions is indicated. The calculation of these values according to protein standards is shown in the Supplementary Information. Fractions collected from extracts with GTSF-1 are shown in (E), and without GTSF-1 are shown in (F). (G-H) Comparison of size exclusion chromatography profiles of 3xFLAG::RRF-3 and GTSF-1. Relative quantification was performed with the Western blot signal using ImageJ. Error bars represent standard deviation of two biological replicates. A.u., arbitrary units. (G) Comparison of profiles of 3xFLAG::RRF-3 in the presence of GTSF-1 (blue line) and in the absence of GTSF-1 (red line). (H) Comparison of profiles of 3xFLAG::RRF-3 (blue line) and GTSF-1 (red line).

hypothesis that GTSF-1 is required to incorporate an RRF-3/ERI-5 pre-complex into ERIC, via an RRF-3/ERI-5/GTSF-1 intermediate.

GTSF-1 and RRF-3 show very similar elution patterns (**Figure II.7E, H**). This indicates that GTSF-1 remains within ERIC, at least for some time after its assembly. Results that we obtained by size-exclusion chromatography on young adult extracts are consistent with the embryo data: in young adults we also find that 3xFLAG::RRF-3 and GTSF-1 display bimodal elution profiles (**Figure II.S6C-F**), with GTSF-1 again being essential to form the ERI complex (**Figure II.S6D-E**). Strikingly, both 3xFLAG::RRF-3 and GTSF-1 show a more pronounced pre-ERI complex peak when compared to embryos (compare **Figure II.S6C, F** with **Figure II.7E, H**), suggesting ERI complex assembly may be less active in the germline. Finally, both in embryos as well as in adults the ratio of pre-ERI complex:ERI complex is consistently higher for GTSF-1 than for RRF-3 (**Figures II.7H** and **II.S6F**). This may indicate that GTSF-1 can dissociate from mature ERI complex to form novel pre-ERI complexes.

Taken together, these data show that GTSF-1 alternates between two states: one associated with the mature ERI complex and another associated with an RRF-3 and ERI-5-containing pre-ERI complex. Also, and most importantly, this pre-complex is required to form a functional ERI complex, competent for driving 26G-RNA biogenesis.

Discussion

Here, we show that GTSF-1 does not participate in TE silencing via the piRNA pathway in *C. elegans*, unlike GTSF-1 orthologs in flies and mice. However, like its orthologs, GTSF-1 is required for normal fertility. Surprisingly, GTSF-1 promotes 26G-RNA biogenesis by incorporating the 26G-RNA generating enzyme RRF-3 into a larger complex known as ERI complex. GTSF-1 thus provides an enticing example of a conserved protein that achieves its function in sRNA pathways via different cofactors in different species, i.e. AGO proteins versus RdRP enzymes. Nevertheless, we propose that the function ascribed to *C. elegans* GTSF-1, of enabling the assembly of larger protein complexes from smaller subunits, may be evolutionarily conserved.

The double CHHC zinc finger as a protein-protein interaction module

Typically, zinc fingers are known to mediate interactions with nucleic acid. Nevertheless, several cases were described in which zinc fingers mediate protein-protein interactions (Gamsjaeger et al., 2007). In some of these cases, zinc fingers of one protein interact directly with the zinc fingers of another protein (*e.g.* like GATA-1 and FOG) (Gamsjaeger et al., 2007).

We found that GTSF-1 interacts with RRF-3 via its tandem CHHC zinc fingers *in vitro* and *in vivo* (**Figure II.6**). Interestingly, the zinc fingers individually could not interact with RRF-3 (**Figure II.6B**). This

suggests the two zinc fingers may function as one structural unit. Mutation of the cysteines of individual zinc fingers reduced but did not completely eliminate the interaction with RRF-3 (**Figure II.6C**, see GST-GTSF-1 znf1- and znf2-). This could point at a certain structural robustness that allows one mutated zinc finger to fold relatively well when adjacent to a wild-type zinc finger. Of note, GTSF-1(znf12-) transgenes could not rescue *gtsf-1* mutant defects (**Figures II.6D** and **II.S5D**), clearly showing that interaction with RRF-3 via its zinc fingers is key for GTSF-1 function *in vivo*.

These results differ from Piwi-Gtsf1 interaction data from *Drosophila* and mouse, in that the C-terminal tail (also referred to as "central region") of Gtsf1 was shown to interact with Piwi, and MIWI2 and MILI, respectively (Dönertas et al., 2013; Yoshimura et al., 2018) Also, Dmel-Gtsf-1 zinc finger mutants were still found to interact with Piwi in cell culture (Ohtani et al., 2013). We note, however, that 1) the zinc fingers of Dmel-Gtsf1 were not tested directly for interaction with Piwi, 2) the four cysteines of both zinc fingers were not simultaneously mutated, unlike our setup (**Figure II.6**), and 3) consistent with our observations, zinc finger mutations are required for Dmel-Gtsf1 function, as assessed by TE derepression (Dönertas et al., 2013; Ohtani et al., 2013). Of note, the function of Dmel-Gtsf1 paralogs has not yet been determined. It may be that these paralog CHHC zinc finger proteins may interact with other proteins via their zinc fingers, and thus have a more similar role to Cel-GTSF-1 in sRNA biology.

For GTSF1L and GTSF2, GTSF1 paralogs in mouse, interaction with Piwi proteins and piRNA pathway cofactors was shown to be complex (Takemoto et al., 2016). For GTSF1L, the double CHHC zinc fingers were shown, by *in vitro* GST pull-downs, to mediate interaction with MIWI and TDRD1. Interaction with MILI seems to be mediated by the "central region" encompassing the conserved acidic residues (**Figure I.10A**). Conversely, GTSF2 interacts with MILI and TDRD1 via its CHHC zinc fingers, while it interacts with MIWI via its "central region". So, like Cel-GTSF-1, the CHHC zinc fingers of these Gtsf1 paralogs also mediate protein-protein interactions, although the relevance of these interactions has not been demonstrated *in vivo*.

It seems that the CHHC zinc fingers present in GTSF proteins are not necessarily interacting with RNA, as was assumed after RNA-interaction was determined for the single CHHC zinc finger of U11-48K proteins (Andreeva and Tidow, 2008; Tidow et al., 2009). Interestingly, GTSF proteins are the only CHHC-containing protein family that has CHHC zinc fingers in tandem. It may be that this particular feature brought about structural possibilities that facilitate specific protein-protein interactions. We hypothesize that the tandem CHHC zinc fingers of GTSF1 homologs may generally function as one structural unit, with different structural characteristics than the individual U11-48K type CHHC zinc finger.

A parallel between GTSF-1 in animals and Stc1 in fission yeast

In *S. pombe*, Stc1 is a protein that is required for sRNA-mediated centromeric heterochromatin formation (Bayne et al., 2010). More concretely, Stc1 bridges the Ago1 RNA-induced transcriptional silencing complex to the Clr4 methyltransferase complex. Although not phylogenetically related to GTSF-1 homologs, Stc1 has astonishingly similar structural features. It has an N-terminal LIM domain (which consists of two tandem zinc fingers) and a very acidic, unstructured, C-terminal domain, much like GTSF-1 (**Figure I.10A**). Structure-function studies indicated that the tandem zinc fingers of Stc1 mediate a direct interaction with Ago1 while its C-terminal tail interacts with Clr4 (He et al., 2013). These modular protein-protein interactions nicely illustrate the bridging functions of Stc1.

As discussed above, it is possible that GTSF1 proteins, including Cel-GTSF-1, may possess multiple interaction surfaces with which they may be able to bring different complexes into close contact. In a similar fashion to Stc1, Cel- GTSF-1 may bridge RRF-3 and the rest of the ERI complex. This would imply that the C-terminal tail of GTSF-1 would interact with another ERI complex factor. We performed mass spectrometry of GST pull-downs of fusion constructs containing the C-terminal tail of GTSF-1. However, these experiments did not enrich for any ERI complex factor, nor for any other plausible candidates (M.V. Almeida and S. Dietz, unpublished observations). It may be that this interaction is too transient to be detected in our experiments. The *in vitro* interaction studies of GTSF1 proteins in mouse, described above, would also lend support to such a bridging function of GTSF1 in animals, i.e. reciprocally bridging MILI and MIWI complexes undergoing the ping-pong cycle.

Also in flies, GTSF1 might function to couple Piwi to downstream effector proteins such as Panoramix (Sienski et al., 2015; Yu et al., 2015). Possibly, this would need to be tested in specific developmental stages, since Piwi activity in flies was proposed to be primarily active in embryos (Akkouche et al., 2017).

GTSF1 homologs and Stc1 are not the sole examples of tandem zinc finger proteins with roles in sRNA pathways. A family of LIM-domain containing proteins in mammals was implicated in miRNA-mediated gene silencing (James et al., 2010). These LIM-domain proteins, LIMD1, Ajuba and WTIP, were found to bridge Ago1/2 with other factors, like eIF4E, in the molecular surroundings of the 5’ Cap structure. This mode of action will ultimately lead to translation inhibition of Ago1/2 targets (James et al., 2010). A more recent study has determined that the LIM domains of LIMD1 are the interaction surface with TNRC6A (Bridge et al., 2017). Moreover, LIMD1 bridges AGO2 to TNRC6A/miRISC (Bridge et al., 2017).

Altogether, it seems likely that small proteins with these structural modules, tandem zinc fingers and unstructured C-terminal domains, have convergently evolved as versatile bridges between different protein complexes with roles in sRNA pathways.

How is the ERI complex recruited to target RNA?

It is still unknown how ERI complex is brought to, or assembled on target mRNA. How are the targets defined in the first place, and which ERI complex component binds the RNA? To answer these questions, efforts should be made to identify the RNA-binding protein(s) involved in the recruitment of the mRNA. This could provide nice insights into the interplay between pre-ERI complex (GTSF-1/ERI-5/RRF-3), ERI complex and target mRNA.

We cannot fully exclude that the zinc fingers of GTSF-1, either together as a unit or individually, interact to some extent with RNA. However, the interaction with RRF-3 is not dependent on RNA (**Figure II.5D**), and *in vitro* crosslinking experiments failed to show significant GTSF-1 association with RNA above background (**Figure II.S5G**). Hence, we believe GTSF-1 is unlikely to be responsible for RNA interaction during ERI complex assembly.

Our FLAG::RRF-3 pull-downs in embryos faithfully retrieved all known ERI complex factors identified previously in other proteomics studies (Duchaine et al., 2006; Thivierge et al., 2012). Interestingly, we also retrieved one new RRF-3-interacting factor, RDE-8. This factor is a paralog of ERI-9 (Gent et al., 2010; Pavelec et al., 2009; Tsai et al., 2015), which was previously shown to interact with other ERI complex factors (Thivierge et al., 2012). RDE-8 and ERI-9 are NYN ribonucleases, and have been previously shown to be involved in RNAi processes, including 26G-RNA biogenesis (Gent et al., 2010; Pavelec et al., 2009; Tsai et al., 2015). Their roles in 26G-RNA biogenesis seem to be independent of nucleic acid cleavage since: 1) ERI-9 lacks the conserved catalytic residues required for nucleic acid cleavage, and 2) RDE-8 transgenes with mutated catalytic residues still accumulate 26G-RNAs. Thus, it was proposed that RDE-8 and ERI-9 may have a structural role within the ERI complex (Tsai et al., 2015). Alternatively, an attractive hypothesis is that RDE-8 and/or ERI-9 may be responsible for target mRNA recognition, or would play a role in stabilizing the ERI complex on its target RNA.

What is the exact molecular function of GTSF-1?

We propose a model in which GTSF-1 and ERI-5 independently associate with RRF-3 to form a pre-ERI complex (**Figure II.8**). This pre-ERI complex is required to build a functional ERI complex that drives 26G-RNA biogenesis. This process seems to be developmentally regulated, since in the young adult germline there seems to be proportionally more GTSF-1/RRF-3 complex than in embryos (compare **Figure II.7E,H** with **Figure II.S6C,F**). This means that this pre-complex may be “packaged” in the young adult germline to promptly initiate 26G-RNA biogenesis during embryonic development. Also, within the pre-ERI complex, GTSF-1 and ERI-5 seem to have diverging roles. Both ERI-5 and GTSF-1 are required for building the ERI complex, but while ERI-5 seems to be required for the stability of GTSF-1 and RRF-3, GTSF-1 does not seem to be required for the stability of RRF-3. We would like to point out that it is unclear why we do not observe all ERI complex components in GTSF-1 IP-MS in

embryos (**Figure II.7B**), given that GTSF-1 does co-fractionate with the mature ERI complex (**Figure II.7E, H**). There may be several reasons for this. For example, it may be that the epitope of GTSF-1 is inaccessible within the ERI complex in embryos, or that GTSF-1 more easily dissociates from the mature ERI complex than from the pre-ERI complex. We do observe some enrichment of PIR-1 and ERI-1 in GTSF-1 IP-MS in embryos, suggesting the latter scenario may indeed apply.

Then, how does GTSF-1 exactly achieve its role? We consider a number of possibilities that are not mutually exclusive. First, GTSF-1 may be influencing the subcellular localization of RRF-3. Second, GTSF-1 may be chaperoning RRF-3 in a way that prompts conformational changes allowing RRF-3 to interact with other proteins. Third, GTSF-1 may allow RRF-3 to interact with target mRNA, which in turn may trigger ERI complex assembly. In order to address these issues and fully understand how RRF-3 works, we will need to develop biochemical assays for ERI complex assembly and function with purified components. Such a system would shed light on the questions stated above and other unresolved mechanistic details, for example whether GTSF-1 remains in the mature complex and how specific target mRNAs are selected.

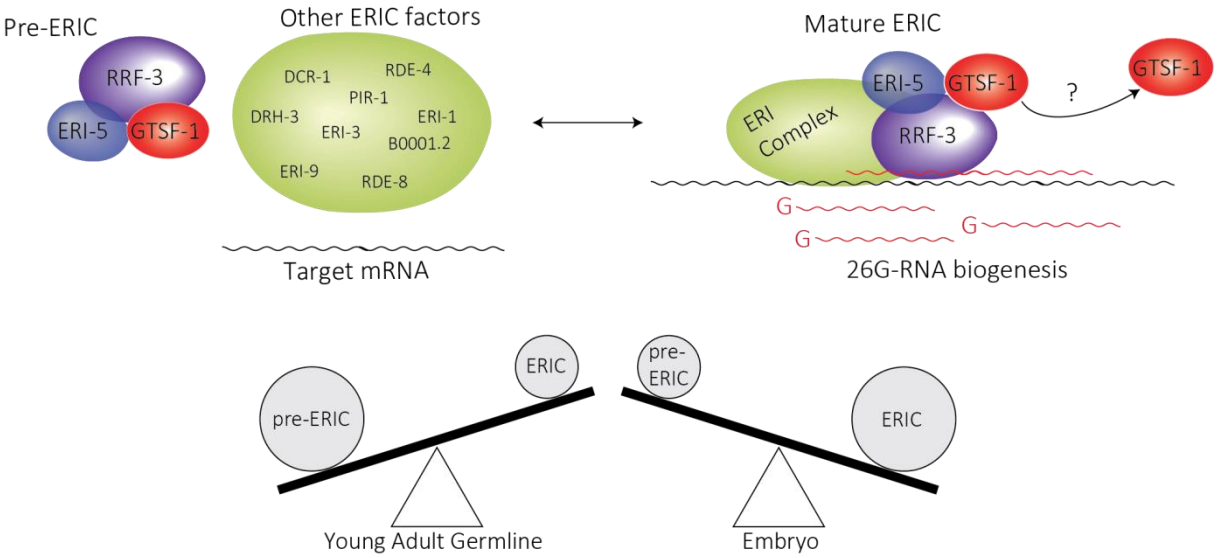


Figure II.8. A model for the function of GTSF-1. GTSF-1 forms a pre-ERI complex together with RRF-3 and ERI-5. GTSF-1 and ERI-5 are both required to incorporate RRF-3 into the ERI complex. Upon deposition of RRF-3 in ERIC, GTSF-1 may dissociate. It is also unclear how stable is the association between GTSF-1 and the mature complex. Also, this process seems to be developmentally regulated. See Discussion for details.

Materials and Methods

C. elegans genetics and culture

C. elegans was cultured on OP50 bacteria according to standard laboratory conditions (Brenner, 1974). Unless otherwise noted, worms were grown at 20 °C. The Bristol strain N2 was used as the standard wild-type strain. The strains used and created in this study are listed in **Table II.S2**, in the **Supplementary Information**.

Creation of *gtsf-1* mutants using CRISPR-Cas9 technology

gtsf-1 mutant alleles were produced as described (Friedland et al., 2013). We successfully targeted the following sequencing on the third exon of *gtsf-1*: (GGAGCCGCTGGAGCTGAACG). Two other targeted sequences, cloned into p46169 in an identical fashion, did not yield any mutants, either alone or in combination: (GATAACATGCCCTTACAATT and GACGTCGGAAATCGAGAAAT).

N2 worms were injected with 150 ng/µl of Cas9 construct p46168 (a gift from John Calarco, Friedland *et al*., 2013), 135 ng/µl of sgRNA construct pRK1134 and 15 ng/µl of co-injection marker pCFJ104 (*Pmyo-3:mCherry:unc-54 3'UTR*, expresses mCherry in body wall muscle). F1 worms positive for mCherry expression in body wall muscle were isolated, allowed to self and then lysed in single worm lysis buffer (5 mM KCl, 2.5 mM $MgCl_2$, 10mM Tris HCl pH=8.3, 0.45% NP40, 0.45% Tween20 and 0.01% gelatin). Subsequently, genotyping was performed with *Taq* Polymerase according to manufacturer's instructions (New England BioLabs, M0273X). After isolation, *gtsf-1* mutant worms were outcrossed five times.

Small RNA library preparation, sequencing and bioinformatics analysis

Detailed procedure for RNA isolation, small RNA enrichment, library preparation, bioinformatics analysis and sequencing statistics can be found in **Table II.S3**, in the **Supplementary Information**.

Antibodies

Custom, affinity-purified rabbit anti-GTSF-1 antibodies were ordered from SDIX. The following protein sequence, comprising the last 91 amino acid residues of GTSF-1 (positions 79-169), was used as an antigen: (KRQSADLRRQLSLEPLELNVAEHLAAQKLRKEYEKDEESLDGSDDSDEDEEEKNLSVTSEIEKSDVEEVEMMLETINR LAYLEMKNDNLIL). The antibody (animal number Q5963) was used in a 1:500 dilution on Western blots and 2 µg were used on Immunoprecipitations. 2 µg of Anti-FLAG antibody (M2 clone, Sigma-Aldrich, F3165) were used on immunoprecipitations and a 1:5000 dilution was used for Western blot. DCR-1 antibody was a kind gift from Thomas Duchaine and it was used in Western blots in dilutions ranging from 1-3000 to 1-5000. More information on this antibody can be found elsewhere (Duchaine et al., 2006; Thivierge et al., 2012). A commercially available, mouse anti-tubulin monoclonal antibody (clone B-5-1-2, Sigma-Aldrich, T6074) was used in Western blots in a 1:10000 dilution to detect *C. elegans* TBA-1 as a loading control. A commercially available, rabbit anti-actin polyclonal antibody (Sigma-Aldrich, A5060) was used in Western blots in a 1:1000 dilution. 30 µL of suspension of EZview™ Red Anti-HA (mouse monoclonal antibody, clone HA-7) Affinity Gel (Sigma, E6779) were used for HA IPs. A

mouse monoclonal Anti-HA antibody (clone HA-7, Sigma, H3663) was used in Western Blots with dilutions ranging from 1-500 to 1-1000.

Mass Spectrometry

Details on worm sample collection and immunoprecipitation can be found in detail in the **Supplementary Experimental Procedures**. Immunoprecipitates were resuspended in NuPAGE LDS Sample Buffer 1X (Life technologies, NP0007) and 0.1 M DTT and heated at 70°C for 10 minutes. The respective samples were separated on a 4%-12% gradient Bis-Tris gel (NuPAGE Bis-Tris gels, 1.0 mm, 10 well, NP0321; Life Technologies) in 1x MOPS (NuPAGE 20x MOPS SDS running buffer, NP0001; Life Technologies) at 180 V for 10 min, afterwards separately processed by in-gel digest (Kappei et al., 2013; Shevchenko et al., 2007) and desalted using a C18 StageTip (Rappsilber et al., 2007).

The digested peptides were separated on a 25-cm reverse-phase capillary (75 µM inner diameter) packed with Reprosil C18 material (Dr. Maisch). Elution of the peptides was done along a 2h gradient from 2%-40% Buffer B (see Stage tip purification) with the EASY-nLC 1000 system (Thermo Scientific). Measurement was done on a Q Exactive Plus mass spectrometer (Thermo Scientific) operated with a Top10 data-dependent MS/MS acquisition method per full scan (Bluhm et al., 2016). The measurements were processed with the Max Quant software, version 1.5.2.8 (Cox and Mann, 2008) against the Uniprot *C. elegans* database (version of May, 2016) for quantitation. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD007665.

Data Availability

Sequencing data have been deposited to the NCBI Gene Expression Omnibus (GEO) and proteomics data are available at the ProteomeXchange Consortium via PRIDE. GEO: GSE103432, PRIDE: PXD007665. Source data is available at:

https://data.mendeley.com/datasets/xyxh6hv3mc/draft?a=03bd1ca8-dd09-44a2-a0a1-3910ad5201c9.

Table II.S1 can be found online at:

http://emboj.embopress.org/content/early/2018/05/16/embj.201899325.

Acknowledgements

We thank all the members of the Ketting laboratory and Hans-Peter Wollscheid for great help and discussion. We thank Yasmin el Sherif, Svenja Hellmann, and Isabel Pötzch for technical assistance. We acknowledge Chung-Ting Han and Hanna Lukas of the IMB Genomics core facility, the IMB

Microscopy core facility, and the IMB Media Lab. We are grateful to Thomas Duchaine and Ahilya Sawh for kindly providing the DCR-1 antibody and the eri-5(tm2528) strain. The authors thank Colin Conine and Craig C. Mello for providing the GFP::ALG-3 strain. Also, the *Caenorhabditis* Genetics Center (CGC), which is funded by NIH Office of Research Infrastructure Programs (P40OD010440), is acknowledged for providing worm strains. This work was supported by a Deutsche Forschungsgemeinschaft grant KE1888/1-1 (Project Funding Programme to R.F.K.) and a ERC-AdG 323179 (to H.D.U.).

References

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., Rodrigues, R.J.C., Bergeijk, P. van, 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.

Andreeva, A., and Tidow, H. (2008). A novel CHHC Zn-finger domain found in spliceosomal proteins and tRNA modifying enzymes. Bioinformatics *24*, 2277–2280.

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.

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.

Bayne, E.H., White, S.A., Kagansky, A., Bijos, D.A., Sanchez-Pulido, L., Hoe, K.-L., Kim, D.-U., Park, H.-O., Ponting, C.P., Rappsilber, J., et al. (2010). Stc1: A Critical Link between RNAi and Chromatin Modification Required for Heterochromatin Integrity. Cell *140*, 666–677.

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., Fischer, S.E.J., and Kim, J.K. (2014). Endogenous RNAi pathways in C. elegans. WormBook 1–49.

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

Boeck, M.E., Huynh, C., Gevirtzman, L., Thompson, O.A., Wang, G., Kasper, D.M., Reinke, V., Hillier, L.W., and Waterston, R.H. (2016). The time-resolved transcriptome of C. elegans. Genome Res. *26*, 1441–1450.

Brenner, S. (1974). The Genetics of Caenorhabditis Elegans. Genetics *77*, 71–94.

Bridge, K.S., Shah, K.M., Li, Y., Foxler, D.E., Wong, S.C.K., Miller, D.C., Davidson, K.M., Foster, J.G., Rose, R., Hodgkinson, M.R., et al. (2017). Argonaute Utilization for miRNA Silencing Is Determined by Phosphorylation-Dependent Recruitment of LIM-Domain-Containing Proteins. Cell Rep. *20*, 173–187.

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.

Burkhart, K.B., Guang, S., Buckley, B.A., Wong, L., Bochner, A.F., and Kennedy, S. (2011). A Pre-mRNA–Associating Factor Links Endogenous siRNAs to Chromatin Regulation. PLOS Genet. *7*, e1002249.

Burton, N.O., Burkhart, K.B., and Kennedy, S. (2011). Nuclear RNAi maintains heritable gene silencing in Caenorhabditis elegans. Proc. Natl. Acad. Sci. *108*, 19683–19688.

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

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.

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

de Albuquerque, B.F.M., Placentino, M., and Ketting, R.F. (2015). Maternal piRNAs Are Essential for Germline Development following De Novo Establishment of Endo-siRNAs in Caenorhabditis elegans. Dev. Cell *34*, 448–456.

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.

Duchaine, T.F., Wohlschlegel, J.A., Kennedy, S., Bei, Y., Conte Jr., 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.

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*, 741–3.

Frøkjær-Jensen, C., Wayne Davis, M., Hopkins, C.E., Newman, B.J., Thummel, J.M., Olesen, S.-P., Grunnet, M., and Jorgensen, E.M. (2008). Single-copy insertion of transgenes in Caenorhabditis elegans. Nat. Genet. *40*, 1375–1383.

Gamsjaeger, R., Liew, C.K., Loughlin, F.E., Crossley, M., and Mackay, J.P. (2007). Sticky fingers: zinc-fingers as protein-recognition motifs. Trends Biochem. Sci. *32*, 63–70.

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.

Gu, W., Shirayama, M., Conte Jr., 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.

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.

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. *106*, 18674–18679.

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

He, C., Pillai, S.S., Taglini, F., Li, F., Ruan, K., Zhang, J., Wu, J., Shi, Y., and Bayne, E.H. (2013). Structural analysis of Stc1 provides insights into the coupling of RNAi and chromatin modification. Proc. Natl. Acad. Sci. *110*, E1879–E1888.

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.

James, V., Zhang, Y., Foxler, D.E., Moor, C.H. de, Kong, Y.W., Webb, T.M., Self, T.J., Feng, Y., Lagos, D., Chu, C.-Y., et al. (2010). LIM-domain proteins, LIMD1, Ajuba, and WTIP are required for microRNA-mediated gene silencing. Proc. Natl. Acad. Sci. *107*, 12499–12504.

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., 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.

Ketting, R.F. (2011). The Many Faces of RNAi. Dev. Cell *20*, 148–161.

Ketting, R.F., Haverkamp, T.H.A., van Luenen, H.G.A.M., and Plasterk, R.H.A. (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.

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 Jr., D., and Mello, C.C. (2012). C. elegans piRNAs Mediate the Genome-wide Surveillance of Germline Transcripts. Cell *150*, 78–87.

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

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Luteijn, M.J., and Ketting, R.F. (2013). PIWI-interacting RNAs: from generation to transgenerational epigenetics. Nat. Rev. Genet. *14*, 523–534.

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.

Mackereth, C.D., Madl, T., Bonnal, S., Simon, B., Zanier, K., Gasch, A., Rybin, V., Valcárcel, J., and Sattler, M. (2011). Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. Nature *475*, 408.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal *17*, 10–12.

Merritt, C., Rasoloson, D., Ko, D., and Seydoux, G. (2008). 3′ UTRs Are the Primary Regulators of Gene Expression in the C. elegans Germline. Curr. Biol. *18*, 1476–1482.

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.

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

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

Ortiz, M. a, Noble, D., Sorokin, E.P., and Kimble, J. (2014). A New Dataset of Spermatogenic vs. Oogenic Transcriptomes in the Nematode Caenorhabditis elegans. G3 Bethesda Md.

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.

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.

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

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.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

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

Shevchenko, A., Tomas, H., Havli, J., Olsen, J.V., and Mann, M. (2007). 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 Jr., D., and Mello, C.C. (2012). piRNAs Initiate an Epigenetic Memory of Nonself RNA in the C. elegans Germline. Cell *150*, 65–77.

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

Sutandy, F.X.R., Ebersberger, S., Huang, L., Busch, A., Bach, M., Kang, H.-S., Fallmann, J., Maticzka, D., Backofen, R., Stadler, P.F., et al. (2018). In vitro iCLIP-based modeling uncovers how the splicing factor U2AF2 relies on regulation by cofactors. Genome Res. *28*, 699–713.

Takemoto, N., Yoshimura, T., Miyazaki, S., Tashiro, F., and Miyazaki, J. (2016). Gtsf1l and Gtsf2 Are Specifically Expressed in Gonocytes and Spermatids but Are Not Essential for Spermatogenesis. PLOS ONE *11*, e0150390.

Thivierge, C., Makil, N., Flamand, M., Vasale, J.J., Mello, C.C., Wohlschlegel, J., Jr, D.C., 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.

Tidow, H., Andreeva, A., Rutherford, T.J., and Fersht, A.R. (2009). Solution Structure of the U11-48K CHHC Zinc-Finger Domain that Specifically Binds the 5′ Splice Site of U12-Type Introns. Structure *17*, 294–302.

Tsai, H.-Y., Chen, C.-C.G., Conte Jr., D., Moresco, J.J., Chaves, D.A., Mitani, S., Yates III, J.R., Tsai, M.-D., and Mello, C.C. (2015). A Ribonuclease Coordinates siRNA Amplification and mRNA Cleavage during RNAi. Cell *160*, 407–419.

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. *107*, 3582–3587.

Yang, W., Dierking, K., and Schulenburg, H. (2016). WormExp: a web-based application for a Caenorhabditis elegans-specific gene expression enrichment analysis. Bioinformatics *32*, 943–945.

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., Miyazaki, T., Toyoda, S., Miyazaki, S., Tashiro, F., Yamato, E., and Miyazaki, J.I. (2007). Gene expression pattern of Cue110: A member of the uncharacterized UPF0224 gene family preferentially expressed in germ cells. Gene Expr. Patterns *8*, 27–35.

Yoshimura, T., Toyoda, S., Kuramochi-Miyagawa, S., Miyazaki, T., Miyazaki, S., Tashiro, F., Yamato, E., Nakano, T., and Miyazaki, J.I. (2009). Gtsf1/Cue110, a gene encoding a protein with two copies of a CHHC Zn-finger motif, is involved in spermatogenesis and retrotransposon suppression in murine testes. Dev. Biol. *335*, 216–227.

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. e42054.

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

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.

Zhuang, J.J., and Hunter, C.P. (2011). Tissue-specificity of Caenorhabditis elegans Enhanced RNAi Mutants. Genetics genetics.111.127209.

Supplementary Information

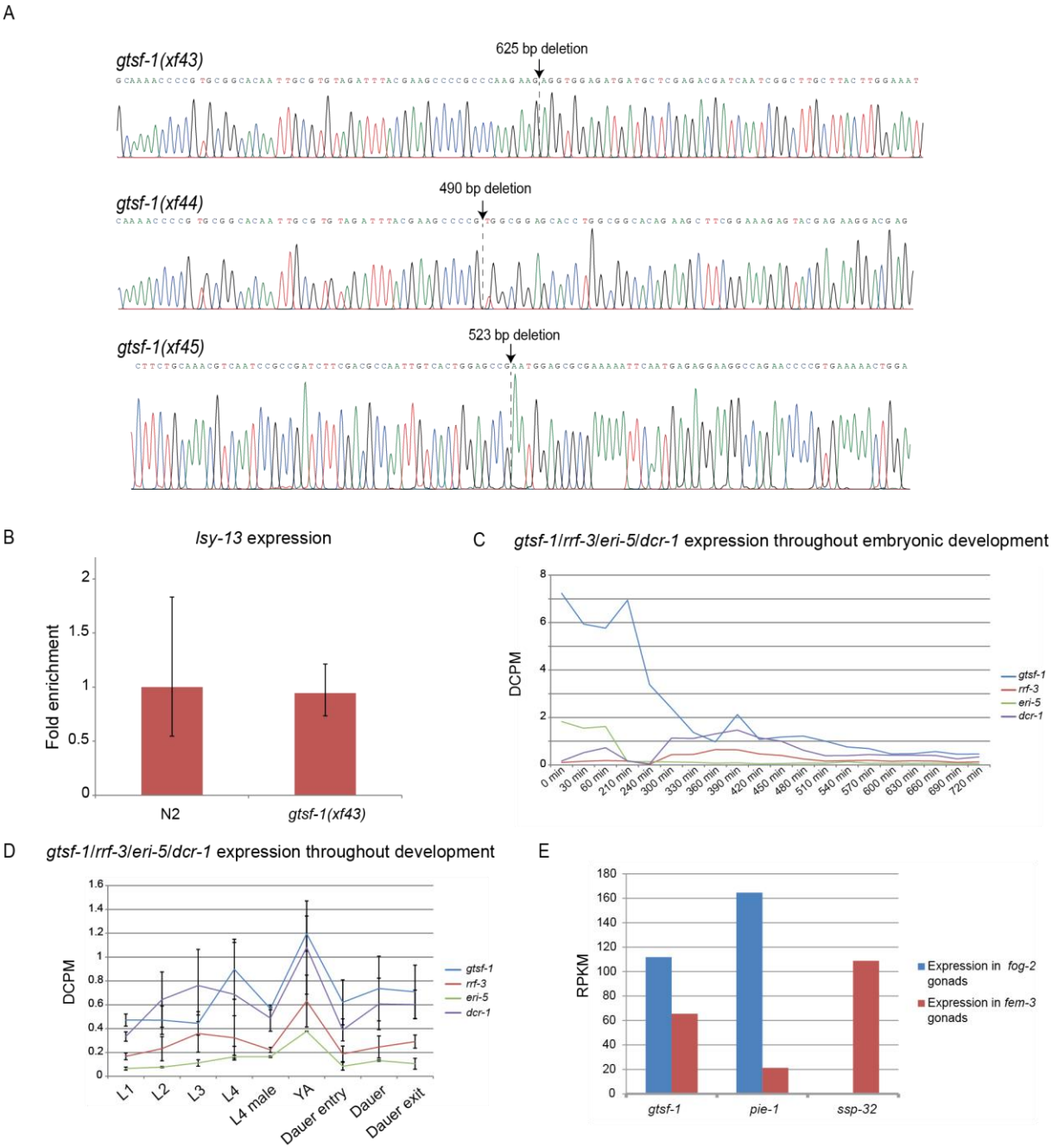


Figure II.S1. T06A10.3/Cel-GTSF-1 is a conserved factor that is expressed during gametogenesis and early development. Related to Figure II.1. (A) Chromatograms of Sanger sequencing of *gtsf-1* mutant alleles. Deletion sites are indicated with arrows. (B) RT-qPCR of *lsy-13* in wild-type and *gtsf-1* mutant embryos. Technical triplicates and biological duplicates were used for this experiment. Error bars represent the standard deviation of two biological replicates. *pmp-3* was used as the normalizing gene. (C) mRNA expression profiles of *gtsf-1, rrf-3, eri-5* and *dcr-1* during embryonic development. A publicly available, Poly-A+ RNA-seq dataset from Boeck et al, 2016 was used. On the y-axis, expression levels are shown in depth of coverage per base per million reads (DCPM). (D) *gtsf-1, rrf-3, eri-5* and *dcr-1* expression profiles throughout larval development and dauer stage, shown in DCPM. Data points represent average between two Poly-A+ RNA-seq datasets from Boeck et al, 2016. Error bars represent the standard deviation of the two replicates. (E) Depiction of expression levels of *gtsf-1*, *pie-1* (an oogenic-enriched gene), and *ssp-32* (a spermatogenic-enriched gene), in both *fog-2* mutant gonads (strictly oogenic) and *fem-3* mutant gonads (strictly spermatogenic), as reported in Ortiz et al, 2014. Expression is shown in reads per kilobase million (RPKM).

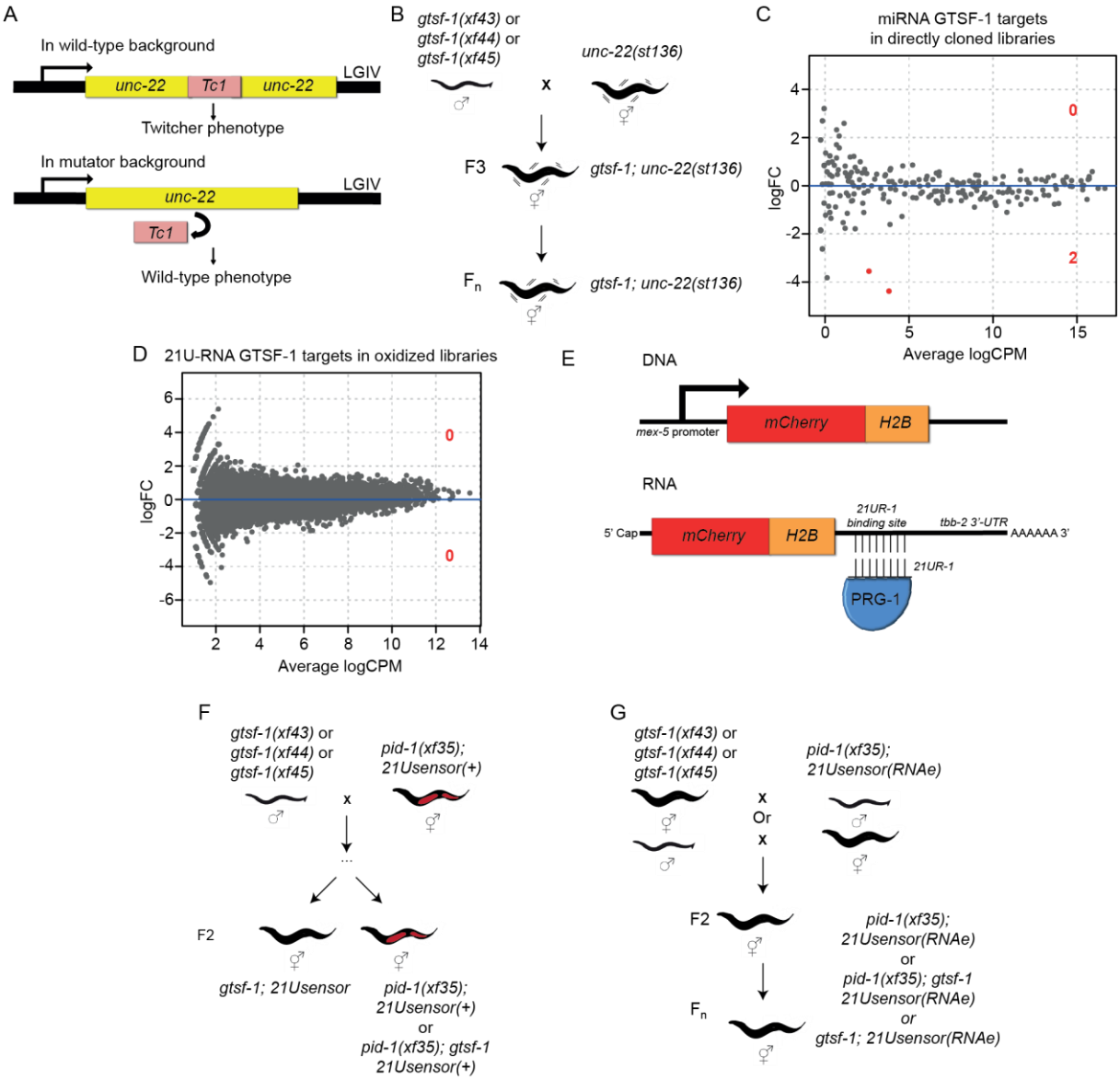


Figure II.S2. *C. elegans* GTSF-1 is not involved in TE silencing, the 21U-RNA, and miRNA pathways. Related to Figure II.2. (A) Schematic of the *unc-22(st136)* allele. (B) Layout of the *unc-22(st136)* x *gtsf-1* crosses to address transposon derepression. No phenotypic reversions to wild-type were observed in ten replicate *gtsf-1;unc-22(st136)* populations, grown in parallel for several generations. (C) Differential gene expression analysis of miRNAs in wild-type versus *gtsf-1* mutant worms, to address whether miRNAs are globally deregulated in *gtsf-1*. Analysis was performed in the directly cloned libraries, given the higher number of miRNA reads observed. Only two miRNAs (mir-260, mir-262) are significantly downregulated (1% FDR) in *gtsf-1* mutants. (D) Differential gene expression analysis of 21U-RNAs in wild-type versus *gtsf-1* mutants. Analysis was performed in the oxidized libraries where a larger number of 21U-RNA reads are found. No significant changes were found at 1% cutoff. (E) Overview of the 21Usensor. It consists of an mCherry-Histone H2B fusion transgene with a binding site for 21U-R1, an abundant 21U-RNA, in the 3'UTR of the transcript. (F) Testing the participation of *gtsf-1* in 21U-RNA-mediated silencing of *C. elegans*. To test this, a non-stably silenced 21Usensor was used and after crossing with *gtsf-1* mutant animals, the F2 of the indicated genotypes was scored for mCherry expression. (G) Schematics of the crosses between *gtsf-1* mutant alleles and an RNAe 21Usensor. No derepression of the 21Usensor was observed in *gtsf-1* mutants in the F2s and in further generations.

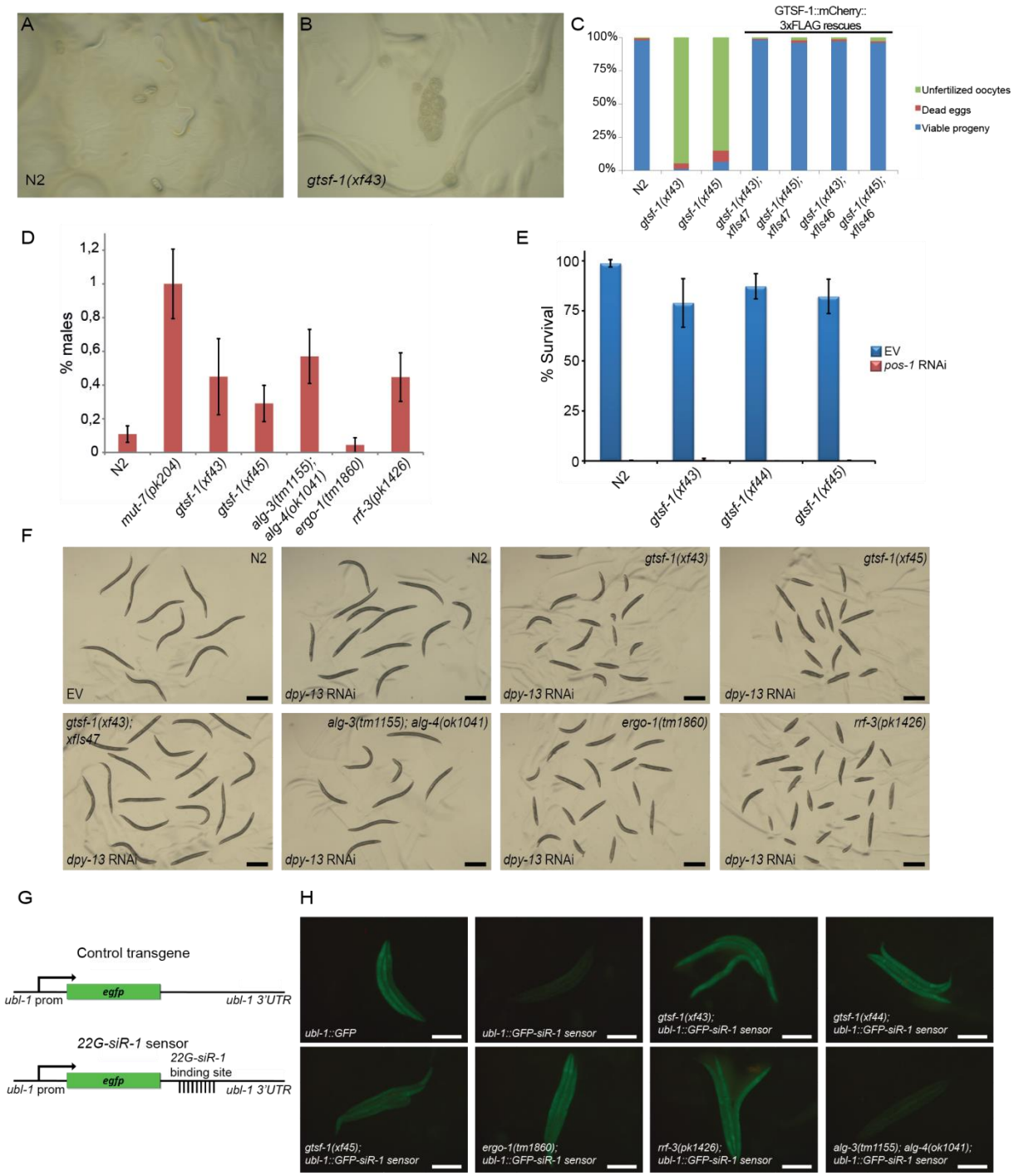


Figure II.S3. *gtsf-1* animals recapitulate the phenotypes of *alg-3/4* and *ergo-1* mutants. Related to Figure II.3. (A-B) Representative pictures of progenies from synchronized parents of wild-type (A) and *gtsf-1* mutants (B) grown at 25 ºC. In (A), L1 larvae and gastrulated embryos can be observed, while in (B) only unfertilized oocytes are identified. (C) Percentage of dead eggs and unfertilized oocytes laid by wild-type and *gtsf-1* mutant worms when grown at 25 ºC. These counts were obtained from the same experiment as in Figure II.3A. n=10. (D) % of males is plotted for the indicated strains. *gtsf-1* mutants show a percentage of males comparable to *alg-3/4* and *rrf-3* mutants, but lower than *mut-7* mutants. Spontaneous male incidence was measured in the progenies of 20 worms. For each plate, percentage of males was calculated and the percentages of the 20 plates were subsequently averaged to obtain the final male percentage value. (E) % Survival of animals of the indicated genotypes on empty vector (EV) and *pos-1* RNAi. Per strain, progenies of n=9-12 worm were scored for embryonic lethality. *gtsf-1* animals respond to *pos-1* RNAi identically to wild-type. (F) This panel shows a representative fraction of animals from the *dpy-13* RNAi experiment (shown in Figure II.3C). Genotypes and treatment with either empty vector (EV) or *dpy-13* RNAi are indicated in the figure. Scale bars represent 0,5 mm. (G) Overview of the 22G transgenes. The control transgene consists of GFP controlled by the *ubl-1* promoter and 3'UTR sequences, which allow ubiquitous expression. The 22G-siR-1 sensor is similar to the control transgene except for a binding site for 22G-siR-1 on the 3'UTR that reports on the activity of ERGO-1-dependent 22G-RNAs. (H) Genetic requirements for 22G-sensor silencing. Representative fluorescence images of animals carrying 22G-siR-1 sensor combined with other mutations. Scale bars represent 0.5 mm.

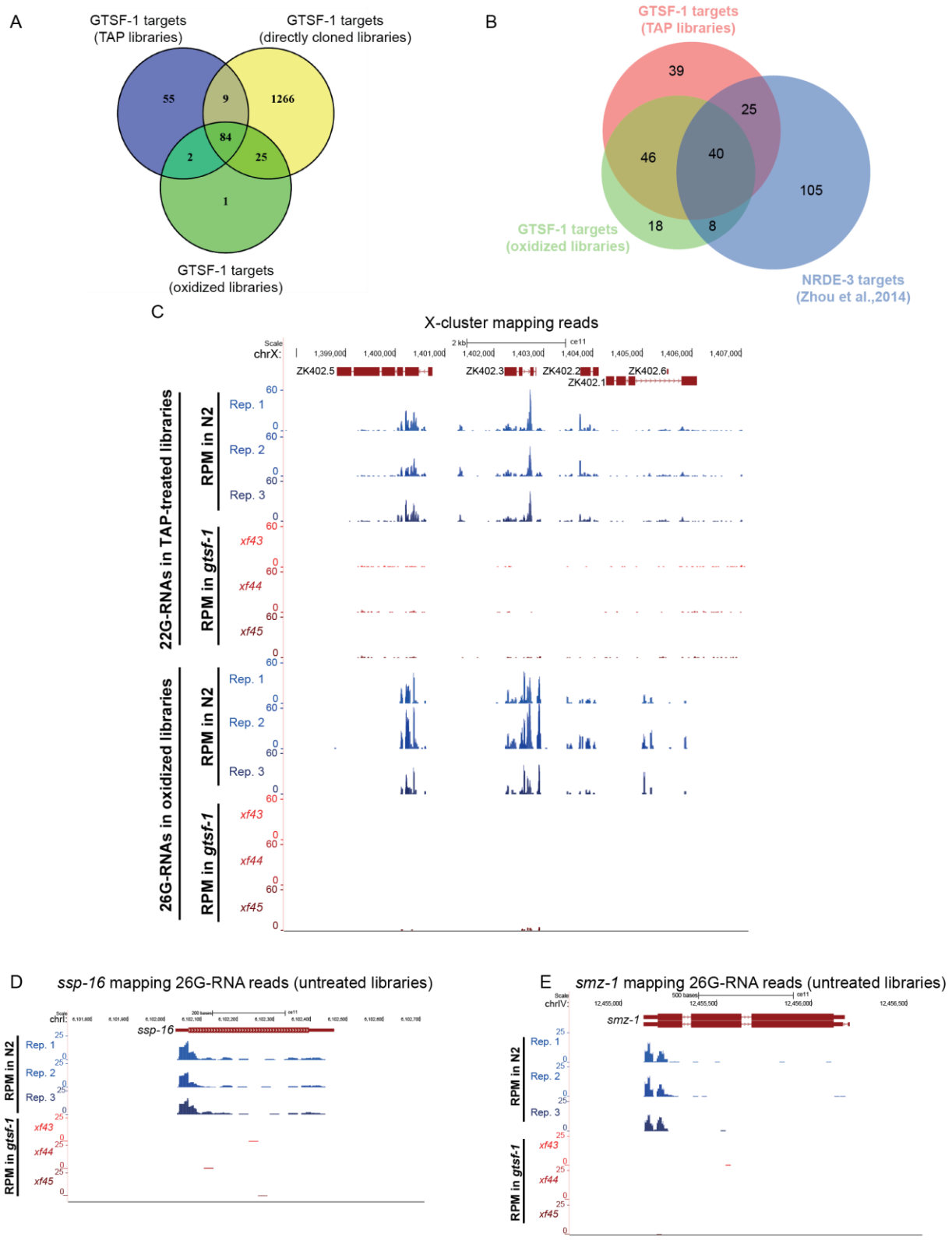


Figure II.S4. 26G-RNAs are severely depleted in *gtsf-1* mutants. Related to Figure II.4. (A) Venn diagram showing overlap between the 1% FDR GTSF-1 targets defined in the three different NGS library preparation conditions. (B) Venn diagram depicting the overlap between the GTSF-1 targets defined in the oxidized and TAP-treated libraries in this study, and NRDE-3 targets as defined elsewhere (Zhou et al., 2014). (C-E) Exemplary UCSC genome browser tracks with RPM-normalized sRNA coverage profiles. Three biological replicates of wild-type N2 worms (Rep. 1 – Rep. 3) and the three *gtsf-1* alleles are shown separately. (C) 22G-RNAs (above) and 26G-RNAs (below) mapping to the 22G-RNA X-cluster. 22G- and 26G-RNA tracks were obtained from TAP-treated and oxidized libraries, respectively. (D-E) 26G-RNA reads mapping to *ssp-16* (D), a sperm-specific family, class P protein, and to *smz-1* (E), a sperm meiosis PDZ domain-containing protein. Directly cloned libraries were used for these tracks.

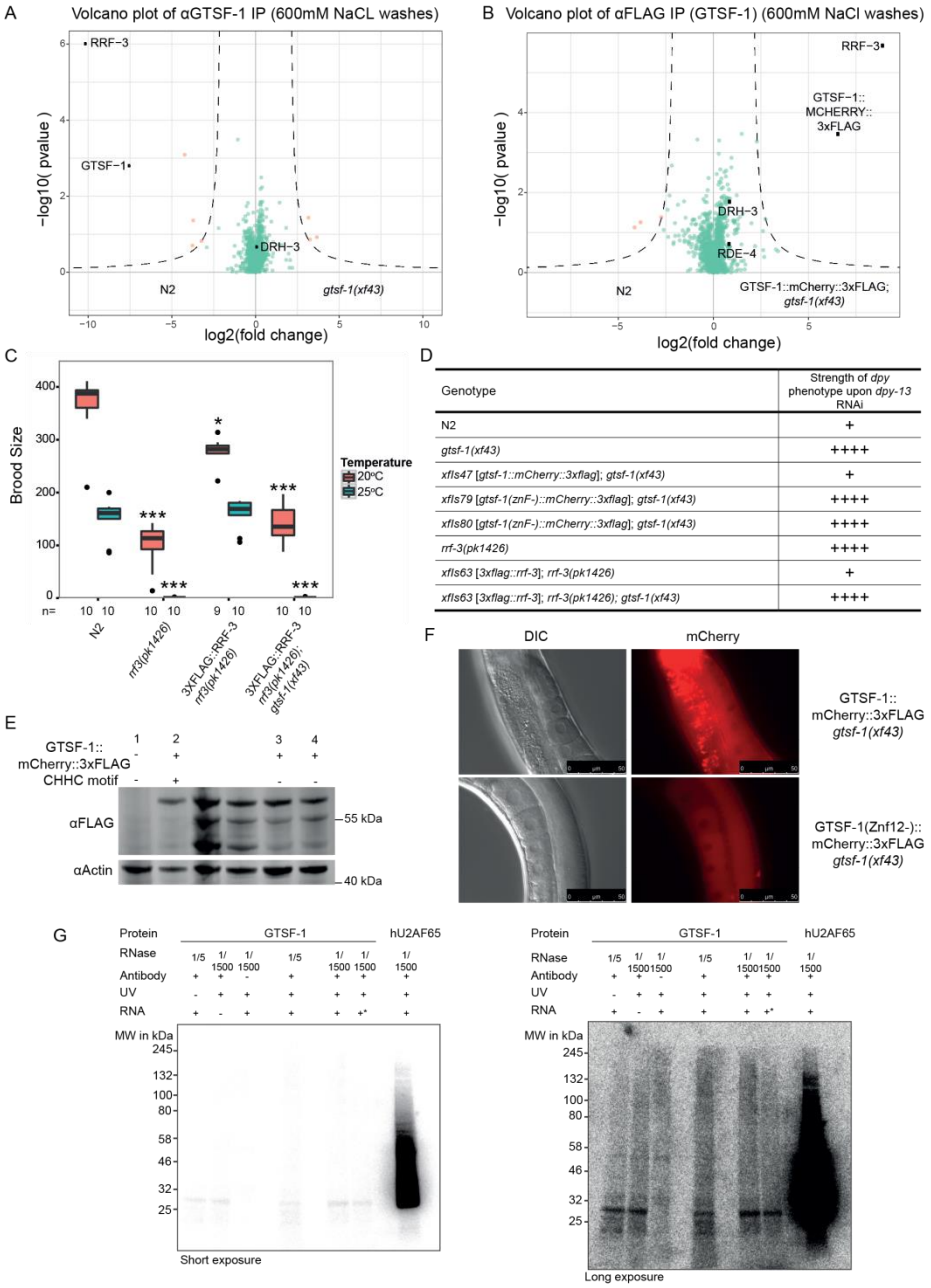


Figure II.S5. GTSF-1 stably interacts with RRF-3 via its tandem CHHC zinc fingers. Related to Figures II.5 and II.6. (A-B) Volcano plots of GTSF-1 IP-mass spectrometry experiments in more stringent wash conditions (600 mM NaCl). Gravid adult worms were used. IPs were performed and measured in quadruplicates. Other than the wash conditions, the setup was identical to Figure II.5A-B. (C) Brood size count of a 3xFLAG::RRF-3 single-copy transgene. This transgene rescues the fertility defects associated with *rrf-3* mutation almost to a complete extent. Asterisk indicates *p-value*= 0.002185; triple asterisk indicates *p-value*<0.0001817. *p-values* calculated with Mann-Whitney-Wilcoxon tests, using N2 brood size as a reference. n is indicated in the x-axis. (D) Overview of *dpy-13* RNAi experiments to worms of the indicated genotypes. A "+" sign indicates that these animals show a normal response to *dpy-13* RNAi, while "++++" animals have an enhanced response to RNAi, having both a more prevalent and stronger *dpy-13* phenotype. (E) Western blot analysis of adult animal populations carrying GTSF-1::mCherry::3xFLAG transgenes. Lane 1, Wild-type, non-transgenic worms. Lane 2, WT GTSF-1 protein. Lanes 3 and 4 represent GTSF-1 zinc finger mutants, wherein the zinc finger cysteines are mutated to alanines. The middle, non-labeled lanes are of other GTSF-1 fusion proteins not discussed in this work. (F) Wide-field DIC and fluorescence microscopy pictures of worms expressing WT GTSF-1::mCherry::3xFLAG (above) and zinc finger mutant GTSF-1::mCherry::3xFLAG (below). The overall localization of mCherry is not dependent on the zinc fingers. Scale bars indicate 50 µm. (G) *in vitro* iCLIP experiment. Purified GTSF-1 protein was incubated with *C. elegans* total RNA from wild-type, UV cross-linked and immunoprecipitated using the GTSF-1 antibody used throughout this study. After the IP, the co-purified RNA was radioactively labeled and the IPs were run on a SDS-PAGE gel followed by membrane transfer. As a positive control, human U2AF65 (hU2AF65) was used (see Sutandy et al., 2018). GTSF-1 protein does not associate with RNA beyond background levels. Also, there are not differences between the no antibody, the no cross-link and the no total RNA controls. Same gel is shown, on the left with short exposure and on the right with longer exposure. +* indicates that *gtsf-1(xf43)* total RNA was used, not wild-type. The rationale behind this was that in the *gtsf-1* mutant background, GTSF-1 targets are upregulated. This was expected to increase the iCLIP signal.

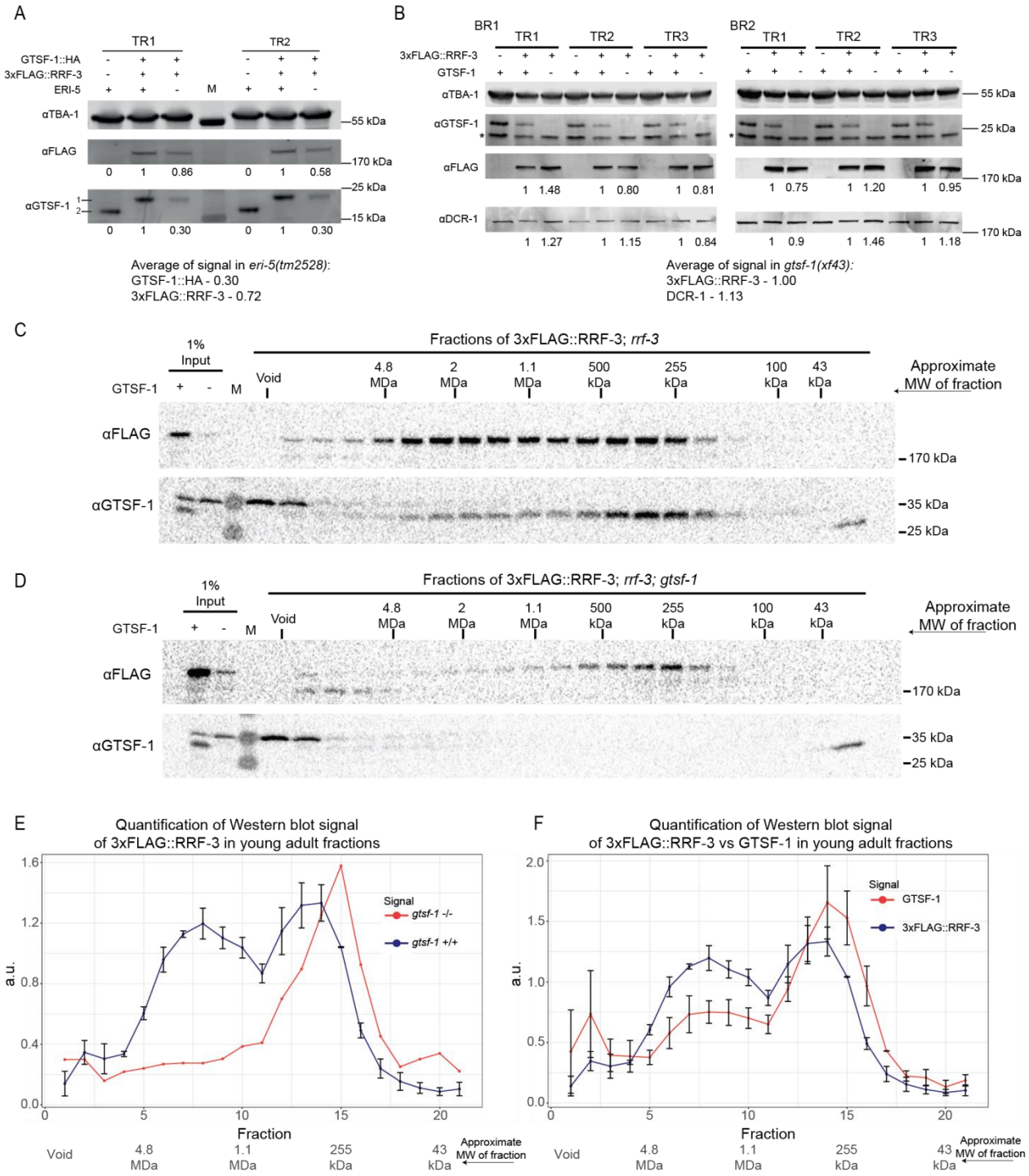


Figure II.S6. Stability of ERI complex factors in different backgrounds and pre-ERI complex/ERI complex profiles in young adults. Related to Figures II.5 and 7. (A) Western blot relative quantification of 3xFLAG::RRF-3 and GTSF-1::HA with and without ERI-5. Absence of ERI-5 (-) indicates the presence of *eri-5(tm2528)* deletion allele. Embryonic extracts were used. Double transgenic, wild-type ERI-5 strain was used as the reference. Two technical replicates are shown. For detection, secondary antibodies compatible with LI-COR Odyssey CLx were used. Quantification was done by gating the region surrounding the bands using Image Studio software (v3.1). M, Molecular weight marker; TR, technical replicate. (B) Relative quantification of Western blot as in (A), but to address the stability of 3xFLAG::RRF-3 and DCR-1 in the absence of GTSF-1 (which indicates the presence of *gtsf-1(xf43)*). Embryonic extracts were used. BR, biological replicate; TR, technical replicate. *indicates unspecific signal. (C-D) Size exclusion chromatography of 3xFLAG::RRF-3-containing young adult extracts. Fractions were collected and Western blot was performed for GTSF-1 and 3xFLAG::RRF-3. Fractions collected from extracts with GTSF-1 are shown in (C), and without GTSF-1 are shown in (D). Approximate Molecular Weight (MW) of the fractions is indicated in the figure. The calculation of these values according to protein standards is shown in the Supplementary Information. (E-F) Comparison of size exclusion chromatography profiles of 3xFLAG::RRF-3 and GTSF-1 in young adult extracts. Relative quantification was performed with the Western blot signal, using ImageJ. Error bars represent standard deviation of two biological replicates, with the exception of 3xFLAG::RRF-3 profile in the absence of GTSF-1 (E), red line. A.u., arbitrary units. (E) Comparison of profiles of 3xFLAG::RRF-3 in the presence of GTSF-1 (blue line) and in the absence of GTSF-1 (red line). (F) Comparison of profiles of 3xFLAG::RRF-3 (blue line) and GTSF-1 (red line).

Worm strains used and produced in this study

Table II.S2. Worm strains used and produced in this study.

Strain reference	Genotype	Available in CGC	Comments
	Wild-type N2	Yes	
RFK607	*gtsf-1(xf43)*	Yes	5x outcrossed
RFK608	*gtsf1(xf44)*	Yes	5x outcrossed
RFK609	*gtsf-1(xf45)*	Yes	5x outcrossed
RFK262	*xfIs46 [gld-1(prm)::gtsf-1::mCherry::3xflag::unc-54(3'UTR)] II; gtsf-1(xf43) IV*		*unc-119* genotype unknown
RFK263	*xfIs46 [gld-1(prm)::gtsf-1::mCherry::3xflag::unc-54(3'UTR)] II; gtsf-1(xf45) IV*		*unc-119* genotype unknown
RFK264	*xfIs47 [gld-1(prm)::gtsf-1::mCherry::3xflag::unc-54(3'UTR)] II; gtsf-1(xf43) IV*		
RFK265	*xfIs47 [gld-1(prm)::gtsf-1::mCherry::3xflag::unc-54(3'UTR)] II; gtsf-1(xf45) IV*		*unc-119* genotype unknown
RFK66	*unc-22(st136) IV*	Yes, as NL3643	
RFK266	*gtsf-1(xf43); unc-22 (st136) IV*		
RFK267	*gtsf-1(xf44); unc-22 (st136) IV*		
RFK268	*gtsf-1(xf45); unc-22 (st136) IV*		
RFK18	*rrf-3(pk1426) II*	Yes, as NL2099	
YY174	*ggIS1[nrde-3p::3xflag::gfp::nrde-3]*	Yes, as YY178 (equivalent)	
RB1079	*alg-4 (ok1041) III*	Yes	
WM300	*alg-3 (tm1155) IV; alg-4 (ok1041) III*	Yes	
RFK271	*alg-3 (tm1155) IV*		
WM158	*ergo-1(tm1860) V*	Yes	
GR1719	*unc-119(ed9) III; mgSi3 [ubl-1p::GFP::ubl-1(3'UTR)] IV*	Yes	
GR1720	*unc-119(ed9) III; mgSi4 [ubl-1p::GFP::siR-1-sensor-ubl-1(3'UTR)] IV*	Yes	
RFK272	*unc-119(ed9) III; mgSi4 [ubl-1p::GFP::siR-1-sensor-ubl-1(3'UTR)] IV; gtsf-1 (xf43) IV*		
RFK273	*rrf-3 (pk1426) II; mgSi4 [ubl-1p::GFP::siR-1-sensor-ubl-1(3'UTR)] IV*		
RFK274	*mgSi4 [ubl-1p::GFP::siR-1-sensor-ubl-1(3'UTR)] IV; ergo-1(tm1860) V*		

RFK275	*unc-119(ed9) III; mgSi4 [ubl-1p::GFP::siR-1-sensor-ubl-1(3'UTR)] IV; gtsf-1 (xf44) IV*		
RFK276	*mgSi4 [ubl-1p::GFP::siR-1-sensor-ubl-1(3'UTR)] IV; gtsf-1 (xf45) IV*		
RFK277	*unc-119(ed9) III; alg-3 (tm1155) IV; mgSi4 [ubl-1p::GFP::siR-1-sensor-ubl-1(3'UTR)] IV*		
RFK278	*alg-4 (ok1041) III; alg-3 (tm1155) IV; mgSi4 [ubl-1p::GFP::siR-1-sensor-ubl-1(3'UTR)] IV*		
RFK279	*alg-4 (ok1041) III; unc-119(ed9) III; mgSi4 [ubl-1p::GFP::siR-1-sensor-ubl-1(3'UTR)] IV*		
WM202	*neIs23[unc-119(+) GFP::ALG-3]* II; *unc-119(ed3)* III; *alg-4(ok1041)* III; *alg- 3(tm1155)* IV		A kind gift from Craig Mello
	xfIs47 [gld-1(prm)::gtsf-1::mCherry::3xflag::unc-54(3'UTR)] II; *neIs23[unc-119(+) GFP::ALG-3]* II; *alg-4(ok1041)* III; *gtsf-1(xf43)* IV; *alg- 3(tm1155)* IV		*unc-119* genotype unknown
	eri-5 (tm2528) IV		A kind gift from Thomas Duchaine
YY158	*nrde-3(gg66) X*	Yes	
RFK458	*ggIS1[nrde-3p::3xflag::gfp::nrde-3]; gtsf-1(xf43)* IV		
RFK463	*ggIS1[nrde-3p::3xflag::gfp::nrde-3]; rrf-3(pk1426)* II		
RFK464	*ggIS1[nrde-3p::3xflag::gfp::nrde-3]*; *alg-4 (ok1041)* III; *alg-3 (tm1155)* IV		
RFK370	*unc-119(ed3)* III; *xfIs63[gld-1(prm)::3xflag::rrf-3::unc-54 (3'UTR)]* IV		
RFK405	*rrf-3 (pk1426)* II; *unc-119(ed3)* III; *xfIs63[gld-1(prm)::3xflag::rrf-3::unc-54 (3'UTR)]* IV		
RFK421	*rrf-3 (pk1426)* II; *xfIs63[gld-1(prm)::3xflag::rrf-3::unc-54 (3'UTR)]* IV; *gtsf-1(xf43)* IV		
RFK432	*xfIs72[gld-1(prm)::gtsf-1::HA::unc-54(3'UTR)]* II; *unc-119(ed9)* III		
RFK450	*xfIs72[gld-1(prm)::gtsf-1::HA::unc-54(3'UTR)]* II; *gtsf-1(xf43)* IV; *xfIs63[gld-1(prm)::3xflag::rrf-3::unc-54 (3'UTR)]* IV		*unc-119* genotype unknown
RFK439	*xfIs79[gld-1(prm)::gtsf-1[C21A; C41A; C58A;*		

	C78A]::mCherry::3xflag::unc-54 3'UTR] II; *unc-119(ed9)* III		
RFK440	*xfIs80[gld-1(prm)::gtsf-1[C21A; C41A; C58A; C78A]::mCherry::3xflag::unc-54 3'UTR]* II; *unc-119(ed9)* III		
RFK448	*xfIs79[gld-1(prm)::gtsf-1[C21A; C41A; C58A; C78A]::mCherry::3xflag::unc-54 3'UTR]* II; *gtsf-1(xf43)* IV		*unc-119* genotype unknown
RFK449	*xfIs80[gld-1(prm)::gtsf-1[C21A; C41A; C58A; C78A]::mCherry::3xflag::unc-54 3'UTR]* II; *gtsf-1(xf43)* IV		*unc-119* genotype unknown
RFK657	*xfIs63[gld-1(prm)::3xflag::rrf-3::unc-54 (3'UTR)]* IV; *gtsf-1(xf43)* IV; *eri-5(tm2528)* IV		
RFK184	*mjSi22 [Pmex-5::mCherry::his-58::21UR-1_as::tbb-2(3'UTR)]* I; *pid-1(xf35)* II		21U sensor ON
RFK422	*mjSi22 [Pmex-5::mCherry::his-58::21UR-1_as::tbb-2(3'UTR)]* I; *pid-1(xf35)* II		21U sensor under RNAe
RFK632	*mjSi22 [Pmex-5::mCherry::his-58::21UR-1_as::tbb-2(3'UTR)]* I; *gtsf-1(xf43)* IV		Originated from 21U sensor ON crosses
RFK633	*mjSi22 [Pmex-5::mCherry::his-58::21UR-1_as::tbb-2(3'UTR)]* I; *pid-1(xf35)* II; *gtsf-1(xf44)* IV		Originated from 21U sensor ON crosses
RFK635	*mjSi22 [Pmex-5::mCherry::his-58::21UR-1_as::tbb-2(3'UTR)]* I; *pid-1(xf35)* II; *gtsf-1(xf43)* IV		Originated from 21U sensor ON crosses
RFK636	*mjSi22 [Pmex-5::mCherry::his-58::21UR-1_as::tbb-2(3'UTR)]* I; *gtsf-1(xf44)* IV		Originated from 21U sensor ON crosses
RFK637	*mjSi22 [Pmex-5::mCherry::his-58::21UR-1_as::tbb-2(3'UTR)]* I; *gtsf-1(xf43)* IV		Originated from 21U sensor RNAe crosses
RFK638	*mjSi22 [Pmex-5::mCherry::his-58::21UR-1_as::tbb-2(3'UTR)]* I; *pid-1(xf35)* II; *gtsf-1(xf43)* IV		Originated from 21U sensor RNAe crosses
RFK639	*mjSi22 [Pmex-5::mCherry::his-58::21UR-1_as::tbb-2(3'UTR)]* I; *gtsf-1(xf44)* IV		Originated from 21U sensor RNAe crosses
RFK640	*mjSi22 [Pmex-5::mCherry::his-58::21UR-1_as::tbb-2(3'UTR)]* I; *pid-1(xf35)* II; *gtsf-1(xf44)* IV		Originated from 21U sensor RNAe crosses

Supplementary Experimental Procedures

Molecular cloning and transgenics. All constructs produced in this study are available upon request.

Cloning sgRNAs. To introduce the targeting sequence in the sgRNA vector p46169 (a gift from John Calarco, Friedland *et al.*, 2013), we performed inverse PCR using always the same reverse primer and a forward primer in which 5' end the desired 20 nucleotide sgRNA sequence was inserted. Q5 High-Fidelity DNA Polymerase (New England BioLabs, M0491L) was used in the reaction according to the manufacturer's instructions. 4 µL of PCR product was re-ligated overnight in a simultaneous reaction (for a total volume of 20 µL) with 10 units T4 PNK (New England BioLabs, M0201S), 400 units of T4 DNA Ligase (New England BioLabs, M0202S) and 20 units DpnI (New England BioLabs, R0176S) to digest the original plasmid and complemented with 1 mM ATP (New England BioLabs, P0756S) and 2 µL of 10X T4 DNA Ligase Buffer. Then, 5 µL of ligation product was transformed into Subcloning Efficiency DH5α chemically competent cells (Life Technologies, 18265-017).

Production of transgenic worms with MosSCI. Transgenic worms were produced via MosSCI, as described (Frøkjær-Jensen et al., 2008). Detailed protocol, reagents and strains can be also found at www.wormbuilder.org. For GTSF-1 fusion transgenes, the *locus ttTi5605* on LGII was targeted. For RRF-3 transgenes, the *cxTi10816 locus* on LGIV was targeted. An injection mix consisting of 50 ng/µl pCFJ601, 50 ng/µl of repair template, 10 ng/µl pMA122, 10 ng/µl pGH8, 2.5 ng/µl pCFJ90 and 5 ng/µl pCFJ104 was injected in the EG6699 and EG6700 strains, which contain a Mos transposon insertion in *ttTi5605* and *cxTi10816*, respectively. After negative selection for the extrachromosomal array, WT moving worms with no detectable mCherry expression from the co-injection markers (pGH8, pCFJ90 and pCFJ104) were isolated. After egg laying, worms were lysed in 5 µL of single worm lysis buffer (5 mM KCl, 2.5 mM $MgCl_2$, 10mM Tris HCl pH=8.3, 0.45% NP40, 0.45% Tween20 and 0.01% gelatin) and genotyped 1) for the wild-type allele with primers flanking the outer portions of the recombination arms and, 2) for the insertions using one primer binding an outer portion of one recombination arm and another primer binding the insert. Genotyping reactions were done using *Taq* Polymerase (New England BioLabs, M0273X) with an extension time of 4 minutes.

RNAi Experiments. dsRNA was supplemented to worms by feeding as described (Kamath et al., 2003). For somatic RNAi, worm populations were synchronized by bleaching and, for each strain, L1 worms were pipetted into each well of two six-well plates. Phenotypes were scored on the second day of adulthood. For germline *pos-1* RNAi, worms were synchronized by bleaching, and L3-L4 stage animals were put on plates with *E. coli* expressing *pos-1* RNAi. Embryonic lethality was scored 3-4 days later, providing enough time for egg laying and hatching.

Fertility and *him* assays. *Brood size counts*. *C. elegans* strains were synchronized by bleaching. Then, late L3 worms were isolated, per strain and per temperature (both 20 °C and 25 °C). Worms were transferred to a new plate every day, until egg laying stopped. For worms grown on 20 °C, the viable number of progeny was assayed. For worms grown on 25 °C (for the experiment reported in **Figure II.3A**) viable progeny as well as dead eggs and unfertilized oocytes were assayed ~24 hours after removing the parent.

him assay. This experiment was performed in parallel with the brood size experiment abovementioned. Strains were synchronized by bleaching and late L3 worms were isolated. The progenies of 20 worms, per strain, were assayed for the prevalence of males at 20 °C. For the N2, *gtsf-1(xf43)* and *gtsf-1(xf45)* strains, the 10 progenies assayed for the brood size in **Figure II.3A**, were also assayed for the number of males. The remaining worms were transferred to new plates every second day. The total number of worms and the number of males were counted 2-3 days after the removal of the parent.

Brood size upon N2 sperm rescue. Crosses were setup as follows: 15 N2 males and 10 late L3/early L4 hermaphrodite worms reared at 20 °C and 25 °C, were mated for around 24 hours in the same temperature they were reared in. After mating, adult hermaphrodites were isolated and allowed to lay eggs. Worms were transferred to new plates every day and the number of viable progeny was counted. Only progenies that had 50% of males, indicative of cross progeny, were taken into account.

Microscopy. Wide-field fluorescence microscopy images shown in **Figures II.2E-F** and **II.S5F** were obtained using a DM6000 B and processed using Leica LAS software and ImageJ. Confocal microscopy micrographs shown in **Figures II.1E-G** and **II.3G** were acquired in a Leica TCS SP5 and processed in Adobe Photoshop. Wide-field photomicrographs of **Figure II.S3** were acquired using a Leica M165FC microscope with a Leica DFC450 C camera, and were processed using Leica LAS software and ImageJ.

Small RNA extraction. *C. elegans* strains, in triplicates (3 times N2 and 1 time each *gtsf-1* mutant allele, were synchronized by bleaching. Following approximately 63 hours of development after L1 arrest, gravid adult hermaphrodites were washed off plate with M9 buffer and lysed in 250 µL of Worm Lysis Buffer (0.2M NaCl, 0.1M Tris pH 8.5, 50 mM EDTA, 0.5% SDS) with 300 µg Proteinase K (Sigma-Aldrich, P2308) for at least 90 minutes at 65°C. Next, 3 volumes of TRIzol LS (Life Technologies, 10296-028) were added and subsequent isolation was as defined by the manufacturer's instructions. Samples were enriched for sRNAs using a mirVana Kit (Life Technologies, AM1561).

Each sRNA enriched sample was treated in three different ways: 1) directly cloned, 2) treated with Tobacco Acid Phosphatase (TAP), and 3) Oxidized. The TAP treatment and oxidation were

previously described (de Albuquerque et al., 2014; Kamminga et al., 2012). 1 µg of RNA was treated (de Albuquerque et al., 2014)with 5 U of tobacco acid phosphatase (Epicenter) at 37°C for 2 h to digest 5' tri- and di-phosphates to mono-phosphates. To oxidate sRNAs, 3 µg of sRNA enriched sample was incubated for 10 minutes at room temperature, with 200 mM NaIO4, 5X borate buffer (148 mM Borax, 148 mM Boric acid and adjust pH to 8.6) and nuclease-free water to a total of 20 µL reaction volume. Then, 2 µL of 100% glycerol were added to stop the reaction for 10 minutes at room temperature.

Purified RNA was precipitated with 100% isopropanol and Glycoblue overnight at -20°C. The pellet was washed once with 75% ethanol and dissolved in nuclease-free water. Then, RNA was size-selected between 18- to 30-nt on 15% TBE-urea gel. Purified fraction was confirmed by Bioanalyzer sRNA chip (Agilent).

Library preparation was based on the NEBNext Multiplex sRNA Library Prep Set for Illumina (New England BioLabs) with slight modifications. To counteract ligation biases, the 3' and 5' adapters contained four random bases at the 5' and 3'-end, respectively, and were chemically synthesized by BioScientific. Adapter-ligated RNA was reverse-transcribed and PCR-amplified for 14 cycles using index primers. The PCR-amplified cDNA construct was purified using AMPure XP beads (Beckman Coulter). The purified PCR reaction was checked on the Bioanalyzer using High Sensitivity DNA chip (Agilent). Size selection of the sRNA library was done on LabChip XT instrument (Perkin Elmer) using DNA 300 assay kit. Only the fraction containing 140-165 bp was pooled in equal molar ratio. The resulting 10 nM pool was denatured to 10 pmol with 5% PhiX spike-in and sequenced as single-read on HiSeq 2500 (Illumina) in rapid mode for 51 cycles (plus 7 cycles index read) using on-board cluster generation. After demultiplexing, on average 8 million passing filter reads were obtained per sample.

Bioinformatic analysis. The raw sequence reads in FastQ format were cleaned from adapter sequences and size-selected in the range 18-30 bases (plus additional 8 random bases) using cutadapt v.1.2.1 (http://cutadapt.readthedocs.org) (Martin, 2011) with parameters '-a AGATCGGAAGAGCACACGTCT -O 10 -m 26 -M 38'. Quality control of the raw and processed data was performed with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Mapping to the *Caenorhabditis elegans* genome (Ensembl WBcel235/ce11 assembly) with concomitant trimming of the 8 random bases was performed using Bowtie v.1.1.2 (http://bowtie-bio.sourceforge.net) (Langmead et al., 2009) with parameters '-p 8 -v 1 -M 1 --best --strata --tryhard --chunkmbs 512 --trim5 4 --trim3 4 –S'. The resulting SAM alignment files were converted into sorted BAM files using Samtools v.1.3.1 (http://www.htslib.org) (Li et al., 2009).

Gene annotation for WBcel235/ce11 in GTF format was downloaded from Ensembl (ftp://ftp.ensembl.org/pub/release-89/gtf/caenorhabditis_elegans/) and annotation for transposable elements (LINE, SINE, LTR, DNA and RC) in BED format was downloaded from the UCSC Table Browser

(genome.ucsc.edu/cgi-bin/hgTables), RepeatMasker rmsk track. Structural reads were considered reads mapping sense to annotated rRNAs, tRNAs, snRNAs or snoRNAs, and these were removed from further analyses using Bedtools v.2.25.0 (http://bedtools.readthedocs.io) (Quinlan and Hall, 2010). miRNA reads were defined as reads from 21 to 24 bases mapping sense to annotated miRNA genes. 21U reads were defined as 21 base reads starting with T and mapping sense to piRNA (21Ur) genes. Transposon-derived 22G-RNA reads were defined as 22 base reads starting with G and aligning antisense to annotated transposons. Gene-derived 22G- and 26G-RNA reads were defined, respectively, as 22 and 26 base reads starting with G and mapping antisense to annotated genes. Read class filtering of the BAM files was performed with Awk, Bedtools and Samtools. The read counts per gene or transposon element were summarised on the meta-feature level for each read class (21U, 22G, 26G, miRNAs), as well as for the entire pool of 18-30 base reads, using featureCounts/Subread v.1.5.1 (http://bioinf.wehi.edu.au/featureCounts/) (Liao et al., 2014) and Ensembl gene annotation or UCSC transposon annotation.

Normalized expression levels of the individual read classes were calculated as reads per million (RPM) based on all 18-30 base non-structural mapped reads in each sample. For UCSC browser visualisation (http://genome.ucsc.edu/), the filtered reads were converted into reads per million (RPM) normalized bigWig coverage tracks using Bedtools and the UCSC utility bedGraphToBigWig (http://hgdownload.soe.ucsc.edu/admin/exe/linux.x86_64). Pairwise (wildtype vs. mutant) differential expression comparisons at 1% FDR cutoff were performed with edgeR v.3.6.8 (https://bioconductor.org/packages/release/bioc/html/edgeR.html) (Robinson et al., 2010). Due to the observed global depletion of 26G-RNA reads in the mutant samples, the edgeR library sizes and normalisation factors were first computed from the entire pool of 18-30 base reads in each sample and were then used in the 26G-RNA differential analyses in order to mitigate TMM normalisation biases. Dataset overlaps were assessed and plotted with Venny (http://bioinfogp.cnb.csic.es/tools/venny/) and BioVenn (http://www.biovenn.nl/). Gene set enrichment analyses of *gtsf-1* target genes were performed with DAVID Bioinformatics Resources v.6.8 (http://david.abcc.ncifcrf.gov) and WormExp v.1.0 (http://wormexp.zoologie.uni-kiel.de/wormexp/) (Yang et al., 2016) at 1% FDR cutoff and using background lists of all genes with 22G- or 26G-RNA read coverage that were subjected to edgeR processing.

Sequencing statistics.

Table II.S3. Sequencing statistics.

Group	Sample	All sequenced reads	18 to 30 nt insert reads	Mapped reads	Non-structural reads
Untreated	N2_1	9,069,236	6,299,087	4,495,427	3,954,482
Untreated	N2_2	8,882,688	6,083,701	5,227,843	4,527,230
Untreated	N2_3	9,757,359	4,941,940	2,963,416	2,445,350
Untreated	*gtsf-1(xf43)*	10,248,545	6,003,447	4,839,343	3,583,860
Untreated	*gtsf-1(xf44)*	9,624,877	6,022,785	4,919,582	3,555,378
Untreated	*gtsf-1(xf45)*	8,654,823	6,144,366	4,989,638	3,850,148
TAP	N2_1 TAP	9,186,976	5,964,680	5,272,783	4,374,892
TAP	N2_2 TAP	9,647,881	7,244,111	4,125,462	3,451,460
TAP	N2_3 TAP	9,336,937	6,106,072	5,073,049	3,833,121
TAP	*gtsf-1(xf43)* TAP	8,591,095	7,115,711	6,430,227	5,261,958
TAP	*gtsf-1(xf44)* TAP	6,960,959	6,026,961	5,502,505	4,500,911
TAP	*gtsf-1(xf45)* TAP	6,581,972	5,282,531	4,488,250	3,616,180
Oxidized	N2_1 Oxi	9,174,231	2,245,943	1,687,068	1,434,120
Oxidized	N2_2 Oxi	8,900,999	3,718,282	1,659,146	1,376,007
Oxidized	N2_3 Oxi	7,853,162	3,353,046	1,058,549	597,434
Oxidized	*gtsf-1(xf43)* Oxi	8,408,350	4,107,568	3,320,877	2,252,014
Oxidized	*gtsf-1(xf44)* Oxi	8,303,337	2,730,473	2,403,390	1,585,142
Oxidized	*gtsf-1(xf45)* Oxi	7,449,672	3,925,982	2,356,583	1,175,778

GST-fusion construct preparation and expression. *Cloning*. *gtsf-1* fragments were amplified using Q5 High-Fidelity DNA Polymerase (New England BioLabs, M0491L), in a 50 µL reaction, from wild-type cDNA (obtained with ProtoScript First Strand cDNA synthesis, E6300, according to manufacturer's instructions). The forward and reverse oligos for amplification contained BamHI and NotI cut sites,

respectively. Fragments were checked on agarose gel, purified, digested with those enzymes (New England BioLabs, R3136 and R3189) for 2 hours at 37°C. Digested fragments were once again purified. Subsequently, these digested and purified *gtsf-1* fragments were ligated with 50 ng of the digested and desphosphorylated destination vector (pGEX-6P-1, pre-digested with BamHI and NotI, desphosphorylated with Antarctic Phosphatase, New England Biolabs, M0289S, and purified on column). A 3 insert:1 backbone molar ratio was used. Ligation with T4 DNA ligase (New England Biolabs, M0202) was performed overnight at 25 °C. Ligation was stopped with by incubating 10 minutes at 65°C. After cooling down, 5 µL of ligation were transformed in Subcloning Efficiency DH5α chemically competent cells (Life Technologies, 18265-017). Positive clones were identified by colony PCR, inoculated in an overnight LB culture, prepped and further confirmed by digestion and Sanger sequencing.

Expression. Positive constructs were retransformed in Rosetta cells (Merck Millipore, 71400-3) and plated in plates with Ampicillin. One colony was inoculated in a 5 mL pre-culture, supplemented with Ampicillin and Chloramphenicol (100 µg/ml and 35 µg/ml, respectively), and allowed to grow overnight at 37°C. The pre-culture was used to inoculate a 250 mL culture of LB, supplemented with Ampicillin and Chloramphenicol (100 µg/ml and 35 µg/ml, respectively) and allowed to grow, at 37°C, up to an OD_{600} of 0.5-0.9. Then, protein expression was induced with 1 mM IPTG (Promega, V3953) and incubated overnight at 18°C. Bacteria were collected by centrifugation (4°C, 4000 rpm, 15 minutes). Supernatant was discarded and the bacteria pellets were frozen at -80°C.

GST-on bead purification. The bacteria pellets were ressuspended in Lysis Buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1 mM DTT, 0.1 mM $ZnCl_2$ and protease inhibitors, cOmplete Mini, EDTA-free, Roche, 11836170001), and subsequently sonicated 3 times for 2 minutes, using output 2 with a Branson Sonifier 450. Debris were pelleted by centrifugation (4°C, 4000 rpm, 30-45 minutes) and discarded, supernatant was filtered through a 0,22 µM filter (Merck Millipore, SLGS033SS) and kept on ice. A 300 µL slurry of Glutathione Sepharose 4 fast flow beads (GE Healthcare, 17513201) was washed 3 times with 1 mL lysis buffer (centrifuge at 800 G for 3 minutes). The cell lysate was then incubated with the beads for 2-3 hours at 4°C with end-to-end mixing. After incubation, beads were washed 6 times with lysis buffer and, after the last wash, suspended 1:1 in lysis buffer. 5% glycerol was added for short-term preservation at 4°C. See next section (Biochemistry, GST-pulldowns) for the detailed GST pulldown protocol.

Biochemistry. *Worm preparations.* Young adult or gravid adult worms were collected and lysed in Lysis Buffer (25 mM Tris HCl pH 7.5, 150 mM NaCl, 1.5 mM $MgCl_2$, 1 mM DTT, 0.1% Triton X-100 and protease inhibitors: cOmplete Mini, EDTA-free, Roche, 11836170001). Lysis was performed either by sonication (10 cycles of 30 seconds) in a Bioruptor Plus (Diagenode), or by grinding frozen worm pellets

in liquid N_2 followed by douncing with 40 strokes, piston B. For embryo collection, large populations of gravid adults were bleached, embryos were washed several times with M9 buffer, and frozen in Lysis Buffer (25 mM Tris HCl pH 7.5, 150 mM NaCl, 1.5 mM $MgCl_2$, 1 mM DTT, 0.1% Triton X-100 and protease inhibitors: cOmplete Mini, EDTA-free, Roche, 11836170001) using liquid N_2. Lysis was performed by grinding frozen embryo pellets and douncing with 40 strokes, piston B. After lysis, lysates were cleared by centrifugation (15 minutes at 21,000 xG, 4°C) and protein concentration was measured using Bradford Protein assay according to manufacturer's instructions (Bio-Rad, 5000006).

Immunoprecipitation. Before IP, input sample was mixed 1:1 with Crack Buffer (0.1 M Tris-HCl pH 6.8, 4% SDS, 20% glycerol, Bromophenol blue, plus supplemented with 200 mM DTT just before usage) and boiled for 10 minutes. 30 µL of beads were washed 3 times with Wash Buffer (25 mM Tris HCl pH 7.5, 150 mM NaCl, 1.5 mM $MgCl_2$, 1 mM DTT and protease inhibitors: cOmplete Mini, EDTA-free, Roche 11836170001). Dynabeads Protein G (Life Technologies, 1004D) were used for FLAG and GTSF-1 IPs, while EZview™ Red Anti-HA Affinity Gel (Sigma, E6779) was used for HA pulldowns. After washing, 2 µg of antibody (in case of IPs with Dynabeads) and 2.5-6 mg of total protein extract were added to the beads and incubated for 3 hours at 4°C, rotating. Before washing, supernatant sample was collected and mixed 1:1 with Crack Buffer. Next, beads were washed 3 times with Wash Buffer, mixed 1:1 with Crack Buffer (0.1 M Tris-HCl pH 6.8, 4% SDS, 20% glycerol, Bromophenol blue, plus supplemented with 200 mM DTT just before usage) and boiled, if used for IP-Western. For mass spectrometry, immunoprecipitates were resuspended in NuPAGE LDS Sample Buffer 1X (Life technologies, NP0007), supplemented with 100 mM DTT, and heated at 70°C for 10 minutes.

GST pulldowns. 5 µg of GST-GTSF-1 beads were washed 3 times with wash buffer (25 mM Tris HCl pH 7.5, 150 mM NaCl, 1.5 mM $MgCl_2$, 1 mM DTT and protease inhibitors: cOmplete Mini, EDTA-free, Roche, 11836170001). Next, 0.5-1 mg of embryo lysate was incubated with the beads for 3 hours, at 4°C, with end-to-end mixing. Beads were washed 3 times and the beads were ressuspended in 10 µL of Crack Buffer (0.1 M Tris-HCl pH 6.8, 4% SDS, 20% glycerol, Bromophenol blue, plus supplemented with 200 mM DTT just before usage), boiled at 95°C for 10 minutes and spun down at 21,000 x G for 5 minutes. Finally, samples were loaded on a 4-12% gradient gel (Invitrogen, NP0321BOX and NP0323BOX) and Western blot was performed as described below.

Western blots. For anti-GTSF-1 blots in **Figure II.1**, 15% Polyacrylamide gels were prepared from a 30% Acrylamide/Bis solution, 29:1 (Bio-Rad, 161-0156) and ran on a Mini-PROTEAN Tetra Cell system (Bio-Rad). Transfer to an Immobilon PVDF, 0.45 µm membrane (Merck Millipore, IPFL00010 for fluorescence detection using LI-COR, or IPVH00010 for ECL detection) was executed on a Trans-Blot SD Semi-Dry Transfer Cell (Bio-Rad) at 20 V for 45 minutes. In other Western blot experiments, 4-12% gradient gels (Invitrogen, NP0321BOX and NP0323BOX) were used, ran on 1X MOPS SDS running buffer (Thermo Fischer Scientific, NP0001) at 150-180 V for 70-90 minutes, in a XCell SureLock Mini-Cell

Electrophoresis System (Thermo Fischer Scientific, EI0001). Wet blotting was subsequently performed overnight at 15 V, using a Bio-Rad setup (Mini-Protean Tetra Cell with a Mini Gel Holder Cassette, 1703931). Following transfer, samples were blocked in Blocking Buffer (PBS, 0.05% Tween and 5% skimmed milk) for 30 minutes at room temperature. Incubation with primary antibodies was performed for 1 hour at room temperature or overnight at 4°C. Prior to adding secondary antibodies, 3 washes with PBS-Tween (0.05%) were done. Secondary antibodies (Goat anti-Rabbit IRDye 800 CW, 926-32211, and Goat anti-Mouse IRDye 680 RD, 926-68070, were used in a 1:15000 dilution, LI-COR) were incubated for 1 hour at room temperature. After 3 more washes with PBS-Tween (0.05%), the membranes were imaged in a LI-COR Odyssey CLx apparatus (LI-COR). For 3xFLAG::RRF-3 blotting after GST-pulldowns, we used a 1:10000 dilution of Anti-mouse IgG, HRP-linked Antibody (Cell Signaling, 7076) and detection was performed with ECL Select Western Blotting Detection Reagent (GE Healthcare, RPN2235) according to manufacturer's instructions.

volume [ml]	log MW	calculated MW [kDa]	96 well
1.0	8.982	960063.591	a1
2.0	8.727	533212.105	a2
3.0	8.472	296141.997	a3
4.0	8.216	164475.040	a4
5.0	7.961	91348.201	a5
6.0	7.705	50734.105	a6
6.5	7.578	37809.419	a7
7.0	7.450	28177.340	**a8**
7.5	7.322	20999.067	a9
8.0	7.195	15649.483	**a10**
8.5	7.067	11662.724	a11
9.0	6.939	8691.605	**a12**
9.5	6.811	6477.389	**b12**
10.0	6.684	4827.252	**b11**
10.5	6.556	3597.493	**b10**
11.0	6.428	2681.020	**b9**
11.5	6.301	1998.021	**b8**
12.0	6.173	1489.018	**b7**
12.5	6.045	1109.686	**b6**
13.0	5.918	826.990	**b5**
13.5	5.790	616.311	**b4**
14.0	5.662	459.304	**b3**
14.5	5.534	342.295	**b2**
15.0	5.407	255.094	**b1**
15.5	5.279	190.108	**c1**
16.0	5.151	141.677	**c2**
16.5	5.024	105.584	**c3**
17.0	4.896	78.686	**c4**
17.5	4.768	58.641	c5
18.0	4.641	43.702	**c6**
18.5	4.513	32.569	c7
19.0	4.385	24.272	**c8**
19.5	4.257	18.088	c9
20.0	4.130	13.480	c10
20.5	4.002	10.046	c11
21.0	3.874	7.487	c12
21.5	3.747	5.580	d12
22.0	3.619	4.158	d11
22.5	3.491	3.099	d10
23.0	3.364	2.309	d9
23.5	3.236	1.721	d8
24.0	3.108	1.283	d7

Table II.S4. Calibration and resolution of the size-exclusion column. See text for details.

Quantification of Western blots. For detection, secondary antibodies compatible with LI-COR Odyssey CLx were used. Quantification was done by gating the region surrounding the bands using Image Studio software (v3.1). Median background was used to correct the signal. Within every technical replicate, the signal of the protein of interest was normalized relative to TBA-1. Next, the normalized signal of either *eri-5* wild-type (in case of **Figure II.S6A**), or *gtsf-1* wild-type (in case of **Figure II.S6B**) were set to 1 for comparison with their mutants.

Size exclusion chromatography-Western blot. Embryo or young adult extracts were prepared, as described above, by grinding and douncing. 500 µL of sample, corresponding to 3.6-4.5 mg 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 fractions were collected (see **Table II.S4**). Included in the red square are the fractions that are covered by the separation range of the column (5-5000 kDa). Included in the green square are the fractions that are covered by the marker run with protein standards (Bio-Rad, Gel Filtration Standard, 1511901). For the fractions within those squares, the calculation of the approximate Molecular Weight is more accurate. Outside the red squares (especially, fractions B11-B5), the extrapolation of the corresponding molecular weight has to

be interpreted with caution. The fractions used for Western blot analysis are in bold (see **Table II.S4**). These fractions were concentrated to ~20 µL using Amicon Ultra 10 kDa cutoff filter units (Merck-Millipore, UFC501096) according to manufacturer's instructions. Samples were then mixed with NuPAGE 1x LDS Buffer (Life technologies, NP0007) supplemented with 100 mM DTT. Next, samples were loaded and ran on a 4-15% Criterion TGX Stain-Free Protein Gel (26 wells, Bio-Rad, 5678085). Transfer onto a Nitrocellulose membrane (Bio-Rad, 1620112) was performed using Trans-Blot Turbo Transfer System (Bio-Rad). Western blot detection was performed as described above, using ECL Select Western Blotting Detection Reagent (GE Healthcare, RPN2235).

Quantification of size exclusion chromatography profiles. Tiff files 3xFLAG::RRF-3 or GTSF-1 Western blot were loaded into ImageJ. A rectangle was drawn encompassing all the bands of interest. That rectangle was then defined as the first lane (Analyze -> Gels -> Select First Lane). Lane was plotted and vertical lines were drawn to separate the signals referring to different wells. Next, the "wand" function was used to select all the wells. The resulting data was exported to excel for analysis. Arbitrary units (a.u.) are defined as [signal band X]/[signal input]. To be clear, all the fractions coming from one 3xFLAG::RRF-3;*rrf-3* biological replicate, were normalized to the input of that particular biological replicate. Standard deviation was calculated for gels of two independent biological replicates.

RT-qPCR. Embryo samples were prepared as described above (see worm preparations in the Biochemistry section). 3 volumes of TRIzol LS (Life Technologies, 10296-028) were added to the samples, thoroughly mixed, then 1 volume of 100% ethanol was added and again thoroughly mixed. This mix was pipetted into a Direct-zol column (Zymo Research, R2070) and manufacturer's instructions were followed (in-column DNase I treatment was included). Reverse transcription was performed in 1 µg of total RNA with ProtoScript First Strand cDNA Synthesis Kit (New England Biolabs, E6300), using the random primer mix, according to manufacturer's instructions. Next, qPCR was performed in a ViiA 7 Real-Time PCR System (Thermo Fisher Scientific) using iTaq Universal SYBR Green Supermix (Bio-Rad, 1725121) according to manufacturer's instructions. 10 µL reactions were prepared in 384 well plates (Applied Biosystems, 4309849), 1/10 of which was cDNA. Primers were used in a final concentration of 300 nM. Cycling conditions are as follows: Standard run, 1.6°C/s of increment in temperature throughout the run; 50°C for 2 minutes, 30 seconds at 95°C; [40 cycles of 95°C for 15 seconds and 60°C for 45 seconds]; melt curve calculation [15 seconds at 95°C, 1 minute at 60°C and 15 seconds at 95°C]. Technical triplicates and biological duplicates were used. Analysis was performed using the ΔΔCT method (Schmittgen and Livak, 2008). *pmp-3* was used as a normalization factor (Hoogewijs et al., 2008). Error bars represent the standard deviation of two biological replicates. A list of used primers is included below, in **Table II.S5**.

Target	Sequence
*pmp-3*_Fw	GTTCCCGTGTTCATCACTCAT
*pmp-3*_Rev	ACACCGTCGAGAAGCTGTAGA
*pgl-3*_Fw	CCCACTGCTCCCTCAAAGCG
*pgl-3*_Rev	CAGTCCTTGGGCGAACTTTTTGAAG
*E01G4.5*_Fw	AAGCTGTGGGCACTTTGAGT
*E01G4.5*_Rev	ACAATCACGGCACACAAAACA
*K02E2.6*_Fw	CGAACGGAACACCCATCTTG
*K02E2.6*_Rev	TTCATAAAAGCCTTTTTCATGAAGTCT
*ZK402.5*_Fw	CGGACGACTTTCTGACACAAGT
*ZK402.5*_Rev	CACTCGAGCTCATTTCAGTTCTC
*Y43F8B.9*_Fw	CGAAGGAGGCATGGACTAAA
*Y43F8B.9*_Rev	GACTCCTTCGACGGATACACA
*Y37E3.30*_Fw	CAACATCTTGATCGGTGTGC
*Y37E3.30*_Rev	TGCATCTGTCACGCAATG
*W04B5.2*_Fw	GAAGGCACAAGGACATGGAT
*W04B5.2*_Rev	GTCTTGGAAGCCGCAAATCT
*Y37E11B.2*_Fw	CCTCATCCGTGAAATCGTCT
*Y37E11B.2*_Rev	GGGTAAGGTTTCAGCGAAGG
*lsy-13*_Fw	ACTGCTGGTGTCAACTGGA
*lsy-13*_Rev	TACTTGCAGCCCGGATTCT

Table II.S5. List of primers used for RT-qPCR.

GTSF-1 *in vitro* iCLIP. Anti-U2AF65 (Sigma, U4758) antibody and anti-GTSF1 antibody were coupled to protein G Dynabeads (Life Technologies, 10004D) and protein A Dynabeads (Life Technologies, 10002D), respectively. We used recombinant wildtype GTSF-1 protein from *C. elegans* (500 nM and 800 nM, respectively), as well as a recombinant peptide comprising the two RRM domains of human U2AF65 (U2AF65^{RRM12}; amino acid residues 148-342; 500 nM), which largely recapitulates the binding characteristics of the full-length protein. The preparation of U2AF65^{RRM12} was done as described in (Mackereth et al., 2011) , with the exception that cells were disrupted in lysis buffer (50 mM Tris pH 7.5, 500 mM NaCl, 5% (v/v) glycerol, 20 mM imidazole, 0.1% Triton X) with glass beads and supernatant was purified with Proteus clarification columns (Generon; GEN-MSF500). For clarification, the lysate was incubated with Ni Sepharose beads (GE Healthcare, 17-5318-01) for 1 h and washed eleven times in wash buffer (50 mM Tris pH 7.5, 500 mM NaCl, 5% (v/v) glycerol, 20 mM imidazole, 1 mM DTT). The proteins were then directly eluted into Proteus clarification columns with elution buffer

(50 mM Tris pH 7.5, 500 mM NaCl, 5% (v/v) glycerol, 500 mM imidazole) in order to clarify the protein extract again. The recombinant proteins were incubated with 30 nM Ribo-Zero-selected RNA (Ribo-Zero rRNA removal kit, MRZH11124, according to manufacturer's instructions) from *C. elegans* for 10 min at 37°C, before they were subjected to UV-C irradiation (100 mJ/cm2, Stratalinker 2400, 254 nm). The protein-RNA mixture was resuspended in lysis buffer (50 mM Tris-HCl pH 7.4, 100 mM NaCl, 1% Igepal CA-630, 0.1% SDS, 0.5% sodium deoxycholate), followed by RNase I (Life Technologies, AM2295) digestion for 3 min at 37°C (1:5 and 1:1500 RNase dilution for high and low RNase treatment, respectively). For immunoprecipitation, the RNase I-digested protein-RNA mixture was added to the beads and incubated for 1 h at 4°C. Afterwards, the samples were washed twice in high-salt wash buffer (50 mM Tris-HCl pH 7.4, 1 M NaCl, 1 mM EDTA, 0.1% SDS, 0.5% sodium deoxycholate, 1% Igepal CA-30) and twice in PNK wash buffer (20 mM Tris-HCl pH 7.4, 10 mM MgCl2, 0.2% Tween-20). Protein-bound RNAs were radioactively labeled with 32P-γ-ATP at their 5' end for 5 min at 37°C with T4 polynucleotide kinase (PNK; New England Biolabs, M0201L). To elute the proteins, the beads were boiled with NuPAGE LDS Sample Buffer (Fisher Scientific, 11549166) for 10 min at 70°C. The samples were run on a NuPAGE Novex 4-12% Bis-Tris protein gel (Life Technologies, NP0322). After electrophoresis, protein-RNA complexes were transferred to a nitrocellulose membrane (VWR, 10401196) by Western blotting. Membranes were exposed to a Fuji imaging plate (Fujifilm, 28956475) for 1 h at 4°C and visualized with a Typhoon phosphorimager.

Chapter III

Maternal and zygotic gene regulatory effects of endogenous RNAi pathways

Miguel Vasconcelos Almeida☯, António Miguel de Jesus Domingues☯, and René F. Ketting

Biology of Non-Coding RNA group, Institute of Molecular Biology, Ackermannweg 4, 55128 Mainz, Germany

☯Equal contribution.

This chapter is a reproduction of the following scientific paper:

Almeida, M.V., Domingues A.M.d.J., and Ketting RF. (2019). Maternal and zygotic gene regulatory effects of endogenous RNAi pathways. PLoS Genetics *15*, e1007784.

Author contributions:

Conceptualization: MVA and RFK; Investigation: MVA; Formal analysis: MVA, AMdJD; Visualization: MVA, AMdJD; Writing - original draft: MVA; Writing - review & editing: all authors contributed; Funding acquisition: RFK.

Abstract

Endogenous sRNAs and AGO proteins are ubiquitous regulators of gene expression in germline and somatic tissues. sRNA-AGO complexes are often expressed in gametes and are consequently inherited by the next generation upon fertilization. In *C. elegans*, 26G-RNAs are primary endogenous sRNAs that trigger the expression of downstream secondary sRNAs. Two subpopulations of 26G-RNAs exist, each of which displaying strongly compartmentalized expression: one is expressed in the spermatogenic gonad and associates with the AGOs ALG-3/4; plus another expressed in oocytes and in embryos, which associates with the AGO ERGO-1. The determinants and dynamics of gene silencing elicited by 26G-RNAs are largely unknown. Here, we provide diverse new insights into these endogenous sRNA pathways of *C. elegans*. Using genetics and deep sequencing, we dissect a maternal effect of the ERGO-1 branch of the 26G-RNA pathway. We find that maternal primary sRNAs can trigger the production of zygotic secondary sRNAs that are able to silence targets, even in the absence of zygotic primary triggers. Thus, the interaction of maternal and zygotic sRNA populations, assures target gene silencing throughout animal development. Furthermore, we explore other facets of 26G-RNA biology related to the ALG-3/4 branch. We find that sRNA abundance, sRNA pattern of origin and the 3' UTR length of target transcripts are predictors of the regulatory outcome by the AGOs ALG-3/4. Lastly, we provide evidence suggesting that ALG-3 and ALG-4 regulate their own mRNAs in a negative feedback loop. Altogether, we provide several new regulatory insights on the dynamics, target regulation and self-regulation of the endogenous RNAi pathways of *C. elegans*.

Author Summary

sRNAs and their partner AGO proteins regulate the expression of target RNAs. When sperm and egg meet upon fertilization, a diverse set of proteins and RNA, including sRNA-AGO complexes, is passed on to the developing progeny. Thus, these two players are important to initiate specific gene expression programs in the next generation. The nematode *C. elegans* expresses several classes of sRNAs. 26G-RNAs are a particular class of sRNAs that are divided into two subpopulations: one expressed in the spermatogenic gonad and another expressed in oocytes and in embryos. In this work, we describe the dynamics whereby oogenic 26G-RNAs setup gene silencing in the next generation. In addition, we show several ways that spermatogenic 26G-RNAs and their partner AGOs, ALG-3 and ALG-4, use to regulate their targets. Finally, we show that ALG-3 and ALG-4 are fine-tuning their own expression, a rare role of AGO proteins. Overall, we provide new insights into how sRNAs and AGOs are regulating gene expression.

Introduction

A plethora of pathways based on non-coding sRNAs regulates gene expression in every domain of life. These are collectively known as RNAi or RNAi-like pathways. In invertebrates, which lack adaptive immune systems and interferon response, RNAi-like pathways fulfill an immune role at the nucleic acid level, by controlling viruses and TEs.

miRNA, piRNA, and endo-siRNA pathways are the better described RNAi-like pathways, which differ in their biogenesis and specialized cofactors. MicroRNAs are commonly found in many, if not all, tissues and broadly regulate gene expression throughout development (Ha and Kim, 2014). piRNAs are typically, but not exclusively, expressed in the metazoan germline, where they assume a central function in TE control (Ernst et al., 2017; Huang et al., 2017; Luteijn and Ketting, 2013; Rojas-Ríos and Simonelig, 2018). Endo-siRNA pathways comprise varied classes of sRNAs expressed in the soma and germline that can regulate the expression of TEs and protein-coding genes (Borges and Martienssen, 2015; Holoch and Moazed, 2015; Kim et al., 2009). A key commonality of RNAi-like pathways is the participation of AGO proteins. These proteins directly associate with sRNAs and AGO-sRNA complexes engage transcripts with sequence complementarity, typically resulting in target silencing. sRNA-directed gene silencing can occur both on the post-transcriptional level, by target RNA cleavage and degradation, and/or on the transcriptional level, via nuclear AGOs that direct heterochromatin formation at target *loci*.

sRNAs can be viewed as genome guardians against “foreign” nucleic acids (Malone and Hannon, 2009). In this light, the germline is an important tissue for sRNA production and function to control the transmission of “non-self” genetic elements to progeny. In multiple animals, Piwi-piRNA complexes have been shown to be maternally deposited into zygotes, where they may initiate TE silencing (Akkouche et al., 2013; Brennecke et al., 2007, 2008; de Albuquerque et al., 2015; Houwing et al., 2007; Kawaoka et al., 2011; Le Thomas et al., 2014; Le Thomas et al., 2014; Ninova et al., 2017; de Vanssay et al., 2012). Endo-siRNAs are abundantly expressed in gametes, being often required to successfully complete gametogenesis. These may also be deposited into embryos and have roles in setting up gene expression in the next generation. For example in plants, TE-derived endo-siRNAs are abundant in male and female gametes (Martinez and Köhler, 2017). Moreover, endo-siRNAs are expressed in *Drosophila* ovaries (Czech et al., 2008) and in mouse oocytes (Tam et al., 2008; Watanabe et al., 2008) to regulate protein-coding genes and TEs. Overall, gamete expression and maternal inheritance of AGO-sRNA complexes seem to be a widespread phenomenon in plants and animals, presumably important to tune gene expression during early development.

RNAi was first identified in the nematode *C. elegans* (Fire et al., 1998). Ever since, *C. elegans* has continuously been an important and fascinating model for studies on RNAi. *C. elegans* has an unprecedented 27 genomically encoded AGO genes, including a whole worm-specific clade of the AGO

protein family (Yigit et al., 2006). Several sRNA species have been identified in worms: miRNAs, 21U-RNAs, 22G- and 26G-RNAs (Ambros et al., 2003; Ruby et al., 2006). 21U-RNAs associate with PRG-1, a Piwi class AGO, in the germline and are therefore considered the piRNAs of *C. elegans* (Batista et al., 2008; Das et al., 2008; Wang and Reinke, 2008). 26G-RNAs can be considered primary endo-siRNAs, in that they elicit production of the overall more abundant secondary endo-siRNA pool, termed 22G-RNAs (Conine et al., 2010; Gent et al., 2010; Vasale et al., 2010).

26G-RNAs are produced by the RdRP RRF-3 (Conine et al., 2010; Gent et al., 2009, 2010; Han et al., 2009; Vasale et al., 2010). The ERI complex is an accessory complex that assists RRF-3 in producing 26G-RNAs (Duchaine et al., 2006; Pavelec et al., 2009; Thivierge et al., 2012). The conserved CHHC zinc finger protein GTSF-1 (see **Chapter II**, especially **Figures II.5-II.8 and II.S5-II.S6**) and the Tudor domain protein ERI-5 form a pre-complex with RRF-3 that is responsible for tethering the RdRP to the ERI complex (Thivierge et al., 2012). Two distinct subpopulations of 26G-RNAs are synthesized in the germline and in embryos. One subpopulation is produced in the spermatogenic gonad in L4 hermaphrodites and in the male gonad, where they associate with the redundantly acting paralog AGOs ALG-3 and ALG-4 (henceforth referred to as ALG-3/4) (Conine et al., 2010; Gent et al., 2009; Han et al., 2009; Pavelec et al., 2009). These 26G-RNAs trigger the biogenesis of secondary 22G-RNAs that have been shown to either promote gene expression through the AGO CSR-1 or to inhibit gene expression through unidentified WAGO proteins (Conine et al., 2010, 2013). Hence, the effects of ALG-3/4-dependent sRNAs on their targets is complex: while some targets appear to be silenced, the expression of others seems to be positively affected. The regulatory effects resulting of the combined action of ALG-3/4 and CSR-1 seem to be more physiologically relevant at elevated temperatures (Conine et al., 2013). The conditions determining regulatory outcome, either silencing or licensing, are still unclear.

In the oogenic hermaphrodite gonad and in embryos another subpopulation of 26G-RNAs is produced. These are 3’ 2’-O-methylated by the conserved RNA methyltransferase HENN-1 (Billi et al., 2012; Kamminga et al., 2012; Montgomery et al., 2012) and bind to the AGO ERGO-1 (Vasale et al., 2010). ERGO-1 targets pseudogenes, recently duplicated genes and lncRNAs (see **Figures II.4**, **II.S4**, and **Table II.S1**; Fischer et al., 2011; Vasale et al., 2010). It has recently been shown that these targets generally have a small number of introns that lack optimal splicing signals (Newman et al., 2018). ERGO-1 may thus serve as a surveillance platform to silence these inefficient transcripts, preventing detrimental accumulation of stalled spliceosomes. Effective silencing of these genes is achieved by secondary 22G-RNAs produced after ERGO-1 target recognition (Gent et al., 2010; Vasale et al., 2010). In turn, these secondary 22G-RNAs may associate with cytoplasmic AGOs that mediate post-transcriptional gene silencing (Vasale et al., 2010), or to the AGO NRDE-3, which is shuttled into

the nucleus and further silences its targets on the transcriptional level (Guang et al., 2008; Zhou et al., 2014).

Depletion of spermatogenic 26G-RNAs, for example in *gtsf-1* (see **Figures II.3** and **II.S3**), *rrf-3*, and *alg-3/4* mutants, results in a range of sperm-derived fertility defects including complete sterility at higher temperatures (Conine et al., 2010; Duchaine et al., 2006; Gent et al., 2009; Han et al., 2009; Pavelec et al., 2009). The elimination of oogenic/embryonic 26G-RNAs, for example by impairment of *gtsf-1* (see **Figures II.3** and **II.S3**), *rrf-3*, and *ergo-1*, gives rise to an Eri phenotype, characterized by a response to exogenous dsRNA that is stronger than in wild-type (Duchaine et al., 2006; Pavelec et al., 2009; Yigit et al., 2006). This phenotype is thought to reflect competition for common factors between exogenous and endogenous RNAi pathways (Duchaine et al., 2006; Lee et al., 2006). However, the Eri phenotype lacks characterization on the molecular level. Furthermore, a strong maternal rescue was reported for Eri factors (Zhuang and Hunter, 2011), suggesting that maternally deposited Eri factors or their dependent sRNAs have an important role in maintaining gene silencing. The basis for this maternal rescue was not further characterized.

In this work, we address a number of gene regulatory aspects of the 26G-RNA pathways in *C. elegans*. First, we genetically dissect a maternal effect displayed by the ERGO-1 branch of the 26G-RNA pathway. Our findings suggest that both maternal and zygotic sRNAs drive gene silencing throughout embryogenesis and larval development until adulthood. Moreover, we interrogate a number of aspects on gene regulation in the ALG-3/4 branch of the 26G-RNA pathway. We report that sRNA abundance, origin of the sRNAs and 3' UTR length of target transcripts are predictors of the regulatory outcome of ALG-3/4 targets. Lastly, we find that the 26G-RNA-binding AGOs ALG-3 and ALG-4 may regulate their own expression in a negative feedback mechanism.

Results

Maternal and zygotic endogenous small RNAs drive RNAi in the soma

rrf-3 and *gtsf-1* mutants lack the two subpopulations of 26G-RNAs and display the phenotypes associated with depletion of both subpopulations: the enhanced RNAi (Eri) phenotype, shared with *ergo-1* mutants (**Figures II.3** and **II.S3**; Duchaine et al, 2006; Pavelec et al., 2009; Yigit et al., 2006), and sperm-derived fertility defects, shared with *alg-3/4* double mutants (**Figures II.3** and **II.S3**; Conine et al., 2010, 2013; Duchaine et al., 2006; Gent et al., 2009; Han et al., 2009; Pavelec et al., 2009). **Figure III.S1A** offers a simplified scheme of these pathways. For clarity, the two subpopulations of 26G-RNAs and downstream 22G-RNAs, dependent on ERGO-1 or ALG-3/4 will be referred to as ERGO-1 branch sRNAs and ALG-3/4 branch sRNAs, respectively.

We have previously shown that germline-specific GTSF-1 transgenes could rescue the enhanced RNAi (Eri) phenotype of *gtsf-1* mutants (**Figures II.3** and **II.S3**). This was an intriguing result, since the Eri phenotype arises after targeting somatically expressed genes with RNAi, indicating that germline-expressed GTSF-1 is able to affect RNAi in the soma, possibly through maternal deposition of GTSF-1 or GTSF-1-dependent sRNAs. We reasoned that if maternal GTSF-1 activity can prime gene silencing in embryos then the transmission of the Eri phenotype should show a maternal rescue. To address this experimentally, we linked *gtsf-1(xf43)* to *dpy-4(e1166)* and crossed the resulting double mutants with wild-type males (**Figure III.1A**). We then allowed for two generations of heterozygosity and assayed for RNAi sensitivity in homozygous *gtsf-1* mutant F1 and F2 generations, scoring for larval arrest triggered by *lir-1* RNAi. Indeed, the Eri phenotype showed a strong maternal effect, arising only in the F2 generation of *gtsf-1* mutants (**Figure III.1A**). This is consistent with a maternal effect reported for other Eri factors (Zhuang and Hunter, 2011). We have previously shown that GTSF-1 is required to silence a GFP transgene reporting on ERGO-1 branch 22G-RNA activity, referred to as 22G sensor (**Figures II.3** and **II.S3**; and Montgomery et al., 2012). Therefore, we also looked at the dynamics of derepression of this transgene upon introduction of *gtsf-1* mutation. We noticed that strong GFP expression appeared only in the second generation of homozygosity of the *gtsf-1* allele (**Figure III.1B-C**). An identical maternal effect on the expression status of this transgene is observed after crossing in *rrf-3*, *ergo-1* and other *gtsf-1* mutant alleles (**Figure III.S1B**). Combined with our previously described rescue of the Eri phenotype using a germline promoter, these results strongly suggest that maternally provided ERGO-1 branch pathway components are sufficient to establish normal RNAi sensitivity in the soma of *C. elegans*.

Although the silencing of the 22G sensor used in our experiments is dependent on ERGO-1, ERGO-1 is not the AGO protein binding to the effector 22G-RNA (Montgomery et al., 2012; Vasale et al., 2010). This has been shown to be driven by the somatically expressed, nuclear AGO protein NRDE-3 (Fischer et al., 2011; Guang et al., 2008) and maybe additional cytoplasmic WAGOs (Vasale et al., 2010) (**Figure III.S1A**). In absence of ERGO-1 and other 26G-RNA pathway factors, NRDE-3 is no longer nuclear, and in *nrde-3* mutants the 22G sensor is activated, indicating that NRDE-3 requires sRNA input from ERGO-1 branch sRNAs (**Figure II.3**; Fischer et al., 2011; Guang et al., 2008; Montgomery et al., 2012). Strikingly, loss of NRDE-3 derepressed the 22G sensor transgene in the first homozygous generation (**Figure III.1D**), showing that in contrast to 26G-RNAs, the downstream 22G-RNA pathway is not maternally provided. MUT-16 is a factor required for the nucleation of mutator foci and 22G-RNA biogenesis (Phillips et al., 2012). Confirming the requirement for zygotically produced 22G-RNAs, absence of MUT-16 derepresses the 22G sensor in the first homozygous mutant generation (**Figure III.1E**). These results suggest a scenario in which 1) NRDE-3 is loaded with zygotically produced

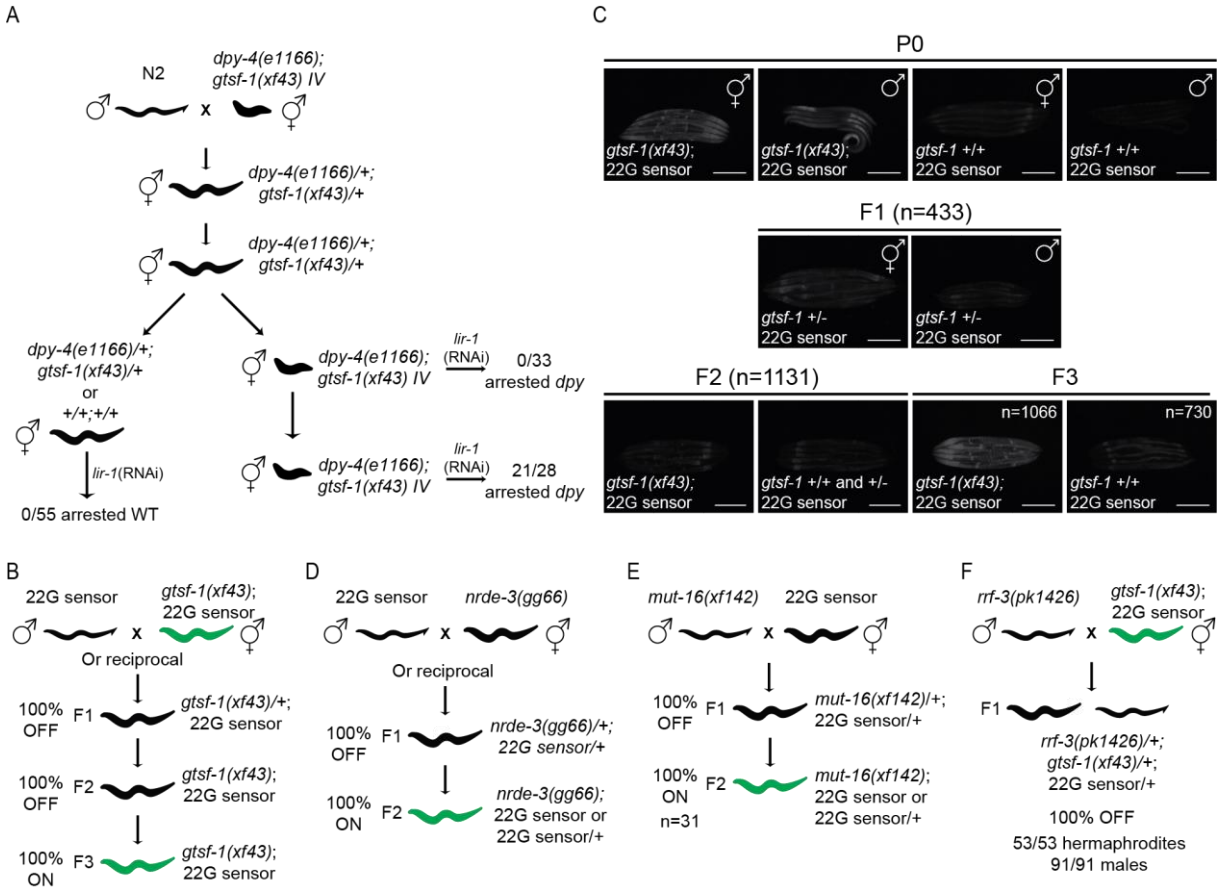


Figure III.1. Maternal and zygotic sRNAs drive RNAi in the soma. (A) Experimental setup to address maternal transmission of the Eri phenotype in *gtsf-1* mutants. Eri phenotype was assessed by transferring worms to plates containing *lir-1* RNAi food and scoring for larval arrest. *dpy-4(e1166)* is weakly semi-dominant. Since the phenotype is mild, for simplicity, we will refer to *dpy-4(e1166)* heterozygotes as "wild-type". (B, D-F) Schematics of genetic crosses using the 22G sensor background. Green worms illustrate ubiquitous derepression of the 22G sensor. Conversely, black worms illustrate repression of the 22G sensor. Unless otherwise noted, for all crosses the number of scored F1s, F2s and F3s was each >50. (C) Related to the cross shown in (B). Wide-field fluorescence microscopy images showing 22G sensor GFP signal. Five representative gravid adult hermaphrodites or adult males from each generation are shown. Of note, some autofluorescence of the gut is observed in gravid adult animals and is especially noticeable in worms with the sensor off. Scale bars represent 0,25 mm.

22G-RNAs that are primed by maternally provided 26G-RNAs and 2) NRDE-3 activity is maintained in somatic tissues until the adult stage, in absence of a zygotic 26G-RNA pathway.

The results presented above show that maternal 26G-RNAs are sufficient for 22G sensor silencing. We also tested whether maternal 26G-RNAs are necessary for 22G sensor silencing by crossing *rrf-3* mutant males with *gtsf-1; 22G sensor* hermaphrodites (**Figure III.1F**). Both of these strains lack 26G-RNAs and their downstream 22G-RNAs, therefore, their progeny will not receive a maternal and/or paternal complement of these sRNAs. The 22G sensor was silenced in all cross progeny, showing that in the absence of maternal 26G-RNAs, zygotic 26G-RNAs can induce production of silencing-competent 22G-RNAs. Thus, maternal 26G-RNAs appear to be sufficient but not necessary for target silencing.

26G-RNA-derived parental effects are restricted to the ERGO-1 branch

The maternal effects described above for the Eri phenotype and for 22G sensor silencing are related to the ERGO-1 branch of the pathway. Next, we wanted to determine if the ALG-3/4 branch also displays such a parental effect. To test this, we assessed the influence of maternal GTSF-1 activity on the temperature-sensitive sterility phenotype. Using the same setup as we used for the Eri experiment (in **Figure III.1A**), we observe that the temperature-sensitive sperm defect of *gtsf-1* mutants was not rescued maternally (**Figure III.S1C**). Given that the ALG-3/4 branch of the 26G-RNA pathway is mostly active during spermatogenesis, next we asked whether a paternal effect is observed for the temperature-sensitive sperm defect. As shown in **Figure III.S1D**, we did not detect any evidence supporting a paternal effect. Overall, these results indicate that 26G-RNA-derived parental effects are likely restricted to the ERGO-1 branch.

Maternal GTSF-1 supports zygotic production of ERGO-1 branch 22G-RNAs

The 22G sensor reports on the silencing activity of a single 22G-RNA that maps to the so-called X-cluster, a known set of targets of ERGO-1 (Montgomery et al., 2012; Vasale et al., 2010). Therefore, the experiments above using this 22G sensor have a limited resolution and our observations may not reflect the silencing status of most ERGO-1 targets. To characterize this maternal effect in more detail and in a broader set of ERGO-1 targets, we decided to analyze sRNA populations in young adult animals. Concretely, we outcrossed *dpy-4*; *gtsf-1* and sequenced sRNAs from wild-type and two consecutive generations of Dpy young adult animals (**Figure III.2A**). First generation *gtsf-1* homozygous mutants will henceforth be addressed as “mutant F1” and second generation *gtsf-1* homozygous mutants as “mutant F2” (**Figure III.2A**). We sequenced young adult animals because they lack embryos, therefore avoiding confounding effects with zygotic sRNAs of the next generation. sRNAs were cloned and sequenced from four biological replicates. The cloning of sRNAs was done either directly (henceforth referred to as untreated samples) or after treatment with the pyrophosphatase RppH (Almeida et al., 2019) before library preparation. The latter enriches for 22G-RNA species that bear a 5’ triphosphate group. Sequenced sRNAs were normalized to all mapped reads excluding structural reads (**Table III.S1**). In our analysis we strictly looked at 26G- and 22G-RNAs that map in antisense orientation to protein-coding and non-coding genes (see **Material and Methods**).

Total 26G-RNA levels are depleted in young adults lacking GTSF-1 (**Figure III.2B**). Mutant F1s have significantly less 26G-RNAs than wild-type worms, while mutant F2s have 26G-RNA levels very close to zero (**Figure III.2B**). For a finer analysis we looked specifically at 26G-RNAs derived from ERGO-1 and ALG-3/4 targets (see **Materials and Methods**). 26G-RNAs mapping to these two sets of targets recapitulate the pattern observed for global 26G-RNAs (**Figure III.2C-D**). The difference between the

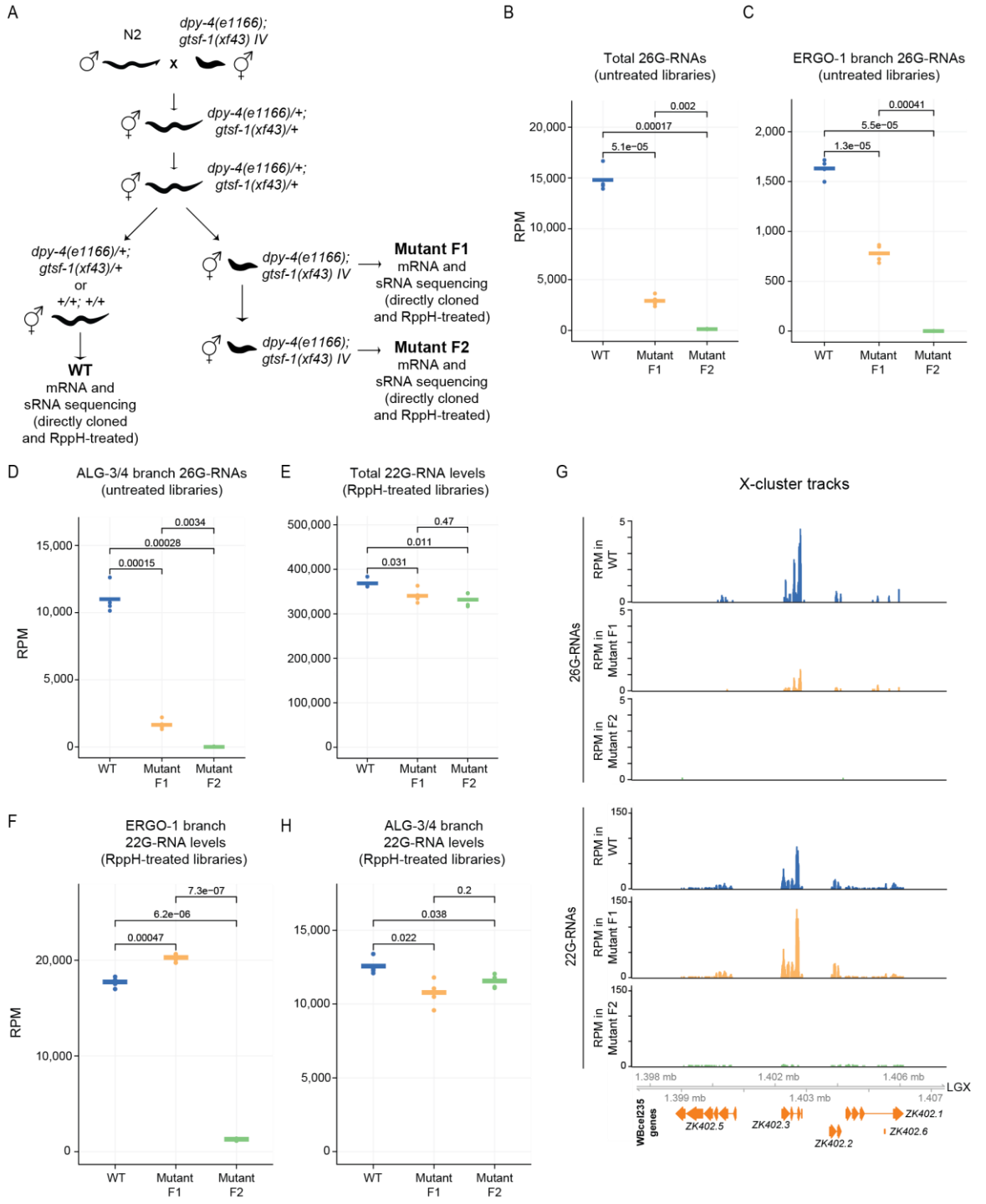


Figure III.2. sRNA dynamics in Eri maternal inheritance. (A) Schematics of the cross setup used to isolate worms of different generations and *gtsf-1* genotypes. *gtsf-1;dpy-4* mutants were outcrossed with N2 males, allowed to self for two generations and then WT, Dpy mutant F1 and Dpy mutant F2 young adult animals were isolated, RNA was extracted, sRNAs and mRNAs were cloned and sequenced. sRNA libraries were either prepared directly or after treatment with RppH. WT, wild-type. (B-F, H) Normalized levels of sRNAs, in RPM (Reads Per Million), per generation/phenotype. Four biological replicates are shown. P-values were calculated with a two-sided unpaired t-test. (B) Total levels of 26G-RNAs in the untreated libraries. (C) Total levels of 26G-RNAs mapping to ERGO-1 targets in the untreated libraries. (D) Total levels of 26G-RNAs mapping to ALG-3/4 targets in the untreated libraries. (E) Total levels of 22G-RNAs in the RppH-treated libraries. (F) Total levels of 22G-RNAs mapping to ERGO-1 targets in the RppH-treated libraries. (G) Genome browser tracks of the X-cluster, a known set of ERGO-1 targets, showing mapped 26G- and 22G-RNAs. 26G- and 22G-RNA tracks were obtained from untreated and RppH-treated libraries, respectively. (H) Total levels of 22G-RNAs mapping to ALG-3/4 targets in RppH-treated libraries.

F1 and F2 mutants might reflect a maternal 26G-RNA pool that is still detectable in the young adult F1, but no longer in the F2. However, we point out that amongst the selected F1 Dpy animals, approximately 5.2% will in fact be *gtsf-1* heterozygous, due to meiotic recombination between *gtsf-1* and *dpy-4* (estimated genetic distance between these two genes is 2.6 map units). Hence, another explanation for the mutant F1 pool of 26G-RNAs may be a contamination of the *gtsf-1* homozygous pool with heterozygous animals. The mutant F2 was isolated from genotyped F1 animals, excluding this confounding effect. We conclude that in young adult mutant F1 animals, maternally provided 26G-RNAs (or 26G-RNAs produced zygotically by maternal proteins) are no longer detectable at significant levels.

Total levels of 22G-RNAs are slightly reduced in mutant F1 and F2 animals (**Figure III.2E**). However, total 22G-RNA levels encompass several distinct subpopulations of 22G-RNAs, including those that do not depend on 26G-RNAs. To have a closer look on 22G-RNAs that are dependent on 26G-RNAs, we focused on 22G-RNAs that map to ERGO-1 and ALG-3/4 targets. Strikingly, compared to wild-type, the 22G-RNA population from ERGO-1 targets is moderately higher in mutant F1 animals and is subsequently depleted in the mutant F2 generation (**Figures III.2F** and **III.S2A**). These effects are not only clear in overall analysis, but also on a well-established set of ERGO-1 branch targets, such as the X-cluster (**Figure III.2G**). Consistent with a role of NRDE-3 downstream of ERGO-1, 22G-RNAs mapping to annotated NRDE-3 targets (Zhou et al., 2014) show the same pattern of depletion as ERGO-1-dependent 22G-RNAs (**Figure III.S2A**). These results are consistent with the idea that the Eri phenotype and 22G sensor derepression are caused by the absence of NRDE-3-bound, secondary 22G-RNAs downstream of 26G-RNAs.

22G-RNAs mapping to ALG-3/4 targets behave differently in this experiment (**Figure III.2H**). Upon disruption of *gtsf-1,* these 22G-RNAs are only slightly affected in both the mutant F1 and F2 (**Figures III.2H** and **III.S2A**), despite the fact that their upstream 26G-RNAs are absent. This is illustrated in **Figure III.S2B** with genome browser tracks of *ssp-16*, a known ALG-3/4 target. We conclude that 26G-RNA-independent mechanisms are in place to drive 22G-RNA production from these genes.

Finally, 21U-RNAs and 22G-RNAs mapping to other known RNAi targets are not affected in this inheritance setup, supporting the notion that *gtsf-1* is not affecting these sRNA species (**Figure III.S2A, C**). One exception are the 22G-RNAs from CSR-1 targets, which seem to be slightly depleted in both the mutant F1 and F2 generations (**Figure III.S2A**). It is not possible to dissect whether this is a direct effect or not, but we note that mRNA levels of CSR-1 targets are slightly downregulated in the analyzed mutants (**Figure III.S2D**). Given that CSR-1 22G-RNAs tend to correlate positively with gene expression (Claycomb et al., 2009), it is conceivable that the reduction of CSR-1 target 22G-RNAs is the result of decreased target gene expression.

ERGO-1 pathway mRNA targets show stronger upregulation in the second *gtsf-1* homozygous mutant generation

The very same samples used for generating sRNA sequencing data were also used for mRNA sequencing (**Figure III.2A** and **Table III.S1**). First, we checked *gtsf-1* expression. As expected, *gtsf-1* is strongly depleted in the mutant samples (**Figure III.S3A**). In the mutant F1 we still observe a low level of *gtsf-1* derived transcripts (about 9.5% of wild-type) that is absent from the mutant F2. These transcripts cover the region deleted in the *gtsf-1(xf43)* mutant allele, indicating they cannot represent zygotically transcribed *gtsf-1* mutant mRNA. Rather, these transcripts likely originate from the above-described contamination of the homozygous F1 population with heterozygous animals.

We hypothesized that ERGO-1 branch 22G-RNAs observed in the mutant F1 generation might be competent to maintain target silencing. If this is true, we should observe strong upregulation of ERGO-1 target mRNAs only the mutant F2 generation. Indeed, the X-cluster is upregulated only in the second mutant generation (**Figure III.3A**). When ERGO-1 targets are analyzed in bulk, we observe the same trend, with stronger upregulation only in the mutant F2, consistent with the maternal effect (**Figure III.3B**). Regarding the mutant F1, we note that the very slight, not statistically significant, upregulation of ERGO-1 target mRNAs may account for the slight increase of 22G-RNAs observed in the mutant F1 (in **Figures III.2F** and **III.S2A**), because of an increased number of molecules available to template RdRP activity. ALG-3/4 targets, as for instance *ssp-16*, were found to be upregulated already in the F1 generation (**Figures III.3B** and **III.S3B**), supporting the notion that the maternal rescue of the 26G-RNA pathways is restricted to the ERGO-1 branch.

Eri targets show stronger expression in embryos

ERGO-1 targets comprise a very diverse set of targets consisting of pseudogenes, fast evolving small genes, paralog genes and lncRNAs (Fischer et al., 2011; Newman et al., 2018; Vasale et al., 2010). Considering the maternal effect described above for ERGO-1-dependent sRNAs and corresponding targets, we postulated that this maternal effect may exist to counteract embryonic expression of ERGO-1 targets. To address this we sequenced mRNA of synchronized populations of all developmental stages (L1, L2, L3, L4, young adult and embryos) of both wild-type (N2) and *rrf-3(pk1426)* mutants (**Table III.S1**). In wild-type worms, ERGO-1 targets are most abundant in embryos (**Figure III.3C**, lower panel, in blue). Moreover, the effect of *rrf-3* mutation on ERGO-1 target expression is stronger in embryos (**Figure III.3C**, lower panel). These results indicate that the maternal effect reported above can reflect deposition of factors which are required to initiate silencing of targets early in development. Differential gene expression data and normalized read counts calculated from the sequencing datasets described in **Figures III.2** and **III.3**, can be found in **Tables III.S2** and **III.S3**.

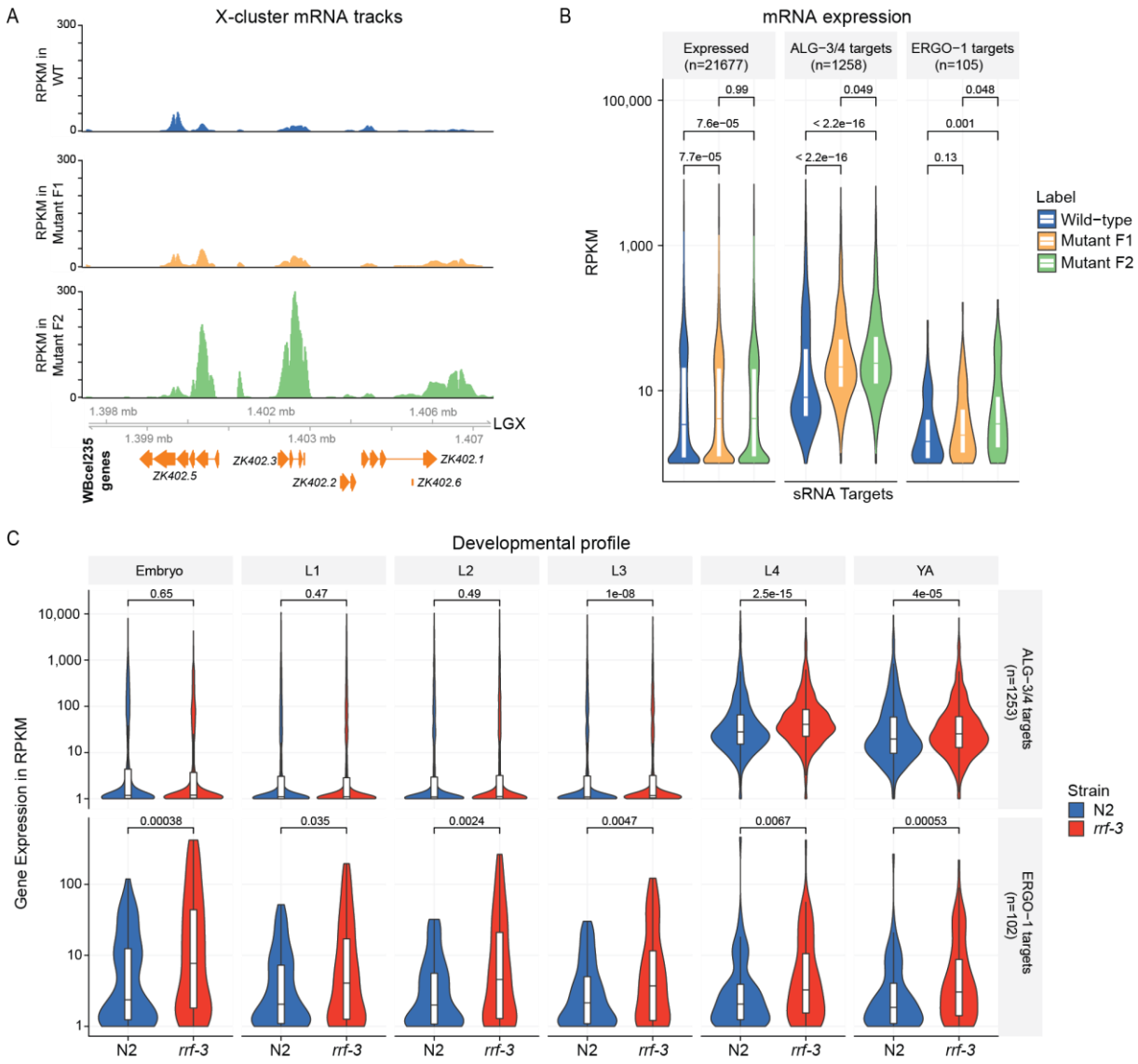


Figure III.3. mRNA dynamics in Eri maternal inheritance. (A) Genome browser tracks showing mRNA, in RPKM (Reads Per Kilobase Million), of X-cluster genes. (B) Distribution of normalized mRNA expression, in RPKM, of all expressed genes, ALG-3/4 targets and ERGO-1 targets in different generations/phenotype. (C) Distribution of normalized mRNA expression, in RPKM, ALG-3/4 targets (upper panel) and ERGO-1 targets (lower panel) throughout development. Expression is shown for wild-type N2 (in blue) and *rrf-3(pk1426)* (in red) animals. YA, young adult. L1-L4, first to fourth larval stages of *C. elegans* development. Violin plots in (B-C) show the distribution density of the underlying data. The top and bottom of the embedded box represent the 75th and the 25th percentile of the distribution, respectively. The line in the box represents the median. P-values were calculated with a two-sided unpaired Mann-Whitney/Wilcoxon rank-sum test.

GTSF-1 is required for sRNA biogenesis and target silencing in adult males

The young adult sequencing datasets we obtained in this study (**Figure III.2A**), as well as previous datasets of gravid adults (**Figures II.4**, **II.S4**, and **Table II.S1**), are not well suited to address ALG-3/4 biology, considering that in these developmental stages ALG-3/4 are not expressed, at least not abundantly. Therefore, in order to further our understanding of the dependency of ALG-3/4 branch sRNAs on GTSF-1, as well as to explore the regulation of ALG-3/4 targets, we generated additional sRNA and mRNA datasets from wild-type and *gtsf-1* male animals grown at 20 °C (**Table III.S1**).

As expected, global 26G-RNA levels are severely affected in *gtsf-1* mutant males, reflecting downregulation of 26G-RNAs from both branches of the pathway (**Figure III.4A-C**). Consistent with the absence of ERGO-1 in adult males, ERGO-1 branch 26G-RNAs are detected in extremely low numbers in wild-type animals (**Figure III.4B**). Global levels of 21U-RNAs seem to be moderately increased (**Figure III.4D**), possibly resulting from the lack of 26G-RNAs in the libraries. Global levels of 22G-RNAs are not affected (**Figure III.4E**), but consistent with a global depletion of 26G-RNAs, 22G-RNAs specifically mapping to ALG-3/4 and ERGO-1 targets are reduced in *gtsf-1* mutant males (**Figure III.4F**). Next, we probed the effects of *gtsf-1* mutation on male gene expression using mRNA sequencing. ALG-3/4 and ERGO-1 targets are both upregulated in *gtsf-1* mutant males (**Figure III.4G**). These changes are illustrated for the X-cluster and *ssp-16* in the genome browser tracks of **Figure III.S4**. **Table III.S4**

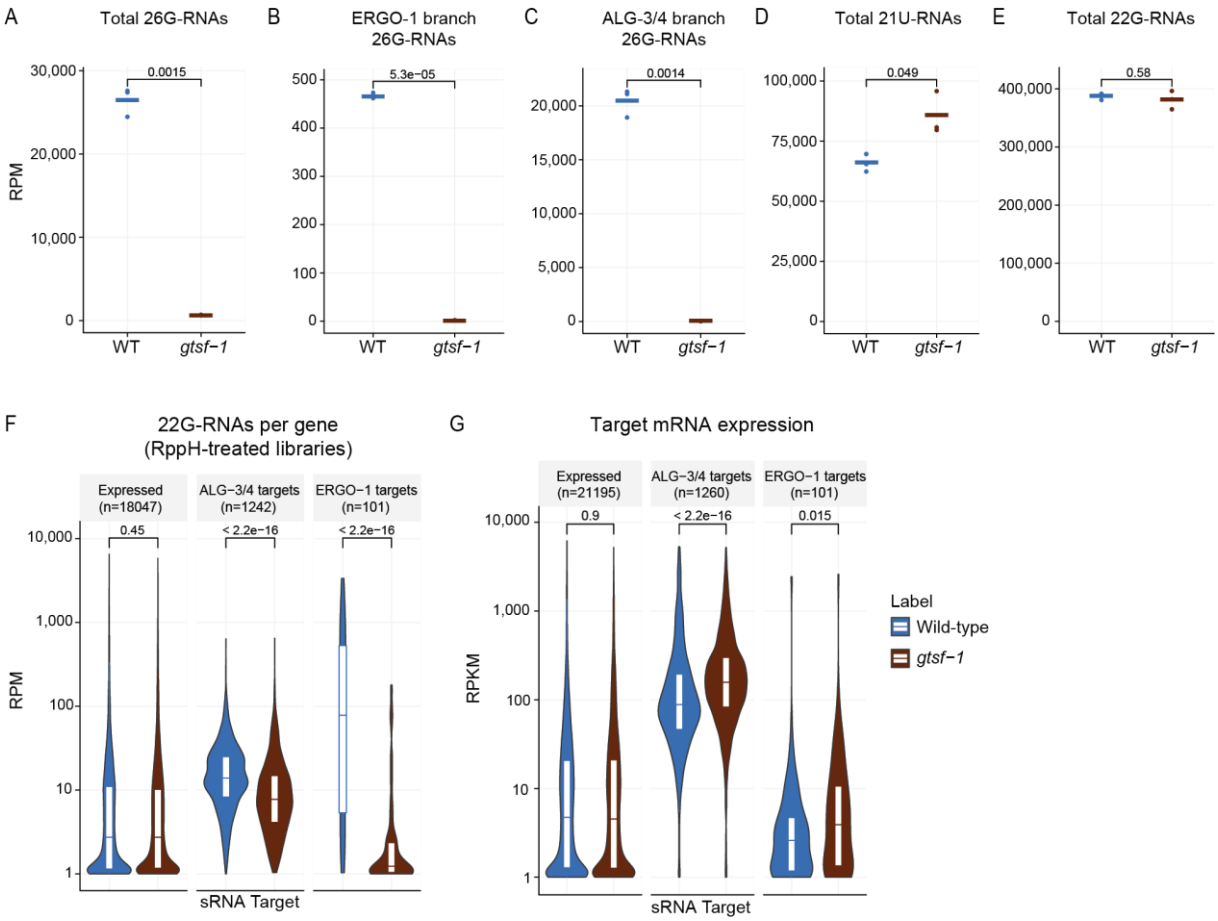


Figure III.4. GTSF-1 is required for sRNA biogenesis and target silencing in adult males. (A-E) Normalized levels of sRNAs in RppH treated libraries, in RPM. Three biological replicates are shown. WT, wild-type. (A) Total levels of 26G-RNAs. (B) 26G-RNAs mapping to ERGO-1 targets. (C) 26G-RNAs mapping to ALG-3/4 targets. (D) total levels of 21U-RNAs. (E) total levels of 22G-RNAs. (F) RPM Levels of sRNAs mapping, per gene, to known targets of ALG-3/4 and ERGO-1. (G) Normalized mRNA expression of ALG-3/4 and ERGO-1 targets, in RPKM. Violin plots in (F-G) show the distribution density of the underlying data. The top and bottom of the embedded box represent the 75th and the 25th percentile of the distribution, respectively. The line in the box represents the median. P-values were calculated either with a two-sided unpaired t-test (A-E), or with a two-sided Mann-Whitney/Wilcoxon rank-sum test (F-G).

includes differential gene expression data and normalized read counts calculated from the sequencing datasets described in **Figure III.4**.

As a final note on the developmental aspects of ALG-3/4 branch, consistent with enrichment in the spermatogenic gonad (Conine et al., 2010, 2013; Gent et al., 2009; Han et al., 2009; Pavelec et al., 2009), ALG-3/4 targets are more highly expressed and more responsive to *rrf-3* mutation in the L4 and young adult stages of hermaphrodite animals (**Figure III.3C**, upper panel). Given that the overall ALG-3/4 target mRNA levels go up upon depletion of *gtsf-1* or *rrf-3* (**Figures III.3C** and **III.4G**), bulk 26G-RNA activity during spermatogenesis seems to be repressive at 20 °C.

We conclude that the activity of GTSF-1 is required in adult males for silencing of 26G-RNA targets by participating in 26G- and 22G-RNA biogenesis.

22G-RNA abundance is a predictor of the regulatory outcome of ALG-3/4 targets

ALG-3/4 were shown to have distinct effects on gene expression, either silencing or licensing (Conine et al., 2010, 2013). However, how these different effects arise is currently unknown. Even though our analysis in males did not reveal a licensing effect of 26G-RNAs, the bulk analysis of targets in **Figures III.3B** and **III.4G** may mask the behavior of distinct target subpopulations. Of note, our sequencing datasets were obtained from animals grown at 20 °C and are therefore blind to the strong positive regulatory effect of ALG-3/4 in gene expression at higher temperatures (Conine et al., 2013).

We reasoned that sRNA abundance may be correlated with different regulatory outcomes. Therefore, we defined ALG-3/4 targets that are upregulated, downregulated, and unaltered upon *gtsf-1* mutation and plotted their 26G-RNA abundance. This reveals a tendency for genes that are

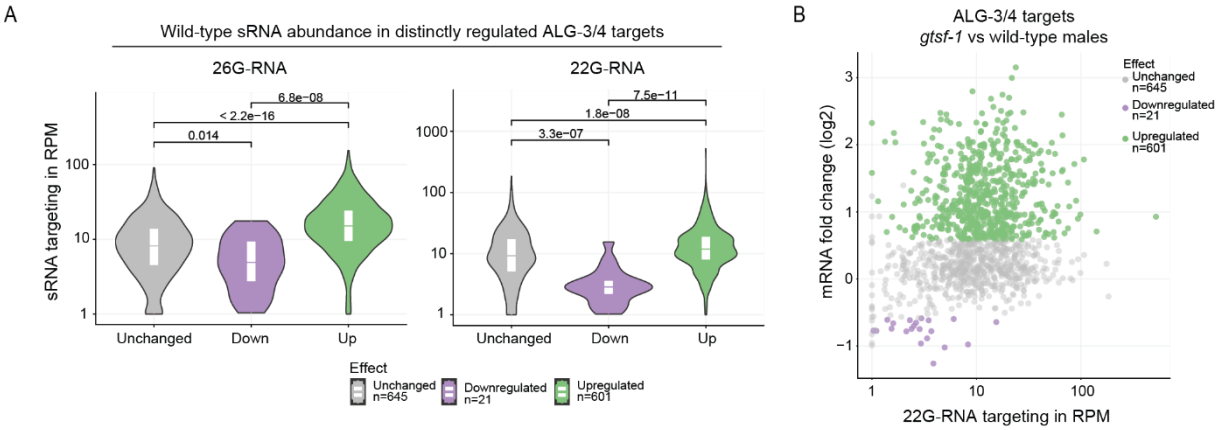


Figure III.5. In adult males, sRNA abundance is a predictor of regulatory outcome by ALG-3/4. (A) Distribution of sRNA levels (26G-RNA on the left panel, 22G-RNA on the right panel) mapping to ALG-3/4 targets that are unchanged, down- or upregulated upon *gtsf-1* mutation. n is indicated below. (B) MA-plot displaying the 22G-RNA levels in respect to regulatory outcome. (B) is another representation of the data shown in the right panel of (A). Violin plots in (A) show the distribution density of the underlying data. The top and bottom of the embedded box represent the 75th and the 25th percentile of the distribution, respectively. The line in the box represents the median. P-values were calculated with a two-sided unpaired Mann-Whitney/Wilcoxon rank-sum test.

upregulated upon loss of GTSF-1 to be more heavily targeted by 26G-RNAs in adult males (**Figure III.5A**, left panel). The same trend is observed for 22G-RNAs: upregulated genes are more heavily covered by 22G-RNAs (**Figure III.5A**, right panel, and **III.5B**). In contrast, ALG-3/4 targets that are downregulated in *gtsf-1* mutant males display a relatively low-level targeting by 22G-RNAs (**Figure III.5A-B**).

We conclude that, in adult males, stronger 26G-RNA targeting promotes stronger 22G-RNA biogenesis and repression of targets, whereas low-level targeting by 26G- and 22G-RNAs does not. Transcripts that are downregulated in absence of GTSF-1 might be licensed for gene expression, but may also respond in a secondary manner to a disturbed 26G-RNA pathway.

ALG-3/4- and ERGO-1-branch 26G-RNA subpopulations display different patterns of origin that influence target expression

It was previously noticed that ALG-3/4-dependent 26G-RNAs mostly map to both the 5' and 3' ends of their targets, and that this may correlate with gene expression changes (Conine et al., 2010). We followed up on this observation by performing metagene analysis of 26G-RNA binding using our broader set of targets. Indeed, ALG-3/4 branch 26G-RNAs display a distinctive pattern with two sharp peaks near the transcription start site (TSS) and transcription end site (TES) (**Figures III.6A** and **III.S5A**, left panels). In contrast, ERGO-1 branch 26G-RNAs map throughout the transcript, with a slight enrichment in the 3' half (**Figure III.6B**, left panel). Contrary to 26G-RNAs, 22G-RNAs from both branches map throughout the transcript (**Figures III.6A-B** and **III.S5A**, right panels). These patterns are consistent with recruitment of RdRPs and production of antisense sRNAs along the full length of the transcript. These findings suggest substantially different regulation modes by ERGO-1- and ALG-3/4-branch 26G-RNAs.

Conine and colleagues reported a correlation between 26G-RNA 5' targeting and negative regulation (Conine et al., 2010). We wanted to address whether our datasets show concrete correlations between the patterns of origin of ALG-3/4-dependent 26G-RNAs and distinct regulatory outcomes. To address this, we ranked genes by 5' and 3' abundance of 26G-RNAs, selected genes predominantly targeted at the 5' or at the 3' ends and plotted their fold change upon *gtsf-1* mutation. In adult males, dominant 5' targeting by 26G-RNAs seems to be correlated with gene silencing (fold change >0 in the mutant, **Figure III.6C**), whereas dominant 3' targeting is accompanied with only weak upregulation and in some cases very mild downregulation (**Figure III.6C-D**). In further support for a non-gene silencing, and potentially licensing role for ALG-3/4 targeting at the 3' end, genes with predominant 3' 26G-RNAs display an overall higher expression than genes predominantly targeted at the 5' region (**Figure III.6E**). The same signatures are found in young adults, with an even stronger signature of the 3' in promoting gene expression (**Figure III.S5B-D**).

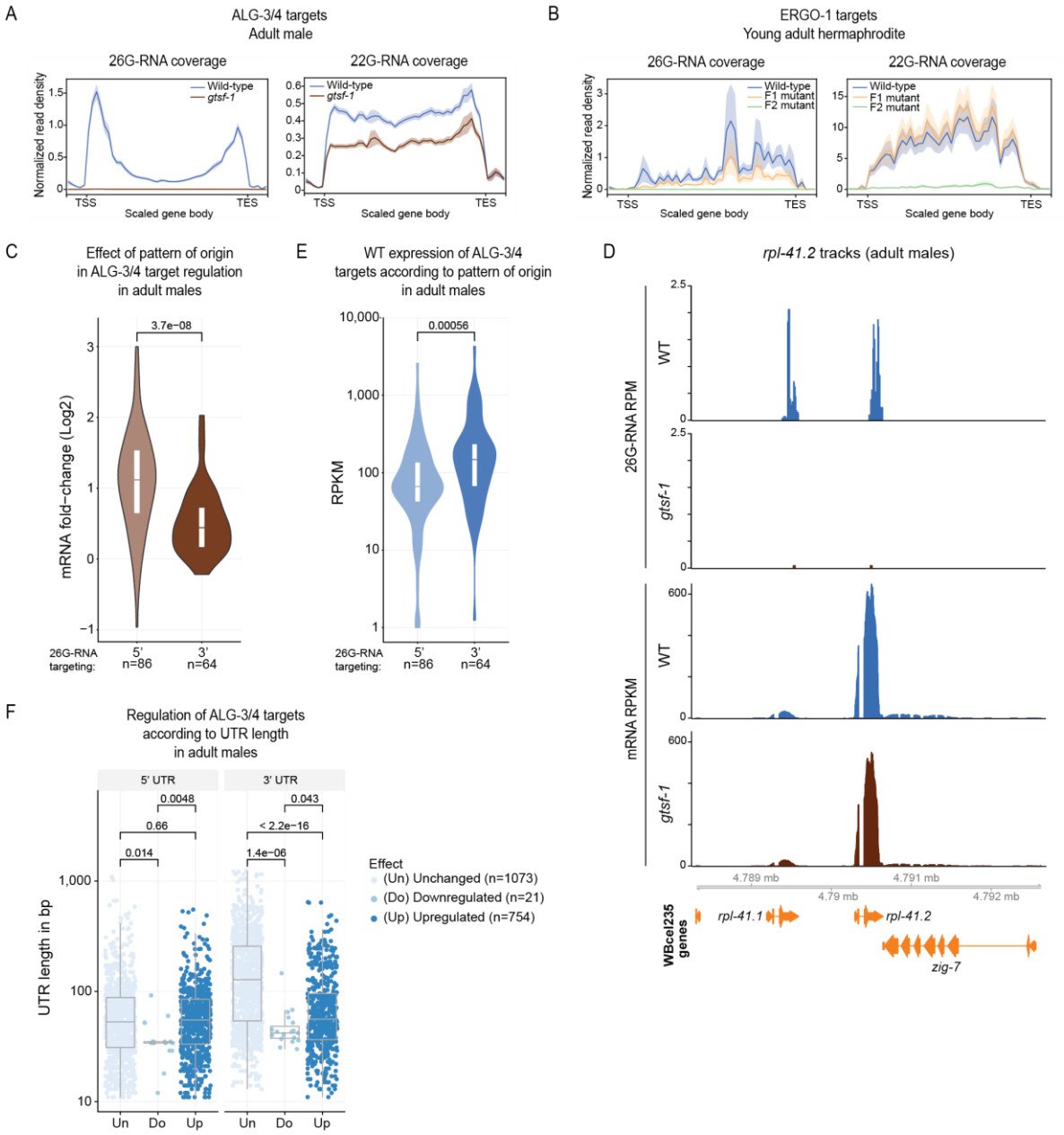


Figure III.6. Predictors of regulatory outcome by ALG-3/4 in males and ERGO-1 branch sRNA metagene analysis. (A-B) Metagene analysis of 26G- (left panel) and 22G-RNAs (right panel) mapping to ALG-3/4 targets (n=1258) in male datasets (A), and to ERGO-1 targets (n=104) in young adult datasets (B), from our maternal effect setup (as in **Figure III.2A**). Target gene body length was scaled between transcription start site (TSS) and transcription end site (TES). Moreover, the regions comprising 250 nucleotides immediately upstream of the TSS and downstream of the TES are also included. (C) Regulation of ALG-3/4 target genes predominantly targeted at the 5' or at the 3' by 26G-RNAs. (D) Genome browser tracks displaying the RPM levels of 26G-RNAs (upper panels) and RPKM mRNA levels (lower panels) mapping to *rpl-41.2*, a gene predominantly targeted by 26G-RNAs at the 3' region, in adult males. WT, wild-type. (E) Wild-type (WT) expression levels, in RPKM, of ALG-3/4 target genes predominantly targeted at the 5' or at the 3' by 26G-RNAs. (F) 5' UTR (left panel) and 3' UTR (right panel) lengths of all the transcript isoforms annotated for ALG-3/4 target genes, according to effect on gene expression. For (C-F) we used male sequencing datasets. Violin plots in (C-D) and the boxplot in (E) show the distribution of the data. The top and bottom of the embedded boxes represent the 75th and the 25th percentile of the distribution, respectively. The line in the box represents the median. P-values were calculated with a two-sided unpaired Mann-Whitney/Wilcoxon rank-sum test.

Finally, we interrogated if the length of 5' and 3' UTRs may be a predictor of regulatory outcome in ALG-3/4 targets. 5' UTR length was not significantly different between unchanged and upregulated genes (**Figure III.6F**, left panel). Downregulated genes have a statistically significant shorter 5' UTR length, but these results should be interpreted with caution due to the low number of transcript isoforms analyzed. In contrast, 3' UTR length is significantly smaller in targets that respond to loss of GTSF-1 in males (**Figure III.6F**, right panel). Interestingly, we find the same and possibly even stronger relation between 3'UTR length and responsiveness to GTSF-1 status in young adult animals (**Figure III.S5E**). Upregulated ERGO-1 targets do not display significantly shorter 3' UTRs (**Figure III.S5F**), indicating that the trend is specific to ALG-3/4 targets.

Altogether, our results suggest that, both in males and young adult hermaphrodites, 3' vs 5' targeting and 3' UTR length are predictors of whether ALG-3/4 targets are silenced or not.

ALG-3 and ALG-4 act in a negative feedback loop

While navigating the lists of GTSF-1 targets defined by differential gene expression analysis, we noticed that *alg-3* and *alg-4* are targets of 26G-RNAs (this study and in **Table II.S1**). These 26G-RNAs are sensitive to oxidation (not enriched in oxidized libraries, see **Table II.S1**) and map predominantly to the extremities of the transcript (**Figure III.7**, upper panels), indicating that these 26G-RNAs share features with ALG-3/4 branch 26G-RNAs. In addition to these 26G-RNAs, significant amounts of 22G-RNAs are found on *alg-3/4* (**Figure III.7**, middle panels). These sRNAs seem to silence gene expression, since mRNA-seq shows that *alg-3* and *alg-4* transcripts are 2-3 fold upregulated in *gtsf-1* mutants (**Figure III.7**, lower panels).

These results strongly suggest that *alg-3/4* are regulating their own expression in a negative feedback loop. Of note, the upregulation of *alg-3* and *alg-4* is in agreement with the results presented above, because these genes are more heavily targeted by 26G-RNAs at their 5' (although *alg-4* also has a sharp 3' 26G-RNA peak, upper panels). Furthermore, these same signatures of negative feedback loop are observed in young adults (**Figure III.S6**).

Discussion

Genetic dissection of a maternal rescue

Animal male and female gametes are rich in RNA. Upon fertilization, several RNA species are thus provided to the zygote. Multiple lines of evidence from several distinct organisms indicate that sRNAs are included in the parental repertoire of inherited RNA. For example, piRNAs have been reported to be maternally deposited in embryos in arthropods, fish and *C. elegans* (Akkouche et al., 2013; Brennecke et al., 2007, 2008; de Albuquerque et al., 2015; Houwing et al., 2007; Kawaoka et al.,

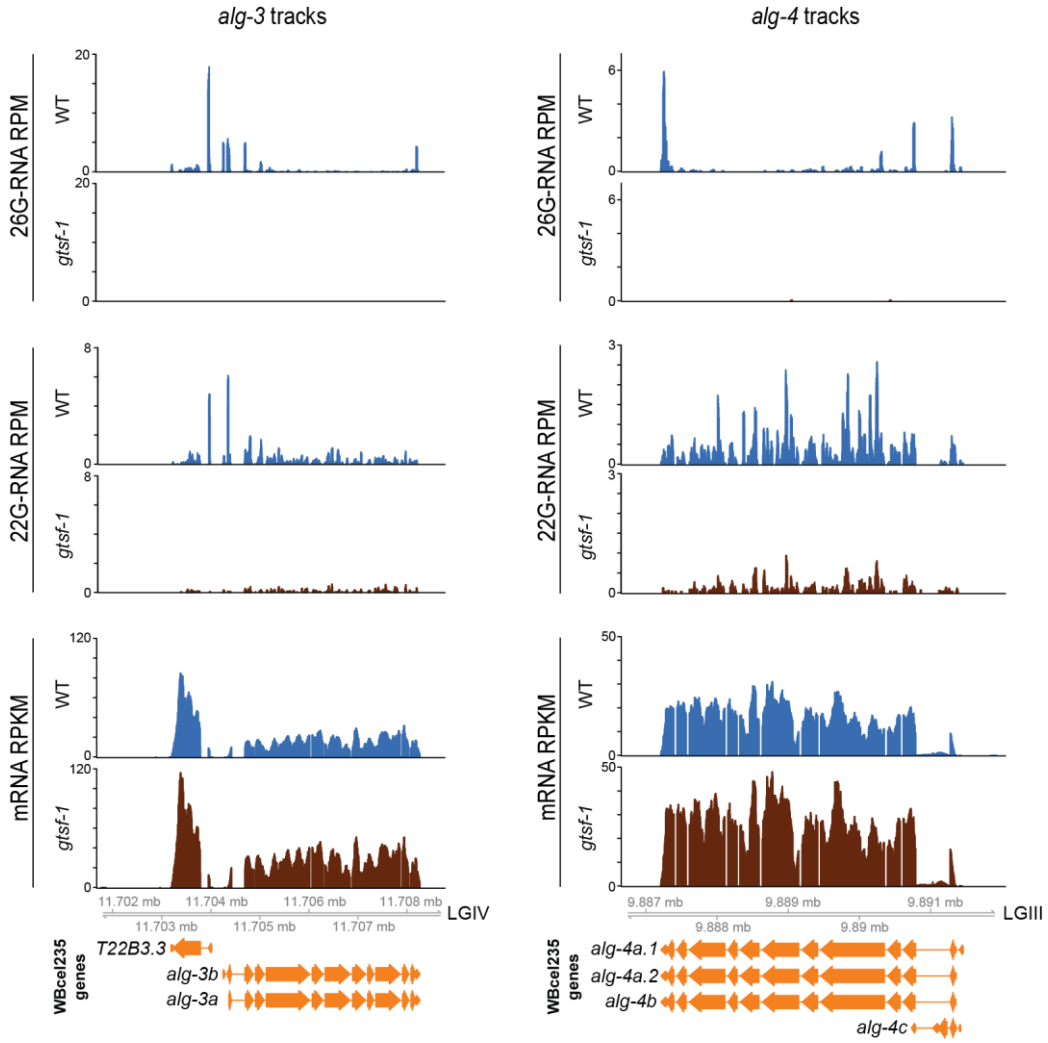


Figure III.7. ALG-3 and ALG-4 are engaged in a negative feedback loop in males. Genome browser tracks showing 26G-RNAs (upper panels) and 22G-RNAs (middle panels) mapping to *alg-3* (left panels) and *alg-4* (right panels), in RPM. Lower panels show the RPKM mRNA levels of *alg-3* (on the left) and *alg-4* (on the right). Sequencing datasets of adult males were used. WT, wild-type.

2011; Le Thomas et al., 2014; Le Thomas et al., 2014; Luteijn et al., 2012; Ninova et al., 2017; Phillips et al., 2015; de Vanssay et al., 2012). In *C. elegans* other endogenous sRNA populations have also been shown to be contributed by the gametes: 1) 26G-RNAs have been shown to be weakly provided by the male, while 22G-RNAs are more abundantly provided (Stoeckius et al., 2014); 2) 26G-RNAs and the AGO ERGO-1 are co-expressed during oogenesis and in embryos (Han et al., 2009; Stoeckius et al., 2009; Vasale et al., 2010); and 3) 22G-RNAs are deposited in embryos via the mother and participate in transgenerational gene silencing (Gu et al., 2009; Luteijn et al., 2012; Minkina and Hunter, 2017; Schott et al., 2014; Wan et al., 2018; Xu et al., 2018).

We describe a maternal effect in the transmission of the Eri phenotype and 22G sensor derepression and characterize the subjacent dynamics of sRNAs and mRNA targets (**Figures III.1-III.3** and **III.S1-III.S3**). We show that both maternal and zygotic 26G-RNAs are sufficient for silencing. Absence of either the maternal or the zygotic pools can thus be compensated, enhancing the

robustness of this system. We note, however, that sufficiency has only been tested with the described 22G sensor. It may be that the silencing of other targets has differential dependencies on maternal and zygotic 26G-RNA populations. The maternal effect was observed in mutants of a variety of Eri genes, including *gtsf-1*, *rrf-3* and *ergo-1*, but not *alg-3/4*. Therefore, these defects are related to impairment of sRNA populations directly associated with and downstream of ERGO-1. These results do not exclude a parental effect for ALG-3/4. In fact, a paternal effect on embryogenesis has been described for *rrf-3* mutants (Gent et al., 2009). Such phenotype most likely arises due to ALG-3/4 branch sRNAs.

Maternal rescue of Eri genes was previously reported (Zhuang and Hunter, 2011), although the genetic basis for this phenomenon was not characterized further. We demonstrate that in the first Eri mutant generation, primary 26G-RNAs are downregulated, while their downstream 22G-RNAs are still present (**Figure III.2**). These ERGO-1-dependent 22G-RNAs, maintained in the absence of their primary triggers, seem to be competent to sustain silencing of ERGO-1 targets throughout life of the animal (**Figure III.3**). Given that 1) ERGO-1 targets display higher expression during embryogenesis; and 2) upon disruption of endogenous RNAi by *rrf-3* mutation, targets become upregulated in all developmental stages (**Figure III.3C**); maternally deposited ERGO-1-dependent factors may be especially required to initiate target silencing during embryogenesis, and to prevent spurious expression throughout development. The ERGO-1-independent maintenance of this silencing response may be mechanistically similar to RNA-induced epigenetic silencing (RNAe), involving a self-perpetuating population of 22G-RNAs (Ashe et al., 2012; Luteijn et al., 2012; Shirayama et al., 2012). Indeed, both processes depend on a nuclear AGO protein: HRDE-1 in RNAe (Ashe et al., 2012; Luteijn et al., 2012; Shirayama et al., 2012) and NRDE-3 for ERGO-1-driven silencing (Guang et al., 2008; Montgomery et al., 2012; Zhou et al., 2014). Self-perpetuating 22G-RNA signals may be also in place in the male germline (see below).

Our genetic experiments and sequencing data are fully consistent with maternal inheritance of 26G-RNAs. However, these may not be the only inherited agent. A non-mutually exclusive idea is that GTSF-1, as well as other ERI complex proteins may be deposited in embryos to initiate production of zygotic sRNAs. In accordance with the latter, we have previously demonstrated that formation of the 26G-RNA-generating ERI complex is developmentally regulated (**Figures II.7** and **II.S6**). While in young adults there is a comparable amount of pre- and mature ERI complex, in embryos there is proportionally more mature ERI complex. These observations suggest that the pre-ERI complex might be deposited in the embryo to swiftly jumpstart zygotic 26G-RNA expression after fertilization.

26G-RNAs act as triggers to induce a self-sustained 22G-RNA-driven silencing in the male germline

We show that GTSF-1 is required in adult males, potentially in the male germline, to produce 26G- and downstream 22G-RNAs (**Figure III.4**) analogous to its role in the hermaphrodite germline and in embryos (as shown in **Figures II.4** and **II.S4**). In addition, the bulk of targets from both 26G-RNA pathway branches seem to be deregulated. Interestingly, we note that although ERGO-1 and its cognate 26G-RNAs are not abundantly expressed in spermatogenic tissues (**Figure III.4B**), *gtsf-1*-dependent, secondary 22G-RNAs mapping to these genes maintain gene silencing in the adult males (**Figure III.4F-G**). In an analogous manner, we find that ALG-3/4 targets maintain 22G-RNAs in gravid adults, even though ALG-3/4 are not expressed at that stage. Mechanistically this may be closely related to how maternal 26G-RNAs can trigger 22G-RNA-driven silencing (see above). NRDE-3 is downstream of ERGO-1 and is likely to silence ERGO-1 targets throughout development. However, the AGOs associated with 22G-RNAs mapping to 1) ERGO-1 targets in the male, and 2) to ALG-3/4 targets in gravid adults have not yet been identified.

Two distinct ALG-3/4 regulatory mechanisms?

ALG-3/4-branch 26G-RNAs map very sharply to the 5’ and 3’ extremities of the targets, very close to the transcription start and end sites. We find that stronger targeting at the 3’ end does not drive robust gene silencing, and may even license expression, while targeting at the 5’ end is associated with stronger gene silencing. Targeting at the 3’ is consistent with RdRP recruitment to synthesize antisense secondary 22G-RNAs throughout the transcript. These may associate with CSR-1 and could have a positive effect on gene expression. The sharp 5’ peak in the metagene analysis could hint at additional regulatory modes, other than 22G-RNA targeting. 5’-end-bound ALG-3/4 could recruit other effector factors, which promote RNA decay or translation inhibition, e.g. by inhibiting the assembly of ribosomes. Of note, when single targets are considered individually, 26G-RNA peaks at 5’ and 3’ can be simultaneously detected (**Figures III.7**, **III.S2B** and **III.S5A**, left panels and **III.S6**). Hence, the resolution of a balance between AGO-sRNA complexes binding at 5’ and 3’ could determine regulatory outcome. Notably, we find shorter 3’ UTRs to be correlated with gene silencing (**Figure III.6F**). In a model where predominant 3’ UTR targeting by AGO-sRNA complexes promotes gene expression, shorter 3’ UTRs and therefore less chance of sRNA binding may shift the balance towards gene silencing. Another possibility may be that longer 3’ UTRs contain binding sites for additional RNA binding proteins that may help to restrict RdRP activity on the transcript in question. Further work will be needed to test such ideas.

An AGO negative feedback loop

In *C. elegans*, primary sRNAs trigger the production of abundant secondary sRNAs. If left uncontrolled, such amplification mechanisms can be detrimental to biological systems. Endogenous and exogenous RNAi pathways in *C. elegans* compete for limiting shared factors and the Eri phenotype is a result of such competition (Duchaine et al., 2006; Lee et al., 2006). Competition for shared factors is in itself a mechanism to limit accumulation of sRNAs. In support of this, exogenous RNAi was shown to affect endogenous sRNA populations, thus restricting the generations over which RNAi effects can be inherited (Houri-Ze'evi et al., 2016).

We find that 26G-RNAs, likely ALG-3/4-bound, as well as 22G-RNAs map to *alg-3* and *alg-4* mRNAs (**Figures III.7** and **III.S6**). In the absence of GTSF-1, a loss of these sRNAs is accompanied by a 2-3 fold upregulation of *alg-3* and *alg-4* on the mRNA level. This means that ALG-3 and ALG-4 may regulate their own expression. In the future, the retrieval of *alg-3* and *alg-4* mRNAs, as well as of 26G-RNAs complementary to their sequence, in immunoprecipitations of ALG-3 or ALG-4 will strongly support this regulatory loop.

Such regulation is not unprecedented. Complementary endo-siRNAs to *ago2* have been described in *Drosophila* S2 cells (Ghildiyal et al., 2008). Since AGO2 is required for the biogenesis and silencing function of endo-siRNAs, it is likely that Ago2 regulates itself in S2 cells. In addition, other studies in *C. elegans* have described cases where sRNAs are regulating the expression of RNAi factors (Gerson-Gurwitz et al., 2016; Houri-Ze'evi et al., 2016; Maniar and Fire, 2011).

Such direct self-regulation of AGO genes may constitute an important mechanism to limit RNAi-related responses, but the biological relevance of this regulation will need to be addressed experimentally. These observations do suggest that the Eri phenotype is but one manifestation of intricate cross-regulation governing the RNAi pathways of *C. elegans*.

Materials and Methods

C. elegans genetics and culture

C. elegans was cultured on OP50 bacteria according to standard laboratory conditions (Brenner, 1974). Unless otherwise noted, worms were grown at 20 ^{o}C. The Bristol strain N2 was used as the standard wild-type strain. All strains used and created in this study are listed in **Table III.S5**.

Microscopy

Wide-field photomicrographs were acquired using a Leica M165FC microscope with a Leica DFC450 C camera, and were processed using Leica LAS software and ImageJ.

Genetic crosses using *dpy-4;gtsf-1* worms

Cross outline. We first linked *gtsf-1(xf43)* and *dpy-4(e1166)*. These genes are 2.62 cM apart, which does not comprise extremely tight linkage. Therefore, throughout the outcrossing scheme, worms were consistently genotyped for *gtsf-1* and phenotyped for *dpy-4*. We started by outcrossing *dpy-4;gtsf-1* hermaphrodites with N2 males (in a 1:2 ratio). *dpy-4(e1166)* is reported as being weakly semi-dominant (https://cgc.umn.edu/strain/CB1166). Indeed, heterozygote worms look only very slightly Dpy, therefore for simplicity, we refer to the heterozygote phenotype as "wild-type" throughout this work. Wild-type looking worms were selected in the F1 and F2 generations. The F2s were allowed to lay embryos for 1-2 days and then were genotyped for *gtsf-1(xf43)* using PCR. Progenies of non-recombined *gtsf-1* heterozygote worms were kept for follow up. F3 progenies that did not segregate *dpy* worms were discarded. F3 *dpys* were isolated, allowed to lay embryos, and genotyped for *gtsf-1(xf43)*. Progenies of non-homozygote mutant *gtsf-1(xf43)* worms were discarded.

RNAi. dsRNA against *lir-1* was supplemented to worms by feeding as described (Kamath et al., 2003). L1 worms were transferred to RNAi plates and larval arrest was scored 2-3 days later. L1 F3 and F4 worms were transferred to RNAi plates blinded to genotype/phenotype (the *dpy* phenotype only shows clearly from L3 onwards).

Temperature-sensitive sterility assay. Single L1 F3 and F4 worms were transferred to OP50 plates, blinded to genotype/phenotype and grown at 25 ^{o}C (the *dpy* phenotype only shows clearly from L3 onwards). Temperature sensitive-sterility was scored on the second day of adulthood and worms with unexpected genotype-phenotype were genotyped for *gtsf-1*.

Animal collection for RNA isolation. Approximately 550 hand-picked wild-type, Dpy F3 (referred in the text as mutant F1) and Dpy F4 (referred in the text as mutant F2) animals were used to isolate RNA (see cross description above, schematics in **Figure III.2A**, and see below for RNA isolation protocol). Four independent outcrosses were performed and independent biological replicates (of wild-type, mutant F1 and mutant F2) were collected from each. Each sample was used to prepare small RNA and mRNA libraries (see below for details on library preparation). Plates with the hand-picked worms were rinsed and washed 4-6 times with M9 supplemented with 0.01% Tween. 50 µL of M9 plus worms were subsequently frozen in dry ice.

Growth and collection of adult males

him-5(e1467) and *him-5(e1467); gtsf-1(xf43)* worm populations were synchronized by bleaching, overnight hatching in M9 and plated on OP50 plates the next day. Worms were grown until adulthood for approximately 73 hours and 400-500 male animals were hand-picked for each sample, in biological triplicates, and used to isolate RNA (see below for RNA isolation protocol). Each sample was used to prepare small RNA and mRNA libraries (see below details on library preparation). Plates

with the hand-picked worms were rinsed and washed 4-6 times with M9 supplemented with 0.01% Tween. 50 µL of M9 plus worms were subsequently frozen in dry ice.

Growth and collection of N2 and *rrf-3* worms

N2 and *rrf-3(pk1426)* animal populations were synchronized by bleaching, overnight hatching in M9 and plated on OP50 plates the next day. L1 animals were allowed to recover from starvation for 5 hours, and then were collected. L2 worms were collected 11 hours after plating. L3 animals were collected 28 hours after plating. L4 animals were collected 50 hours after plating, and young adults were collected 56 hours after plating. Animals were rinsed off plates and washed 4-6 times with M9 supplemented with 0.01% Tween. 50 µL of M9 plus worms were subsequently frozen in dry ice. Embryo samples were collected from bleached gravid adult animals, followed by thorough washes with M9. Samples were collected in triplicate and RNA isolation proceeded as described below.

RNA isolation

Worm aliquots were thawed and 500 µL of Trizol LS (Life Technologies, 10296-028) was added and mixed vigorously. Next, we employed six freeze-thaw cycles to dissolve the worms: tubes were frozen in liquid nitrogen for 30 seconds, thawed in a 37 ^{o}C water bath for 2 minutes, and mixed vigorously. Following the sixth freeze-thaw cycle, 1 volume of 100% ethanol was added to the samples and mixed vigorously. Then, we added these mixtures onto Direct-zol columns (Zymo Research, R2070) and manufacturer's instructions were followed (in-column DNase I treatment was included).

Library preparation for mRNA sequencing

NGS library prep was performed with Illumina's TruSeq stranded mRNA LT Sample Prep Kit following Illumina's standard protocol (Part # 15031047 Rev. E). Starting amounts of RNA used for library preparation, as well as the number of PCR cycles used in amplification, are indicated in **Table III.S6**. 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). Number of pooled samples, Flowcell, type of run and number of cycles used in the different experiments are all indicated in **Table III.S6**.

RppH treatment and library preparation for small RNA sequencing

For maternal effect sequencing, RNA was directly used for library preparation, or treated with RppH prior to library preparation. RppH treatment was performed as described in (Almeida et al., 2019) with slight modifications. In short, 500 ng of RNA were incubated with 5 units of RppH and 10x NEB Buffer 2 for 1 hour at 37^{o}C. Reaction was stopped by incubating the samples with 500 mM EDTA for 5

minutes at 65°C. RNA was reprecipitated in 100% Isopropanol and ressuspended in nuclease-free water. 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). Both directly cloned and RppH-treated libraries were prepared with a starting amount of 200ng and amplified in 16 PCR cycles. Amplified libraries were purified by running an 8% TBE gel and size-selected for 18 – 40 nts. 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). All 24 samples were pooled in equimolar ratio and sequenced on 1 NextSeq 500/550 High-Output Flowcell, SR for 1x 75 cycles plus 6 cycles for the index read.

RNA from adult males was RppH-treated as described above with the difference that 800 ng of RNA were used for RppH treatment. Library preparation of these samples was performed exactly as described above with the following modifications: starting amount of 460 ng; and amplification in 15 PCR cycles.

Bioinformatic analysis

A summary of the sequencing output can be found in **Table III.S1**.

Small RNA read processing and mapping. Illumina adapters were removed with cutadapt v1.9 (Martin, 2011) (-a TGGAATTCTCGGGTGCCAAGG -O 5 -m 26 -M 38) and reads with low-quality calls were filtered out with fastq_quality_filter (-q 20 -p 100 -Q 33) from the FASTX-Toolkit v0.0.14. Using information from unique molecule identifiers (UMIs) added during library preparation, reads with the same sequence (including UMIs) were collapsed to remove putative PCR duplicates using a custom script. Prior to mapping, UMIs were trimmed (seqtk trimfq -b 4 -e 4) and reads shorter than 15 nucleotides (nts) were discarded (seqtk seq -L 15). Library quality was assessed with FastQC twice, for the raw and for the processed reads. Processed reads were aligned against the *C. elegans* genome assembly WBcel235 with bowtie v0.12.8 (Langmead et al., 2009) (–tryhard –best –strata -v 0 -M 1). Reads mapping to structural genes were filtered out (r/t/s/sn/snoRNA) using Bedtools 2.25.0 (Quinlan and Hall, 2010) (bedtools intersect -v -s -f 0.9) and further analysis was performed using non-structural RNAs.

Small RNA class definition and quantification. Gene annotation was retrieved from Ensembl (release-38). Transposon coordinates were retrieved from wormbase (PRJNA13758.WS264) and added to the ensembl gene annotation to create a custom annotation used for further analysis. To define RNAs as belonging to particular classes of small RNA, mapped reads were categorized as follows: 21U-RNAs (piRNAs) are considered those sequences that are 21 nt long, and map sense to annotated piRNA loci; 22G-RNAs are those whose sequence is exactly 20-23 nts, have a guanine at their 5’ and map antisense to annotated protein-coding/pseudogenes/lincRNA/transposons; 26G-RNAs, are those

which are 26 nt, and map antisense to annotated protein-coding/pseudogenes/lincRNA. Read filtering was done with a python script available at https://github.com/adomingues/filterReads/blob/master/filterReads/filterSmallRNAclasses.py which relies on pysam v0.8.1 an htslib wrapper (Li et al., 2009), in combination with Bedtools intersect. Reads fulfilling these definitions were then counted for each library (total levels). Genome browser tracks were created using Bedtools (genomeCoverageBed -bg -split -scale -ibam -g), to summarize genome coverage normalized to mapped reads * 1 million (Reads Per Million or RPM), followed by bedGraphToBigWig to create the bigwig track. To quantify the effects on small RNAs of particular branches/pathways, we collected lists of genes previously identified as being targeted by these pathways: CSR-1 (Conine et al., 2013); NRDE-3 (Zhou et al., 2014); Mutators (Phillips et al., 2014); and WAGO-1 (Gu et al., 2009). ERGO-1 targets were defined as genes that lose oxidation-resistant 26G-RNAs (that are 3' 2'-O-methylated) upon *gtsf-1* mutation (**Table II.S1**, sheet 1.2). ALG-3/4 targets are defined as genes that lose 26G-RNAs upon *gtsf-1* mutation (**Table II.S1**, sheet 1.1), excluding ERGO-1 targets. The genomic locations of 22G- and 26G-RNAs was then intersected with that of the genes, and counted for each library.

mRNA read processing and mapping. Library quality was assessed with FastQC before being aligned against the *C. elegans* genome assembly WBcel235 and a custom GTF, which included transposon coordinates (described above) with STAR v2.5.2b (–runMode alignReads –outSAMattributes Standard –outSJfilterReads Unique –outSAMunmapped Within –outReadsUnmapped None –outFilterMismatchNmax 2 –outFilterMultimapNmax 10 –alignIntronMin 21 –sjdbOverhang 79). Reads mapping to annotated features in the custom GTF were counted with subread featureCounts v1.5.1 (-s 2 -p -F GTF –donotsort -t exon -g gene_id). Coverage tracks were generated with deepTools v2.4.3 (bamCoverage –smoothLength 60 –binSize 20 –normalizeUsingRPKM) (Ramírez et al., 2016).

Differential expression/small RNA targeting. Reads mapping to annotated features in the custom GTF were counted with htseq-count v0.9.0 (Anders et al., 2014)(htseq-count -f bam -m intersection-nonempty -s reverse) for sRNA-seq data, and with subread featureCounts v1.5.1 (Liao et al., 2014) (-s 2 -p -F GTF –donotsort -t exon -g gene_id) for mRNA-seq. Pairwise differential expression comparisons were performed with DESeq2 v.1.18.1 (Love et al., 2014). For the selection of genes differentially targeted (sRNA) or expressed (mRNA), a p-adjusted value of less than 0.1 and cut-off of at least 1.5 fold-change difference between conditions was applied. As previously reported (see **Chapter II**, **Supplementary experimental procedures**), due to the observed global depletion of 26G-RNA reads in some samples (sRNA), DESeq2 library sizes computed from all 18-30 nt reads in each sample were used for 26G-RNA differential analyses. Gene expression in RPKM (Reads Per Kilobase Million) was calculated by retrieving the fragments/counts per million mapped fragments from the

DEseq2 object (fpm(object, robust = TRUE)) and normalizing to gene length. Genes with a replicate average RPKM or RPM of more than zero were considered as expressed.

Metagene analysis. The average coverage at each gene from a particular branch was determined with deepTools v2.4.3 (computeMatrix scale-regions –metagene –missingDataAsZero –b 250 -a 250 –regionBodyLength 2000 –binSize 50 –averageTypeBins median), using the transcript locations of each gene, and plotted with plotProfile –plotType se –averageType mean –perGroup.

UTR targeting by 26G-RNAs. To identify genes predominantly targeted at their 5' or 3', coverage values of scaled genes were obtained with deepTools, as done for the metagene analysis (see above), with the difference that only the WT track was used, and options -a and -b were set to 0. That is, only the scaled body regions were used. 5' and 3' sRNA targeting was defined for each gene based on the coverage at the first or last 25% of the scaled gene body. The genes were then classified in low, medium or high targeting if they were in the 0-25, 25-75, or 75-100 percentile of the sRNA coverage distribution for either the 5' or the 3'. Primarily 5' or 3' targeted genes were further defined if they were in the 5' high and 3' low category (5' targeted), or high in the 3' and low in the 5' (3' targeted).

Accession numbers

All sequencing data has been submitted to SRA, accession number SRP166194.

Acknowledgements

We thank all the members of the Ketting lab for great help and discussion. A special thanks to Yasmin el Sherif and Svenja Hellmann for excellent technical assistance, and to Jan Schreier for producing the *mut-16(xf142)* mutant. The authors are grateful to Hanna Lukas, Clara Werner, and Maria Mendez-Lago of the IMB genomics core facility for library preparation. We thank the IMB Media Lab for consumables. We also acknowledge the *Caenorhabditis* Genetics Center (CGC) for providing worm strains.

References

Akkouche, A., Grentzinger, T., Fablet, M., Armenise, C., Burlet, N., Braman, V., Chambeyron, S., and Vieira, C. (2013). Maternally deposited germline piRNAs silence the tirant retrotransposon in somatic cells. EMBO Rep. *14*, 458–464.

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

Ambros, V., Lee, R.C., Lavanway, A., Williams, P.T., and Jewell, D. (2003). MicroRNAs and Other Tiny Endogenous RNAs in C. elegans. Curr. Biol. *13*, 807–818.

Anders, S., Pyl, P.T., and Huber, W. (2014). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics btu638.

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.

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.

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.

Borges, F., and Martienssen, R.A. (2015). The expanding world of small RNAs in plants. Nat. Rev. Mol. Cell Biol. *16*, 727–741.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. Cell *128*, 1089–1103.

Brennecke, J., Malone, C.D., Aravin, A.A., Sachidanandam, R., Stark, A., and Hannon, G.J. (2008). An Epigenetic Role for Maternally Inherited piRNAs in Transposon Silencing. Science *322*, 1387–1392.

Brenner, S. (1974). The Genetics of Caenorhabditis Elegans. Genetics *77*, 71–94.

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.

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

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.

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., and Ketting, R.F. (2015). Maternal piRNAs Are Essential for Germline Development following De Novo Establishment of Endo-siRNAs in Caenorhabditis elegans. Dev. Cell *34*, 448–456.

Duchaine, T.F., Wohlschlegel, J.A., Kennedy, S., Bei, Y., Conte Jr., 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.

Ernst, C., Odom, D.T., and Kutter, C. (2017). The emergence of piRNAs against transposon invasion to preserve mammalian genome integrity. Nat. Commun. *8*, 1411.

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, 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*.

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.

Gu, W., Shirayama, M., Conte Jr., 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.

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.

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

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. *106*, 18674–18679.

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

Houri-Ze'evi, L., Korem, Y., Sheftel, H., Faigenbloom, L., Toker, I.A., Dagan, Y., Awad, L., Degani, L., Alon, U., and Rechavi, O. (2016). A Tunable Mechanism Determines the Duration of the Transgenerational Small RNA Inheritance in C. elegans. Cell *165*, 88–99.

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., Fejes Tóth, K., and Aravin, A.A. (2017). piRNA Biogenesis in Drosophila melanogaster. Trends Genet. *33*, 882–894.

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., 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.

Kawaoka, S., Arai, Y., Kadota, K., Suzuki, Y., Hara, K., Sugano, S., Shimizu, K., Tomari, Y., Shimada, T., and Katsuma, S. (2011). Zygotic amplification of secondary piRNAs during silkworm embryogenesis. RNA.

Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. Nat. Rev. Mol. Cell Biol. *10*, 126–139.

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.

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

Lee, R.C., Hammell, C.M., and Ambros, V. (2006). Interacting endogenous and exogenous RNAi pathways in Caenorhabditis elegans. RNA *12*, 589–597.

Le Thomas, A., Marinov, G.K., and Aravin, A.A. (2014). A Transgenerational Process Defines piRNA Biogenesis in Drosophila virilis. Cell Rep. *8*, 1617–1623.

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

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

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*.

Luteijn, M.J., and Ketting, R.F. (2013). PIWI-interacting RNAs: from generation to transgenerational epigenetics. Nat. Rev. Genet. *14*, 523–534.

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.

Malone, C.D., and Hannon, G.J. (2009). Small RNAs as Guardians of the Genome. Cell *136*, 656–668.

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.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal *17*, 10–12.

Martinez, G., and Köhler, C. (2017). Role of small RNAs in epigenetic reprogramming during plant sexual reproduction. Curr. Opin. Plant Biol. *36*, 22–28.

Minkina, O., and Hunter, C.P. (2017). Stable Heritable Germline Silencing Directs Somatic Silencing at an Endogenous Locus. Mol. Cell *65*, 659-670.e5.

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.

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.

Ninova, M., Griffiths-Jones, S., and Ronshaugen, M. (2017). Abundant expression of somatic transposon-derived piRNAs throughout Tribolium castaneum embryogenesis. Genome Biol. *18*, 184.

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.

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.

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

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–W165.

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

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.

Schott, D., Yanai, I., and Hunter, C.P. (2014). Natural RNA interference directs a heritable response to the environment. Sci. Rep. *4*, 7387.

Shirayama, M., Seth, M., Lee, H.-C., Gu, W., Ishidate, T., Conte Jr., D., and Mello, C.C. (2012). piRNAs Initiate an Epigenetic Memory of Nonself RNA in the C. elegans Germline. Cell *150*, 65–77.

Stoeckius, M., Maaskola, J., Colombo, T., Rahn, H.-P., Friedländer, M.R., Li, N., Chen, W., Piano, F., and Rajewsky, N. (2009). Large-scale sorting of *C. elegans* embryos reveals the dynamics of small RNA expression. Nat. Methods *6*, 745–751.

Stoeckius, M., Grün, D., and Rajewsky, N. (2014). Paternal RNA contributions in the Caenorhabditis elegans zygote. EMBO J. *33*, 1740–1750.

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.

Thivierge, C., Makil, N., Flamand, M., Vasale, J.J., Mello, C.C., Wohlschlegel, J., Jr, D.C., 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.

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

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. *107*, 3582–3587.

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 1.

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

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.

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.

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.

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.

Zhuang, J.J., and Hunter, C.P. (2011). Tissue-specificity of Caenorhabditis elegans Enhanced RNAi Mutants. Genetics genetics.111.127209.

Supplementary Information

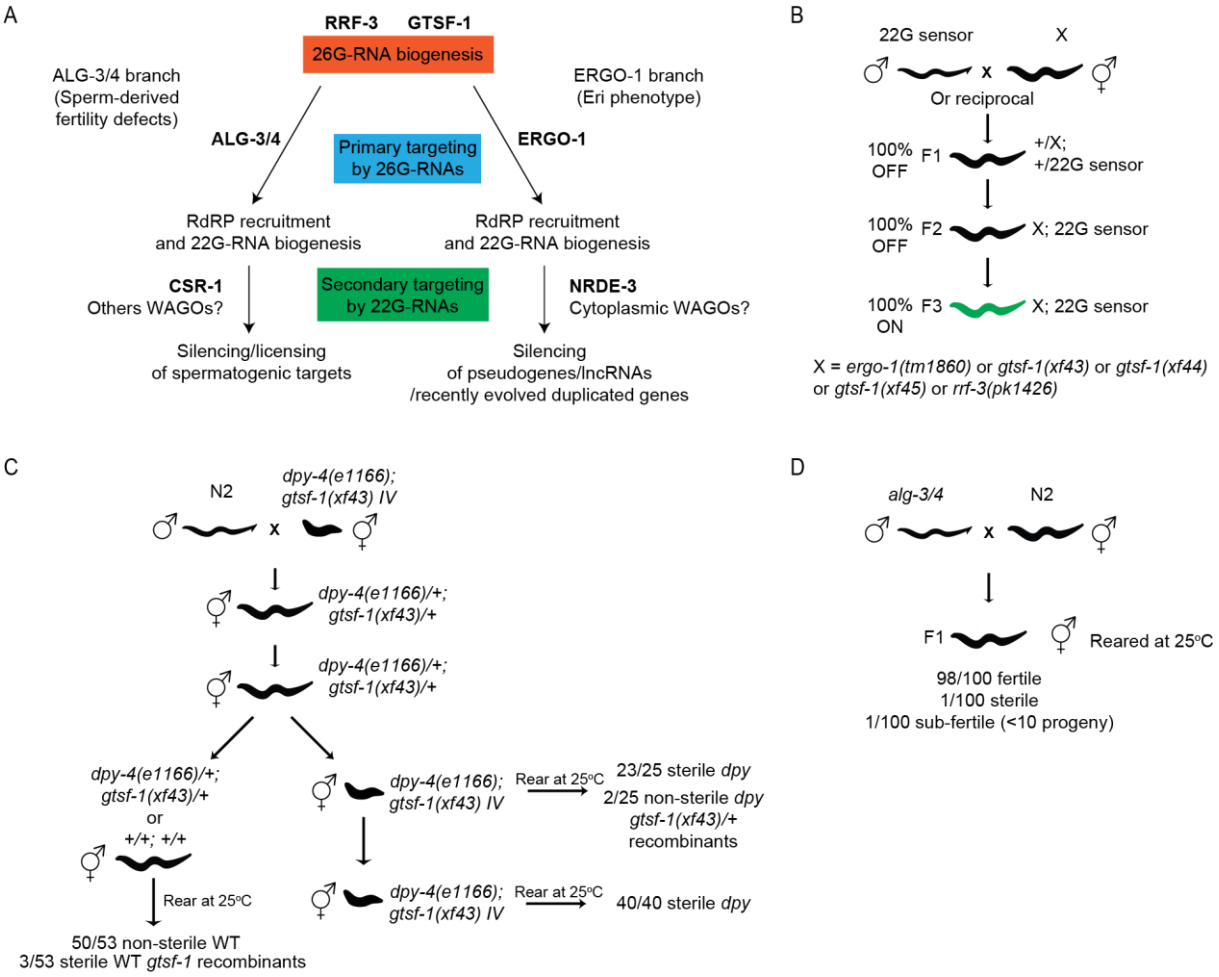


Figure III.S1. Parental effects of the 26G-RNA pathway. (A) Illustration of the current understanding of 26G-RNA pathways. 26G-RNAs are produced by RRF-3, assisted by GTSF-1 and other accessory factors. 26G-RNAs can associate with ALG-3/4 in the spermatogenic gonad (ALG-3/4 branch) or with ERGO-1 in oocytes and embryos (ERGO-1 branch). Upon target binding, RdRPs are recruited and synthesize secondary 22G-RNAs. NRDE-3 binds ERGO-1 branch 22G-RNAs, while CSR-1 is downstream of ALG-3/4 branch 26G-RNAs. Other unidentified AGOs may play a role in these pathways. (B) Schematics of genetic crosses of mutant strains with the 22G sensor. Green worms illustrate derepression of the 22G sensor. Black worms depict repression of the 22G sensor. "X" corresponds to different mutant alleles that share the same maternal rescue. (C) Experimental setup to address the maternal transmission of the temperature-sensitive sterility phenotype at 25 ºC. Worms were constantly grown at 20 ºC until transfer to 25 ºC to assay sterility. L2-L3 worms were transferred to 25 ºC. (D) Experimental setup to test paternal effect by ALG-3/4 branch 26G-RNAs. *alg-3/4* mutant males were crossed with wild-type hermaphrodites at 20 ºC. Cross progeny were isolated to fresh plates as L2-L3, transferred to and grown at 25 ºC. Fertility was assessed on the 3rd day of adulthood.

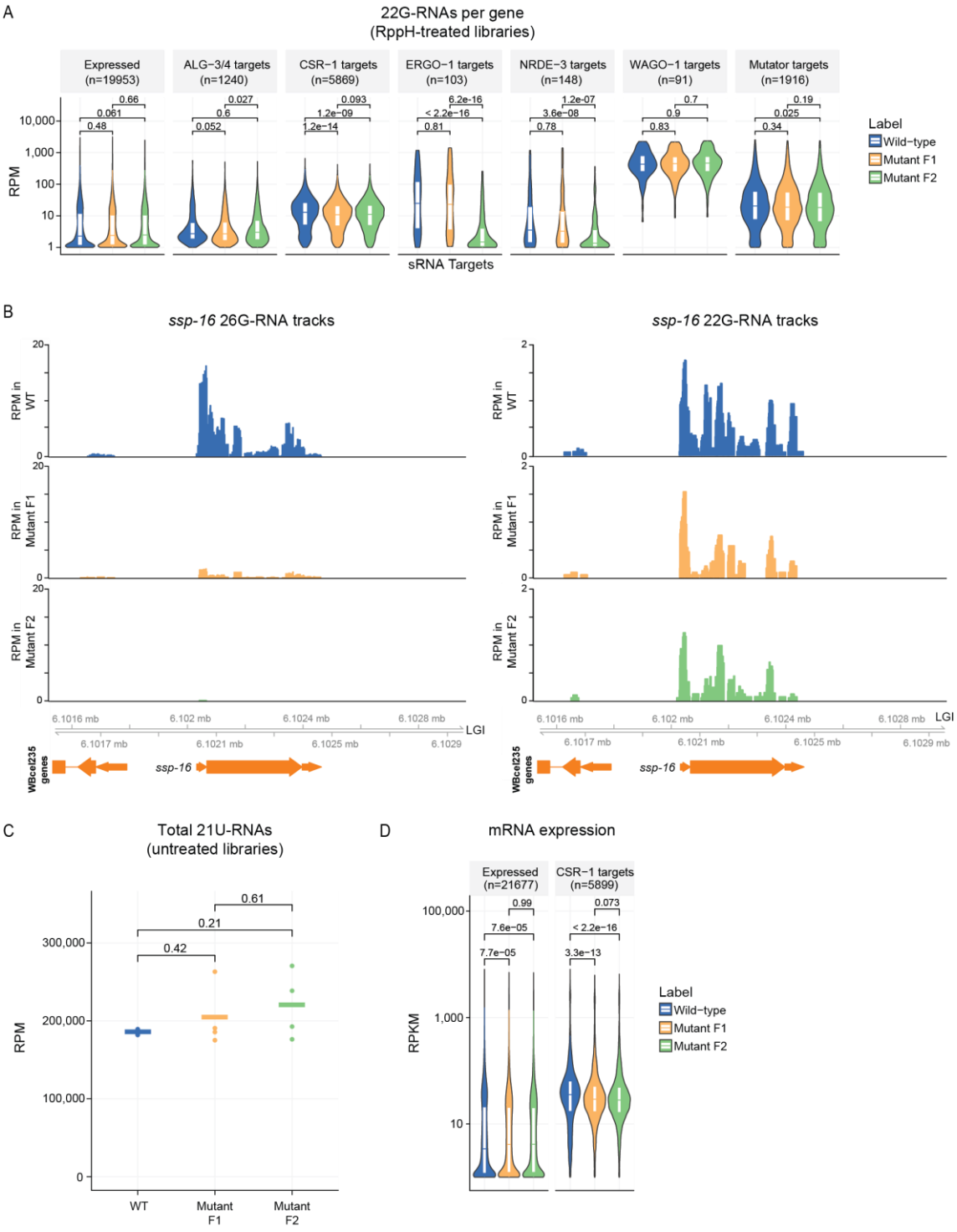


Figure III.S2. Dynamics of sRNA expression upon *gtsf-1* mutation. (A) Violin plot showing the distribution of RPM levels of 22G-RNAs mapping, per gene, to known targets of diverse sRNA pathways. Left-most panel shows the distribution of RPM levels in all genes with mapped 22G-RNAs. RPM values calculated from the RppH-treated libraries. (B) Genome browser tracks of *ssp-16*, a known ALG-3/4 target, showing mapped 26G- (left panels) and 22G-RNAs (right panels). 26G- and 22G-RNA tracks were obtained from untreated and RppH-treated libraries, respectively. (C) Total 21U-RNA levels in different generations/phenotype, in RPM. (D) Distribution of normalized mRNA expression of all expressed genes and CSR-1 targets, in RPKM. Violin plots in (A) and (D) show the distribution density of the underlying data. The top and bottom of the embedded box represent the 75th and the 25th percentile of the distribution, respectively. The line in the box represents the median. P-values were calculated with a two-sided unpaired Mann-Whitney/Wilcoxon rank-sum test.

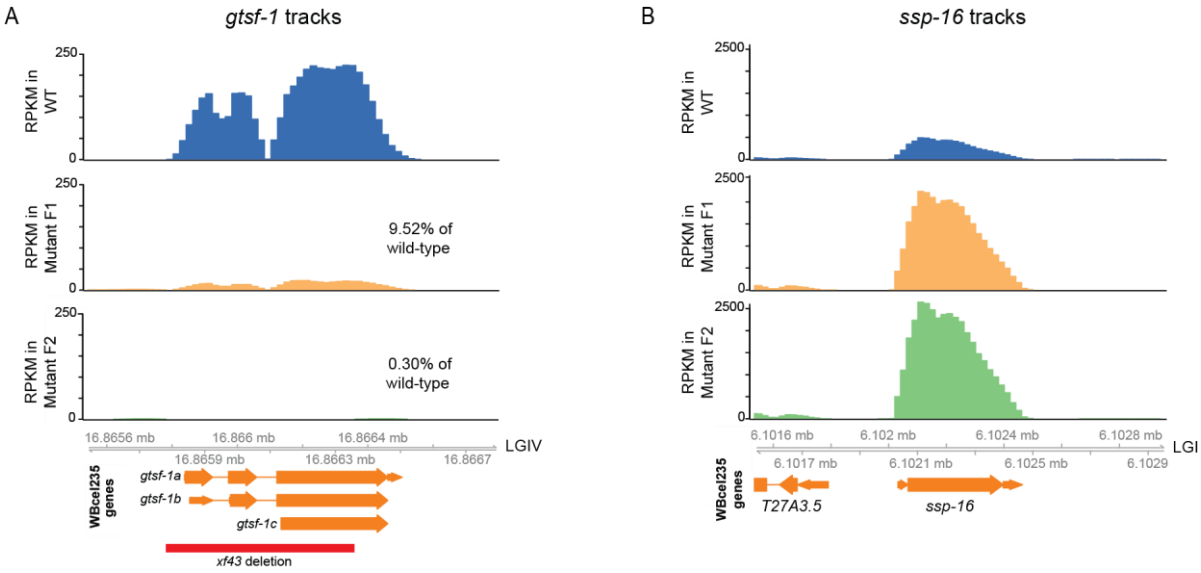


Figure III.S3. mRNA dynamics upon *gtsf-1* mutation. (A) Genome browser tracks displaying *gtsf-1* mRNA levels in RPKM. The *gtsf-1(xf43)* deletion allele is represented below. *gtsf-1* levels in the mutant F1 cover the *xf43* deletion sequence, thereby indicating contamination with Dpy worms that recombined a wild-type copy of *gtsf-1*. The mutant F2 was isolated from mutant F1 Dpy whose *gtsf-1* genotype was confirmed. Therefore, as expected, the only observed reads are flanking the *xf43* deletion. (B) Genome browser tracks with the mRNA levels, in RPKM, of *ssp-16*. Upregulation occurs immediately in the F1, indicating no maternal effect.

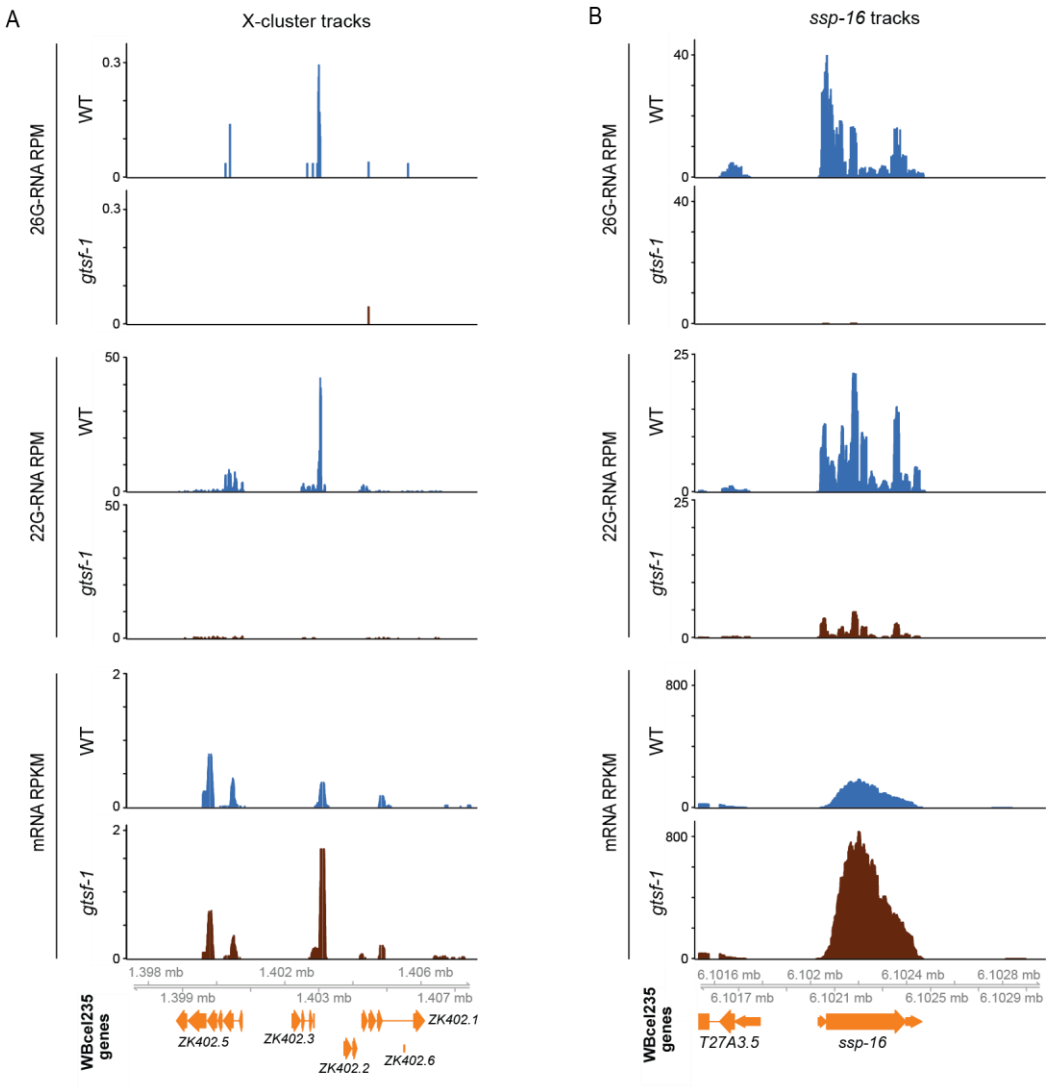


Figure III.S4. sRNA and mRNA profiles of ERGO-1 and ALG-3/4 targets in males. (A-B) RPM levels of 26G-RNAs (upper panels) and 22G-RNAs (middle panels) mapping to the X-cluster (A) and *ssp-16* (B). Lower panels show RPKM mRNA levels of these targets. WT, wild-type.

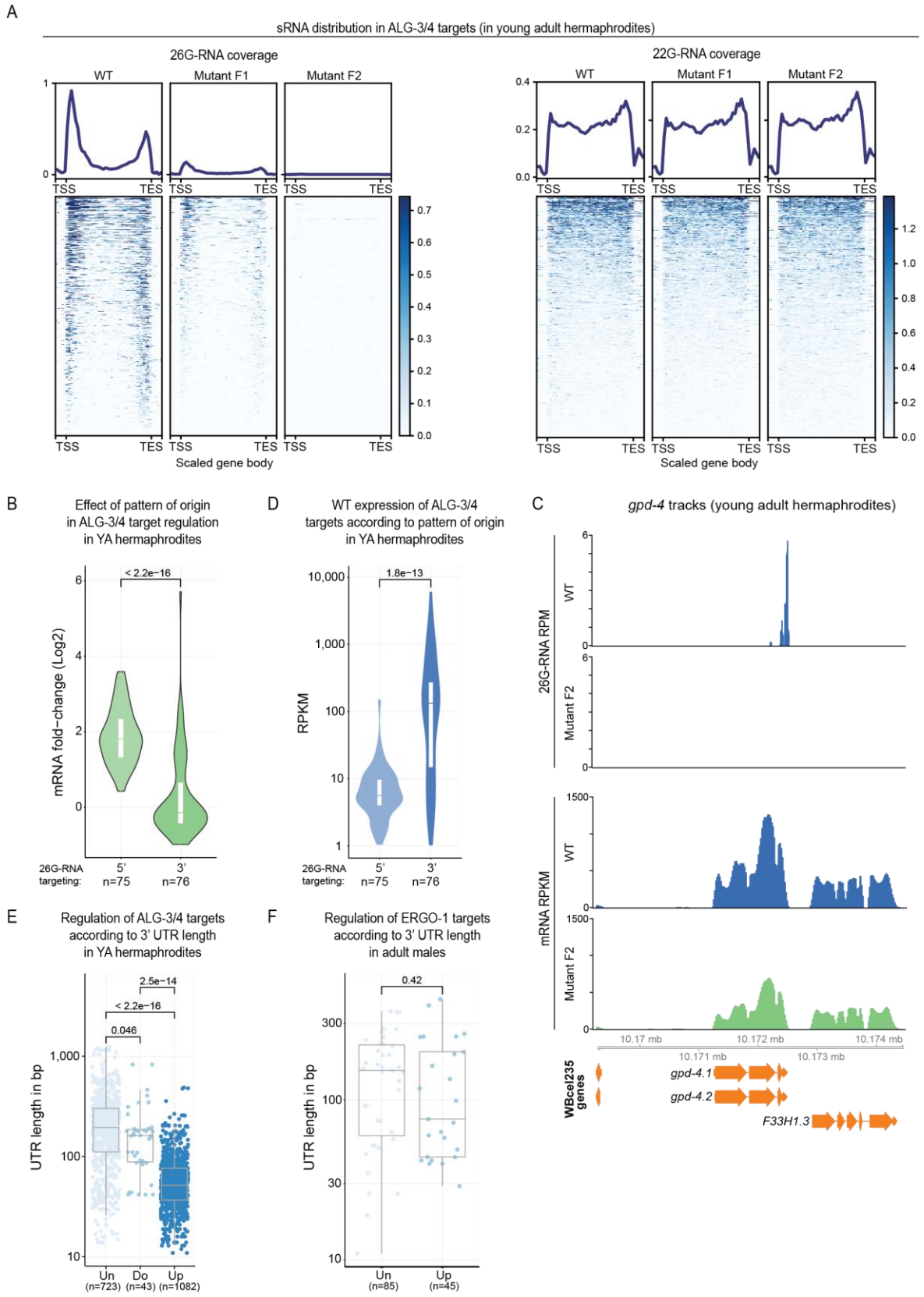

A
sRNA distribution in ALG-3/4 targets (in young adult hermaphrodites)
26G-RNA coverage
22G-RNA coverage
WT
Mutant F1
Mutant F2
TSS
TES
Scaled gene body
B
Effect of pattern of origin in ALG-3/4 target regulation in YA hermaphrodites
mRNA fold-change (Log2)
< 2.2e−16
26G-RNA targeting:
5' n=75
3' n=76
D
WT expression of ALG-3/4 targets according to pattern of origin in YA hermaphrodites
RPKM
1.8e−13
C
gpd-4 tracks (young adult hermaphrodites)
26G-RNA RPM
mRNA RPKM
10.17 mb
10.171 mb
10.172 mb
10.173 mb
10.174 mb
WBcel235 genes
gpd-4.1
gpd-4.2
F33H1.3
E
Regulation of ALG-3/4 targets according to 3' UTR length in YA hermaphrodites
UTR length in bp
2.5e−14
< 2.2e−16
0.046
Un (n=723)
Do (n=43)
Up (n=1082)
F
Regulation of ERGO-1 targets according to 3' UTR length in adult males
0.42
Un (n=85)
Up (n=45)

Figure III.S5. Predictors of regulatory outcome by ALG-3/4 in young adults. (A) Metagene analysis of 26G- (left panel) and 22G-RNAs (right panel) mapping to ALG-3/4 targets in young adult datasets, from our maternal effect setup (as in **Figure III.2A**). On the upper part of each panel is the mean coverage profile for sRNA species in every generation. On the lower part of each panel, the heatmaps show the density across individual targets. Target gene body length was scaled between transcription start site (TSS) and transcription end site (TES). Moreover, the regions comprising 250 nucleotides immediately upstream of the TSS and downstream of the TES are also included. Simultaneous 26G-RNA targeting in the 5' and 3' can be observed in some genes. (B) Violin plot depicting the regulation of ALG-3/4 target genes predominantly targeted at the 5' or at the 3' by 26G-RNAs. (C) Genome browser tracks displaying the RPM levels of 26G-RNAs (upper panels) and RPKM mRNA levels (lower panels) mapping to *gpd-4*, a gene predominantly targeted by 26G-RNAs at its 3' end, in young adult hermaphrodites. WT, wild-type. (D) Violin plot showing the wild-type expression levels of ALG-3/4 target genes predominantly targeted at the 5' or at the 3' by 26G-RNAs. (E) 3' UTR lengths of all the transcript isoforms annotated for ALG-3/4 target genes, according to effect on gene expression. (F) 3' UTR lengths of all the transcript isoforms annotated for ERGO-1 target genes, according to effect on gene expression. Fold-change data was obtained from adult male sequencing datasets. With the exception of (F), all the panels of this figure were prepared using young adult hermaphrodite sequencing datasets from our maternal effect experiments. In B and E, regulatory outcome was defined as differential gene expression between the wild-type and *gtsf-1* mutant F2. In (E-F), Un refers to genes with unchanged gene expression; Do means genes downregulated in the mutant; and Up refers to genes upregulated in the mutant. Violin plots in (B and D) and the boxplots in (E-F) show the distribution of the data. The top and bottom of the embedded boxes represent the 75th and the 25th percentile of the distribution, respectively. The lines in the boxes represent the median. P-values were calculated with a two-sided unpaired Mann-Whitney/Wilcoxon rank-sum test.

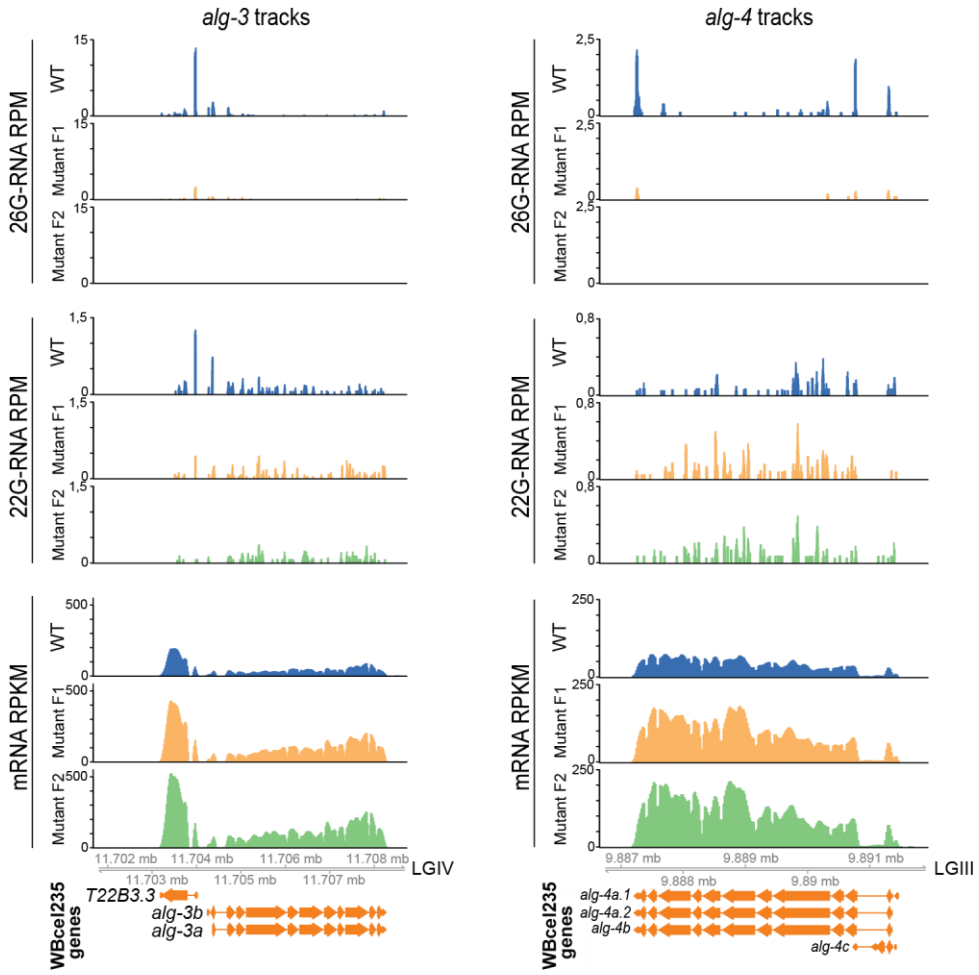


Figure III.S6. ALG-3 and ALG-4 are engaged in a negative feedback loop in young adults. Genome browser tracks showing 26G-RNAs (upper panels) and 22G-RNAs (middle panels) mapping to *alg-3* (left panels) and *alg-4* (right panels), in RPM. Lower panels show the RPKM mRNA levels of *alg-3* (on the left) and *alg-4* (on the right). Sequencing datasets of young adult hermaphrodites from our maternal effect setup were used.

Supplementary tables can be found at:

https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1007784#sec033

Table III.S1. Summary of sequencing output.

Table III.S2. Supplementary data related to Figures III.2, III.3, III.S2 and III.S3.

Table III.S3. Supplementary data related to Figure III.3.

Table III.S4. Supplementary data related to Figures III.4 and III.S4.

Table III.S5. Strains used in this study.

Table III.S6. Specifics of library preparation and sequencing.

Chapter IV

Comparative studies of nematode *gtsf-1* genes

Miguel Vasconcelos Almeida[1], António Miguel de Jesus Domingues[1], Hahn Witte[2], Ralf J. Sommer[2], and René F. Ketting[1]

[1]Biology of Non-Coding RNA group, Institute of Molecular Biology, Ackermannweg 4, 55128 Mainz, Germany

[2]Max Planck Institute for Developmental Biology, Department of Evolutionary Biology, Max-Planck-Ring 5, 72076 Tübingen, Germany

Author contributions:

Conceptualization: MVA and RFK; Investigation: MVA and HW; Formal Analysis: MVA and AMdJD; Visualization: MVA and AMdJD; Writing – original draft: MVA; Writing – review & editing: MVA and RFK; Funding acquisition: RJS and RFK.

Abstract

C. elegans is by far the most heavily studied nematode model organism. However, given the incredible ecological and evolutionary diversity of the phylum Nematoda, *C. elegans* is unlikely to represent a universal model for nematode biology. Therefore, it is imperative to extend functional studies to satellite nematode models. As described in **Chapter II**, in *C. elegans*, GTSF-1 is required to assemble a protein complex that synthesizes a class of endogenous small RNAs, termed 26G-RNAs. In stark contrast, the function of GTSF1 proteins in flies and mouse is to assist Piwi AGO proteins in eliciting target gene silencing. This evolutionary plasticity of GTSF-1 proteins prompted us to extend our studies on GTSF-1 function to other nematode species. First, we identified *gtsf-1* ortholog genes in *Caenorhabditis inopinata*, *Caenorhabditis briggsae*, and *Pristionchus pacificus*. Then, using CRISPR/Cas9 technology, we created mutant alleles of the *C. briggsae* and *P. pacificus* orthologs and found that *gtsf-1* mutants in these organisms display a temperature-sensitive sterility phenotype. However, while *gtsf-1* mutation abolishes 26G-RNA biogenesis in *P. pacificus* embryos, it is surprisingly dispensable for embryonic 26G-RNA biogenesis in *C. briggsae*. Furthermore, we show that the *P. pacificus* genome encodes a gene with similarity to both *pir-1* and *gtsf-1* genes of *C. elegans*. We provide evidence for the existence of a fused *pir-1::gtsf-1* transcript that suggests distinct mechanisms of 26G-RNA biogenesis in *P. pacificus*. Altogether, in this chapter we provide data supporting the fast evolution of *gtsf-1* genes, even in closely related nematode species.

Introduction

Nematodes comprise an extremely diverse group of animals with approximately 23,000 described species, and an estimated number of more than 1 million of total species (Blaxter, 2011). Nematodes can be found in terrestrial or aquatic environments, can be free-living or parasitic to plants and animals, and display different modes of reproduction, including gonochorism, hermaphroditism and parthenogenesis. Given this astonishing diversity of life styles, one can easily apprehend that *C. elegans* biology is by no means representative of the whole phylum Nematoda. Phylogenetic analysis have led to the division of the Nematoda phylum into three classes: Dorylaimia (clade I), Enoplia (clade II), and Chromadoria (clades III-V) (Blaxter, 2011). The phylogenetic relationship of these classes is still under debate, as their branching order remains unclear. The *Caenorhabditis* genus is part of the Chromadoria, clade V group of nematodes.

Homologous RNAi factors and sRNA populations have been profiled in a few members of each class. Although functional studies are lacking, the available studies already shed some light on the evolution of nematode RNAi factors and sRNA classes. These support the notion that nematode RNAi-like pathways are evolving fast (Palmer et al., 2018).

miRNA factors are ubiquitous in nematodes, contrary to many other factors (**Table IV.1**). Indeed, miRNAs and their associated machinery, for example ALG-1-like AGOs and DCR-1 enzymes are well conserved across all nematode clades (Dalzell et al., 2011; Holz and Streit, 2017; Sarkies et al., 2015; Shi et al., 2013; Wang et al., 2011; de Wit et al., 2009). Therefore, the ancestral nematode must have expressed miRNAs and their ancillary factors.

Clade V nematodes have similar RNAi-like pathways to *C. elegans*. Three additional species of the *Caenorhabditis* genus (*C. briggsae*, *C. remanei* and *C. brenneri*), and *Pristionchus pacificus* had their sRNAs profiled (Holz and Streit, 2017; Shi et al., 2013; Tu et al., 2014; de Wit et al., 2009). Excluding miRNAs, these species express 26G-, 22G- and 21U-RNAs with largely identical features as their *C. elegans* counterparts. However, specific individual targets and sRNA sequences are not conserved. This scenario is consistent with fast evolving sRNA pathways targeting foreign, fast evolving sequences. An expected exception are CSR-1 targets, which are bona fide germline expressed genes and therefore show a higher degree of conservation than WAGO targets across *Caenorhabditis* (Shi et al., 2013; Tu et al., 2014).

		Clade I-II	Clade III	Clade IV	Clade V
	DCR-1				
RdRPs	RRF-3				
	EGO-1/RRF-1				
AGOs	ALG-1/2				
	CSR-1				
	HRDE-1				
	NRDE-3				
	ERGO-1				
	ALG-3				
	PRG-1				
	HENN-1				
sRNA classes	21U-RNA				
	22G-RNA			*	
	26G-RNA				
	miRNA				

Table IV.1. Conservation of RNAi factors throughout nematodes. Green and gray colored box indicate the existence or absence, respectively, of clear homologs in the designated nematode clade. This table is based on published data (Dalzell et al., 2011; Holz and Streit, 2017; Sarkies et al., 2015). * The Strongyloididae family has lost 5' triphosphorylated sRNAs.

Interestingly, some associations have been established between RNAi biology and reproductive life-style. *C. remanei* and *C. brenneri*, which are both gonochoristic, have 2-3 times more 21U-RNAs than the hermaphroditic *C. elegans* and *C. briggsae* (Shi et al., 2013). Moreover, the genomes of *C. remanei* and *C. briggsae* encode a total of 37 and 46 AGOs, respectively, almost doubling those of *C. elegans* and *C. briggsae*. The authors proposed that such an expansion may have arisen to counteract increased exposure to more diverse TEs, brought about by obligatory mating (Shi et al., 2013), consistent with a "Red Queen" arms race scenario.

Amongst genes involved in RNAi, AGO proteins are perhaps the set of factors showing higher evolutionary plasticity (Buck and Blaxter, 2013). Although other nematodes have a high number of genomically encoded AGO proteins similar to *C. elegans*, it is often hard to find 1-to-1 orthologues, especially within the WAGO clade (Holz and Streit, 2017; Wang et al., 2011). Clear ALG-3-like orthologues can be found across all clades, whereas ERGO-1-like AGOs seem to be restricted to clade V (**Table IV.1**) (Buck and Blaxter, 2013; Sarkies et al., 2015). Notably, PRG-1-like Piwi AGOs, HENN-1 enzymes, 21U-RNAs, and Ruby motifs are largely absent outside clade V (Holz and Streit, 2017; Sarkies et al., 2015; Wang et al., 2011). However, other sRNA species were found to target TEs in clades I-IV: the 5' monophosphorylated, 23-25 nucleotide long sRNAs of clade I *Trichinella spiralis* (Sarkies et al., 2015); the 22G-RNAs of clade IV *Globodera pallida* and clade III *Brugia malayi* (Sarkies et al., 2015); and 27 nucleotide long sRNAs with a bias for 5' triphosphate guanosine or adenosine (or 27GA-RNAs), expressed in three species belonging to the clade IV Strongyloididae family, a particularly relevant group of animal parasites (Holz and Streit, 2017).

RRF-3-like RdRP orthologues are distributed throughout the nematode phylum, suggesting that ancestral nematode sRNA biogenesis was perhaps similar to RRF-3-mediated 26G-RNA biogenesis in *C. elegans* (**Table IV.1**) (Sarkies et al., 2015). However, sRNA species clearly analogous to *C. elegans* 5'-monophosphorylated 26G-RNAs were not detected outside clade V (Holz and Streit, 2017; Sarkies et al., 2015), with the exception of the 26G-RNAs of clade III *Ascaris suum* (Wang et al., 2011). The 26G-RNAs of *A. suum* are particularly abundant in testes and target spermatogenesis-specific genes, much like ALG-3/4 branch 26G-RNAs in *C. elegans*. Accordingly, three ALG-3/4-like AGOs were identified in *A. suum* (Wang et al., 2011). Contrary to *C. elegans*, the *A. suum* spermatogenesis-enriched 26G-RNAs do not show the overall biased distribution toward the 5' and 3' ends of target transcripts (**Figures III.6** and **III.S5**; Conine et al., 2010). Altogether, these observations suggest high plasticity of sRNA biogenesis mechanisms by RRF-3-like RdRPs.

The RRF-1 and EGO-1 family of RdRPs is a novelty of clades III-V, indicating that 5'-triphosphorylated sRNAs are not expressed in clades I-II (**Table IV.1**) (Dalzell et al., 2011; Sarkies et al., 2015). Indeed, 5'-triphosphorylated sRNAs have not been detected in clade I *T. spiralis*, and in clade II *Enopolus brevis* and *Odonotophora rectangular* (Sarkies et al., 2015). Absence of 5'-triphosphorylated sRNAs is mirrored by lack of NRDE-3 and HRDE-1 AGOs (Dalzell et al., 2011; Sarkies et al., 2015; Tu et al., 2014), implying absence of AGO-driven TGS, or the existence of other alternative TGS mechanisms in these nematodes. Supporting the latter, DNA methylation was detected in clade I *T. spiralis* (Gao et al., 2012) and a strong correlation was observed between methylated regions and sRNA abundance (Sarkies et al., 2015). This may hint at an ancestral nematode mechanism of RNA-directed DNA methylation (Sarkies et al., 2015), reminiscent of what is observed in plants (Borges and Martienssen, 2015).

CSR-1 homologs have been identified in clades III and V (**Table IV.1**) (Dalzell et al., 2011; Tu et al., 2014). The presence of CSR-1 genes in clade III is enigmatic because these species lack PRG-1 homologs. It will be interesting to determine whether clade III CSR-1 gene display a positive gene regulatory function in the absence of PRG-1.

We were fascinated by the striking evolutionary co-option of GTSF-1 proteins by different sRNA pathways: in *Drosophila melanogaster*, GTSF1 interacts with Piwi in the nucleus, contributing to TGS of TEs (Dönertas et al., 2013; Ohtani et al., 2013), while *C. elegans* GTSF-1 is required for sRNA biogenesis in the cytoplasm, by interacting with the RdRP RRF-3 (as described in **Chapter II**). We wondered whether the interaction with RRF-3 is a *C. elegans* oddity or more broadly observed in clade V nematodes. To do this, we started growing different nematode species in the lab, in order to perform comparative studies of *gtsf-1* genes. Our nematode collection includes *P. pacificus*, *C. briggsae* and *C. inopinata*, recently identified as the sister species of *C. elegans* (Kanzaki et al., 2018). Using CRISPR/Cas9 technology, we have created *gtsf-1* mutants in *P. pacificus* and *C. briggsae*. These mutant animals display complete sterility when grown at elevated temperatures, much like *C. elegans*. However, 26G-RNAs are only lacking in *P. pacificus gtsf-1* mutants. Also, we report an intriguing gene fusion of *pir-1* and *gtsf-1* in *P. pacificus* with potential consequences for ERI complex assembly and function.

Results

Identification of *gtsf-1* orthologs in nematodes

To perform comparative studies of nematode *gtsf-1* genes, we first acquired and grew *P. pacificus*, *C. briggsae*, and *C. inopinata* in the lab. These species share morphological differences, as can be appreciated in **Figure IV.1A**. All are free-living nematodes, but have divergent reproductive modes. *P. pacificus* and *C. briggsae* have mostly hermaphroditic reproduction. Much like *C. elegans*, they have androdioecious populations, i.e. with a great majority of hermaphrodites and rare males, which can mate with hermaphrodites. In contrast, *C. inopinata* has gonochoristic reproduction, involving male and female animals.

The genomes of these three species have been sequenced (Dieterich et al., 2008; Kanzaki et al., 2018; Stein et al., 2003). Molecular analysis allowed for the estimation of divergence times (**Figure IV.1B**). The *Caenorhabditis* group and *P. pacificus* share a common ancestor that lived between 200-300 million years ago (**Figure IV.1B**) (Dieterich et al., 2008). *C. briggsae* and *C. elegans* diverged approximately 100 million years ago, whereas the common ancestor of *C. inopinata* and *C. elegans*, lived around 10.5 million years ago (**Figure IV.1B**).

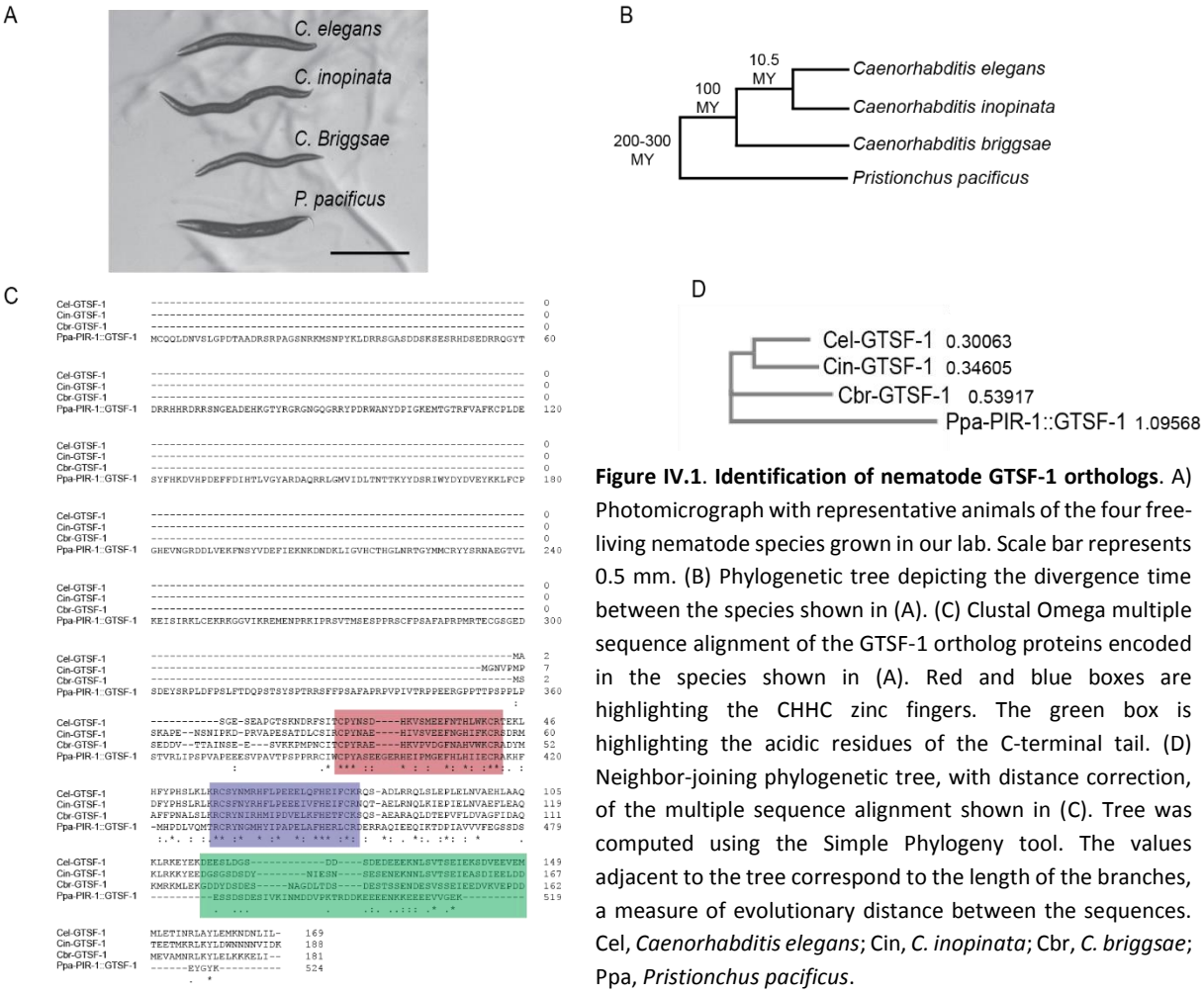


Figure IV.1. **Identification of nematode GTSF-1 orthologs**. A) Photomicrograph with representative animals of the four free-living nematode species grown in our lab. Scale bar represents 0.5 mm. (B) Phylogenetic tree depicting the divergence time between the species shown in (A). (C) Clustal Omega multiple sequence alignment of the GTSF-1 ortholog proteins encoded in the species shown in (A). Red and blue boxes are highlighting the CHHC zinc fingers. The green box is highlighting the acidic residues of the C-terminal tail. (D) Neighbor-joining phylogenetic tree, with distance correction, of the multiple sequence alignment shown in (C). Tree was computed using the Simple Phylogeny tool. The values adjacent to the tree correspond to the length of the branches, a measure of evolutionary distance between the sequences. Cel, *Caenorhabditis elegans*; Cin, *C. inopinata*; Cbr, *C. briggsae*; Ppa, *Pristionchus pacificus*.

We started by identifying *gtsf-1* orthologs in each of these species by reciprocal BLAST. The genomes of *P. pacificus*, *C. briggsae* and *C. inopinata* encode one unambiguous *gtsf-1* ortholog each. We will henceforth refer to the distinct orthologs using a species prefix: Cel for *C. elegans*; Cin for *C. inopinata*; Cbr for *C. briggsae*; and Ppa for *P. pacificus*. Next, we performed a multiple sequence alignment using the predicted protein sequence of GTSF-1 orthologs (**Figure IV.1C**). Several insights can be extracted from the alignment: 1) the GTSF-1 protein of *P. pacificus* is more than triple the size of the *Caenorhabditis* GTSF-1 proteins; 2) the cysteine and histidine amino acid residues of the two tandem CHHC zinc fingers are perfectly conserved in all species (**Figure IV.1C**, red and blue boxes); 3) the C-terminal tail of all GTSF-1 proteins is rich in acidic amino acid residues (**Figure IV.1C**, green box); and 4) the first CHHC zinc finger of *P. pacificus* GTSF-1 has a four amino acid extension between the first cysteine and the first histidine, potentially precluding proper zinc finger folding and function. Otherwise, based on sequenced conservation, all other zinc fingers have the potential to be functional. Of note, the multiple sequence alignment recapitulates the species divergence (compare **Figure IV.1B** and **IV.1D**).

We conclude that *P. pacificus*, *C. briggsae* and *C. inopinata* have one, potentially functional, *gtsf-1* ortholog gene each.

Cbr-GTSF-1 is required for normal fertility in *C. briggsae*

We used CRISPR/Cas9 technology to create *gtsf-1* mutants in *C. briggsae*. After several attempts, we isolated one mutant allele consisting of an in-frame, clean deletion of 6 bps (**Figure IV.2A**). The two removed amino acids are not conserved, but are between the highly conserved CPY and H residues of the first CHHC zinc finger of *gtsf-1* (**Figure IV.2B**). *Cbr-gtsf-1* mutants are viable and show no other obvious morphological defects. Therefore, *Cbr-gtsf-1* is not an essential gene. However, *Cbr-gtsf-1* mutant animals grown at 30 ºC, an elevated growth temperature for *C. briggsae*, became completely sterile (**Figure IV.2C**). The phenotype above is very similar to the temperature-sensitive sterile phenotype of *Cel-gtsf-1* mutants (**Figure II.3** and **II.S3**).

The sterility observed at higher temperatures prompted us to determine whether *Cbr-gtsf-1* is required for 26G-RNA biogenesis, identically to *Cel-gtsf-1*. To address this, we sequenced sRNAs of wild-type (AF16) and *gtsf-1* embryos. Reads were mapped to the *C. briggsae* genome and normalized to the total number of reads (see **Materials and Methods**). Surprisingly, no single sRNA population seems to be depleted in *Cbr-gtsf-1* mutants (**Figure IV.3A-D**). In *C. elegans*, disruption of *gtsf-1* affects both primary 26G-RNAs and secondary 22G-RNAs To address whether 22G-RNAs downstream of 26G-RNAs are particularly affected by *Cbr-gtsf-1*, we performed a metagene analysis of 22G-RNA coverage around 26G-RNA mapping sites. 22G-RNAs expressed in the vicinity of 26G-RNAs are not affected by *Cbr-gtsf-1(xf207)* (**Figure IV.3E**). Manual navigation through genome browser tracks confirmed that no sRNA population is affected by *Cbr-gtsf-1(xf207)*. Targets of *C. elegans* embryonic, ERGO-1-bound, 26G-RNAs seem to be comprised by recently duplicated and fast evolving genes (and thus non-conserved), including, for example, F-box genes (Fischer et al., 2011; Vasale et al., 2010). The embryonic 26G-RNAs of *C. briggsae* seem to map to a similar set of targets, as shown before (Shi et al.,

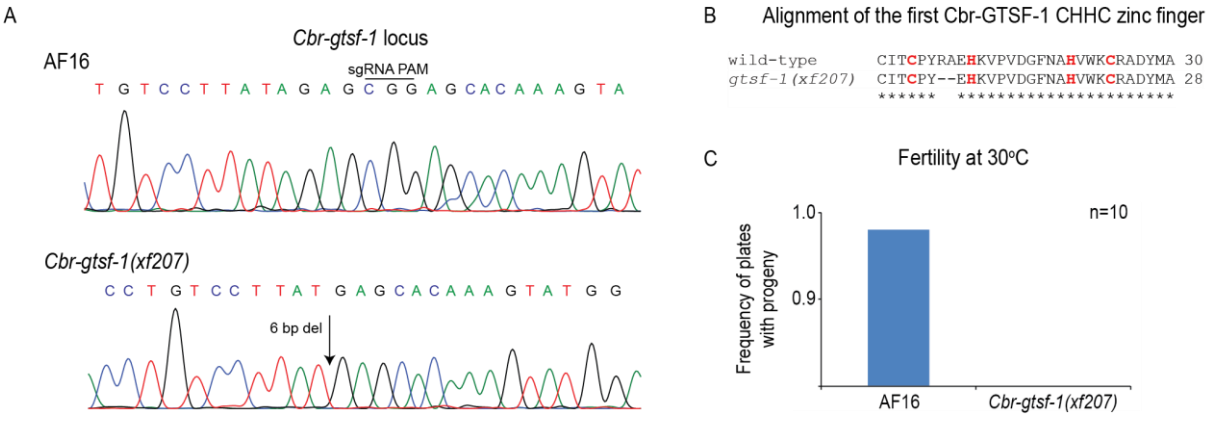


Figure IV.2. ***Cbr-gtsf-1* mutants are temperature-sensitive sterile**. (A) Chromatograms depicting the wild-type (above), and *gtsf-1* mutant (below) sequences. The protospacer adjacent motif (PAM) of the sgRNA is indicated in the wild-type sequence. The arrow indicates the site of the deletion in the mutant. (B) Clustal Omega multiple sequence alignment of the first zinc finger of Cbr-GTSF-1 in the wild-type and in the mutant. (C) Relative frequency of fertile animals grown at 30 ºC, an elevated temperature for *C. briggsae*. n=10 animals.

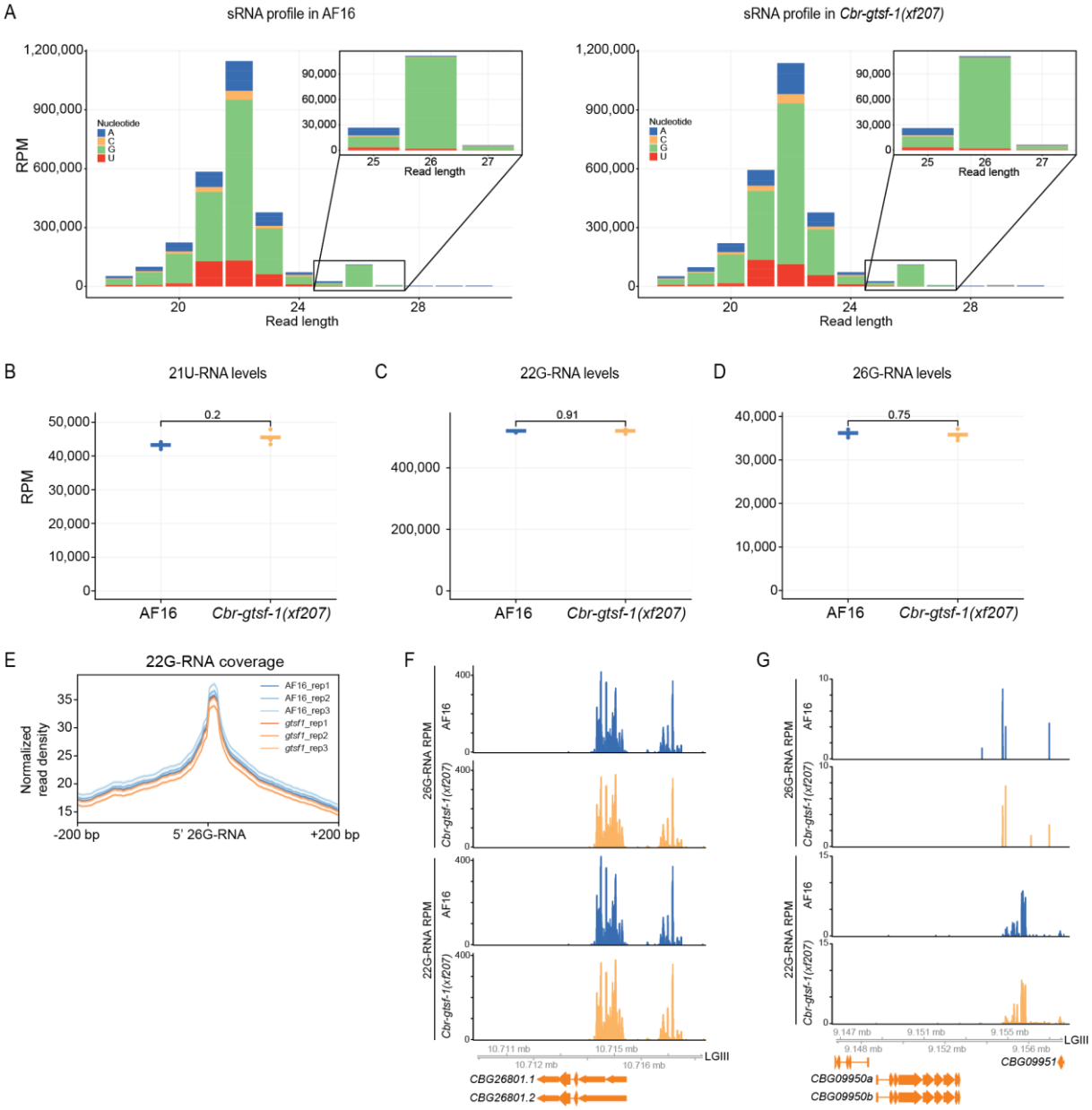


Figure IV.3. sRNA sequencing of wild-type and *Cbr-gtsf-1* mutant embryos. (A) Normalized levels of sRNAs, in RPM (Reads Per Million), summarizing the length and 5' ribonucleotide bias in wild-type AF16 (left panel) and *Cbr-gtsf-1* mutants (right panel). The 5' ribonucleotide bias is color-coded. Given the low overall abundance of 26G-RNAs, the inset displays a rescaled view of the sRNAs that are 25-27 ribonucleotides long. (B-D) Normalized levels of specific sRNA species, in RPM, in AF16 and *Cbr-gtsf-1* mutants. Three biological replicates are shown. P-values were calculated with a two-sided unpaired t-test. (B) Total levels of 21U-RNAs. (C) Total levels of 22G-RNAs. (D) Total levels of 26G-RNAs. (E) Metagene analysis of 22G-RNA levels at, and around 26G-RNA sites in wild-type AF16 (blue lines) and *Cbr-gtsf-1* mutant (orange lines) embryos. (F-G) Genome browser tracks displaying the normalized 26G- (panels above) and 22G-RNA (panels below) levels mapping to CBG26801 (F), CBG09950 and CBG09951 (G), both in wild-type (in blue) and in the mutant (in orange).

2013). As an example, an F-box gene that is targeted by 26G- and 22G-RNAs in *C. briggsae* is shown in **Figure IV.3F**.

In **Chapter III** (**Figures III.7** and **III.S6**), we have provided evidence suggesting that Cel-ALG-3/4 and cognate 26G-RNAs may regulate the mRNAs of *alg-3* and *alg-4* in a negative feedback loop. *C. briggsae* has one single ortholog of *alg-3/4*, called *CBG09950*, which is not significantly targeted by

26G- and 22G-RNAs (**Figure IV.3G**). 26G- and 22G-RNA signal is observed only 3 kb downstream of *CBG09950*. Hence, it is unlikely that such sRNAs could regulate *CB09950*.

Altogether, we conclude that *Cbr-gtsf-1* is required for fertility at higher temperatures, but is dispensable for 26G-RNA biogenesis. Interestingly, the *alg-3/4* negative feedback loop, reported in **Chapter III**, does not seem to be evolutionarily conserved.

gtsf-1 and *pir-1* are fused in *P. pacificus*

BLAST of the longer Ppa-GTSF-1 to the *C. elegans* proteome revealed a surprising alignment with Cel-PIR-1 (**Figure IV.1C**). Cel-PIR-1 is an RNA phosphatase that is part of the ERI complex (Chaves, 2015; Duchaine et al., 2006) (**Figure II.7**). Upon closer inspection, we observed that *pir-1* and *gtsf-1* sequences seem to be fused, yielding one predicted transcript and one predicted protein, which we refer to as Ppa-PIR-1::GTSF-1 (**Figure IV.4A**). The amino acid residues of Cel-PIR-1 presumably required for its catalytic activity (Chaves, 2015) are well conserved in Ppa-PIR-1::GTSF-1 (**Figure IV.4B**). Moreover, the key amino acid residues of the tandem CHHC zinc fingers of Cel-GTSF-1 are also well conserved in Ppa-PIR-1::GTSF-1 (**Figure IV.4B**). These observations indicate that Ppa-PIR-1::GTSF-1 may encode a fusion protein with functional domains.

To exclude any annotation issues, and to confirm that *pir-1::gtsf-1* is a single bona fide transcript, we performed RT-PCRs on wild-type *P. pacificus* young adults and embryos. Several hybrid amplicons could be amplified in both tissues (**Figure IV.4C**), suggesting that *pir-1::gtsf-1* indeed comprises a fused transcript. Of note, the complete predicted transcript could be amplified in embryos, but not in young adults, indicating that *pir-1::gtsf-1* may be more abundant in embryonic tissues. These results suggest that Cel-PIR-1 and Cel-GTSF-1 have one single fused ortholog gene in *P. pacificus*.

Ppa-GTSF-1 is required for normal fertility and 26G-RNA biogenesis in *P. pacificus*

To investigate the function of Ppa-PIR-1::GTSF-1, we created mutants using CRISPR/Cas9 technology. We targeted the 11th exon to produce indel alleles in the sequence homologous to *Cel-gtsf-1*, while leaving the sequence homologous to *Cel-pir-1* intact. We isolated two *Ppa-pir-1::gtsf-1* alleles, each of which is predicted to create frameshifts and premature stop codons (**Figure IV.5A**). *Ppa-gtsf-1* mutants are viable and show no obvious morphological defects. When grown at 27 ^{o}C, an elevated growth temperature for *P. pacificus*, *gtsf-1* mutants are completely sterile (**Figure IV.5B**). Therefore, depletion of *gtsf-1* in *P. pacificus* results in temperature-sensitive sterility, similarly to *C. elegans*.

Next, given the temperature-sensitive sterility of *Ppa-pir-1::gtsf-1*, similar to the defects displayed by *Cel-gtsf-1* mutants, we wished to understand the contribution of Ppa-PIR-1::GTSF-1 to the

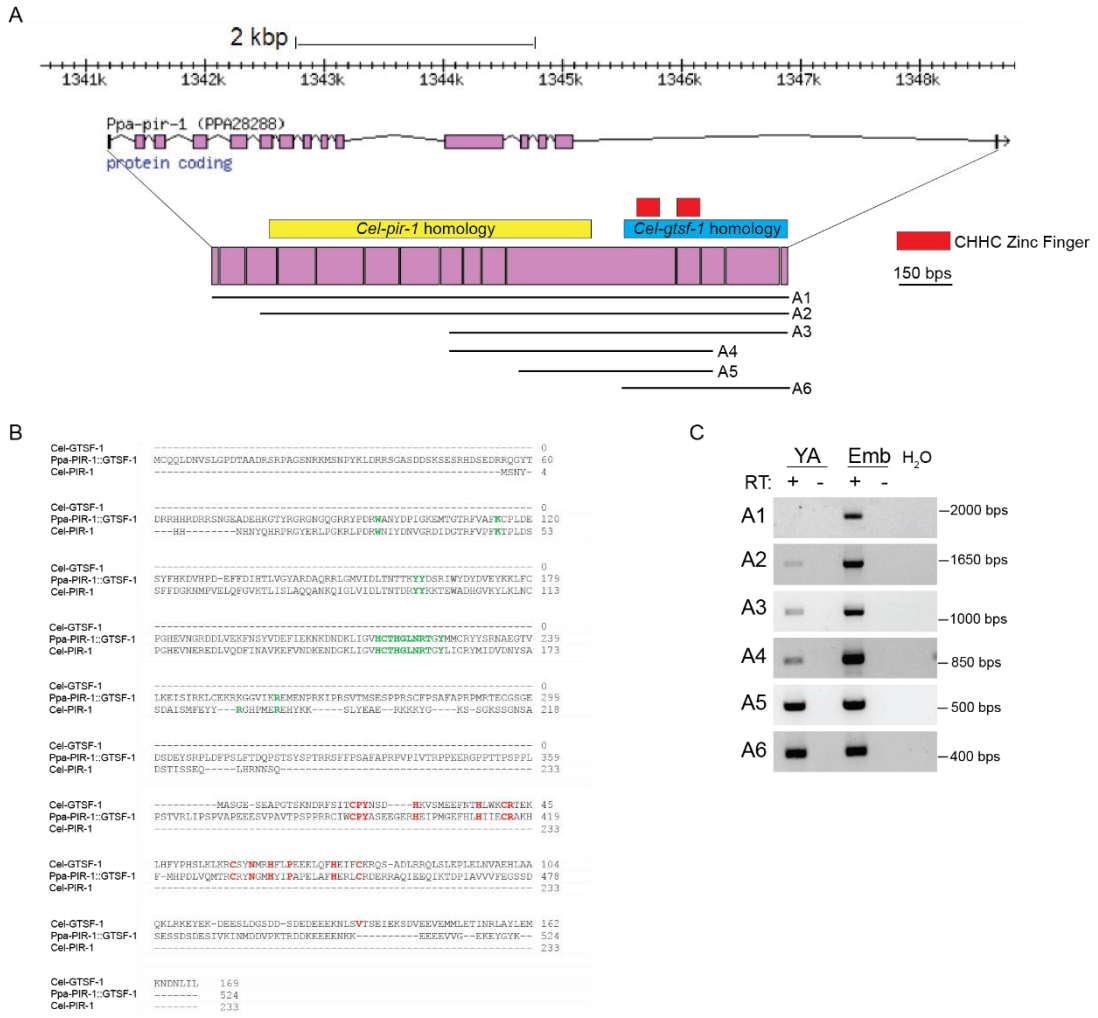


Figure IV.4. **One single gene in *P. pacificus* is a fused homolog of *Cel-pir-1* and *Cel-gtsf-1***. (A) GBrowse snapshot of the *Ppa-pir-1* locus, including a schematic of the spliced transcript, highlighting the regions similar to *Cel-pir-1* (yellow box) and *Cel-gtsf-1* (blue box). The two red boxes represent the conserved CHHC zinc fingers. A1-A6 are the amplicons targeted by RT-PCR, see (C). (B) Clustal Omega multiple sequence alignment of Cel-GTSF-1, Ppa-PIR-1::GTSF-1 and Cel-GTSF-1. Amino acid residues presumably required for the catalytic activity of PIR-1 are showed in green. The conserved amino acid residues of the CHHC zinc fingers of GTSF-1 proteins are shown in red. (C) RT-PCRs in embryos (Emb) and in young adults (YA) of the amplicons shown in (A).

sRNA pathways of *P. pacificus*. To do this, we sequenced sRNAs of embryos of wild-type (PS312) and *Ppa-pir-1::gtsf-1* mutants. Reads were mapped to the *P. pacificus* genome and normalized to the total number of reads (see **Materials and Methods**). *Ppa-pir-1::gtsf-1* mutant embryos strongly lack 26G-RNAs and have (non-statistically significant) mild decrease of 22G-RNAs and increase of 21U-RNAs (**Figure IV.6A-D**). The strong lack of 26G-RNAs is consistent with a function for Ppa-PIR-1::GTSF-1 in 26G-RNA biogenesis. It is not known whether 26G-RNAs trigger the production of 22G-RNAs in *P. pacificus*. To probe into this, we looked for 22G-RNA expression around 26G-RNA sites. In the wild-type, 22G-RNAs are indeed accumulated at, and around, 26G-RNA sites (**Figure IV.6E**, in blue). Moreover, since 22G-RNAs are mostly absent at, and around 26G-RNA sites in *Ppa-pir-1::gtsf-1* mutants (**Figure IV.6E**, in orange), we conclude that the accumulation of 22G-RNAs at 26G-RNA sites is: 1) dependent on 26G-RNAs; and 2) dependent on Ppa-PIR-1::GTSF-1. **Figure IV.6F-G** display genome

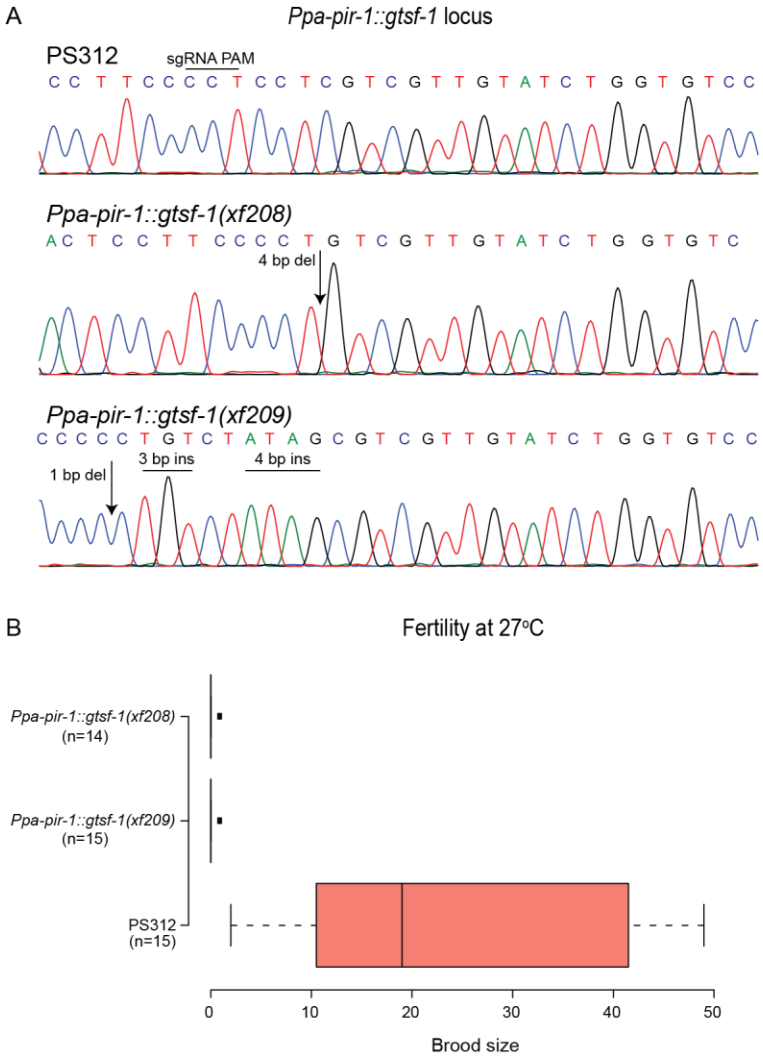


Figure IV.5. ***Ppa-pir-1::gtsf-1*** **mutants are temperature-sensitive sterile**. (A) Chromatograms depicting the wild-type (upper panel), and *Ppa-pir-1::gtsf-1* mutant (middle and lower panels) sequences. The protospacer adjacent motif (PAM) of the sgRNA is indicated in the wild-type sequence. The arrows indicates the site of the deletion in the mutant. (B) Brood size of animals grown at 27 ºC, an elevated temperature for *P. pacificus*. PS312 is the wild-type strain. The n is indicated in the figure.

browser tracks of selected target genes of 26G-RNAs and downstream 22G-RNAs. Neither of these genes is conserved beyond *P. pacificus*, again consistent with 26G-RNAs targeting fast evolving sequences. Lastly, the single *alg-3/4* ortholog of *P. pacificus*, gene PPA09361, is not targeted by 26G-RNAs above background levels (**Figure IV.6H**).

These results suggest that PIR-1::GTSF-1 is required for normal fertility and 26G-RNA biogenesis in *P. pacificus*, much like GTSF-1 in *C. elegans*. Like for *C. briggsae*, we find no evidence of self-regulation of an ALG-3/4-like gene by 26G-RNAs.

Caenorhabditis gtsf-1 genes are evolving fast

Evolutionary analysis of protein sequence is often based on ratios of non-synonymous to synonymous amino acid substitutions (dN/dS ratios) (Obbard et al., 2009; Palmer et al., 2018). A lower dN/dS ratio is generally interpreted as evidence of purifying selection that constrains the protein sequence. Conversely, a higher dN/dS ratio is typically considered as a signature of rapid, positive natural selection (McDonald and Kreitman, 1991; Obbard et al., 2009; Palmer et al., 2018). Of note, another interpretation is that more non-synonymous mutations are also found in proteins that have low selective constraints, allowing for fixation of specific variants by genetic drift (McDonald and Kreitman, 1991).

Such evolutionary analysis were performed in several arthropod and nematode species for genes involved in RNAi pathways (Obbard et al., 2009; Palmer et al., 2018). The higher dN/dS ratios of

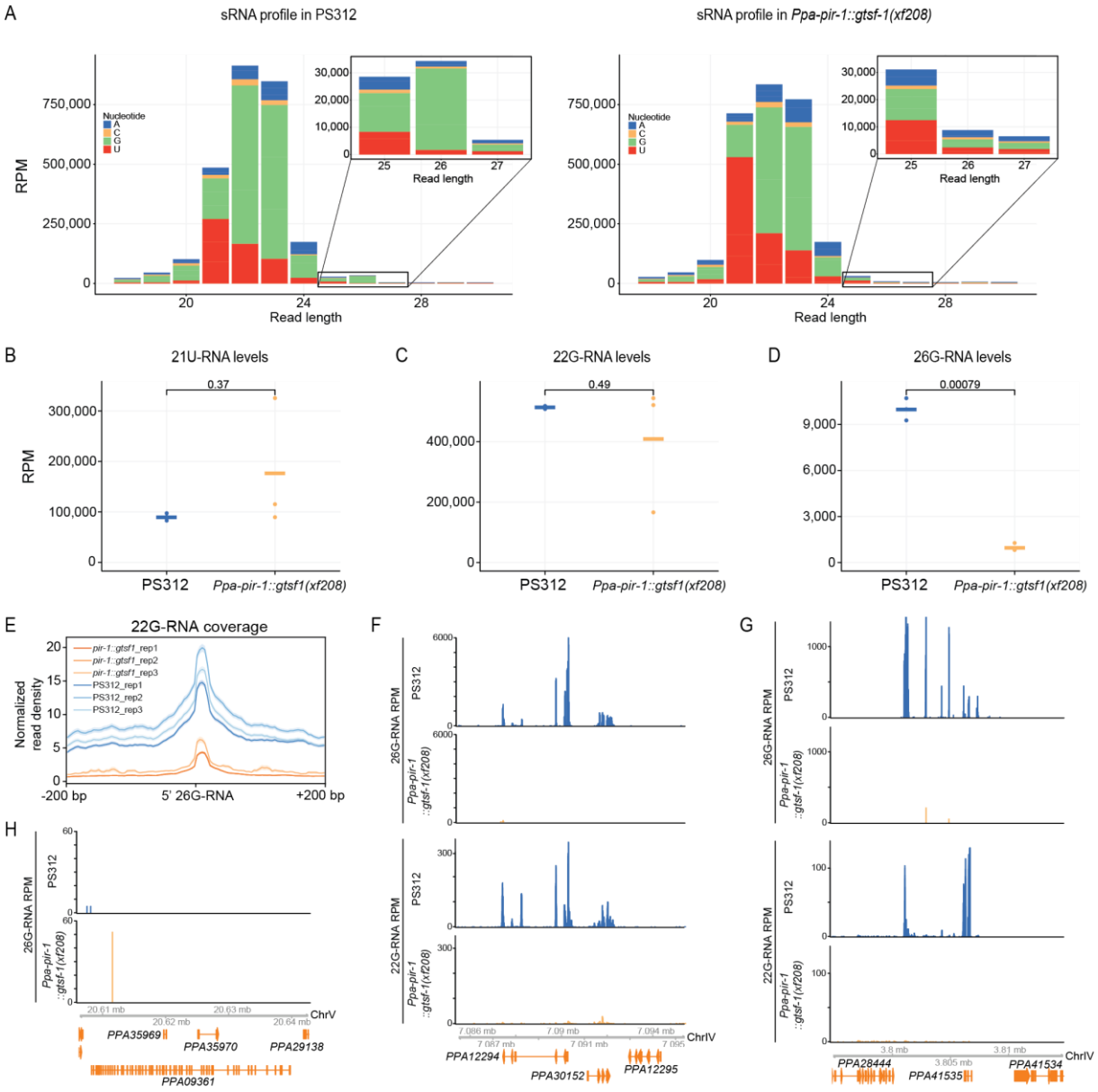


Figure IV.6. sRNA sequencing of wild-type and *Ppa-pir-1::gtsf-1* mutant embryos. (A) Normalized levels of sRNAs, in RPM (Reads Per Million), summarizing the length and 5' ribonucleotide bias in wild-type PS312 (left panel) and *Ppa-pir-1::gtsf-1* mutants (right panel). The 5' ribonucleotide bias is color-coded. Given the low overall abundance of 26G-RNAs, the inset displays a rescaled view of the sRNAs that are 25-27 ribonucleotides long. (B-D) Normalized levels of specific sRNA species, in RPM, in PS312 and *Ppa-pir-1::gtsf-1* mutants. Three biological replicates are shown. P-values were calculated with a two-sided unpaired t-test. (B) Total levels of 21U-RNAs. (C) Total levels of 22G-RNAs. (D) Total levels of 26G-RNAs. (E) Metagene analysis of 22G-RNA levels at, and around 26G-RNA sites in wild-type PS312 (blue lines) and *Ppa-pir-1::gtsf-1* mutant (orange lines) embryos. (F-G) Genome browser tracks displaying the normalized 26G- (panels above) and 22G-RNA (panels below) levels mapping to *PPA12294* and *PPA30152* (F), and *PPA41535* (G), both in wild-type (in blue) and in the mutant (in orange). (H) Genome browser tracks displaying 26G-RNA levels, in RPM, mapping to the single *Cel-alg-3/4* ortholog of *P. pacificus*, gene *PPA09361*.

RNAi genes between species, when compared to the background proteome, led to the conclusion that these genes display signatures of positive selection and may therefore be evolving fast.

Given the astonishing functional plasticity of GTSF-1 proteins, we sought to determine whether nematode GTSF-1 proteins are evolving fast. To address this, we performed dN/dS analysis of GTSF-1 protein sequences in four *Caenorhabditis* nematode species: *C. elegans*, *C. briggsae*, *C. brenneri* and

	C. elegans GTSF-1	*C. briggsae* CBG13683	*C. brenneri* CBN18835	*C. remanei* CRE30384
C. elegans GTSF-1		1.25 (0.21)	1.10 (0.14)	1.06 (0.22)
C. briggsae CBG13683			1.06 (0.16)	0.71 (0.27)
C. brenneri CBN18835				1.08 (0.20)

Table IV.2. dN/dS ratios of *Caenorhabditis* GTSF-1 proteins. Ratios between two species are indicated, followed by a background value, between parentheses, of the average dN/dS ratio for 20 genes not related to RNAi.

C. remanei. *gtsf-1* ortholog genes were identified in *C. brenneri* and *C. remanei* by reciprocal BLAST. Indeed, GTSF-1 proteins have higher dN/dS ratios than the background (average dN/dS of 1.04, versus a background average of 0.20) (**Table IV.2**). Moreover, we find similarly high dN/dS values for nematode homologs of RRF-3 (average dN/dS of 1.16, versus a background average of 0.20). To conclude, a crude evolutionary analysis supports the fast evolution of GTSF-1 and RRF-3 proteins in nematodes.

Discussion

The results described in this chapter are the first steps into an extension of *gtsf-1* studies to additional nematode species. To this end, we created and studied *gtsf-1* mutants in satellite nematode models. We find that, much like in *C. elegans*, *gtsf-1* is required for fertility at higher temperatures in *C. briggsae* and *P. pacificus* (**Figures IV.2C** and **IV.5B**), thereby displaying a conserved phenotype throughout 200-300 million years of evolutionary time. *Cel-gtsf-1* shares its fertility defects, which are sperm-derived and temperature-sensitive, with *alg-3/4* and *rrf-3*, amongst others. In *Cel-alg-3/4* mutants, sperm is incapable of proper activation, precluding successful fertilization of oocytes, especially at higher temperatures (Conine et al., 2010, 2013). Future experiments, should determine whether the temperature-sensitive sterility of *gtsf-1* mutants in *P. pacificus* and *C. briggsae* is due to defective sperm.

Surprisingly, we find that only Ppa-PIR-1::GTSF-1 is required for the biogenesis of 26G-RNAs and of downstream 22G-RNAs, in embryos (**Figures IV.6**). *P. pacificus* has two annotated genes orthologous to *Cel-ergo-1*. It will be interesting to determine if Ppa-ERGO-1 paralogs associate with embryonic 26G-RNAs, and if there are signs of functional diversification between the two. Also, *P. pacificus* animals undergoing spermatogenesis should be sequenced in order to address the existence of spermatogenic 26G-RNAs, and to define the contribution of Ppa-PIR-1::GTSF-1 to their

biogenesis and function. To our knowledge, this is the first described RNAi mutant gene in *P. pacificus*. More functional studies, by creating and characterizing mutant alleles, are required to better understand the sRNA pathways of *P. pacificus*.

A previous study has sequenced sRNAs from *C. briggsae* embryos and detected 26G-RNAs (Shi et al., 2013). Moreover, the same study has identified four *Cbr-ergo-1* ortholog genes. However, the genetic requirements for 26G-RNA biogenesis were not defined. We have shown that Cbr-GTSF-1 is not required for global 26G-RNA biogenesis in embryos (**Figure IV.3**). This is hard to reconcile with the requirement of Ppa-PIR-1::GTSF-1 and Cel-GTSF-1 for embryonic 26G-RNAs, especially since *C. elegans* and *C. briggsae* are more closely related than *C. elegans* and *P. pacificus*. We envision three possible explanations. A first possibility is that Cbr-GTSF-1 specialized in the biogenesis of spermatogenic 26G-RNAs, since in *C. elegans* it is the lack of spermatogenic, not embryonic, 26G-RNAs that underlies the temperature-sensitive fertility defects. Profiling sRNA populations in wild-type and *Cbr-gtsf-1* mutant animals undergoing spermatogenesis should shed light on this aspect. Second, the *Cbr-gtsf-1* mutant allele we isolated may be an hypormorph. *Cbr-gtsf-1(xf207)* consists of an in-frame deletion of 6 bp in the first CHHC zinc finger (**Figure IV.2A-B**). The two deleted amino acids are directly C-terminal to the highly conserved CPY residues, and directly N-terminal to the first histidine residue. As demonstrated in **Chapter II** (**Figure II.6**), interaction with RRF-3 is only fully abolished if the two CHHC zinc fingers are simultaneously mutated, suggestive of structural and functional robustness between the two zinc fingers. It is possible to envision that at 20 ^{o}C, a non-stressful temperature, 26G-RNA biogenesis still ensues aided by a Cbr-GTSF-1 protein with only one fully functional zinc finger. However, when grown at 30 ^{o}C, a stressful temperature, mutant Cbr-GTSF-1 may not be able to initiate 26G-RNA production, therefore leading to sterility. To test this possibility, fertility at 20 ^{o}C should be quantified in wild-type and mutant *C. briggsae*, and other stronger, frameshift mutant alleles should be isolated and characterized. Lastly, Cbr-GTSF-1 may have a function unrelated to the 26G-RNA pathway. Although intriguing, considering the role in 26G-RNA biogenesis in *P. pacificus*, this last possibility is the most exciting. Since GTSF-1 proteins seem to be evolving fast (**Table IV.2**), it may be that Cbr-GTSF-1 was co-opted by another sRNA pathway, or by an unrelated biological process. Identification of proteins interacting with Cbr-GTSF-1 may shed some light on this aspect. To this end, we attempted an immunopurification of Cbr-GTSF-1 using our anti-Cel-GTSF-1 polyclonal antibody (used in **Chapter II**), with no success (data not shown). Assuming a fraction of the four Cbr-ERGO-1 proteins has been co-opted in other sRNA pathway(s), it is possible that Cbr-GTSF-1 was similarly co-opted.

Interestingly, we demonstrate that orthologs of *Cel-pir-1* and *Cel-gtsf-1* are fused in *P. pacificus* (**Figure IV.4**). However, at this point we have no means of confirming the existence of a fused protein. We tried Western blot analysis of PIR-1::GTSF-1 by using our polyclonal anti-Cel-GTSF-1 antibody.

Unfortunately, the antibody did not unambiguously recognize a specific band (data not shown). If true, a PIR-1::GTSF-1 fusion protein has potential implications for 26G-RNA pathway biology and ERI complex formation. In *C. elegans*, GTSF-1 is in a precursor ERI complex with ERI-5 and RRF-3 (see **Chapter II** and Thivierge et al., 2012). GTSF-1 and ERI-5 tether RRF-3 to the remaining ERI factors to create a mature ERI complex that can initiate 26G-RNA biogenesis. Cel-PIR-1 was not found to strongly associate with GTSF-1 in the pre-ERI complex, neither in embryos, nor in adult animals (**Chapter II, Figures II.5**, **II.7** and **II.S5**). Future experiments should test whether Ppa-PIR-1::GTSF-1 maintains a tethering function of RRF-3 to the ERI complex. Also, the relative amounts of Ppa-PIR-1::GTSF-1 present in a precursor complex and mature complex should be determined.

There is a low level of synteny between *C. elegans* and *P. pacificus* (Lee et al., 2003), making it hard to formulate an hypothesis regarding the generation of the fused locus. However, of note, we detected an open reading frame coding for a transposase, flanked by inverted repeats, that is approximately 1 kb upstream of Ppa-PIR-1::GTSF-1. We hypothesize that this TE may have dragged the sequence homologous to *pir-1* together with the *gtsf-1* sequence. Unfortunately, other than comparing this region with a syntenic region in other closely related nematodes of the *Pristionchus* genus, we envision no other experiments that could test this hypothesis.

We extend reports of previous studies, reporting that genes involved in RNAi are fast evolving (Obbard et al., 2009; Palmer et al., 2018), to GTSF-1 proteins of *Caenorhabditis* nematodes (**Table IV.2**). The fast evolution of RNAi genes may be easily interpreted in the light of the “Red Queen” hypothesis, as a result of an evolutionary arms race between the RNAi machinery and the target genetic element (Obbard et al., 2009; Palmer et al., 2018; Parhad et al., 2017; Pearson, 2001; Simkin et al., 2013; Vermaak et al., 2005). The observed higher dN/dS ratios than non-RNAi genes is likely to reflect a signature of positive selection. We note that the analysis showed in **Table IV.2** is very low throughput. Future analysis should be more thorough, include GTSF-1 orthologs of more nematode species and the entire proteome as background.

Alas, we could not extend our functional studies to *C. inopinata*, the recently identified sister species of *C. elegans* (Kanzaki et al., 2018). *C. briggsae* was until recently the species of choice for comparative studies with *C. elegans*. But *C. briggsae* and *C. elegans* have diverged approximately 100 million years ago, a divergence time similar to that of mouse and human (Stein et al., 2003). Taking into account its evolutionary closeness to *C. elegans*, its amenability for transgenesis and exogenous RNAi as a knock-down tool (Kanzaki et al., 2018), *C. inopinata* will undoubtedly be a treasure trove of important comparative biology studies in the coming years. The RNAi field is no exception, particularly taking into account the astonishing genomic expansion of TEs, and the absence of an *ergo-1* gene (Kanzaki et al., 2018). The latter makes it even more enticing to functionally characterize the 26G-RNA pathway of this nematode.

BLAST searches throughout the available genomes/transcriptomes of flatworms (Platyhelminthes) did not detect any GTSF-1 orthologs. In nematodes, we could not detect GTSF-1 orthologs in clade I and II species. However, outside clade V, we could identify by reciprocal BLAST *gtsf-1* genes in many clade III and IV species, including for example, *Ascaris suum* and *B. malayi*, which are parasites of mammals. Ultimately, extension of studies of RNAi genes to other nematode clades may help in the development of effective gene therapies to control parasitic species with negative clinical and economic impact.

Materials and methods

Nematode genetics and culture

C. elegans was cultured on OP50 bacteria according to standard laboratory conditions (Brenner, 1974). Animal populations were maintained at 20 °C. The AF16 and PS312 strains were used as the standard wild-type strain of *C. briggsae* and *P. pacificus*, respectively. All strains used and created in this study are listed in **Table IV.3**.

Strain reference	Genotype	Available in CGC	Comments
	Wild-type N2	Yes	
AF16	*C. briggsae* wild isolate	Yes	
RFK948	*Cbr-gtsf-1(xf207)*		
PS312	*Pristionchus pacificus* wild isolate	Yes	
RFK949	*Ppa-pir-1::gtsf-1(xf208)*		Mutations in the sequence homologous to *Cel-gtsf-1*
RFK950	*Ppa-pir-1::gtsf-1(xf209)*		
	sp.34 wild isolate		A kind gift by Taisei Kikuchi

Table IV.3. List of nematode species and strains used/produced in this study

Microscopy

Wide-field photomicrograph were acquired using a Leica M165FC microscope with a Leica DFC450 C camera, and were processed using Leica LAS software and ImageJ.

CRISPR/Cas9 in *C. briggsae*

We constructed a sgRNA that successfully targeted the following sequence in *Cbr-gtsf-1* (*CBG13683*): ATCACCTGTCCTTATAGAG. Adult AF16 worms were injected with a mix of 50 ng/µl of Cas9 construct pMVA280 (Peft-3::Cas9::tbb-2 3'UTR; PU6::sgRNA) and 5 ng/µl of co-injection marker pCFJ90

(Pmyo-2::mCherry:unc-54 3' UTR, mCherry expression in the pharynx). F1 worms positive for mCherry expression in the pharynx were isolated, allowed to self, and then lysed in single worm lysis buffer (5 mM KCl, 2.5 mM MgCl2, 10 mM Tris–HCl pH = 8.3, 0.45% NP-40, 0.45% Tween-20, and 0.01% gelatin). Subsequently, PCR was performed with Taq Polymerase according to manufacturer's instructions (New England BioLabs, M0273X), to amplify the region targeted by the sgRNA. Molecular lesions were detected by Sanger sequencing.

CRISPR/Cas9 in *P. pacificus*

To hybridize target-specific crRNA and universal tracrRNA (IDT, ALT-R product line), 10 µL of the 100 µM stock of each molecule are combined, denatured at 95 °C for 5 min and allowed to cool down and anneal at room temperature for 5 min. 5 µL of the hybridization product are combined with 1 µL of 62 µM Cas9 protein (IDT Cat.#1081058) and incubated at room temperature for 5 min. Then 1 µL of 100 µM repair oligo is added to the mixture. The mixture is diluted with TE buffer to the total volume of 25 µL and injected in the gonad rachis in young hermaphrodites that have 0-2 eggs in the uterus. The crRNAs are designed to target 20 bp upstream of PAMs and BLASTed against the El Paco assembly to verify the absence of off-target sites. The length of the homology arms may vary between 35 and 50 bp. Molecular lesions were detected by Sanger sequencing.

Brood size experiments

L2-L3 *C. briggsae* worms grown at 20 °C were isolated and moved to 30 °C. For *P. pacificus*, J2-J3 animals grown at 20 °C were isolated and moved to 27 °C. Animals were transferred to a new plate every day, until egg laying stopped. Number of viable progeny were counted ~24 hours after removing the parent.

Animal growth and collection, and RNA isolation

For small RNA sequencing and RT-PCR, embryo samples were collected from bleached gravid adult animals, followed by thorough washes with M9. For RT-PCR, young adults were isolated from animal populations synchronized by bleaching. Samples were collected in triplicate. 50 µL of M9 plus worms/embryos were subsequently frozen in dry ice. For RNA isolation frozen young adult/embryo aliquots were thawed, 500 µL of Trizol LS (Life Technologies, 10296-028) was added and mixed vigorously. Next, we employed six freeze-thaw cycles to dissolve the embryos/young adult animals: tubes were frozen in liquid nitrogen for 30 seconds, thawed in a 37 °C water bath for 2 minutes, and mixed vigorously. Following the sixth freeze-thaw cycle, 1 volume of 100% ethanol was added to the samples and mixed vigorously. Then, we added these mixtures onto Direct-zol columns (Zymo

Research, R2070) and manufacturer's instructions were followed (in-column DNase I treatment was included).

RT-PCR

Reverse transcription was performed in 1 µg of total RNA with ProtoScript First Strand cDNA Synthesis Kit (New England Biolabs, E6300), using the random primer mix, according to manufacturer's instructions. PCR was performed with Q5 High-Fidelity DNA Polymerase (New England Biolabs, M0491) according to manufacturer's instructions. Primers were used in a final concentration of 200 nM. Cycling conditions were as follows: 2 minutes at 95°C; X cycles of [95 °C for 30 seconds, 60 °C for 30 seconds and 68 °C for a duration according with the amplicon length – using a 1kb per 60 seconds reference]; and 68 °C for 5 minutes. Only one of two technical duplicates are represented in **Figure 4**. Technical duplicates produced consistent results. **Table IV.4** includes the primers and the number of cycles used for amplification.

RT-PCR oligo_amplicon(s)	**Sequence**
Fw1_A1	ATGTGTCAGCAGTTAGACAACGT
Rev1_A1_A2_A3_A6	TTATTTGTATCCGTACTCCTTTTCTCC
Fw2_A2	ATCGGGCGCATCTGATGATT
Fw3_A3_A4	TGTCCACTGCACTCATGGTC
Rev2_A4_A5	CTGCTCCTCAATTTGCGCTC
Fw4_A5	GAATCCCCTCCTCGAAGCTG
Fw5_A6	TCCCCTCCTCGTCGTTGTAT
Number of cycles used for amplification:	
Amplicon 1 – 40 cycles	
Amplicon 2 – 35 cycles	
Amplicon 3 – 35 cycles	
Amplicon 4 – 35 cycles	
Amplicon 5 – 33 cycles	
Amplicon 6 – 33 cycles	

Table IV.4. List of the oligos used for RT-PCR and the number of PCR cycles employed per amplicon.

RppH treatment and library preparation for small RNA sequencing

For maternal effect sequencing, RNA was directly used for library preparation, or treated with RppH prior to library preparation. RppH treatment was performed as described in (Almeida et al., 2019) with slight modifications. 500 ng of *C. briggsae* RNA and 1 µg of *P. pacificus* RNA were used for RppH

treatment. In short, RNA was incubated with 5 units of RppH and 10x NEB Buffer 2 for 1 hour at 37°C. Reaction was stopped by incubating the samples with 500 mM EDTA for 5 minutes at 65°C. RNA was reprecipitated in 100% Isopropanol and ressuspended in nuclease-free water. 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). *C. briggsae* libraries were prepared with a starting amount of 250 ng and amplified in 16 PCR cycles. *P. pacificus* libraries were prepared with a starting amount of 500 ng and amplified in 15 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). All 12 samples (6 samples of *C. briggsae* and 6 samples of *P. pacificus*) were pooled together in an equimolar ratio and sequenced on 1 NextSeq 500/550 Midoutput Flowcell, SR for 1x 75 cycles plus 7 cycles for the index read.

Bioinformatic analysis

Small RNA read processing and mapping. Illumina adapters were removed with cutadapt v1.9 (Martin, 2011)(-a TGGAATTCTCGGGTGCCAAGG -O 5 -m 26 -M 38) and reads with low-quality calls were filtered out with fastq_quality_filter (-q 20 -p 100 -Q 33) from the FASTX-Toolkit v0.0.14. Using information from unique molecule identifiers (UMIs) added during library preparation, reads with the same sequence (including UMIs) were collapsed to removed putative PCR duplicates using a custom script. Prior to mapping, UMIs were trimmed (seqtk trimfq -b 4 -e 4) and reads shorter than 15 nucleotides (nt) were discarded (seqtk seq -L 15). Library quality was assessed with FastQC twice, for the raw and for the processed reads. Processed reads were aligned against either the *C. briggsae* or the *P. pacificus* genome assembly WBPS11 with bowtie v0.12.8 (Langmead et al., 2009) (–tryhard –best –strata -v 1 -M 1). Genome assemblies and gene annotations were obtained from Wormbase (Parasite).

Small RNA class definition and quantification. To define particular classes of small RNA, mapped reads were categorized as follows: 21U-RNAs are considered those sequences that are exactly 21 nt long, and have a uridine at their 5'; 22G-RNAs are those whose sequence is 20-23 nt and have a guanine at their 5'; 26G-RNAs are those which are 26 nt, and have a guanine at their 5'. Read filtering was done with a python script available at https://github.com/adomingues/filterReads/blob/master/filterReads/filterSmallRNAclasses.py which relies on pysam v0.8.1 an htslib wrapper (Li et al., 2009). Reads fulfilling these definitions were then counted for each library (total levels). Genome browser tracks were created using Bedtools (genomeCoverageBed -bg -split -scale -ibam -g), to summarize genome coverage normalized to

mapped reads * 1 million (RPM), followed by bedGraphToBigWig to create the bigwig track. The three replicates were considered cumulatively for the sRNA length profile.

Metagene analysis. Putative 26G-RNA target sites were identified by taking the 5' of each 26G-RNA in the WT, defined as described above, keeping only the position with at least 5 supporting reads. The average 22G-RNA coverage around each of these sites was determined with deepTools v2.4.3 (computeMatrix reference-point --missingDataAsZero --b 250 -a 250 --binSize 5 --averageTypeBins median) and plotted with plotProfile --plotType se --averageType mean --perGroup.

dN/dS analysis of GTSF-1 protein sequences

Using the BLAST tool in wormbase.org, we identified *gtsf-1* and *rrf-3* ortholog genes in *C. briggsae*, *C. brenneri* and *C. remanei* by reciprocal BLAST. *C. remanei* has three paralog *gtsf-1* genes. Two shorter paralogs, mapping exclusively to the C-terminal portion of Cel-GTSF-1 were excluded from this analysis. Only the long paralog, which displays sequence similarity with the entire Cel-GTSF-1 protein, was used. In addition, when more than one protein isoform are predicted, we used the longest isoform for the analysis. With the protein sequences in hand, we performed a multiple sequence alignment in MUSCLE (at https://www.ebi.ac.uk/Tools/msa/muscle/) using the default parameters. The resulting alignment was subsequently used as input in SNAP (Synonymous Non-synonymous Analysis Program) v2.1.1., which computed the dN/dS ratios. Background genes were selected from the entire ortholog gene list, obtained with InParanoid (at http://inparanoid.sbc.su.se/cgi-bin/index.cgi). Targets were randomly chosen, but size was taken into account: proteins of all sizes were chosen, but a preference was given to small proteins with a size more similar to that of Cel-GTSF-1 (approximately half of the proteins). Genes with multiple paralogs were avoided. The 20 non-RNAi genes chosen as the background were the following: ACL-5, NRA-4, LIN-52, CLEC-112, HLH-13, GRSP-3, RPL-34, PRK-1, PQN-88, SHN-1, FIS-1, ALH-10, MES-3, RPN-5, APC-11, TOS-1, ZOO-1, SRD-5, SMA-1, ZYG-9.

Acknowledgements

We thank all the members of the Ketting lab for all the help and good discussions, in particular Svenja Hellmann for excellent technical assistance. We are very grateful to Miguel Andrade for the great input and suggestions regarding the dN/dS analysis. We would like to express our gratitude to Hanna Lukas, Clara Werner and Maria Mendez-Lago of the IMB genomics core facility for library preparation. We thank the IMB Media Lab for consumables. We also acknowledge the *Caenorhabditis* Genetics Center (CGC), which is funded by NIH Office of Research Infrastructure Programs (P40

OD010440), for providing worm strains. This work was supported by a Deutsche Forschungsgemeinschaft grant KE 1888/1-1 (Project Funding Programme to R.F.K.).

References

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

Blaxter, M. (2011). Nematodes: The Worm and Its Relatives. PLOS Biol. *9*, e1001050.

Borges, F., and Martienssen, R.A. (2015). The expanding world of small RNAs in plants. Nat. Rev. Mol. Cell Biol. *16*, 727–741.

Brenner, S. (1974). The Genetics of Caenorhabditis Elegans. Genetics *77*, 71–94.

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

Chaves, D.M. de A. de M. (2015). The RNA 5' phosphatase PIR-1 cooperates with dicer to produce endogenous small RNAs and suppress viral replication in C. elegans.

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

Dalzell, J.J., McVeigh, P., Warnock, N.D., Mitreva, M., Bird, D.M., Abad, P., Fleming, C.C., Day, T.A., Mousley, A., Marks, N.J., et al. (2011). RNAi Effector Diversity in Nematodes. PLoS Negl. Trop. Dis. *5*, e1176.

Dieterich, C., Clifton, S.W., Schuster, L.N., Chinwalla, A., Delehaunty, K., Dinkelacker, I., Fulton, L., Fulton, R., Godfrey, J., Minx, P., et al. (2008). The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism. Nat. Genet. *40*, 1193–1198.

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.

Duchaine, T.F., Wohlschlegel, J.A., Kennedy, S., Bei, Y., Conte Jr., 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.

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*.

Gao, F., Liu, X., Wu, X.-P., Wang, X.-L., Gong, D., Lu, H., Xia, Y., Song, Y., Wang, J., Du, J., et al. (2012). Differential DNA methylation in discrete developmental stages of the parasitic nematode Trichinella spiralis. Genome Biol. *13*, R100.

Holz, A., and Streit, A. (2017). Gain and Loss of Small RNA Classes—Characterization of Small RNAs in the Parasitic Nematode Family Strongyloididae. Genome Biol. Evol. *9*, 2826–2843.

Kanzaki, N., Tsai, I.J., Tanaka, R., Hunt, V.L., Liu, D., Tsuyama, K., Maeda, Y., Namai, S., Kumagai, R., Tracey, A., et al. (2018). Biology and genome of a newly discovered sibling species of Caenorhabditis elegans. Nat. Commun. *9*, 3216.

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, K.-Z., Eizinger, A., Nandakumar, R., Schuster, S.C., and Sommer, R.J. (2003). Limited microsynteny between the genomes of Pristionchus pacificus and Caenorhabditis elegans. Nucleic Acids Res. *31*, 2553–2560.

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

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal *17*, 10–12.

McDonald, J.H., and Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in Drosophila. Nature *351*, 652–654.

Obbard, D.J., Gordon, K.H.J., Buck, A.H., and Jiggins, F.M. (2009). The evolution of RNAi as a defence against viruses and transposable elements. Philos. Trans. R. Soc. Lond. B Biol. Sci. *364*, 99–115.

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

Palmer, W.H., Hadfield, J.D., and Obbard, D.J. (2018). RNA-Interference Pathways Display High Rates of Adaptive Protein Evolution in Multiple Invertebrates. Genetics *208*, 1585–1599.

Parhad, S.S., Tu, S., Weng, Z., and Theurkauf, W.E. (2017). Adaptive Evolution Leads to Cross-Species Incompatibility in the piRNA Transposon Silencing Machinery. Dev. Cell *43*, 60-70.e5.

Pearson, P.N. (2001). Red Queen Hypothesis. Encycl. Life Sci. 1–4.

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.

Shi, Z., Montgomery, T.A., Qi, Y., and Ruvkun, G. (2013). High-throughput sequencing reveals extraordinary fluidity of miRNA, piRNA, and siRNA pathways in nematodes. Genome Res. *23*, 497–508.

Simkin, A., Wong, A., Poh, Y.-P., Theurkauf, W.E., and Jensen, J.D. (2013). Recurrent and recent selective sweeps in the piRNA pathway. Evolution *67*, 1081–1090.

Stein, L.D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M.R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., et al. (2003). The Genome Sequence of Caenorhabditis briggsae: A Platform for Comparative Genomics. PLOS Biol. *1*, e45.

Thivierge, C., Makil, N., Flamand, M., Vasale, J.J., Mello, C.C., Wohlschlegel, J., Jr, D.C., 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.

Tu, S., Wu, M.Z., Wang, J., Cutter, A.D., Weng, Z., and Claycomb, J.M. (2014). Comparative functional characterization of the CSR-1 22G-RNA pathway in Caenorhabditis nematodes. Nucleic Acids Res. gku1308.

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. *107*, 3582–3587.

Vermaak, D., Henikoff, S., and Malik, H.S. (2005). Positive Selection Drives the Evolution of rhino, a Member of the Heterochromatin Protein 1 Family in Drosophila. PLOS Genet. *1*, e9.

Wang, J., Czech, B., Crunk, A., Wallace, A., Mitreva, M., Hannon, G.J., and Davis, R.E. (2011). Deep small RNA sequencing from the nematode Ascaris reveals conservation, functional diversification, and novel developmental profiles. Genome Res.

de Wit, E., Linsen, S.E.V., Cuppen, E., and Berezikov, E. (2009). Repertoire and evolution of miRNA genes in four divergent nematode species. Genome Res.

Chapter V

Discussion

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

Almeida, M.V., Andrade-Navarro M., and Ketting R.F. (2019). Function and evolution of nematode RNAi pathways. Noncoding RNA *5*, 1-24.

Miguel Andrade-Navarro constructed the phylogenetic trees shown in **Figure V.4**.

GTSF1 proteins: small builders of protein complexes

In **Chapter II**, the core work of this thesis, we provide mechanistical insights into the function of Cel-GTSF-1. We have shown that Cel-GTSF-1, a small protein with two CHHC zinc fingers and an acidic C-terminal tail, is in a precursor complex with the RdRP RRF-3 and the Tudor domain protein ERI-5. To allow for functional 26G-RNA biogenesis, GTSF-1 and ERI-5 bring RRF-3 together with the remaining ERI complex. Moreover, we show that the tandem CHHC zinc fingers of GTSF-1, unlike previously assumed, provide a robust module for protein-protein interactions.

The mode of action and function of Cel-GTSF-1 differs substantially from that of Dmel-Gtsf1. Cel-GTSF-1 is in the cytoplasm, interacting with an RdRP to elicit small RNA biogenesis, while Dmel-Gtsf1 is in the nucleus, interacting with a Piwi AGO to elicit TGS of TEs. Despite these two very striking scenarios, which demonstrate the functional plasticity of GTSF1 proteins, we postulate that a common denominator of GTSF1 function exists. We propose a universal function of GTSF1 proteins in assembling protein complexes with effector function (**Figure V.1A**): GTSF1 proteins use their CHHC zinc fingers and/or C-terminal tail to interact with proteins involved in sRNA pathways and guide those to other accessory factors, thereby promoting maturation of complexes and execution of effector functions. Irrespective of their exact function, the protein complexes involved associate with RNA. Whether GTSF1 proteins bind RNA directly and how target RNA affects the GTSF1-dependent complex assembly remains to be determined. Our efforts in *C. elegans* did not unveil any connection with RNA, but additional *in vitro* experiments, like electrophoretic mobility shift assays, could shed some light on this aspect.

In the context of RNAi-driven heterochromatin formation of fission yeast, Stc1 has a similar function as Cel-GTSF-1, not by assembling protein complexes per se, but by bridging distinct complexes (Bayne et al., 2010; He et al., 2013). Stc1 displays features similar to GTSF1 proteins: two N-terminal tandem zinc fingers arranged in a LIM domain, and an unstructured, acidic C-terminal tail. Stc1 was shown to interact with Ago1, via its two zinc fingers, and to Clr4 via its C-terminal tail (Bayne et al., 2010; He et al., 2013). With its two interaction surfaces, Stc1 can bring together the RITS and the CLRC complexes, allowing for the establishment of a positive feedback loop that ultimately consolidates heterochromatin at pericentromeric regions. Besides the CHHC zinc fingers of Cel-GTSF-1, we found no evidence in line with the C-terminal tail providing an additional protein-protein interaction surface. This suggests that GTSF-1 may not act in a manner perfectly analogous to Stc1. Exactly how Cel-GTSF-1 tethers RRF-3 to the remaining ERI complex requires further dissection, especially to investigate the role of the C-terminal tail. Conversely, similar to Stc1, GTSF-1 proteins in mouse were shown to interact with distinct piRNA pathway factors via their CHHC zinc fingers and C-terminal tail (Takemoto et al.,

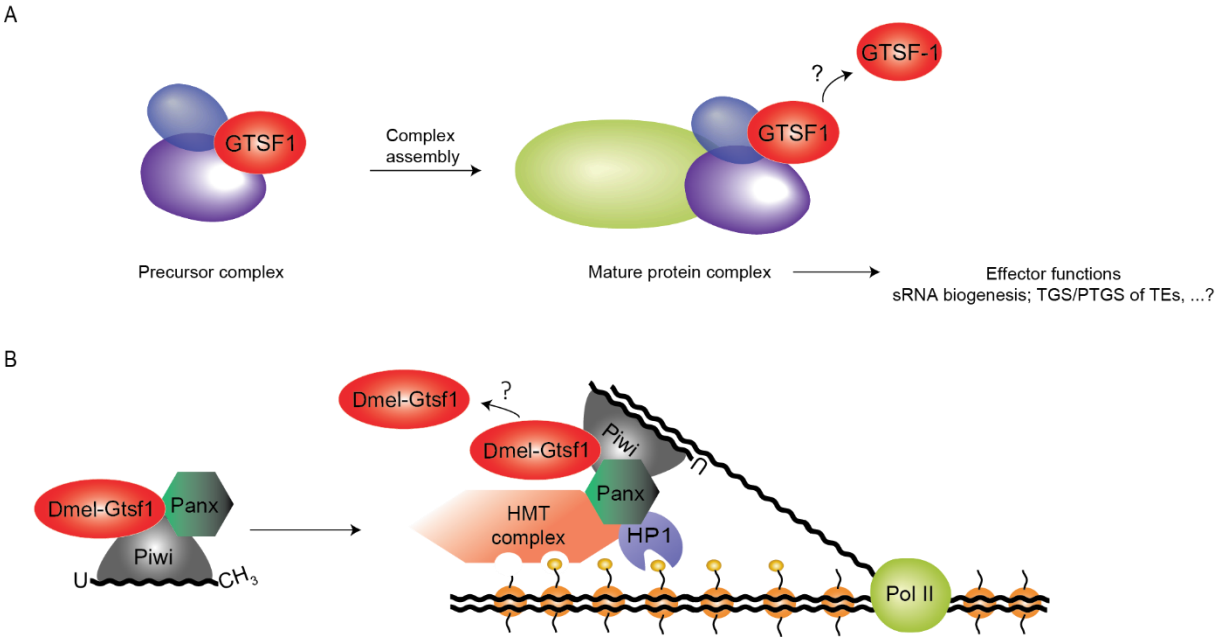


Figure V.1. **A unified model of the function of GTSF1 proteins**. (A) Proposed unified model. GTSF1 proteins interact with multiple proteins, thereby bringing together distinct subunits of protein complexes. Alternatively, GTSF1 may not directly interact with distinct protein modules, but it may, for example, cause a conformational change of its interactors, allowing further interactions and complex assembly. A fraction of GTSF1 proteins may dissociate from the mature complex. It is currently unknown if target RNA is required for GTSF1-dependent complex assembly. (B) A working model of Dmel-Gtsf1 function in *D. melanogaster* embryos. Within the nucleus, Dmel-Gtsf1 assists Piwi and Panx in assembling a complex that will ultimately silence euchromatic transposon insertions on the transcriptional level. sRNA, small RNA; TGS, transcriptional gene silencing; PTGS, post-transcriptional gene silencing; TEs, transposable elements; Dm, *Drosophila melanogaster*; Pol II, RNA Polymerase II; HMT, histone methyltransferase.

2016; Yoshimura et al., 2018). However, these interaction data were obtained using only *in vitro* assays and therefore should be interpreted with caution, as the *in vivo* physiological relevance of these interactions was not determined. Further biochemical examination of the functions of mouse GTSF-1 is required. Likewise, future biochemical experiments should address whether GTSF1 assembles a complex in *D. melanogaster*. We propose that Dmel-GTSF1 and Panoramix may be required to place Piwi within an effector complex with HMTs and HP1 proteins (**Figure V.1B**). In parallel to *C. elegans*, we speculate that Dmel-GTSF1 activity is especially relevant during early development. In accordance, the activity of Piwi is required during embryogenesis to elicit TE silencing later in development (Akkouche et al., 2017). Interactomics studies in *Drosophila* embryos would clarify this hypothesis.

Beyond model organisms

Historical contingency determined the initial collection and adoption of the *model organisms* we have nowadays. Sydney Brenner chose the *C. elegans* N2 Bristol strain instead of N1, or perhaps instead of another nematode species. Likewise, Thomas Hunt Morgan chose *Drosophila* instead of another arthropod. Indeed, these organisms display certain key biological features that make them more experimentally tractable and adaptable to lab conditions. However, nowadays we no longer have

the same technological and experimental constraints of the past. We are witnessing the age of genome editing, brought about by CRISPR/Cas9 technology. Nowadays, every organism is a potential model organism. And indeed, the deeper we delve into the biology of non-model organisms, the more we understand that 1) our model organisms are not representative of their whole class; and 2) that every organism is extremely derived. Consider, for example, the early development of *Drosophila*. Its study has brought great insight into basic developmental processes. However, *Drosophila* undergo long germ band embryogenesis, unlike most arthropods, which undergo short germ band development (Davis and Patel, 2002). Hence, *Drosophila* early development is poorly representative of arthropod early development. Many other examples exist, including, in line with the subject of this thesis, the unprecedented number of AGO proteins of *Caenorhabditis* nematodes (Shi et al., 2013; Yigit et al., 2006). Data obtained in one organism are often extrapolated as being representative of an entire class. In many cases, this may be misleading.

Ideally, to have a holistic understanding of the biology of a specific protein or pathway, one should also take into account the evolutionary history of the factors at play. This logic is all the more important when considering proteins that are evolving fast, as is the case of RNAi factors in general, and of GTSF1 proteins specifically. With this in mind, in **Chapter IV** we initiated the study of fast evolving nematode GTSF1 proteins of species closely related to *C. elegans*. Adding to previous studies of GTSF1 proteins, performed only in evolutionarily distant organisms, the findings reported in **Chapter IV** underscore the functional plasticity of GTSF-1 proteins and highlight the need for further functional studies. For instance, we have shown that Cbr-GTSF-1, surprisingly, is not required for 26G-RNA biogenesis in *C. briggsae* embryos (**Figure IV.3**). Future continuation of these studies should reveal the function of Cbr-GTSF-1. Moreover, we report that in *P. pacificus* the ortholog of Cel-GTSF-1 is expressed as a fused transcript with an ortholog of Cel-PIR-1. Cel-PIR-1 is a phosphatase of unclear function that integrates the ERI complex (Duchaine et al., 2006). The Ppa-PIR-1::GTSF-1 fusion is fascinating and implies a mechanism of ERI complex assembly in *P. pacificus* distinct to that of *C. elegans,* because Cel-GTSF-1 and Cel-PIR-1 are not in a pre-ERI complex (**Figures II.5A-B**, **II.6E**, and **II.7B**). In sum, an important lesson from this work is that the conclusions of studies on fast evolving factors using common model organisms, such as mouse and fruit fly, should not be extrapolated to other systems.

The parental gift of small RNAs

In general, sRNAs and AGOs are most important for gametogenesis and embryonic development, and are often deposited into embryos by the parental gametes (Akkouche et al., 2017; Brennecke et al., 2007, 2008; Czech et al., 2008; Houwing et al., 2007; Kawaoka et al., 2011; Le Thomas et al., 2014; Le Thomas et al., 2014; Martinez and Köhler, 2017; Ninova et al., 2017; Tam et al., 2008;

de Vanssay et al., 2012; Watanabe et al., 2008). In *C. elegans*, interestingly, all sRNA classes are parentally contributed to the next generation. 21U- and 22G-RNAs are maternally and paternally deposited in embryos (de Albuquerque et al., 2015; Gu et al., 2009; Luteijn et al., 2012; Minkina and Hunter, 2017; Phillips et al., 2015; Schott et al., 2014; Stoeckius et al., 2014; Wan et al., 2018; Xu et al., 2018). Likewise, 26G-RNAs are maternally and paternally provided, the latter in low numbers (Chapter III and (Han et al., 2009; Stoeckius et al., 2009, 2014; Vasale et al., 2010). In **Chapter III**, we show that maternal contribution of 26G-RNAs is restricted to the oogenic ERGO-1 branch (**Figures III.1** and **III.S1**). Eri mutants lacking ERGO-1 branch 26G- and 22G-RNAs display a maternal rescue of the Eri phenotype and ERGO-1 target silencing (**Figures III.1** and **III.S1**) (Zhuang and Hunter, 2011). In **Chapter III**, we have elucidated the dynamic interplay between maternal and zygotic ERGO-1 branch sRNA populations in establishing gene silencing of targets throughout development (**Figures III.1-III.3** and **III.S1-III.S3**). Maternally inherited 26G-RNAs trigger biogenesis of zygotic 22G-RNAs in the embryo, which can establish target gene silencing. Curiously, in the absence of maternal 26G-RNAs, zygotic 26G-RNAs can still trigger 22G-RNA biogenesis and establish gene silencing, attesting for the robustness of this system (**Figure III.1F**).

Parentally provided sRNAs were also shown to play an important role in relation to the function of CSR-1 in protecting against erroneous gene silencing. It was demonstrated that in the simultaneous absence of parental 21U-RNAs and RNAe memory, in the form of 22G-RNAs and/or histone tail modifications, animals that can produce 22G-RNAs are sterile (de Albuquerque et al., 2015; Phillips et al., 2015). Further dissection revealed that this synthetic sterility arises because WAGOs typically involved, for example, in TE silencing, such as HRDE-1, start to silence typical CSR-1 targets (de Albuquerque et al., 2015; Phillips et al., 2015). These studies illustrate that a memory of gene expression has to be transmitted to the progeny via sRNAs in order to ensure bona fide gene expression and proper development.

What can explain the ubiquity of sRNA inheritance in plants and animals? A possible answer may lie in the expression of non-self sRNA targets. For example, TEs often have expression restricted to the germline and to embryonic tissues (Chuong et al., 2017), because such expression pattern will favor vertical dissemination of TEs, in contrast to somatic expression. One can thus envision that selective pressure to control TEs, or other non-self genetic elements, during germline and embryonic development may have favored sRNA expression in gametes and embryos. This may indeed be the case for ERGO-1 branch sRNAs and respective targets, since the latter are more highly expressed in embryos (**Figure III.3C**, lower panels). Overall, sRNAs likely represent an important adaptive parental gift to the next generation in order to resume the repression of non-self genetic elements potentially deleterious to genome stability.

Target regulation by ALG-3/4 and ERGO-1

Perhaps due to their low abundance, it remains largely unknown how 26G-RNAs and associated factors act to regulate their targets. Also, it is completely unknown how the ERI complex is positioned at target mRNAs and whether its assembly occurs directly at the target RNA.

The answer may partially be ERI-6/7, a homolog of the MOV10L1 and Armitage helicases, which is required for the accumulation of ERGO-1 branch 26G- and 22G-RNAs, but not of ALG-3/4 branch 26G-RNAs (Fischer et al., 2008, 2011). In accordance, *eri-6/7* mutants share the Eri phenotype with *ergo-1*, *rrf-3,* and other Eri genes (Fischer et al., 2011). Interestingly, ERI-6/7 was not found to physically associate with RRF-3, GTSF-1, DCR-1, ERI-1, or ERI-5, indicating that this factor is not an integral part of the ERI complex (**Figures II.5A-B** and **II.7B-C**) (Duchaine et al., 2006; Thivierge et al., 2012). Mouse MOV10L1 and *Drosophila* Armitage recruit piRNA precursors to initiate piRNA biogenesis (Pandey et al., 2017; Vourekas et al., 2015). Therefore, an attractive hypothesis based on ortholog gene function would be a role for ERI-6/7 in defining target transcripts upon which the ERI complex can be loaded to drive 26G-RNA biogenesis. Artificially tethering ERI-6/7 to non-ERI complex targets and probing for *de novo* 26G-RNA biogenesis would shed light on this hypothesis. However, it should be noted that since ERI-6/7 affect only ERGO-1 branch 26G- and 22G-RNAs, it would only account for the definition of ERGO-1 branch targets. Other yet unidentified factors may account for definition of ALG-3/4 targets.

An alternative hypothesis, not necessarily mutually exclusive with a role of ERI-6/7 in selecting ERGO-1 targets, consists of target definition based on non-optimal splicing. A recent study showed that ERGO-1 targets are overall small, poorly conserved genes with few introns (Newman et al., 2018). Furthermore, the introns of these genes tend to have poor splicing consensus sequences, in comparison to endogenous genes. The authors went on to show that spliceosomes are enriched on these target transcripts (Newman et al., 2018). These observations contributed to a working model whereby ERGO-1 target transcripts are signaled as foreign by the lack of optimal splicing signals, and concomitantly accumulate stalled spliceosomes. A similar phenomenon was reported in the yeast *Criptococcus neoformans* (Dumesic et al., 2013). It should be noted that *C. neoformans* has nuclear RNAi factors that directly interact with spliceosomes, but it is unclear how cytoplasmic ERGO-1 activity would connect to nuclear spliceosomes in *C. elegans*. ALG-3/4 targets are also overall small genes with a low number of introns (see for example *ssp-16* gene browser tracks in **Figures II.S4D**, **III.S2B** and **III.S3B**), but Newman and colleagues did not experimentally address if spliceosomes are enriched in these genes. Future studies should concretely address whether small, fast evolving genes with non-optimal splicing signals may accumulate spliceosomes and stimulate 26G-RNA biogenesis.

The regulatory effects of ALG-3/4 on their targets is complex and dependent on temperature. At 20 ^{o}C, a temperature not stressful for *C. elegans*, the regulation of ALG-3/4 targets seems to be

predominantly repressive (**Figure III.3B-C, III.S3B, III.4G** and **III.5**) (Conine et al., 2010). However, ALG-3/4 appear to function in promoting gene expression at elevated temperatures in the male germline (**Figure V.2A**) (Conine et al., 2013). Interestingly, such target regulation by ALG-3/4 was shown to be linked to CSR-1 and may occur on the transcriptional level (Conine et al., 2013). The authors proposed a model in which spermatogenic ALG-3/4-RISCs trigger biogenesis of downstream 22G-RNAs, which in turn associate with CSR-1. These CSR-1-RISCs would be paternally transmitted to the next generation, thereby providing a paternal memory of germline gene expression (Conine et al., 2013). Of note, dependence of spermatogenesis on temperature is not specific to *C. elegans*. Instead it appears to be a recurring phenomenon throughout animal evolution (Wallach et al., 1988), and it will be interesting to know if these temperature effects are more generally linked to sRNA pathways.

What dictates the regulatory outcome of genes targeted by ALG-3/4? So far, this is not clear. As shown in **Figure III.5**, abundance of mapped 26G- and 22G-RNAs may dictate the regulatory outcome of a target (**Figure V.2B**). Moreover, the very striking patterns of origin of ALG-3/4 branch 26G-RNAs may contribute to the regulatory outcome (**Figures III.6** and **III.S5**) (Conine et al., 2010), with an enrichment on the 5' and 3' terminal regions of target transcripts. Individual ALG-3/4 targets can have predominant 26G-RNA peaks on the 5' or on the 3', or both peaks equally abundant. We have shown in **Chapter III** that negatively regulated ALG-3/4 targets have more abundant levels of 26G- and 22G-RNAs, as well as predominant 5' 26G-RNA targeting (**Figures V.2C, III.5**-**III.6**, and **III.S5**). Predominant 3' targeting by 26G-RNAs tends to be correlated with weaker silencing or even licensing of gene expression. In accordance, ALG-3/4 targets that have predominant 3' 26G-RNA targeting were found to be more highly expressed than those targeted predominantly on the 5'. These observations are consistent with a model in which 5' versus 3' end targeting by ALG-3/4-RISCs is somehow coupled to the regulatory outcome. Clearly, this understanding is incomplete. For instance, an effect of 3' UTR length on ALG-3/4-mediated regulation was also found (**Figures III.6F** and **III.S5E**) (**Figure V.2D**). More studies are needed to thoroughly dissect the apparent positive and negative target regulatory effects of ALG-3/4 and how they are coupled to temperature.

Self-perpetuation of 26G-RNA-dependent 22G-RNAs

21U-RNAs can initiate a very stable form of gene silencing termed RNAe (Ashe et al., 2012; Luteijn et al., 2012; Shirayama et al., 2012). After establishment, RNAe is most notable and considered truly epigenetic, because it can self-perpetuate even in the absence of the initial 21U-RNA trigger. Interestingly, as we show in **Chapter III**, targets of 26G-RNAs can also maintain 22G-RNAs in the absence of the primary 26G-RNA trigger, in what may be RNAe-like mechanisms. ALG-3/4 branch 22G-RNAs are partially depleted in adult males in response to *gtsf-1* mutation, while in young adults their levels are unaffected (**Figures V.3A, III.2H**, and **III.4F**). Furthermore, maternal 26G-RNAs elicit

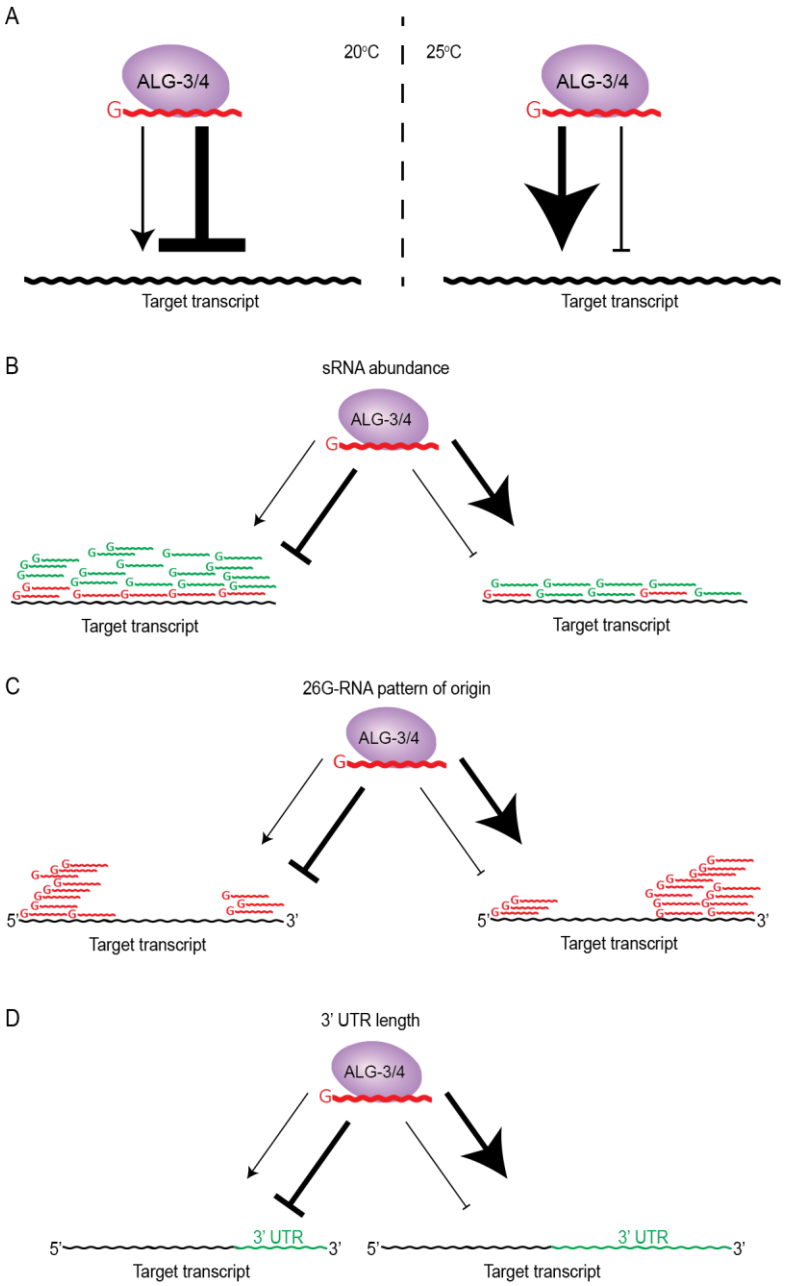


Figure V.2. **Determinants of target regulation by ALG-3/4**. (A) Distinct regulatory programs are in place at different temperatures. At 20 ºC, ALG-3/4 have predominantly repressive activity, while at 25 ºC, an elevated growth temperature for *C. elegans*, ALG-3/4 show a stronger tendency to positively regulate target gene expression. sRNA abundance (B), pattern of origin of 26G-RNAs (C), and 3' UTR length (D) influence the regulatory outcome.

zygotic production of 22G-RNAs, which in turn are maintained throughout development, silencing their targets even in the absence of the primary maternal 26G-RNA trigger (**Figures V.3B**, **III.1-III.3**, and **III.S1-III.S3**). Lastly, although ERGO-1 branch 26G-RNAs are expressed only in oogenesis and in embryos, ERGO-1 targets are still targeted by relatively abundant 22G-RNA populations in adult males, suggesting that these 22G-RNAs are maintained in the absence of primary 26G-RNA triggers (**Figures V.3B** and **III.4F**). NRDE-3 is a good candidate AGO to carry on silencing of ERGO-1 targets in the adult male. However, expression and function of NRDE-3 in the male soma were not addressed thus far.

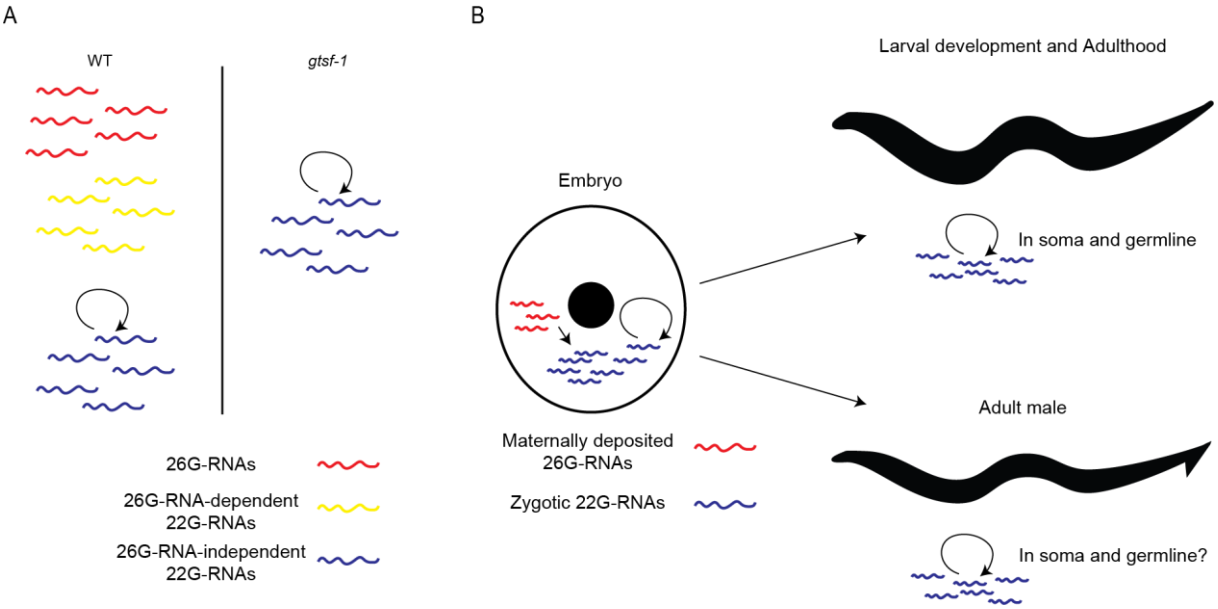


Figure V.3. **Self-perpetuation of 26G-RNA-dependent 22G-RNAs**. (A) Specific subpopulations of 22G-RNAs do not respond to *gtsf-1* mutation, suggesting the existence of additional mechanisms of 22G-RNA maintenance. (B) Maternal 26G-RNAs trigger the biogenesis of secondary 22G-RNAs which, in turn, self-perpetuate in the absence of the initial primary trigger. Self-perpetuation occurs both throughout hermaphrodite and male development.

Altogether, self-perpetuating 22G-RNA populations in the absence of the primary trigger appear to be a recurring theme in *C. elegans* RNAi-like pathways, but the underlying molecular mechanisms are unresolved. We envision two non-mutually exclusive mechanistic bases for the self-perpetuation of sRNA populations. First, 22G-RNAs can serve as primary triggers that stimulate the production of more 22G-RNAs. Although such mechanistic basis may be possible for endo-RNAi, it should be noted that 22G-RNAs secondary to primary exogenous dsRNA triggers have been shown not to induce further 22G-RNA production (Pak et al., 2012). Second, since secondary, but not primary, RNAi pathways involve a nuclear RNAi component, the chromatin environment of target *loci* may be key in the self-perpetuation of 22G-RNAs. This model implies that transcripts originating from silenced target *loci* will somehow be marked, e.g. by specific proteins, and upon export into the cytoplasm, the marks could promote further RdRPs activity.

Regulating the regulators: cross- and self-regulation of RNAi-like pathways

RNAi in *C. elegans* typically involves rare primary sRNA triggers that elicit the production of abundant secondary 22G-RNAs in a self-reinforcing positive feedback loop. Since such positive feedback amplification mechanisms can exert a considerable strain on biological systems, there must be mechanisms to limit their unbridled continuation. As previously mentioned, the Eri phenotype is likely a reflection of competition of exogenous and endogenous RNAi pathways for shared limiting factors (Duchaine et al., 2006; Lee et al., 2006). It is plausible that such extensive interactions evolved

to limit uncontrolled amplification of RNAi responses. In support of this, exogenous RNAi was shown to affect endogenous sRNA populations and their inheritance (Houri-Ze'evi et al., 2016). Over a period of 2-3 generations, the endogenous RNAi machinery counteracts and ultimately dilutes the effect of exogenous RNAi. Also, secondary 22G-RNAs arising from exogenous RNAi do not trigger further RdRP activity (Pak et al., 2012). Why are such mechanisms employed to control exogenous RNAi responses, but self-perpetuation of 22G-RNAs against endogenous genetic elements still exist? An attractive hypothesis is that exogenous RNAi must be especially limited in order to allow flexibility and incorporation of new environmental inputs in a rapidly changing environment. An observation reported in **Chapter III** adds one more regulatory mechanism to AGOs. ALG-3 and ALG-4 were found to bind 26G-RNAs that target and dampen the expression of their own mRNAs (**Figures III.7** and **III.S6**). This is the first example in *C. elegans* of AGOs directly regulating their own expression. In sum, the extensive AGO repertoire of *C. elegans* and its intricate RNAi pathways cross-regulate and self-regulate in a myriad of ways, allowing robust, but finite responses to developmental and environmental cues. It should be noted that cross-regulation of non-coding RNA pathways is not fully unprecedented outside nematodes. For example, a recent study from the Bartel lab described a regulatory network in the mammalian brain involving two miRNAs, a lncRNA, and a circular RNA (Kleaveland et al., 2018).

To be, or not to be, a nematode piRNA

The putative common ancestor to all nematodes most likely expressed miRNAs and a cognate ALG-1/2-like AGO, DCR-1, Piwi AGO(s), endogenous sRNAs made by an RRF-3-like RdRP, RNA-directed heterochromatin formation, and DNA methylation (Sarkies et al., 2015). As a result of broad loss of Piwi genes but overall AGO family expansion in nematodes, we argue that multiple AGOs and sRNA species adopted piRNA-like features. We will substantiate this claim by drawing parallels between nematode RNAi pathways and metazoan Piwi/piRNA pathways, both in mechanistic and functional terms, highlighting common ground and dissimilarities.

Germline expression

Metazoan Piwi-RISCs are highly expressed in germ cells (Huang et al., 2017; Iwasaki et al., 2015). Hence, germline-specificity could be an argument for defining the Piwi pathway of *C. elegans*. However, several findings argue against such simple classification. First, many *C. elegans* pathways appear to be highly germ cell specific (**Table V.1**). Indeed, 21U-RNAs and PRG-1 are highly enriched in the germline and are required for normal fertility, but also ALG-3/4, ERGO-1, and their cognate 26G-RNAs, as well as other ERI complex factors, are strongly enriched in the spermatogenic and oogenic germlines, respectively (**Figure II.1D-G**) (Conine et al., 2010; Han et al., 2009; Vasale et al.,

		Metazoan piRNAs	21U-RNAs	26G-RNAs
	Expression	Predominantly germline and embryos	Germline and embryos	Germline and embryos
	Length (in nucleotides)	23-35	21	~26
	5’ Bias	U	U	G
	Phenotype	Mutants are sterile	Viable; transgenerational germline mortality	Some mutants cause sterility at higher temperatures
Cofactors	Piwi clade Argonautes			
	Hen1 enzymes			
	Gtsf1 proteins			
	Armitage/MOV10L1/ ERI-6/7			

Table V.1. Comparison between metazoan piRNA pathways and main *C. elegans* RNAi-like pathways. Green and gray colored boxes indicate the existence or absence, respectively, of interaction of the cofactors with the corresponding sRNA class.

2010). Second, a number of WAGO proteins has been shown to be specifically expressed in germ cells (e.g. HRDE-1, WAGO-1, and WAGO-4) (Buckley et al., 2012; Gu et al., 2009; Ishidate et al., 2018; Xu et al., 2018). Third, in further support of the idea that tissue specificity does not provide a handle for defining a Piwi pathway, somatic Piwi/piRNA expression has been convincingly demonstrated in many arthropods and mollusks (Jehn et al., 2018; Lewis et al., 2017). In fact, such studies hint that the ancestral metazoan used somatic Piwi/piRNA pathways that became increasingly compartmentalized in the germline. We conclude that tissue specificity does not provide any support for classifying a particular *C. elegans* RNAi-like pathway as “the” Piwi/piRNApathway.

Function

Piwi/piRNA pathways comprise specialized RNAi-like pathways that recognize and silence the non-self, most notably TEs (Huang et al., 2017; Iwasaki et al., 2015; Luteijn and Ketting, 2013). In this light, piRNA pathways can be thought of as immune systems, given their ability to recognize the non-self, initiate a response, amplify the response, and keep memory of the contact, thereby safeguarding future contacts. However, this aspect does not provide a foothold to unambiguously define a *C. elegans* piRNA pathway, as both 21U- and 26G-RNA pathways target distinct sets of non-self transcripts and share these recognition/amplification/memory features that are key principles of immune systems.

Mutation of Piwi in flies and mice leads to a range of gametogenesis defects that result in sterility (Huang et al., 2017; Iwasaki et al., 2015). Conversely, *prg-1* mutant worms are viable but have

reduced brood sizes (Batista et al., 2008; Das et al., 2008; Wang and Reinke, 2008). As mentioned in the **Introduction** of **Chapter IV**, PRG-1 orthologs and 21U-RNAs have been lost in multiple nematode lineages (Holz and Streit, 2017; Sarkies et al., 2015; Wang et al., 2011). These observations suggest that PRG-1/21U-RNA pathways may be less essential to viability than other piRNA systems (**Table V.1**). Conversely, the strongly conserved ALG-3/4-like AGOs are required for normal fertility in *C. elegans* (Conine et al., 2010). Therefore, no *C. elegans* RNAi pathway strictly displays complete sterility, again blurring a parallel with other metazoan piRNA pathways.

Lastly, piRNAs are required in embryos to prime gene silencing that is maintained until adulthood (Akkouche et al., 2017). Similarly, the embryonic activity of both 21U-RNAs and 26G-RNAs is required during embryogenesis (**Figures III.1-III.3**) (de Albuquerque et al., 2015; Luteijn et al., 2012; Phillips et al., 2015). When taking all these arguments into account, no single nematode RNAi pathway is a good direct counterpart of metazoan Piwi/piRNA pathways.

sRNA features

Are there sRNA features that can help us define "the" worm piRNA pathway? 21U-RNAs have a distinct 5' uridine bias which is shared by 21U-RNAs (Batista et al., 2008; Das et al., 2008; Huang et al., 2017; Iwasaki et al., 2015; Ruby et al., 2006). Indeed, PRG-1-bound sRNAs have a very strong 5' uridine bias. However, the length profile of 21U-RNAs, which are almost uniquely 21 nucleotides long, differs from typical piRNAs. The latter mostly span from 26-32 nucleotides and the populations show a characteristic bell-shaped length distribution (**Table V.1**). 26G-RNAs have a more piRNA-like length distribution but display a different 5' nucleotide bias. Moreover, 26G-RNAs are produced by an RdRP (Conine et al., 2010; Han et al., 2009; Vasale et al., 2010), not by RNA Pol II as typical piRNA precursors (Huang et al., 2017; Iwasaki et al., 2015). Metazoan piRNAs are commonly not defined by individual promoters, a feature shared by 26G- and 22G-RNAs. In contrast, type-I 21U-RNAs are individually transcribed from their own Ruby motif.

piRNAs are 3' 2'-O-methylated by Hen1 proteins (Billi et al., 2012; Horwich et al., 2007; Kamminga et al., 2010, 2012; Lim et al., 2015; Montgomery et al., 2012; Saito et al., 2007). Again, this is not a feature unique to one *C. elegans* pathway, as both 21U-RNAs and ERGO-1 class 26G-RNAs are similarly methylated (Billi et al., 2012; Kamminga et al., 2012; Montgomery et al., 2012). Overall, a single worm pathway homologous to other metazoan piRNA pathways cannot be identified based on sRNA features.

Sequence homology of AGO proteins

21U-RNAs were initially classified as piRNAs precisely because of their interaction with PRG-1 (Batista et al., 2008; Das et al., 2008; Wang and Reinke, 2008). PRG-1 is without a doubt most closely

related to other metazoan Piwi clade AGOs. However, ERGO-1 was also shown to be relatively closely related to the Piwi clade phylogenetically (Billi et al., 2012; Montgomery et al., 2012; Yigit et al., 2006). Our analysis of full length AGO proteins, rooted with a Prokaryotic AGO, shows that ERGO-1 is still part of the Ago clade, although very basal (**Figure I.3C**). AGOs have several distinct domains with defined functions: the MID domain, which binds the 5' extremity of the sRNA; the PAZ domain, which accommodates the 3' end of the sRNA; and the Piwi domain, which is the catalytic domain of AGOs that mediates cleavage. To understand the phylogenetic relationship of ERGO-1 with Piwi clade AGOs, we performed phylogenetic analysis of AGO proteins by domain.

This deeper analysis shows that the MID domain of ERGO-1 does not cluster with any of the eukaryotic AGO clades (**Figure V.4A**). We could not detect homologs for this region of ERGO-1 outside *Caenorhabditis*, even using iterative searches (neither with PSIBLAST nor with HMMER). In fact, although ERGO-1 has to retain the ability to bind the 5' end of the sRNA, its MID domain seems to be quite degenerate (data not shown). The PAZ domain of ERGO-1 is more closely related to Wago clade AGOs than with the Piwi clade, even though it has to accommodate 3' 2'-O-methylated sRNAs, much like the PAZ domain of Piwi AGOs (**Figure V.4B**). Moreover, the PIWI domain of ERGO-1 is strongly related to that of Ago clade AGOs (**Figure V.4C**). Our phylogenetic analysis does not support previous classifications of ERGO-1 as a Piwi protein. We argue that ERGO-1 is part of the Ago clade and has a Wago-like PAZ domain. Hence, given that the domains of AGO proteins have non-overlapping functions, establishing phylogenetic relationships based on full sequence alignments may mask domain-specific information. Future phylogenetic studies should therefore analyze the different domains separately.

Evolutionarily conserved cofactors

Another argument that could be used to define one *C. elegans* RNAi-related pathway as the main equivalent to a metazoan piRNA pathway is the existence of shared cofactors. However, as we will show below, many examples exist that argue against this possibility (**Table V.1**).

Members of the GTSF1 protein family were found to interact with Piwi proteins in flies (Dönertas et al., 2013; Ohtani et al., 2013) and mice (Takemoto et al., 2016; Yoshimura et al., 2018). *C. elegans* has just one GTSF1 ortholog, which does not interact with PRG-1 and 21U-RNAs (**Table V.1**). Instead, as shown in **Chapter II**, GTSF-1 was co-opted in *C. elegans* for the biogenesis of both classes of 26G-RNAs, by interacting with the RdRP RRF-3.

A similar scenario holds true for the *C. elegans* helicase ERI-6/7, which is required for the accumulation of ERGO-1-class 26G-RNAs (**Table V.1**) (Fischer et al., 2008, 2011). Its homologs in flies and mice, Armitage and MOV10L1, respectively, are piRNA pathway factors (Saito et al., 2010; Vourekas et al., 2015).

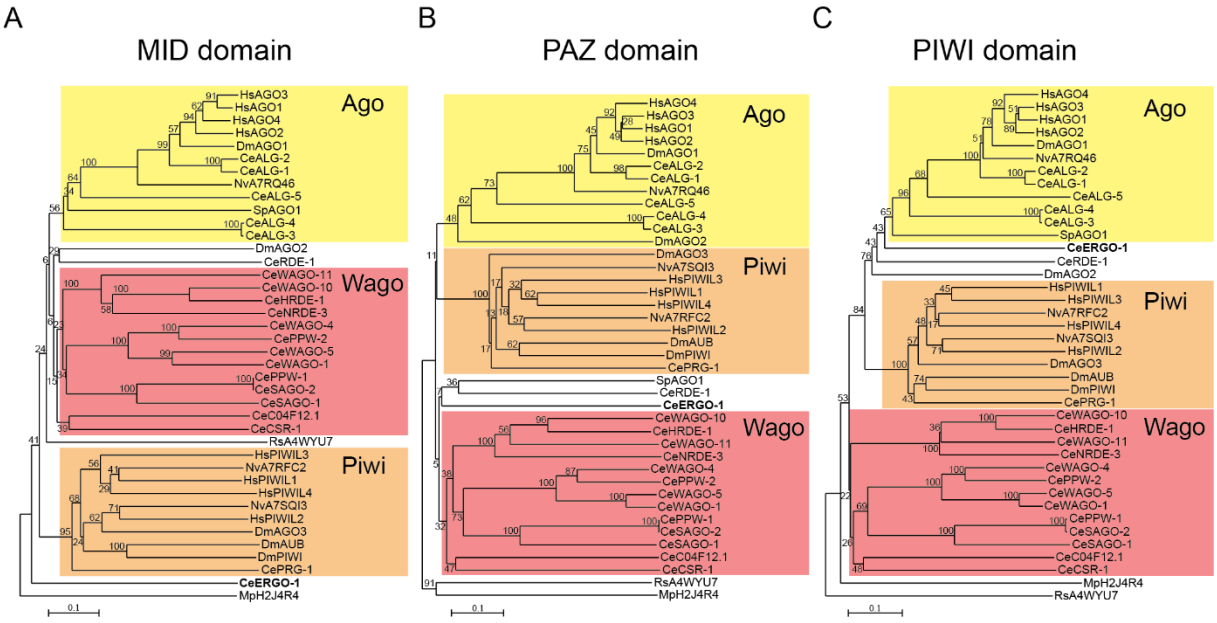


Figure V.4. Phylogenetic analysis of the MID, PAZ and PIWI domains of AGO proteins. Phylogenetic trees of the MID (A), PAZ (B) and PIWI (C) domains of AGO proteins. ERGO-1 is represented in bold. The multiple sequence alignment used in Chapter I was trimmed to the coordinates of either the MID, the PAZ or the PIWI domains, respectively, and each tree was constructed as in Figure I.3C. AGOs are represented by a two letter prefix indicating the species, followed by the AGO name or UniProt ID. Monophyletic groups including most of the members of one clade are grouped by color. Species legend: Ce, *Caenorhabditis elegans*; Dm, *Drosophila melanogaster*; Hs, *Homo sapiens*; Mp, *Marinitoga piezophila*; Nv, *Nematostella vectensis*; Rs, *Rhodobacter sphaeroides*; Sp, *Schizosaccharomyces pombe*.

Finally, metazoan piRNA pathways typically involve proteins with Tudor domains (Iwasaki et al., 2015; Ketting, 2011). In *C. elegans*, two Tudor domain-containing proteins, ERI-5 and EKL-1 were implicated in 26G-RNA biogenesis and CSR-1 function, respectively (Claycomb et al., 2009; Thivierge et al., 2012). In contrast, no Tudor domain protein has yet been described to act in the PRG-1 pathway.

Similar to the other discussed features, conservation of cofactors does not help to define a single sRNA pathway of *C. elegans* as the main counterpart of metazoan piRNA pathways. Instead, different aspects are represented in different pathways, although most of these tend to cluster in the 21U- and 26G-RNA pathways.

Historically, an evident first step in understanding nature is to categorize organisms, genes, and pathways into distinct classes. While such classification is certainly very useful, it can also be artificial and not fully reflect evolutionary trajectory. We reason that nematode RNAi pathways have blurred borders separating what we define as piRNAs and siRNAs in other metazoans. We believe that referring exclusively to 21U-RNAs as piRNAs is too simplistic and that it may in some cases even be misleading. We propose using the already existing nematode-specific nomenclature to indicate specific sRNA pathways in worms: 21U-, 26G- and 22G-RNA pathways and to refrain from talking about a *C. elegans* piRNA pathway. In fact, a less general terminology overall may be called for in piRNA studies, since numerous differences exist between fly and mouse piRNA pathways, making it dangerous to promptly generalize results.

Genetic conflict and beyond in the non-self perspective

This thesis focused on the function and evolution of RNAi pathways. The tone assumed may have inadvertently seemed one-sided, portraying the so-called non-self genetic elements as the proverbial "bad guys". As a final note, I will discuss: 1) how the non-self fights back; 2) how the non-self can have positive effects on organismal evolution; and 3) how the terms of self and non-self may have a tenuous distinction in some situations.

Host RNAi machinery and non-self genetic elements are engaged in an evolutionary arms race. In this light, many viruses have evolved the capacity to suppress host RNAi (Li and Ding, 2006; Obbard et al., 2009). Thus, one can envision the arms race as an unrelenting dance between RNAi adaptation to evade viral suppression and subsequent counter-adaptation by the viral suppression machinery. Viral suppression of RNAi is common in animals and is especially widespread in plants (Obbard et al., 2009). Viral suppressors of RNAi are typically proteins, but viral RNA molecules have also been implicated in suppression. Many mechanisms are employed to suppress key steps in RNAi pathways like, for example, inhibition of sRNA biogenesis and sRNA sequestration (Li and Ding, 2006; Obbard et al., 2009). For unknown reasons, unlike viruses, there are no documented cases in the literature of TEs suppressing RNAi machinery.

I highlighted in **Chapter I** the need to control non-self genetic elements like TEs. Indeed, the mobility of TEs generates genome instability by creating double-strand breaks (Chuong et al., 2017). Also, TEs may insert in or near protein-coding genes, influencing their expression. Both possibilities can have detrimental effects on the host organism, thus affecting its fitness and, if occurring in the germline, the fitness of the next generation. Besides the fertility defects commonly observed in Piwi mutants correlating with TE mobilization, TEs have been shown to be misregulated in disease in numerous cases (Chuong et al., 2017). However, it is still unclear whether TE mobilization is a cause or a consequence of the disease state. Notwithstanding the negative roles, I will now argue that TEs are extremely abundant elements in eukaryotic genomes that are increasingly recognized as major players in shaping genomic and transcriptomic landscapes (Chuong et al., 2017). In this light, TEs can be beneficial by providing raw genetic variation, or by changing pre-existing expression patterns/levels.

TE elements adapt to host transcription and translation machinery by including cis-regulatory sequences that are read by host machinery. For example, several TE elements, like Long-terminal repeat (LTR) elements and Long interspersed nuclear elements (LINEs), carry their own RNA Pol II promoters. This means that LTRs and LINEs, via their RNA Pol II promoters, have the potential to generate new, or influence pre-existing transcriptional units (Chuong et al., 2017). In support of this claim, TEs underlie the origin of a large number of lncRNAs (Chuong et al., 2017; Kapusta et al., 2013).

Many examples can be found in the literature where TEs provide adaptive *in cis* changes to gene expression. In mice and rats, a truncated, oocyte-specific Dicer isoform, termed Dicer^{O}, is created

by alternative splicing of an exon derived from an LTR element (Flemr et al., 2013). This change makes Dicer^{O} more efficient in processing dsRNA into siRNAs, turning Dicer^{O}-dependent endo-siRNAs into the predominant RNAi pathway in oocytes, in lieu of the piRNA pathway. Another fascinating example comes from a recent study elucidating the genetic basis of a textbook case-study of adaptation: the industrial melanism of the English peppered moth (van't Hof et al., 2016). The authors showed that a TE insertion within an intron of a protein-coding gene causes the appearance of the darker moth phenotype. In addition, TEs are often repurposed as enhancers and insulators and may thereby assume key roles in the formation of large genome domains (Choudhary et al., 2018; Chuong et al., 2017; Kruse et al., 2019). Taken together, the observations above demonstrate that TEs can provide adaptive changes to gene expression beneficial to hosts. Also, TEs not only affect gene expression *in cis*, but in fact have much deeper functions in shaping genome structure throughout evolution.

The dichotomic view presented in this thesis, for example in **Chapter I**, regarding self and non-self is reductionist at best. TEs and sequences of retroviral origin can become completely "domesticated", at which point these logically become part of the "self". Thus, one can envision a spectrum of genetic elements ranging from completely "foreign" to completely "domesticated". Importantly, the factors and mechanisms that determine silencing or licensing of TEs need to be clarified in the future.

Concluding remarks

RNAi-like pathways are broadly used by animals, plants, and fungi in the de-escalation of genetic conflict between hosts and non-self sequences. As a result of this relentless arms race, RNAi-like pathways are evolving fast. The existence of several species- or genus-specific RNAi factors and RNAi factor interactions supports such view. The extremely diverse phylum Nematoda epitomizes this evolutionary fluidity, with its many AGOs and sRNA classes interacting in complex ways, both inside and outside the germline, to regulate gene expression in a myriad of processes, including gametogenesis and embryogenesis. Further studies on non-model nematodes, including parasitic species of medical and veterinary relevance, are required to elucidate more aspects on the biology of these species, including how RNAi pathways have been, and are being evolved.

References

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. Molecular Cell *66*, 411-419.e4.

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.

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. Molecular Cell *31*, 67–78.

Bayne, E.H., White, S.A., Kagansky, A., Bijos, D.A., Sanchez-Pulido, L., Hoe, K.-L., Kim, D.-U., Park, H.-O., Ponting, C.P., Rappsilber, J., et al. (2010). Stc1: A Critical Link between RNAi and Chromatin Modification Required for Heterochromatin Integrity. Cell *140*, 666–677.

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.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. Cell *128*, 1089–1103.

Brennecke, J., Malone, C.D., Aravin, A.A., Sachidanandam, R., Stark, A., and Hannon, G.J. (2008). An Epigenetic Role for Maternally Inherited piRNAs in Transposon Silencing. Science *322*, 1387–1392.

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.

Choudhary, M.N., Friedman, R.Z., Wang, J.T., Jang, H.S., Zhuo, X., and Wang, T. (2018). Co-opted transposons help perpetuate conserved higher-order chromosomal structures. BioRxiv 485342.

Chuong, E.B., Elde, N.C., and Feschotte, C. (2017). Regulatory activities of transposable elements: from conflicts to benefits. Nature Reviews Genetics *18*, 71–86.

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.

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. Proceedings of the National Academy of Sciences of the United States of America *107*, 3588–3593.

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

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.

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. Molecular Cell *31*, 79–90.

Davis, G.K., and Patel, N.H. (2002). Short, Long, and Beyond: Molecular and Embryological Approaches to Insect Segmentation. Annual Review of Entomology *47*, 669–699.

de Albuquerque, B.F.M., Placentino, M., and Ketting, R.F. (2015). Maternal piRNAs Are Essential for Germline Development following De Novo Establishment of Endo-siRNAs in Caenorhabditis elegans. Developmental Cell *34*, 448–456.

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

Duchaine, T.F., Wohlschlegel, J.A., Kennedy, S., Bei, Y., Conte Jr., 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.

Dumesic, P.A., Natarajan, P., Chen, C., Drinnenberg, I.A., Schiller, B.J., Thompson, J., Moresco, J.J., Yates, J.R., Bartel, D.P., and Madhani, H.D. (2013). Stalled Spliceosomes Are a Signal for RNAi-Mediated Genome Defense. Cell *152*, 957–968.

Fischer, S.E.J., Butler, M.D., Pan, Q., and Ruvkun, G. (2008). Trans-splicing in C. elegans generates the negative RNAi regulator ERI-6/7. Nature *455*, 491–496.

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 Genetics *7*.

Flemr, M., Malik, R., Franke, V., Nejepinska, J., Sedlacek, R., Vlahovicek, K., and Svoboda, P. (2013). A Retrotransposon-Driven Dicer Isoform Directs Endogenous Small Interfering RNA Production in Mouse Oocytes. Cell *155*, 807–816.

Gu, W., Shirayama, M., Conte Jr., 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. Molecular Cell *36*, 231–244.

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. PNAS *106*, 18674–18679.

He, C., Pillai, S.S., Taglini, F., Li, F., Ruan, K., Zhang, J., Wu, J., Shi, Y., and Bayne, E.H. (2013). Structural analysis of Stc1 provides insights into the coupling of RNAi and chromatin modification. PNAS *110*, E1879–E1888.

van't Hof, A.E., Campagne, P., Rigden, D.J., Yung, C.J., Lingley, J., Quail, M.A., Hall, N., Darby, A.C., and Saccheri, I.J. (2016). The industrial melanism mutation in British peppered moths is a transposable element. Nature *534*, 102–105.

Holz, A., and Streit, A. (2017). Gain and Loss of Small RNA Classes—Characterization of Small RNAs in the Parasitic Nematode Family Strongyloididae. Genome Biol Evol *9*, 2826–2843.

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. Current Biology *17*, 1265–1272.

Houri-Ze'evi, L., Korem, Y., Sheftel, H., Faigenbloom, L., Toker, I.A., Dagan, Y., Awad, L., Degani, L., Alon, U., and Rechavi, O. (2016). A Tunable Mechanism Determines the Duration of the Transgenerational Small RNA Inheritance in C. elegans. Cell *165*, 88–99.

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., Fejes Tóth, K., and Aravin, A.A. (2017). piRNA Biogenesis in Drosophila melanogaster. Trends in Genetics *33*, 882–894.

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. Molecular Cell *70*, 639-649.e6.

Iwasaki, Y.W., Siomi, M.C., and Siomi, H. (2015). PIWI-Interacting RNA: Its Biogenesis and Functions. Annual Review of Biochemistry *84*, 405–433.

Jehn, J., Gebert, D., Pipilescu, F., Stern, S., Kiefer, J.S.T., Hewel, C., and Rosenkranz, D. (2018). PIWI genes and piRNAs are ubiquitously expressed in mollusks and show patterns of lineage-specific adaptation. Communications Biology *1*, 137.

Kamminga, L.M., Luteijn, M.J., Broeder, M.J. den, 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. The EMBO Journal *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.

Kapusta, A., Kronenberg, Z., Lynch, V.J., Zhuo, X., Ramsay, L., Bourque, G., Yandell, M., and Feschotte, C. (2013). Transposable Elements Are Major Contributors to the Origin, Diversification, and Regulation of Vertebrate Long Noncoding RNAs. PLOS Genetics *9*, e1003470.

Kawaoka, S., Arai, Y., Kadota, K., Suzuki, Y., Hara, K., Sugano, S., Shimizu, K., Tomari, Y., Shimada, T., and Katsuma, S. (2011). Zygotic amplification of secondary piRNAs during silkworm embryogenesis. RNA.

Ketting, R.F. (2011). The Many Faces of RNAi. Developmental Cell *20*, 148–161.

Kleaveland, B., Shi, C.Y., Stefano, J., and Bartel, D.P. (2018). A Network of Noncoding Regulatory RNAs Acts in the Mammalian Brain. Cell *174*, 350-362.e17.

Kruse, K., Diaz, N., Enriquez-Gasca, R., Gaume, X., Torres-Padilla, M.-E., and Vaquerizas, J.M. (2019). Transposable elements drive reorganisation of 3D chromatin during early embryogenesis. BioRxiv 523712.

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

Lee, R.C., Hammell, C.M., and Ambros, V. (2006). Interacting endogenous and exogenous RNAi pathways in Caenorhabditis elegans. RNA *12*, 589–597.

Le Thomas, A., Marinov, G.K., and Aravin, A.A. (2014). A Transgenerational Process Defines piRNA Biogenesis in Drosophila virilis. Cell Reports *8*, 1617–1623.

Lewis, S.H., Quarles, K.A., Yang, Y., Tanguy, M., Frézal, L., Smith, S.A., Sharma, P.P., Cordaux, R., Gilbert, C., Giraud, I., et al. (2017). Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. Nature Ecology & Evolution 1.

Li, F., and Ding, S.-W. (2006). Virus Counterdefense: Diverse Strategies for Evading the RNA-Silencing Immunity. Annual Review of Microbiology *60*, 503–531.

Lim, S.L., Qu, Z.P., Kortschak, R.D., Lawrence, D.M., Geoghegan, J., Hempfling, A.-L., Bergmann, M., Goodnow, C.C., Ormandy, C.J., Wong, L., et al. (2015). HENMT1 and piRNA Stability Are Required for Adult Male Germ Cell Transposon Repression and to Define the Spermatogenic Program in the Mouse. PLOS Genetics *11*, e1005620.

Luteijn, M.J., and Ketting, R.F. (2013). PIWI-interacting RNAs: from generation to transgenerational epigenetics. Nat Rev Genet *14*, 523–534.

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. The EMBO Journal *31*, 3422–3430.

Martinez, G., and Köhler, C. (2017). Role of small RNAs in epigenetic reprogramming during plant sexual reproduction. Current Opinion in Plant Biology *36*, 22–28.

Matsumoto, N., Nishimasu, H., Sakakibara, K., Nishida, K.M., Hirano, T., Ishitani, R., Siomi, H., Siomi, M.C., and Nureki, O. (2016). Crystal Structure of Silkworm PIWI-Clade Argonaute Siwi Bound to piRNA. Cell *167*, 484-497.e9.

Minkina, O., and Hunter, C.P. (2017). Stable Heritable Germline Silencing Directs Somatic Silencing at an Endogenous Locus. Molecular Cell *65*, 659-670.e5.

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.

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.

Ninova, M., Griffiths-Jones, S., and Ronshaugen, M. (2017). Abundant expression of somatic transposon-derived piRNAs throughout Tribolium castaneum embryogenesis. Genome Biology *18*, 184.

Obbard, D.J., Gordon, K.H.J., Buck, A.H., and Jiggins, F.M. (2009). The evolution of RNAi as a defence against viruses and transposable elements. Philosophical Transactions of the Royal Society of London B: Biological Sciences *364*, 99–115.

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

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.

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

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. Developmental Cell *34*, 457–465.

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.

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

Saito, K., Ishizu, H., Komai, M., Kotani, H., Kawamura, Y., Nishida, K.M., Siomi, H., and Siomi, M.C. (2010). Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in Drosophila. Genes Dev. *24*, 2493–2498.

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.

Schott, D., Yanai, I., and Hunter, C.P. (2014). Natural RNA interference directs a heritable response to the environment. Scientific Reports *4*, 7387.

Shi, Z., Montgomery, T.A., Qi, Y., and Ruvkun, G. (2013). High-throughput sequencing reveals extraordinary fluidity of miRNA, piRNA, and siRNA pathways in nematodes. Genome Res. *23*, 497–508.

Shirayama, M., Seth, M., Lee, H.-C., Gu, W., Ishidate, T., Conte Jr., D., and Mello, C.C. (2012). piRNAs Initiate an Epigenetic Memory of Nonself RNA in the C. elegans Germline. Cell *150*, 65–77.

Stoeckius, M., Maaskola, J., Colombo, T., Rahn, H.-P., Friedländer, M.R., Li, N., Chen, W., Piano, F., and Rajewsky, N. (2009). Large-scale sorting of *C. elegans* embryos reveals the dynamics of small RNA expression. Nature Methods *6*, 745–751.

Stoeckius, M., Grün, D., and Rajewsky, N. (2014). Paternal RNA contributions in the Caenorhabditis elegans zygote. The EMBO Journal *33*, 1740–1750.

Takemoto, N., Yoshimura, T., Miyazaki, S., Tashiro, F., and Miyazaki, J. (2016). Gtsf1l and Gtsf2 Are Specifically Expressed in Gonocytes and Spermatids but Are Not Essential for Spermatogenesis. PLOS ONE *11*, e0150390.

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.

Thivierge, C., Makil, N., Flamand, M., Vasale, J.J., Mello, C.C., Wohlschlegel, J., Jr, D.C., 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.

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

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. PNAS *107*, 3582–3587.

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

Wallach, E.E., Kandeel, F.R., and Swerdloff, R.S. (1988). Role of temperature in regulation of spermatogenesis and the use of heating as a method for contraception. Fertility and Sterility *49*, 1–23.

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 1.

Wang, G., and Reinke, V. (2008). A C. elegans Piwi, PRG-1, Regulates 21U-RNAs during Spermatogenesis. Current Biology *18*, 861–867.

Wang, J., Czech, B., Crunk, A., Wallace, A., Mitreva, M., Hannon, G.J., and Davis, R.E. (2011). Deep small RNA sequencing from the nematode Ascaris reveals conservation, functional diversification, and novel developmental profiles. Genome Res.

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.

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 Reports *23*, 2482–2494.

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 Reports e42054.

Zhuang, J.J., and Hunter, C.P. (2011). Tissue-specificity of Caenorhabditis elegans Enhanced RNAi Mutants. Genetics genetics.111.127209.