INTRONS SAFEGUARD mRNA EXPRESSION IN THE C. ELEGANS GERMLINE AGAINST MULTIPLE SURVEILLANCE MECHANISMS

A dissertation presented by

YEKATERINA V. MAKEYEVA

Submitted to the faculty of the University of Massachusetts Morningside Graduate School of Biomedical Sciences, Worcester in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

November 11, 2022

Interdisciplinary Graduate Program
INTRONS SAFEGUARD mRNA EXPRESSION IN THE C. ELEGANS GERMLINE AGAINST MULTIPLE SURVEILLANCE MECHANISMS

A dissertation presented by
Yekaterina Vitalyevna Makeyeva

Dissertation Examination Committee:
Craig Mello, Ph.D., Thesis Advisor
William Theurkauf, Ph.D., Chair of Committee
Eric Miska, Ph.D., Member of Committee
Joel Richter, Ph.D., Member of Committee
Sean Ryder, Ph.D., Member of Committee
Erik Sontheimer, Ph.D., Member of Committee

Mary Ellen Lane, Ph.D.,
Dean of the Morningside Graduate School of Biomedical Sciences

November 11, 2022
Interdisciplinary Graduate Program
ACKNOWLEDGEMENTS

When I came to UMass for graduate school, I thought I would join a neuroscience lab. However, in the first couple months, I met Craig when he was teaching a module on small RNAs. I admired the way he thought through scientific questions and placed them in a larger context, as well as the way he could tie it all together to tell a captivating story. Hoping to learn some of those skills, I asked Craig to rotate in his lab – and the rest is history.

I am grateful to my mentor for the freedom to try virtually any experiment to probe a scientific inquiry. Our brainstorming sessions aimed at developing novel experimental designs to address unanswered questions, zealous discussions about possible hypotheses and models, and Craig’s mentorship in story-telling shaped me into the scientist I am today. Apart from scientific training, Craig’s concern and understanding made it possible for me to successfully complete my graduate school journey.

Another person who significantly contributed to my training is Masaki Shirayama. As we collaborated on a number of projects, we held countless scientific discussions, shared discoveries, and, ultimately, became good friends. I am thankful for Masaki’s invigorating attitude and that he was always there for me.

I am also deeply appreciative of every single member of the Mello lab – they are an inseparable part of my memorable graduate school journey. I would like to thank Takao Ishidate, Heesun Kim, Ahmet Ozturk, and Yuehe Ding for their boundless technical expertise and inspiration. I am equally thankful for the support and thought-provoking scientific and non-scientific conversations with Krishna
Ghanta, Dan Durning, Wendy Tan, Berto Ochoa, Maya Spichal, April Travers, Siyuan Dai, Cole Pero, Johan Girgenrath, Bing Sun, and Gangming Zhang. I would like to express my gratitude to Deb Bordne, Nikki Grace, and Marcy Kelley for helping me navigate the bureaucratic side of science, and Darryl Conte and Emily Haberlin for spending an immeasurable amount of time and effort to help me develop scientific writing skills in the process of grant applications.

I would like to thank my TRAC committee: Bill Theurkauf (Chair), Joel Richter, Sean Ryder, and Athma Pai – who ensured I was hitting graduate training milestones and provided invaluable career advice along the way. I am grateful to Erik Sontheimer and Eric Miska for finding time to join my Dissertation Examination Committee, and, as other members, taking the time to read and evaluate this thesis.

I cherish the UMass community for the collaborative spirit and ample opportunities that have enriched my training. I am thankful for my classmates, many of whom will hopefully become my lifelong friends. From the qualifying exam to defense, we supported each other during the tough times and celebrated our victories and life-changing events.

Finally, I would like to thank my friends and family without whom this thesis would not have been possible. My parents – for fostering the pursuit of excellence in my upbringing and insisting that I follow my passion and dreams. My in-laws and siblings – for their unwavering support and patience. My younger sister – for being proud of me, letting me share my passion for science with her, and occasionally being my lab tech. And, most importantly, my husband – my pillar of sanity, patience, love, and support.
ABSTRACT

Organisms employ sophisticated systems for genome defense against foreign and potentially harmful elements, while leaving room for gene adaptation. In animals, conserved PIWI Argonautes use genomically encoded small RNA guides (called piRNAs) to detect and silence foreign nucleotide sequences, such as transposons. In *Caenorhabditis elegans*, the detection of foreign transcripts by PIWI triggers the production of a second class of antisense small RNAs (called 22G-RNAs), which guide worm-specific Argonautes (WAGOs) to direct transcriptional and posttranscriptional silencing. PIWI-piRNA complexes recognize targets via imperfect base-pairing, which could threaten the expression of endogenous host genes. Nevertheless, worms use yet a third small RNA pathway involving the Argonaute CSR-1 to license endogenous germline gene expression and prevent inappropriate silencing by the PIWI pathway. How and why certain genes are licensed remains unknown.

Here I show that introns and, by inference, mRNA splicing protect messenger RNAs from germline silencing. Intronless reporters encounter 22G-RNA-dependent and -independent silencing mechanisms, which we collectively termed “intronless silencing.” Genetic studies revealed that primary Argonautes, e.g., PIWI, are not required for the 22G-RNA-dependent intronless silencing mechanism, suggesting that intronless reporters are silenced by default. Nuclear and cytoplasmic WAGOs enabled the transmission of silencing from an intronless allele to a homologous intron-containing allele. The 22G-RNA-independent mechanism not only reduced intronless reporter mRNA levels, compared to the homologous intron-containing genes, but also prevented polyadenylation and nuclear export. Cis-acting elements that promote export from the nucleus nevertheless failed to fully activate expression of intronless reporters, suggesting additional layers of regulation in the small RNA-independent mechanism of intronless silencing. These findings suggest that multiple germline surveillance systems monitor transcript splicing, reveal a protective role of splicing in transcript licensing, and provide evidence for a splicing-dependent, sequence-independent mode of Argonaute programming.
Table of Contents

CHAPTER I: Distinction of self vs. non-self: Interplay of regulatory small RNA pathways in the C. elegans germline ................................................................. 1

INTRODUCTION ........................................................................................................... 2

Sequence-specific resistance to genome parasites ....................................................... 3
Sequence-independent detection of parasitic nucleic acid ........................................... 4

SMALL RNA PATHWAYS – GUARDIANS OF THE C. ELEGANS GERMLINE .................. 6

History of germline transgene silencing .................................................................. 6
A short note on Argonautes ....................................................................................... 8
PIWI/piRNA pathway ................................................................................................. 9
Additional primary Argonautes that detect foreign RNA ......................................... 15
CSR-1 and licensing of endogenous mRNA ............................................................... 17
Introns impart protection against germline silencing ............................................... 23
Final note on regulation of gene expression .............................................................. 26

CHAPTER II: Cues from mRNA splicing prevent default Argonaute silencing in C. elegans .... 27

ABSTRACT ...................................................................................................................... 28

INTRODUCTION ........................................................................................................... 29

RESULTS ....................................................................................................................... 32

Intron removal prevents gene expression in the germline but not the soma ............ 32
Intronless gene silencing involves small RNA-dependent and -independent pathways .... 34
Intronless gene silencing does not require the piRNA pathway ................................ 38
Endogenous genes are sensitive to intronless silencing .......................................... 39
Regulatory elements of a naturally intronless gene can partially suppress intronless silencing. 44
Intronless silencing requires WAGO Argonautes but is initiated independently of known primary Argonautes ................................................................. 48

Trans-acting signals from endogenous genes with intact splicing protect against small RNA silencing ......................................................................................... 53

DISCUSSION ................................................................................................................ 55

LIMITATIONS OF THE STUDY ................................................................................. 62

MATERIALS AND METHODS ...................................................................................... 63

CHAPTER III: Intronless silencing – small RNA-independent branch .......................... 74

ABSTRACT ...................................................................................................................... 75

INTRODUCTION ........................................................................................................... 76
RESULTS ................................................................................................................................. 78
  Intronless reporter mRNA is absent from the cytoplasm and likely suppressed transcriptionally ........................................................................................................................................... 78
  Cis-mode of intronless silencing can repress cytoplasmic mRNA post-transcriptionally.............. 83
  mRNA polyadenylation of intronless oma-1* reporter was not detected........................................ 87
  Introns in partially intronless reporters prevent trans-silencing but not cis-silencing............... 91
  “Long” terminal exon can trigger cis-silencing........................................................................ 95
DISCUSSION ............................................................................................................................ 98
  Cis-silencing acts on transcriptional and post-transcriptional levels ........................................ 98
  Lack of polyadenylation as a feature of cis-silencing............................................................... 99
  PIWI-independent trans-silencing and cis-silencing triggered by distinct cues ....................... 101
METHODS .................................................................................................................................. 103
CHAPTER IV: Discussion and future work.................................................................................. 110
  Cases of PIWI-independent WAGO-mediated silencing in the C. elegans germline.................. 111
  What triggers default small RNA silencing? ............................................................................ 112
  Screen strategy to identify cis-silencing factors...................................................................... 115
  A word on transitive protection............................................................................................. 119
  Parallel with humans............................................................................................................. 120
BIBLIOGRAPHY ...................................................................................................................... 121
List of Figures

Figure 1.1. Small RNA-mediated silencing: initiation and maintenance................................. 14
Figure 1.2. Schematic of RNA-induced epigenetic gene activation (RNAa). .............................. 21
Figure 2.1. Intron removal leads to germline gene silencing that acts in trans ...................... 33
Figure 2.2. Intronless cdk-1*::gfp* is silenced by a default WAGO-dependent small RNA pathway and by a cis-acting small RNA-independent pathway .................................................. 36
Figure 2.3. Endogenous genes are sensitive to intronless silencing ...................................... 41
Figure 2.4. Low mRNA abundance supports absence of intronless reporter expression in the germline ..................................................................................................................... 43
Figure 2.5. Regulatory elements of a naturally intronless gene can partially suppress intronless silencing .................................................................................................................................. 47
Figure 2.6. Small RNA-mediated intronless silencing is initiated independently of known primary Argonautes ...................................................................................................................................... 51
Figure 2.7. Intronless reporters are not targeted by the 26G-RNA pathways ......................... 52
Figure 2.8. Nuclear RNAi pathway contributes to the small RNA-mediated silencing of intronless reporters ...................................................................................................................................... 52
Figure 2.9. Trans-acting signals from endogenous loci protect against small RNA-mediated silencing. ........................................................................................................................................ 54
Figure 2.10. Splicing machinery does not promote default intronless silencing ...................... 61
Figure 3.1. Intronless silencing affects oma-1 mRNA distribution ........................................ 81
Figure 3.2 Spatial distribution of intronless gfp*::his-61 reporter in the germline ..................... 85
Figure 3.3 Intronless reporter transcript is not polyadenylated .............................................. 89
Figure 3.4 Partially intronless transgenes: removal of 5’ and 3’-introns leads to contrasting effects on cis-silencing ....................................................................................................................... 93
Figure 3.5. Working model for intronless silencing .................................................................. 96
Figure 4.1 Small RNA sequencing of the intronless screen strain ........................................ 118
Preface

Parts of this dissertation have been published as:

CHAPTER I

Distinction of self vs. non-self: Interplay of regulatory small RNA pathways in the *C. elegans* germline
INTRODUCTION

Acquisition of new genomic information is the driving force of evolution. Underlying DNA variation can result from events such as random mutations, recombination, or the activity of viruses and mobile genetic elements (MGE). Viruses can disrupt the host genome by directly integrating genomic material into host DNA, while transposable and retroviral elements, already encoded in host DNA, can mobilize causing additional DNA alterations. DNA variation can be beneficial to the host – in fact, many human long non-coding RNA (lincRNA) and proteins appear to have MGE origin (Britten, 2006; Grishok et al., 2001; Kannan et al., 2015; Smit, 1999). For example, a neuronal gene Arc, essential for long-term memory in mammals, bears structural and functional homology with Ty3/gypsy retrotransposons (Pastuzyn et al., 2018). Centromere protein B (CENP-B), a DNA binding protein of a pogo DNA transposon origin, facilitates centromere formation and confers chromosome stability (Morozov et al., 2017; Smit and Riggs, 1996). However, despite reported adaptive cases of mobile element activity, more frequently, viral and MGE-mediated deletions, insertions, and re-arrangements lead to diseases like cancer, hemophilia, and Duchenne muscular dystrophy (Hancks and Kazazian, 2016; Kazazian and Moran, 2017). Therefore, to safeguard the genome from the deleterious activity of DNA parasites, organisms have developed various surveillance systems to detect and suppress “foreign” or “non-self” DNA and RNA, while granting the expression of host’s “self” genes required for the development and function of an organism. Organisms recognize foreign nucleic acid in sequence-specific and sequence-
independent manner. Upon detection, selfish genes are suppressed in various ways, depending on tissue type and host organism.

**Sequence-specific resistance to genome parasites**

Some of the sequence-specific defense mechanisms include CRISPR-Cas and KRAB/KAP1. Bacteria and archaea employ a system known as clustered regularly interspaced short palindromic repeats (CRISPR)-Cas for defense against viruses and MGEs (Barrangou et al., 2007; Ishino et al., 1987; Mojica et al., 1993). In short, a piece of invader’s genetic material is incorporated in the host genome to be transcribed and used as an RNA guide (termed CRISPR RNA or crRNA) to detect and fend off recurring offenses by the same invader. When invader’s DNA or RNA encounters a complementary crRNA, it is cleaved by a nuclease associated with the crRNA, thereby debilitating the mobility of viruses and MGEs (reviewed in Newsom et al., 2020). The ability to expand, modify and pass down the DNA library of invaders’ snippets lends to the adaptability of the CRISPR-Cas system. To avoid mistargeting of endogenous bacterial and archaeal genes required for normal organism function, nucleolytic systems rely on two factors. First, compulsory complementarity between the crRNA and the target greatly narrows down the pool of potential endogenous targets (Deveau et al., 2008). Second, depending on the CRISPR-Cas system type, the additional prerequisite for cleavage is either (1) the absence of extensive complementarity region between the guide and the target (to prevent cutting host DNA locus carrying crRNAs) (Marraffini and Sontheimer, 2010), or (2) the presence of specific motifs in invaders’ nucleic acid sequences –
protospacer adjacent motif (PAM) or protospacer flanking sequence (PFS) (Abudayyeh et al., 2016; Deveau et al., 2008; Westra et al., 2013; Zhang et al., 2012b).

Another way to achieve sequence specificity in nucleic acid target recognition is via zinc-finger (ZF) motifs. Each ZF motif binds a defined three nucleotide sequence; the combination of multiple ZF domains provides more extensive sequence specificity (Pavletich and Pabo, 1991). Numerous RNA- and DNA-binding proteins are reported to have zinc fingers (Cassandri et al., 2017). To repress the expression of distinct gene groups, cells employ proteins containing DNA-binding domain (such as ZF-containing) and a repressor domain. Kruppel-associated box (KRAB) is one of the common repressor domains, associated with about one third of Kruppel ZF-containing proteins forming KRAB-ZFP. Some MGEs exist as multicopy elements in a single host’s genome, thus it is not surprising that a dedicated KRAB-ZFP is used to silence endogenous retroviruses (ERV) in embryonic stem cells. KRAB domain of KRAB-ZFP recruits KRAB-associated protein-1 (KAP-1) that serves as a scaffold for heterochromatin-inducing factors to deposit silencing histone H3 lysine 9 trimethylation marks in the promoter region of ERVs (Groner et al., 2010). Due to the sequence-specificity of KAP1/KRAB-ZFP epigenetic silencing, endogenous genes are spared from the repressive effects of this system.

**Sequence-independent detection of parasitic nucleic acid**

Akin to innate immunity, organisms have developed ways to recognize pathogenic patterns in nucleic acid structures. Double stranded RNA (dsRNA) is a characteristic
feature of many viral invaders, which is formed when viral genetic material is reverse transcribed (for RNA viruses) or bidirectionally transcribed (DNA viruses) by viral RNA polymerases (Weber et al., 2006). Toll-like receptors (TLRs) and retinoic acid-inducible gene I (RIG-I)-like receptors detect dsRNA and trigger signal transduction cascade resulting in type I interferon (IFN) response and expression of antiviral factors, which target almost every step in viral life cycle (Ohtake et al., 1990; Rehwinkel and Gack, 2020). Not every organism and tissue mediate type I IFN response; alternatively, antiviral immunity is achieved by RNA interference (RNAi) (Fire et al., 1998). Type III ribonuclease DICER 1 binds and cleaves long dsRNAs, generating small interfering RNA (siRNA) species that associate with Argonaute protein and guide it to complementary viral RNAs for destruction via endonucleolytic cleavage (Bernstein et al., 2001; Hammond et al., 2000; Hammond et al., 2001; Ketting et al., 2001; Zamore et al., 2000). In some organisms, RNAi can also lead to transcriptional repression (Mette et al., 2000; Volpe et al., 2002).

In eukaryotes, DNA is restricted to the nuclear compartment, and its occurrence in the cytosol acts as a pathogen alert. Cyclic GMP-AMP (cGAMP) cyclase (cGAS) recognizes cytosolic DNA and activates the stimulator of the interferon genes (STING) adaptor protein to induce antiviral IFN response (Ishikawa and Barber, 2008; Ishikawa et al., 2009; Sun et al., 2013).

As a result of this thesis work, we have uncovered yet another sequence-independent silencing pathway. Using the model organism Caenorhabditis elegans, we found that the absence of introns and, by inference, splicing serve as a strong silencing trigger for transcriptional and post-transcriptional gene suppression. We
argue that genes without introns and thus resolved splicing may be viewed by the host as foreign. Surprisingly, silencing of intronless reporters is restricted to the germline and is mediated by at least two pathways: small RNA-dependent and -independent mechanisms. In subsequent chapters, I will describe the initial discovery of the intronless silencing phenomenon, which resulted in a peer-reviewed publication (Chapter II), and include the latest work probing the underlying silencing mechanism (Chapter III). First, however, I will discuss well-known pathways that nematodes implement to detect non-self/mobile genetic elements, which threaten germline integrity.

**SMALL RNA PATHWAYS – GUARDIANS OF THE *C. ELEGANS* GERMLINE**

**History of germline transgene silencing**

The germline of the nematode *Caenorhabditis elegans* is virtually resistant to viruses and it keeps mobile genetic elements (MGEs) under tight control utilizing both sequence-specific and sequence-independent mechanisms (Batista et al., 2008; Das et al., 2008; Emmons and Yesner, 1984; Felix et al., 2011; Ketting et al., 1999; Tabara et al., 1999; Vastenhouw et al., 2003). A robust system of small RNA pathways, reliant on the association between a ~20-30 nt-long non-coding RNA and an Argonaute protein, enables the worm to amplify precise silencing signals to effectively control viral and MGE activity transcriptionally and post-transcriptionally (Gu et al., 2009; Sijen and Plasterk, 2003; Vastenhouw et al., 2003; Yigit et al., 2006; Youngman and Claycomb, 2014). Furthermore, the same surveillance system frequently prevents expression of transgenes in the *C. elegans* germline, especially
if they contain repetitive fragments and/or an unfamiliar to the worm nucleotide sequence, presumably because the host recognizes these transgenes as foreign. In 1997, Kelly et al. reported that a simple extrachromosomal array (an artificial heritable DNA structure comprised of 70-300 tandem construct repeats (Stinchcomb et al., 1985)) carrying a transgenic fusion of *C. elegans* *let-858* gene and *gfp* (green fluorescent protein), failed to express in the worm germline, while exhibiting a bright GFP signal in somatic tissues (Kelly et al., 1997). The group also showed that adding sequence complexity to extrachromosomal arrays could attenuate germline silencing, suggesting that initial simple extrachromosomal array construction may be reminiscent of tandem repeat elements that originate from transposable elements (Paco et al., 2019). In support of that conclusion, Praitis and colleagues demonstrated that when several copies of a GFP reporter were integrated into the worm genome in an interspersed manner, transgene silencing in the germline was mitigated (Praitis et al., 2001).

A clue about the transgene silencing pathway came from studies exploring the mechanisms and genetic requirements for RNA interference (RNAi) (Ketting et al., 1999; Sijen et al., 2001; Tabara et al., 1999). Analysis of small RNAs showed that, in *C. elegans*, RNA-dependent RNA polymerases (RdRPs) amplify RNAi signal via secondary small RNA synthesis (Sijen et al., 2001). Coupled with the finding that RNAi-deficient mutants exhibit transposon activation and, in select cases, rescue germline transgene expression (Tabara et al., 1999), these observations suggested that suppression of foreign elements in the worm germline was mediated by a small RNA pathway. In the next couple decades, the field of small RNAs in worms and
other organisms rapidly and vastly expanded. A conserved function of the small RNA pathways appears to be defense against non-self-elements, but instances of self-gene regulation are being uncovered in a variety of organisms to this day (Kuscu et al., 2018; Quarato et al., 2021; Vargas-Asencio and Perry, 2019; Zagoskin et al., 2022), suggesting there is much more to learn about the roles of these fascinating pathways.

In the following sections, I will discuss how small RNA pathways recognize and silence foreign nucleic acid sequences in *C. elegans*.

**A short note on Argonautes**

To carry out its function and avoid nucleolytic degradation, a small RNA species requires a protein counterpart. Argonautes were initially discovered in plant *Arabidopsis thaliana* as factors required for leaf development (Bohmert et al., 1998). The appearance of mutant plants resembled that of a squid, which is how the gene, and, consequently, the entire protein family, got its name – Argonaute – after the octopus *Argonauta argo*. Since then, Argonautes were found in all domains of life and determined to be well-conserved structurally (Catalanotto et al., 2000; Cerutti et al., 2000; Cogoni and Macino, 1997; Cox et al., 1998; Lin and Spradling, 1997; Tabara et al., 1999). Crystal structure and biochemical studies reveal that Argonautes have several prominent domains (Song et al., 2004; Yuan et al., 2005). The MID (middle) module anchors the 5′ end of the small RNA guide, while the PAZ (PIWI/Argonaute/Zwille) domain accommodates its 3′ end. The PIWI (P-element-induced wimpy testis) domain can cleave or “slice” target RNA using RNase H-like
D(E)DE/H motif, provided the catalytic triad or tetrad is intact (Liu et al., 2004; Nakanishi et al., 2012; Rivas et al., 2005; Song et al., 2004; Yuan et al., 2005). A less-conserved N-terminal domain was reported to affect Argonaute folding and catalytic activity (Hauptmann et al., 2013; Schurmann et al., 2013; Song et al., 2004).

Argonautes are classified into two major clades based on the type of the associated small RNA guide: AGO and PIWI (Matsumoto et al., 2016; Youngman and Claycomb, 2014). PIWI Argonautes are loaded with PIWI-interacting RNAs (piRNAs, discussed in the next sub-section), while AGO Argonautes associate with microRNAs (miRNAs) and small interfering RNAs (siRNAs) (Matsumoto et al., 2016). In nematodes, a worm-specific Argonaute clade (WAGO) comprises Argonautes that associate with secondary siRNAs produced by RdRP (Yigit et al., 2006; Youngman and Claycomb, 2014). WAGOs lack the catalytic cleavage triad (Yigit et al., 2006), but were reported to mediate transcriptional (Ashe et al., 2012; Buckley et al., 2012; Guang et al., 2010; Guang et al., 2008; Luteijn et al., 2012; Shirayama et al., 2012) and post-transcriptional gene silencing (TGS and PTGS, respectively) (Fire et al., 1998; Montgomery et al., 1998; Yang et al., 2012).

**PIWI/piRNA pathway**

Coincidentally, the first evidence of both the PIWI Argonaute and piRNAs came from the work in *Drosophila melanogaster*. Lin and Spradling identified *piwi* as a factor required for stem cell proliferation and fertility in male and female organisms (Lin and Spradling, 1997). Several years later, Aravin and colleagues reported that
downregulation of the multi-copy \textit{Stellate} (Ste) gene by the \textit{Suppressor of Stellate} (Su[Ste]) repeats was mediated by an endogenous RNAi-like mechanism, whereby Su[Ste] produced sense and antisense RNA homologous to the \textit{Stellate} gene (Aravin et al., 2001). Surprisingly, however, short RNA species produced by the Su[Ste] were longer than canonical siRNAs (25-27-nt vs. 21-23-nt, respectively) (Aravin et al., 2004). In addition, the group revealed that the synthesis of Su[Ste] short RNAs required factors that were dispensable for siRNA production, further supporting the hypothesis that a novel and distinct pathway was at play (Aravin et al., 2004). Subsequent work in the field uncovered a link between silencing homologous 25-27-nt short RNAs and \textit{piwi}, leading to the naming of such species as piRNAs (Sarot et al., 2004).

The piRNA pathway is conserved among metazoans, and ample evidence from various organisms indicates that piRNAs suppress foreign and transposable element activity, thereby safeguarding animal genomes (Aravin et al., 2006; Aravin et al., 2007; Batista et al., 2008; Das et al., 2008; Girard et al., 2006; Grivna et al., 2006; Houwing et al., 2007; Lau et al., 2006; Vagin et al., 2006). More recent studies report piRNAs also regulate host gene expression (Chen et al., 2021; Gou et al., 2014; Tang et al., 2018) and emerge as potential biomarkers and therapeutic targets for cancer (Cheng et al., 2011; Cheng et al., 2019; He et al., 2015; Liu et al., 2019). Despite several mechanistic differences in the PIWI pathway, the effector ribonucleoprotein complex is conserved. PIWI Argonaute, loaded with a 21-34nt-long single stranded non-coding piRNA guide, surveys the transcriptome, detects target RNA based on complementary base-pairing, and triggers transcriptional and
post-transcriptional silencing (reviewed in Ozata et al., 2019). In general, piRNA guides are transcribed from the host genome by RNA polymerase II as 7-methylguanosine capped single-stranded precursors that are later processed at both 5’ and 3’ ends: final piRNA species bear monophosphate at the 5´ end and 2´-O-methyl at the 3´ terminus. The piRNA pathway mainly operates in the germ cells, ovaries, and developing embryos to ensure fertility and genome stability.

In *C. elegans*, piRNAs are classified into two types based on their genomic organization. Type I piRNAs come from two large transposon-rich clusters (Ruby et al., 2006), while type II piRNAs are derived from the transcription start sites of germline-expressed genes (Gu et al., 2012b). In both cases, each piRNA is transcribed separately (Cecere et al., 2012; Gu et al., 2012b; Ruby et al., 2006), unlike in other metazoans, where long piRNA precursor transcripts are processed to yield multiple piRNA species (Brennecke et al., 2007).

PIWI Argonaute PRG-1 associates with 21-nt-long RNA species that begin with uridine (21U-RNA) (Batista et al., 2008; Gu et al., 2012b; Ruby et al., 2006), scans every germline transcript for non-worm sequences, and silences foreign RNA by secondary cytoplasmic and nuclear Argonautes (WAGOs), with the latter establishing heritable transcriptional suppression (Ashe et al., 2012; Bagijn et al., 2012; Das et al., 2008; Gu et al., 2009; Luteijn et al., 2012; Shirayama et al., 2012). According to immunofluorescence imaging studies, PRG-1 resides at perinuclear structures called P granules (Batista et al., 2008), which are named after the P cell lineage that gives rise to germ cells in *C. elegans*. P granules are phase-separation droplets analogous to RNA and protein containing germ granules found in
vertebrates and insects. In germ cells, P granules associate with nuclear pore complexes – a co-localization that allows PRG-1 to scan the majority of transcripts exported from the nucleus before they are translated (Pitt et al., 2000; Sheth et al., 2010). As shown in Figure 1.1, in worms, foreign RNA sequences recognized by PIWI/piRNA complexes are recruited as templates to RdRPs (Das et al., 2008; Shirayama et al., 2012), which are located in mutator foci structures adjacent to P granules (Phillips et al., 2012). To amplify the silencing signal, RRF-1 and EGO-1 RdRPs produce more target-specific secondary guides - antisense siRNAs that are perfectly complementary to the target mRNA and distributed along the entirety of the transcript unlike the original piRNA (Das et al., 2008; Gu et al., 2009; Maniar and Fire, 2011). In *C. elegans*, these secondary siRNAs are called 22G-RNAs because of their 22-nt length and guanidine bias at the 5´ terminus (Gu et al., 2009; Maniar and Fire, 2011). A recent study showed that sense templates for 22G-RNA generation are appended with non-templated alternating uridine and guanosine ribonucleotides (termed pUG tails) by nucleotidyltransferase RDE-3 (Shukla et al., 2020). After synthesis, 22G-RNAs are loaded onto WAGO clade Argonautes, some of which are cytoplasmic, like WAGO-1, and mediate silencing post-transcriptionally; while others, like WAGO-9, act in the nucleus to pinpoint target DNA locus and suppress the gene co-transcriptionally via heterochromatin formation established in concert with other nuclear factors (Ashe et al., 2012; Buckley et al., 2012; Burkhart et al., 2011; Gu et al., 2009; Guang et al., 2010; Guang et al., 2008; Shirayama et al., 2012).
Shirayama and colleagues showed that once WAGO\(s\) establish target silencing, the PIWI/piRNA system becomes dispensable for target suppression, suggesting WAGO\(s\) can autonomously re-enforce and maintain silencing (Shirayama et al., 2012). However, PRG-1 is still required to recognize the target mRNA originally in order to engage WAGO\(s\) and initiate target silencing.

When the group mutated prg-1 in the strain carrying a silenced gfp::cdk-1 transgene (a fusion between GFP and nematode cyclin-dependent kinase), the worm germline remained devoid of GFP signal and abundant with antisense 22G-RNAs mapping to the gfp sequence, unlike with wago-1 and wago-9 mutations. However, when gfp::cdk-1 was instead introduced into prg-1-mutant animals, GFP expression was clearly visible in worm gonads, supporting the role of PRG-1 in the initiation of silencing instead of maintenance (Figure 1.1). Notably, foreign transcript recognition is based on a partial complementarity between the target mRNA and the piRNA (Shen et al., 2018; Zhang et al., 2018). None of the C. elegans piRNA species bear full complementarity to the gfp sequence, according to the genome-wide piRNA sequence analysis and PRG-1 immunoprecipitation followed by RNA sequencing (IP-seq) studies. Indeed, exploration of complementarity requirements for piRNA silencing showed that base-pairing at positions 2 through 8 (“the seed region”) and some additional pairing between positions 14-19 are important for target mRNA recognition and suppression (Shen et al., 2018; Zhang et al., 2018). Based on the silencing rules, the Li lab created a pirScan tool to help avoid piRNA-mediated gene silencing. Slightly altering nucleotide sequence of the gfp\(^\text{piRNA}\)::cdk-1 transgene to avoid predicted base-pairing with five worm piRNAs resulted in visible
transgenic expression (Wu et al., 2018). Intriguingly, the same silencing rules show that piRNAs should be targeting endogenous worm germline genes as well (Zhang et al., 2018), in agreement with an in-vivo crosslinking, ligation, and sequencing of hybrids (CLASH) approach designed to capture piRNA:target mRNA pairs (Shen et al., 2018). Yet small RNA sequencing and IP-seq studies fail to show that endogenous germline genes are targeted by antisense WAGO 22G-RNAs, and mRNA sequencing data from gonad tissue shows substantial levels of germline transcripts that should be silenced by the piRNA pathway (Gu et al., 2009). These findings suggest that perhaps the RdRP-mediated silencing cascade is not triggered upon base-pairing between a piRNA and germline self mRNA.

**Figure 1.1**

Small RNA-mediated silencing: initiation and maintenance.

Small RNA silencing entails two stages: initiation and maintenance. During initiation, a primary Argonaute (i.e. PRG-1, RDE-1, ERGO-1, or ALG-3/4) complexed with a small RNA guide detects target mRNA via guide:mRNA basepairing (labeled as step “1”). Next, target mRNA is tailed with alternating uridine and guanosine ribonucleotides (pUG) by nucleotidyltransferase RDE-3 (step “2”). RNA-dependent RNA polymerases are recruited to pUG-tailed templates and synthesize antisense 22G-RNAs that are in turn loaded into worm-specific Argonautes (WAGO) (depicted
in step “2”). Armed with 22G-RNAs, WAGOs locate their targets (step “3”) and mediate transcriptional and post-transcriptional silencing (not depicted). In addition, WAGOs can re-inforce pUG-tailing of target RNA (step “4”) thus amplifying silencing signal and promoting gene suppression loop even in the absence of an initiating primary Argonaute – a phenomenon that accounts for the maintenance stage of small RNA silencing.

**Additional primary Argonautes that detect foreign RNA**

WAGO 22G-RNA amplification system effectively suppresses foreign and some host genes in the *C. elegans* germline and somatic tissues (Gu et al., 2009; Yigit et al., 2006). However, previous studies show that prior to the secondary siRNA synthesis, a target RNA needs to be recognized and presented to the RdRP complex (Figure 1.1). As was described in the previous section, PIWI Argonaute PRG-1 is one factor that fulfills that role. However, *C. elegans* utilizes other primary Argonautes as well. Argonautes RDE-1, ERGO-1, and ALG-3/4 associate with small RNA guides and scan worm transcriptome for target mRNA populations. Unlike PRG-1-guides, RDE-1, ERGO-1, and ALG-4/3 guides are not produced from dedicated genomic loci.

RDE-1 is a ubiquitous Argonaut essential for RNAi, and it associates with guides derived from dsRNA duplexes processed by DCR-1 (Tabara et al., 1999; Zou et al., 2019). Upon encountering a complementary transcript, similar to PRG-1 (see section above), RDE-1 nucleoprotein complex triggers a secondary siRNA production (Yigit et al., 2006). Resulting WAGO 22G-RNAs mediate silencing via cytoplasmic (Gu et al., 2009; Sendoel et al., 2019) and nuclear WAGO Argonautes (Buckley et al., 2012), and establish chromatin-level silencing that can be transmitted over generations even after the initial stimulus (such as exogenous RNAi) is removed (Buckley et al., 2012; Burton et al., 2011; Fire et al., 1998; Gu et
al., 2012a; Wan et al., 2018; Xu et al., 2018). In the absence of exogenous RNAi, RDE-1 binds endogenous RNAi products and more rarely miRNAs, and silences corresponding targets via the activity of WAGO/22G-RNA complexes (Correa et al., 2010; Tabara et al., 1999; Yigit et al., 2006). Some transposable elements, owing to their repeat-rich nature and/or bidirectional transcription, form dsRNA structures that are processed by DCR-1 akin to exogenous RNAi (Ketting et al., 1999; Sijen and Plasterk, 2003; Tabara et al., 1999). In case of viral infection, dsRNA comes from viral RNA bearing hairpin structures or RdRP activity that generates dsRNA while amplifying viral RNA for reproduction. miRNAs constitute another source of RDE-1 guides (Correa et al., 2010; Grishok et al., 2001). miRNA maturation involves dsRNA intermediate and DCR-1 processing (Grishok et al., 2001; Hutvagner et al., 2001; Ketting et al., 2001). Typically miRNAs are loaded onto ALG-1 and ALG-2 Argonautes and detect target RNAs via partial complementarity (Grishok et al., 2001; Steiner et al., 2007), but in cases when a miRNA is fully complementary to a transcript, RDE-1 loading has been observed (Correa et al., 2010; Steiner et al., 2007).

ERGO-1 is a primary Argonaute that suppresses foreign elements like retrotransposons, recently duplicated genes and pseudogenes via WAGO 22G-RNA amplification system (Fischer et al., 2011; Fischer and Ruvkun, 2020; Vasale et al., 2010). Although ERGO-1 targets some embryogenesis genes as well (Gent et al., 2010; Vasale et al., 2010), its biased target repertoire suggests ERGO-1 is more select toward non-\textit{C. elegans} genes. Expressed in the oocytes and early embryos (Han et al., 2009; Vasale et al., 2010), ERGO-1 associates with 26G-RNA species,
which are 26 nucleotides in length, start with guanosine, and generated by RNA-dependent RNA polymerase (RdRP) rrf-3 (Gent et al., 2010; Han et al., 2009; Vasale et al., 2010). RRF-3 interaction with DCR-1 revealed by immunoprecipitation studies (Duchaine et al., 2006), dependence of 26G-RNA on DCR-1 function (Vasale et al., 2010), and 26G-RNA phasing shown by a computational small RNA analysis (Blumenfeld and Jose, 2016) suggest DCR-1 may be cleaving RNA synthesized by RRF-3 to generate 26G-RNAs (Blumenfeld and Jose, 2016). Even though multiple studies delve into the specifics of 26G biogenesis and stability (Billi et al., 2012; Chaves et al., 2021; Fischer et al., 2011; Kamminga et al., 2012; Montgomery et al., 2012), the initial triggering event of 26G-RNA synthesis remains unknown.

*C. elegans* also utilizes ALG-3 and ALG-4 primary Argonautes guided by a sperm-specific population of 26G-RNAs, but the targets of this regulation mainly include spermatogenesis-related genes and not mobile genetic elements (Conine et al., 2010; Han et al., 2009; Pavelec et al., 2009). For a similar reason, miRNA-associated ALG-1, ALG-2, and, to a smaller degree, ALG-5, that regulate expression of somatic and germline *C. elegans* genes via a WAGO 22G-RNA-independent mechanism are not further discussed in the current work (Brown et al., 2017; Grishok et al., 2001; Vasquez-Rifo et al., 2012).

**CSR-1 and licensing of endogenous mRNA**

The abundance of silencing mechanisms, some of which operate via partial sequence-complementarity, begs the question – how can the host distinguish
between “self” and “non-self”-genes? In *C. elegans*, target transcript detection by PIWI Argonaute PRG-1 tolerates multiple mismatches between a piRNA and its target (Shen et al., 2018; Zhang et al., 2018). Several sequencing and computational studies show that piRNA sequences are not specific to non-self-genes and should suppress self-genes as well (McEnany et al., 2022; Shen et al., 2018).

One of the candidates for guarding the host transcriptome from the silencing activity of the PIWI pathway is the Argonaute CSR-1. In the germline, CSR-1 localizes to P-granules and oocyte nuclei (Claycomb et al., 2009; Gerson-Gurwitz et al., 2016), and associates with 22G-RNAs generated by RdRP EGO-1, dicer-related helicase DRH-3, and EKL-1 (Gu et al., 2009). Argonaute Immunoprecipitation (IP) studies followed by small RNA sequencing show that CSR-1 is loaded with 22G-RNA species mapping antisense to germline-expressed genes (Claycomb et al., 2009), whereas WAGO Argonautes target a distinct group of genes, which includes silenced transposons and pseudogenes (Gu et al., 2009). While it is widely accepted that WAGO targeting enforces gene silencing, when it comes to CSR-1, there are reports of both downregulation and upregulation of gene expression, suggesting additional factors determine how CSR-1 affects its targets.

Gerson-Gurwitz and colleagues show that CSR-1 tunes germline gene expression via its catalytic cleavage activity: a group of genes targeted the most by CSR-1 22G-RNAs undergo a ~1.4-2 fold mRNA increase in *csr-1* cleavage mutants (Gerson-Gurwitz et al., 2016). The majority of the upregulated gene targets are expressed in the germline and early embryos. Early embryonic lethality of *csr-1* mutants and abnormal embryonic phenotypes associated with some of the
upregulated gene targets led the authors to speculate that CSR-1 tunes the transcriptome of maternal oocytes to ensure early embryo development. While the role of CSR-1 in oocytes is yet to be firmly established, Argonaute’s role in gene downregulation is supported by the findings from the Cecere group. CSR-1, guided by homologous small RNAs, depletes a subset of maternally deposited mRNAs in early embryos to promote embryogenesis (Quarato et al., 2021). Other studies describe the role of CSR-1 in preventing transcription of sperm-specific genes in female germline (Campbell and Updike, 2015) and suppressing precocious embryonic genome activation (EGA) in oocytes (Fassnacht et al., 2018), but, in contrast to the first two examples, the discussed gene targets do not seem to correspond to the repertoire of CSR-1 22G-RNAs, suggesting the CSR-1 effect may be indirect.

A common theme in studies showing that CSR-1 positively regulates germline gene expression is promoting male and female germline identities. Spermatocyte-specific CSR-1a isoform positively regulates mRNA levels of spermatogenic genes (Nguyen and Phillips, 2021). Global Run-On sequencing (GRO-seq) and Pol II ChIP studies from L4 hermaphrodites reveal that mutations in csr-1 or drh-3 result in decreased transcription of CSR-1 target genes and upregulation of silent and low-expressing non-CSR-1 gene targets (Cecere et al., 2014). Based on the DNA-independent interaction between CSR-1 and Pol II, and increase in global antisense transcription in csr-1 mutant animals, the group proposed that CSR-1 interacts with nascent transcripts via 22G-RNAs thereby stabilizing sense-oriented RNA pol II and restricting transcription to CSR-1 targets. In addition, csr-1 mutants exhibit a drop in
core histones mRNA and protein level (Avgousti et al., 2012). The group proposes that CSR-1 aids histone processing via cleavage at the transcript’s 3’ UTR. Further experiments will determine whether CSR-1 promotes histone expression directly or indirectly, through an intermediate.

While the studies above present diligent analyses of CSR-1 effect on transcript regulation in various developmental stages and tissues, they do not directly investigate the relationship between CSR-1 and silencing WAGO pathways, perhaps because CSR-1 depletion does not trigger silencing of CSR-1 targets by WAGO 22G-RNAs (Claycomb et al., 2009). Yet, some reports demonstrate that CSR-1 plays a protective role in counteracting PIWI silencing. Shirayama et al. showed that, in select cases, a silenced PIWI target can be transitively activated by an expressing gene provided the two share nucleotide sequence similarity (Shirayama et al., 2012). When animals carrying a PIWI-silenced gfp::cdk-1 transgene were crossed with worms bearing an expressing oma-1::gfp transgene, the resulting F1 progeny exhibited not only OMA-1::GFP expression but also GFP::CDK-1, indicating that oma-1::gfp transitively activated gfp::cdk-1 – a phenomenon that was termed RNA-induced epigenetic gene activation or “RNAa” (Figure 1.2). Later work revealed that CSR-1 is required for RNAa, as exposure of newly hatched F1 progeny to csr-1 RNAi or csr-1 heterozygosity in one of the parents prevented transitive activation of gfp::cdk-1 (Seth et al., 2013). In animals carrying oma-1::gfp, but not silenced gfp::cdk-1 construct, CSR-1 associates with 22G-RNA species mapping antisense to gfp. Notably, active status of gfp::cdk-1 can be maintained over several generations after outcrossing oma-1::gfp, provided
gfp::cdk-1 in the parental strain has been exposed to the RNAa allele for multiple
generations – suggesting memory of active expression can be formed and passed
down transgenerationally. However, CSR-1 is not required for the maintenance of
that memory, as depletion of CSR-1 in the activated gfp::cdk-1(+) strain does not
abolish germline GFP::CDK-1 expression (personal correspondence with M.
Shirayama), which indicates that CSR-1 is essential to establish transitive activation
but not to maintain it.

Figure 1.2

(A) Genetic cross between strains carrying either an actively expressing oma-1::gfp
transgene or a silenced gfp::cdk-1 construct. In F1 offspring, a zoomed-in cartoon
view of the germline (indicated with dashed lines) depicts oocyte expression of both
GFP::CDK-1 (nuclear) and OMA-1::GFP (cytoplasmic), revealing activation of
gfp::cdk-1.

(B) Cartoon representation of RNAa on molecular level. CSR-1 loaded with
antisense 22G-RNAs templated from oma-1::gfp transitably protects gfp::cdk-1 from
PIWI-induced small RNA silencing by halting biogenesis of WAGO 22G-RNAs
complementary to gfp (red oligo stretches).
In another study probing CSR-1-mediated resistance to PIWI silencing, the Claycomb group showed that tethering CSR-1 to piRNA target transcript leads to an increase of mRNA at the transcriptional and post-transcriptional levels (Wedeles et al., 2013). Recruited CSR-1 may be upregulating expression via chromatin remodeling – ChIP assays reveal CSR-1 enrichment at target gene loci (Claycomb et al., 2009; Wedeles et al., 2013). Alternatively, tethering may create an artificial effect, such as transcript sequestration in P-granules, which could not only block mRNA from the silencing machinery but also prevent its export and translation. Protein expression of the piRNA target gene with tethered CSR-1 would help address that possibility.

How does CSR-1 protect mRNA from PIWI silencing? It either prevents detection by the primary Argonaute PRG-1 or counteracts downstream events such as WAGO 22G-RNA amplification and silencing. Studies suggest it is the latter: \texttt{gfp::cdk-1} allele in the maintenance phase of PIWI-triggered silencing (Figure 1.2) is transitively activated by \texttt{oma-1::gfp} even in the absence of PRG-1, which suggests that CSR-1 counteracts WAGO 22G-RNA-mediated silencing and not PRG-1 directly (Seth et al., 2013). In a CLASH study, Shen et al. capture piRNAs scanning and/or binding expressing germline genes, indicating that protection does not preclude the scanning of endogenous genes by PRG-1 and perhaps suggesting that CSR-1 functions downstream of PRG-1 (Shen et al., 2018).

While (1) CSR-1 is required to establish transitive activation, (2) its mRNA targets are associated with expressing genes, and (3) CSR-1 gene targets are
distinct from WAGO targets, its role in the protection of the worm germline genes against small RNA silencing remains to be further investigated.

**Introns impart protection against germline silencing**

Another candidate for the protection of germline genes is not an Argonaute, but instead, introns. Introns are intervening, non-coding parts of the gene that are excised from nascent transcripts by the spliceosome – a complex comprised of a number of protein and RNA factors (Shi, 2017). Intron splicing promotes transcription (Fong and Zhou, 2001; Furger et al., 2002), 3′-end processing (Chiou et al., 1991; Cooke et al., 1999; Nesic et al., 1993; Niwa et al., 1990), nuclear export (Luo et al., 2001; Valencia et al., 2008) and ribosomal recruitment and translation (Braddock et al., 1994; Lu and Cullen, 2003; Matsumoto et al., 1998; Nott et al., 2004; Nott et al., 2003). The benefits of introns in gene expression have been known for a while, and the phenomenon is collectively termed intron-mediated enhancement (IME) (Gallegos and Rose, 2015). The majority of *C. elegans* genes contain introns, while many viruses and transposable elements are either intronless or carry introns with suboptimal splicing motifs (Newman et al., 2018). This difference in gene architecture presents an opportunity for the host to use intron content as a feature to detect non-self genes without prior exposure.

In yeast, stalled spliceosomes serve as a beacon for the small RNA silencing machinery to initiate the suppression of defective mRNA and the corresponding gene locus (Dumesic et al., 2013). The Ruvkun group showed that some small RNA silencing targets, like transposon and unconserved endogenous genes, exhibit
retention on the spliceosome in *C. elegans* embryos (Newman et al., 2018). In plants, intronless transgenes are more susceptible to RdRP-mediated silencing compared to homologous intron-containing genes (Christie et al., 2011). In *C. elegans*, various splicing machinery components promote activity of WAGO 22G-RNA pathways (Akay et al., 2017; Jiao et al., 2019; Newman et al., 2018). In our studies, we show that transgenic constructs without introns induce heritable, PIWI-independent, WAGO-mediated silencing in *C. elegans* germline but not somatic tissues – a phenomenon we termed “intronless silencing” (see Chapter II or Makeyeva et al., 2021). In contrast, homologous intron-containing transgenes exhibit visible germline expression, suggesting that introns and/or mRNA splicing counteract small RNA-mediated gene suppression. In agreement with that, Akay et al. demonstrate that decreasing the intron number from three to one in a PIWI reporter leads to a 5-6-fold increase in antisense small RNA accumulation (Akay et al., 2017). Intronless silencing is not limited to transgenic constructs and reporters: we find that removal of introns from endogenous germline genes also triggers robust small RNA-mediated silencing (Chapter II or Makeyeva et al., 2021). In addition, endogenous intron-containing genes can exert transitive protection on homologous sequences within intronless transgenes, resembling RNAa phenomenon.

Why and how exactly do introns impart gene licensing? Frøkjær-Jensen et al. show that introns with a higher density of 10-basepair periodic An/Tn-clusters (PATC) positively correlate with transgenic gene expression in the germline, while PATC in gene coding regions do not exhibit the same trend (Frokjaer-Jensen et al., 2016). The Lee group also showed that PATC-rich introns can counteract piRNA
silencing (Zhang et al., 2018). Given PATC periodicity and association with open chromatin, the clusters are thought to oppose WAGO 22G-RNA silencing via chromatin modulation (Fire et al., 2006; Frokjaer-Jensen et al., 2016; Zhang et al., 2018). However, both groups note that many expressing germline genes are not PATC-rich, indicating there may be yet another licensing feature. In light of our findings, where removal of PATC-poor introns from an endogenous oma-1 gene triggers rde-3-dependent small RNA accumulation, we propose that introns impart additional protection independent of PATC.

Splicing certainly presents an obvious possibility, as cues deposited on spliced transcripts (e.g. exon junction complex) could serve as licensing marks that prevent the recruitment of silencing machinery. In fact, observed polyadenylation failure of intronless reporter transcripts (Chapter III) lends support to the splicing hypothesis, based on the studies illustrating that splicing promotes polyadenylation (Chiou et al., 1991; Huang and Gorman, 1990; Niwa et al., 1990). In turn, unpolyadenylated RNA may be recruited to RdRPs in a guide-independent manner, initiating or contributing to intronless silencing. Work in plants shows that unpolyadenylated, improperly terminated RNAs become RdRP substrates (Luo and Chen, 2007). The finding that rde-3-mediated silencing targets intronless reporters independently of primary Argonautes suggests silencing may be induced by default (Chapter II).

Apart from small RNA-mediated silencing, intronless reporters undergo a second mode of suppression that does not require WAGO 22G-RNAs (Chapter II). This “cis-mode” of intronless silencing cannot transitively suppress other cognate
genes and likely operates on a transcriptional level (Chapter III). Aberrant mRNA processing may contribute to cis-silencing, as histone-based intronless reporters, carrying special regulatory elements that enable splicing-independent mRNA metabolism, partially bypass cis-silencing (Chapter II). In human cell lines, the HUSH complex transcriptionally suppresses intronless genes (Seczynska et al., 2022), suggesting the phenomenon may be conserved among organisms. Future studies disabling the cis-mode of intronless silencing will determine if it also serves as a small RNA pathway trigger. In addition, a FISH assay with a histone-based intronless reporter revealed that cis-silencing can act post-transcriptionally as well (Chapter III). While it is clear that introns protect genes from various silencing pathways across distant organisms, the mechanistic underpinnings, as well as factors that confine intronless silencing to specific tissues, are yet to be determined.

The proceeding chapters of this thesis will provide evidence for intron-mediated protection and default silencing of intronless genes (Chapter II). Chapter III will describe preliminary findings of the intronless silencing mechanism and introduce an overall framework of PIWI-independent default silencing.

**Final note on regulation of gene expression**

Silencing and protection rules are not absolute. Instead, gene expression is determined by an intricate architecture of numerous silencing and protective signals, with a dash of randomness. Without such complexity, genomes would not be flexible, and genome flexibility is paramount to evolution.
CHAPTER II

Cues from mRNA splicing prevent default Argonaute silencing in *C. elegans*
ABSTRACT

In animals Argonaute small-RNA pathways scan germline transcripts to silence self-replicating genetic elements. Little is known however about how endogenous gene expression is recognized and licensed. Here we show that the presence of introns and by inference the process of mRNA splicing prevents default Argonaute-mediated silencing in the C. elegans germline. Silencing of intronless genes is initiated independently of the piRNA pathway but nevertheless engages multiple components of the downstream amplification and maintenance mechanisms that mediate transgenerational silencing including both nuclear and cytoplasmic members of the worm-specific Argonaute gene family (WAGO). Small RNAs amplified from intronless mRNAs can trans-silence cognate intron-containing genes. Interestingly, a second small RNA-independent cis-acting mode of silencing also acts on intronless mRNAs. Our findings suggest that cues put in place during mRNA splicing license germline gene expression and provide evidence for a splicing-dependent and dsRNA- and piRNA-independent mechanism that can program Argonaute silencing.
INTRODUCTION

The importance of guided search in biology is now widely appreciated thanks to the discoveries of RNA interference (RNAi) and CRISPR/Cas systems (Barrangou et al., 2007; Fire et al., 1998; Jansen et al., 2002). In both systems, Argonautes and Cas proteins become programmed with 20- to 30-nucleotide (nt) guide RNAs, and use the sequence information in their guides to find and regulate genetic targets (Brouns et al., 2008; Elbashir et al., 2001; Hammond et al., 2001; Marraffini and Sontheimer, 2008). One obvious ancestral function of these guided-search mechanisms is to defend against pathogenic nucleic acids, such as viruses and other mobile genetic elements, and several insights have been made into how pathogenic activity leads to programming of the respective search mechanisms (see reviews by Ding, 2010; Hille et al., 2018; Malone and Hannon, 2009). Anti-viral Argonautes, for example, are programmed by accessory factors that recognize and process virus-derived cytoplasmic double-stranded (ds)RNA into short interfering (si)RNA guides (Bernstein et al., 2001; reviewed in Wilson and Doudna, 2013). Interestingly, dsRNA-initiated programming of Argonautes has been co-opted to regulate endogenous gene expression by processing genomically encoded stem-loop RNAs into guides called micro-RNAs (miRNAs)(reviewed in Carthew and Sontheimer, 2009; Lee et al., 1993).

In animal germlines transposon surveillance is mediated by the Piwi Argonaute system which also silences transgenes and a number of other cellular genes (see review by Ozata et al., 2019). However, PIWI-interacting (pi)RNA guides are not processed from dsRNA but from single-stranded precursors transcribed by RNA polymerase II (Aravin et al., 2006; Brennecke et al., 2007; Girard et al., 2006; Grivna
et al., 2006; Gu et al., 2012b; Lau et al., 2006). Several components of the machinery that processes piRNA precursors are conserved, however the organization of precursor genes differ widely among organisms and the precise cues that cause piRNA precursors to be recognized by the Piwi Argonaute programming machinery are not understood.

The C. elegans germline is an excellent model in which to study mechanisms of Argonaute programming. In addition to the canonical RDE-1 Argonaute system that detects foreign dsRNA (Steiner et al., 2009; Tabara et al., 1999; Yigit et al., 2006), the C. elegans germline expresses multiple Argonaute systems that collectively engage nearly all germline mRNAs (Batista et al., 2008; Claycomb et al., 2009; Conine et al., 2010; Conine et al., 2013; Das et al., 2008; Gent et al., 2010; Gu et al., 2009; Han et al., 2009; Vasale et al., 2010). The C. elegans Piwi, PRG-1, engages tens of thousands of 21-nt 5’ U piRNA species (also called 21U-RNAs) (Batista et al., 2008; Gu et al., 2012b; Ruby et al., 2006). PRG-1 can initiate heritable silencing on its targets by recruiting RNA-dependent RNA polymerase, RdRP, which amplifies 22-nt antisense RNAs that most frequently initiate with a 5’ G residue (22G-RNAs) and are loaded onto members of an expanded group of Worm Argonautes, WAGO (Ashe et al., 2012; Bagijn et al., 2012; Das et al., 2008; Gu et al., 2009; Luteijn et al., 2012; Shirayama et al., 2012). A second RdRP system programs the Argonaute CSR-1 with 22G-RNAs that target germline-expressed mRNAs and are thought to protect them from piRNA-dependent induction of WAGO silencing (Seth et al., 2013; Wedeles et al., 2013). The WAGO and CSR-1 Argonautes and their respective repertoires of 22G-
RNAs are inherited transgenerationally via sperm and egg (Conine et al., 2013; Phillips et al., 2015; Seth et al., 2013; Shirayama et al., 2012).

In many organisms, mRNA splicing communicates with downstream events in mRNA expression, including mRNA 3’-end formation (Chiou et al., 1991; Cooke et al., 1999; Nesic et al., 1993; Niwa et al., 1990), nuclear export (Luo et al., 2001; Valencia et al., 2008), and mRNA translation on the ribosome (Braddock et al., 1994; Lu and Cullen, 2003; Matsumoto et al., 1998; Nott et al., 2004; Nott et al., 2003). Genetic studies in C. elegans have identified conserved components of the splicing machinery required for Argonaute-mediated silencing in C. elegans (Akay et al., 2017; Jiao et al., 2019; Newman et al., 2018). However, whether these factors directly or indirectly participate in silencing remains elusive.

Here we explore the relationship between splicing and Argonaute surveillance in the germline of the nematode C. elegans. We show that genes lacking introns, including endogenous genes from which introns were removed by precision genome editing, become default targets for Argonaute-mediated silencing and that the resulting amplified small RNAs can act in trans to silence cognate intron-containing genes. Moreover, intron-containing regions of an endogenous gene can also act in trans to protect homologous regions of an intronless transgene from small RNA targeting. We show that the small RNA-mediated arm of intronless silencing depends on the WAGO pathway but is not initiated by piRNAs.

Interestingly, intronless genes failed to express in the germline even when Argonaute small RNA pathways were disarmed by mutation, indicating that a small RNA-independent, cis-acting pathway acts in parallel to silence intronless genes in
the germline. Together, our findings support a model in which RNA splicing, and/or other splicing-dependent RNA processing mechanisms, impart signals on nascent transcripts that prevent their default recognition as templates for Argonaute guide-RNA programming.

RESULTS

Intron removal prevents gene expression in the germline but not the soma

To investigate how mRNA splicing affects Argonaute small RNA surveillance in *C. elegans*, we first compared the expression of intron-containing *cdk-1::gfp* and intronless *cdk-1*::*gfp* transgenes (Figure 2.1A). Aside from the presence or absence of introns, the transgenes have the same regulatory sequences and were inserted into identical locations on chromosome II using Mos1-mediated single-copy insertion (MosSCI) (Frokjaer-Jensen et al., 2008). The intron-containing version of *cdk-1::gfp* was robustly expressed in both the soma and germline (Figure 2.1B, C; (Shirayama et al., 2012). By contrast, the intronless *cdk-1*::*gfp* transgene was weakly expressed in the soma (Figure 2.1B) and completely silenced in the germline of three independent strains for over 10 generations (Figure 2.1C), suggesting that introns are essential for germline expression of *cdk-1::gfp*. 

32
Figure 2.1

(A) Schematic structures of intron-containing *cdk-1::gfp* and intronless *cdk-1*::*gfp* transgenes. Gray boxes, *cdk-1* coding sequence; green boxes, *gfp* coding sequence; V-shaped lines, introns.

(B and C) Fluorescence (upper) and DIC micrographs (lower) of (B) L3/L4-stage vulva in representative *cdk-1::gfp*, *cdk-1*::*gfp*+, and wild-type (WT) larvae, and (C) adult oocytes in the indicated strains. Scale bars, 20μm.

(D and E) Genetic crosses and representative epifluorescence images testing the ability of intronless *cdk-1*::*gfp*+ to trans-silence (D) *cdk-1::gfp* or (E) *oma-1::gfp*. Percent GFP+ (ON) or GFP− (OFF) worms and the number of worms analyzed in each generation is indicated. In (D), control (left) and test (right) crosses are shown. WT indicated non-transgenic worms. Images show presence (ON) or absence (OFF) of CDK-1::GFP signal in oocyte nuclei. In (E), *oma-1::gfp* is gradually trans-silenced by the intronless *cdk-1*::*gfp*+ transgene. Images show presence or absence of OMA-1::GFP signal in cytoplasm of oocytes. White dashed circles indicate location of nuclei in GFP− oocytes.
Intronless gene silencing involves small RNA-dependent and -independent pathways

Transgenes that are heritably silenced by small RNAs can trans-silence homologous transgenes (Shirayama et al., 2012). When crossed to non-transgenic wild-type worms, cdk-1::gfp transgenic animals produced offspring that exhibited bright and easily detected GFP signals (Figure 2.1D). By contrast, when crossed to worms bearing the cdk-1*::gfp* transgene, cdk-1::gfp transgenic animals produced offspring that were completely silenced (Figure 2.1D). Moreover, the cdk-1::gfp transgene remained silent in subsequent generations, even after segregating away the intronless allele (Figure 2.1D). A previous study showed that the oma-1::gfp transgene is resistant to transitive silencing (Seth et al., 2013). However, although the oma-1::gfp transgene was active in F1 progeny after crossing to worms bearing the cdk-1*::gfp* transgene, OMA-1::GFP expression gradually declined and ultimately became fully silenced over the course of five generations when propagated in the presence of the intronless cdk-1*::gfp* transgene (Figure 2.1E).

Trans-silencing of homologous transgenes by the intronless cdk-1*::gfp* transgene suggests that silencing is maintained by the RDE-3 and WAGO Argonaute-dependent small RNA effector pathway (Ashe et al., 2012; Bagijn et al., 2012; Shirayama et al., 2012). Small RNA-sequencing analyses suggested this to be the case. Compared to the active intron-containing cdk-1::gfp transgene, which produced few antisense small RNAs (Figure 2.2A) (Shirayama et al., 2012), the cdk-1*::gfp* transgene was robustly targeted by small RNAs that map antisense to the gfp* sequences of the transgene (Figure 2.2B).
RDE-3 is required for the production of WAGO-dependent small RNAs and thus for heritable silencing by the WAGO effector pathway (Chen et al., 2005; Shirayama et al., 2012). To ask if silencing of cdk-1*::gfp* requires RDE-3 activity we crossed the intronless transgene into an rde-3(ne3370) null mutant. We found that cdk-1*::gfp* remained silent—even over multiple generations—in the homozygous rde-3(ne3370) null mutant background (n=50) (Figure 2.2C). To ask if transitive silencing induced by cdk-1*::gfp* requires RDE-3 we crossed rde-3(ne3370) bearing the intronless transgene into a second homozygous rde-3(ne3370) strain expressing the intron-containing cdk-1::gfp transgene. Consistent with a requirement for RDE-3 activity, trans-silencing failed to occur in the rde-3(ne3370)-mutant background, even after propagating the transgenes together for several generations (Figure 2.2C). To ask if RDE-3 is required to maintain the silencing of a transi
tively silenced intron-containing oma-1::gfp transgene, we used precision genome editing to delete rde-3 in a >5th generation silenced strain (Figure 2.2D). We found that 100% of the F2 rde-3 homozygous progeny generated in three independent CRISPR lines exhibited bright GFP expression in proximal oocytes indicative of reactivation of the transi
tively silenced gene (Figure 2.2D). CRISPR editing in this (Figure 2.2D, legend) and all subsequent experiments was confirmed by genomic PCR and Sanger sequencing. As expected, 22G-RNAs antisense to gfp* failed to accumulate in the rde-3 mutants (Figure 2.2E). Thus although intronless cdk-1*::gfp* is targeted by an RDE-3-dependent small RNA pathway, it also appears to be silenced by a cis-
acting mechanism, independent of small RNAs.
Figure 2.2

(A and B) Plots showing the density of antisense small RNAs mapping along the coding regions of (A) cdk-1::gfp or (B) cdk-1*:gfp*. Positions of exon junctions in cdk-1::gfp (and corresponding positions in intronless cdk-1*:gfp*) are indicated by broken vertical lines in gene cartoons. The height of each bar represents the number of reads that begin at that position per million total reads.

(C and D) Genetic analyses and representative epifluorescence images testing the role of RDE-3 in intronless cdk-1*:gfp* silencing. Percentage of worms with GFP+ (ON) or GFP− (OFF) germ cells and the number of worms analyzed in each generation is indicated. In (C), intronless cdk-1*:gfp* remained silenced when crossed with an (C) rde-3(ne3370) null mutant. Crosses reveal that cdk-1*:gfp* does not trans-silence cdk-1::gfp in the rde-3 mutant. In (D), CRISPR was used to delete rde-3 (ne4865) in
worms in which oma-1::gfp is trans-silenced by cdk-1*::gfp*. OMA-1::GFP signal was restored after homozygosing rde-3(ne4865). CRISPR gene editing events were validated by genomic PCR and Sanger sequencing.

(E) Density of antisense small RNAs mapping along cdk-1*::gfp* in the rde-3(ne3370) mutant.

(F) Intronless cdk-1*::gfp* was silenced when introduced into a prg-1(tm872) null mutant. Crosses reveal that cdk-1*::gfp* trans-silences cdk-1::gfp in the prg-1 mutant.

(G) Density of antisense small RNAs mapping along cdk-1*::gfp* in the prg-1(tm872) mutant.
**Intronless gene silencing does not require the piRNA pathway**

The silencing of foreign sequences by the RDE-3/WAGO-dependent small RNA pathway is initiated by the PIWI Argonaute PRG-1 and piRNAs (Ashe et al., 2012; Bagijn et al., 2012; Lee et al., 2012; Shirayama et al., 2012). We therefore tested if PRG-1 initiates silencing of the intronless *cdk-1*::*gfp* transgene. Using MosSCI, we inserted the *cdk-1*::*gfp* transgene directly into a *prg-1(tm872)* null mutant (at the same chromosomal site used above) and obtained four independent insertion lines (Frokjaer-Jensen et al., 2008). None of these expressed the intronless *cdk-1*::*gfp* transgene in the germline (Figure 2.2F), indicating that PRG-1 is not required for silencing of the intronless *cdk-1*::*gfp* transgene. Moreover, in the complete absence of PRG-1, we found that the intronless *cdk-1*::*gfp* transgene was still able to trans-silence an intron-containing *cdk-1::gfp* transgene (100%, n=20; Figure 2.2F). As in wild type, trans-silencing in the *prg-1* mutant background was maintained for multiple generations after segregation away from the intronless allele (Figure 2.2F). These results indicate that PRG-1 is not required to initiate small RNA-dependent silencing induced by the intronless *cdk-1*::*gfp* transgene. Indeed, small RNAs antisense to the *gfp* transgene accumulated to higher levels in *prg-1* mutants than in wild-type worms (Figure 2.2G), suggesting that the piRNA pathway competes with the intronless silencing pathway for RDE-3 or other small-RNA silencing components. Thus, piRNAs do not initiate small RNA-mediated silencing of the intronless *cdk-1*::*gfp* transgene, nor are piRNAs required for cis-silencing of the intronless transgene.
Endogenous genes are sensitive to intronless silencing

The finding that 22G-RNAs are limited to the *gfp* sequences implies a difference of some kind between endogenous and foreign sequences in the transgene. To test if endogenous sequences can elicit intronless silencing, we used CRISPR to first delete the majority of the coding region—including the introns—of the non-essential *oma-1* gene and then to insert either the intron-containing *oma-1* gene or the intronless *oma-1* cDNA (Figure 2.3A). To assess the function of the edited *oma-1* alleles, we used RNAi to knock down *oma-2*. The *oma-1* and *oma-2* genes are redundantly required for oocyte maturation (Detwiler et al., 2001; Shimada et al., 2002). In worms with the restored, intron-containing *oma-1* gene, we found that *oma-2*(RNAi) resulted in viable and fertile progeny (n=41), indicating that the *oma-1* gene, when restored with introns, is functional. By contrast, *oma-2*(RNAi) resulted in sterile progeny (n=39), in all three strains edited with intronless *oma-1* *, indicating that the intronless allele is silenced.

RT-qPCR analysis supported *oma-1* *silencing and showed that the mRNA produced by the intron containing allele was about 90 times more abundant compared to the mRNA produced by the *oma-1* *allele (Figure 2.4A). Small RNA sequencing revealed a marked accumulation of 22G-RNAs targeting *oma-1* * (Figure 2.3B–2.3C). The majority of 22G-RNAs mapped to the 3'-untranslated region immediately after the stop codon with additional peaks in the 5' half of the coding region (Figure 2.3C). As was true for the *cdk-1* *::gfp* *transgene, the 22G-RNAs targeting *oma-1* * were absent in *rde-3* mutant worms and were increased in *prg-1* mutant worms (Figure 2.3D–2.3E). Thus, an intronless *oma-1* *allele at the endogenous *oma-1* locus triggers an RDE-3-dependent response that is independent of piRNAs.
To determine if \textit{oma-1}* is silenced by a small RNA-independent mechanism, we used CRISPR HDR to precisely delete the \textit{oma-2} gene in the \textit{rde-3} mutant background. In this experiment, RNAi knockdown of \textit{oma-2} could not be used because \textit{rde-3} mutants are RNAi deficient. As a control, we also deleted \textit{oma-2} in a strain where \textit{oma-1} was restored with introns (Figure 2.3A). For each genetic background, three independent \textit{oma-2} deletions were generated and were confirmed by genomic PCR and Sanger sequencing. In strains where \textit{oma-1} was restored with introns, F2 animals homozygous for the \textit{oma-2} deletion were 100% fertile (Figure 2.3A). By contrast, in strains restored with intronless \textit{oma-1}, 100% of the F2 worms homozygous for the \textit{oma-2} deletions were sterile (Figure 2.3A). Thus, the intronless \textit{oma-1}* allele generates RDE-3-dependent small RNAs but is also silenced by an additional \textit{rde-3}-independent mechanism. In agreement with this conclusion, although RT-qPCR analysis showed that \textit{oma-1}* mRNA levels were slightly increased in \textit{rde-3} mutants, they were nevertheless more than 50-fold lower than the levels of the mRNA produced by the intron-containing \textit{oma-1} gene (Figure 2.4B).
Figure 2.3. Endogenous genes are sensitive to intronless silencing.
(A) Schematic outlining the removal of endogenous oma-1 coding region and insertion of intron-containing oma-1 or intronless oma-1* alleles. CRISPR was used to modify the endogenous oma-1 locus. An in-frame 3xflag sequence was first inserted at the 3' end of the oma-1 gene. The oma-1 gene from the middle of exon 1 to the end of exon 6 was replaced with a short stuffer containing a new CRISPR guide site (marked with a black dashed rectangle) to allow subsequent insertion of the wild-type oma-1 sequence with or without (oma-1*) introns. Silencing of intronless oma-1* was confirmed by oma-2(RNAi). To test whether intronless oma-1* is also silenced by the cis-silencing pathway, CRISPR was used to delete rde-3 and then oma-2. The oma-2 genotype of F2 progeny (n=48) was determined by PCR. The percentage of each genotype is indicated, and the percentage of fertile or sterile worms is indicated. Every CRISPR gene editing event was validated by genomic PCR and Sanger sequencing.

(B and C) Plots showing the density of antisense small RNAs mapping along the coding regions of (B) oma-1::3xflag or (C) oma-1*::3xflag. Positions of exon junctions in oma-1::3xflag (and corresponding locations in intronless oma-1*) indicated by broken vertical lines in gene cartoon. The height of each bar represents the number of reads that begin at that position per million total reads.

(D and E) Plots showing the density of antisense small RNAs mapping along oma-1*::3xflag in (D) rde-3(ne4871) or (E) prg-1(ne4844) mutants.
**Figure 2.4**

(A, B) RT-qPCR of intronless and intron-containing *oma-1* alleles from one-day old adults. Whole worm pellet was used for total RNA extraction. *oma-1* expression was normalized by actin. (A) Fold change relative to intronless *oma-1* is shown (p = 0.00189). Panel (B) shows that intronless *oma-1* expression in *rde-3(ne4871)* mutant animals is slightly higher compared to the strain with an intact *rde-3* (p = 0.00254).

In (C), RT-qPCR was performed on adult gonadal tissue to compare expression of intronless and intron-containing *gfp::his-61*. RNA polymerase II subunit *ama-1* was used as an endogenous control. Fold change relative to the intronless *gfp*::*his-61* is shown (p = 1.1e-6).

Bars indicate mean fold change, points represent technical replicates, and error bars are based on standard deviation.
Regulatory elements of a naturally intronless gene can partially suppress intronless silencing

Replication-dependent histones comprise a group of intronless genes that are robustly expressed in dividing somatic and germline tissues (Keall et al., 2007; Pettitt et al., 2002; Robbins and Borun, 1967). To explore the sensitivity of a histone gene to intronless silencing, we used CRISPR to insert gfp coding sequences—with or without introns—immediately after the start codon in the endogenous his-61 locus (Figure 2.5A). Both types of the tagged histone gene were robustly expressed in somatic tissues (including embryos; Figure 2.5A). Surprisingly, however, intronless gfp*::his-61—but not the intron-containing gfp::his-61—was completely silenced throughout the germline, including in oocytes, and in mitotic and pachytene nuclei of the adult germline (Figure 2.5A). Congruently, RT-qPCR analysis of dissected gonads showed that gfp*::his-61 mRNA level was about 30 times lower compared to the cognate intron-containing allele (Figure 2.4C). Small RNA sequencing revealed an abundant accumulation of 22G-RNAs targeting the gfp sequences of gfp*::his-61 (Figure 2.5B, upper right panel). Moreover, as with other intronless transgenes, the accumulation of gfp* 22G-RNAs depended on RDE-3 but not PRG-1 (Figure 2.5B).

Interestingly, whereas the intronless cdk-1*::gfp* and oma-1* genes remained completely silenced in rde-3-mutant germlines, we observed that gfp*::his-61 was expressed in the distal mitotic region in rde-3 germlines (Figure 2.5A, bottom panel). This result suggests that intronless gfp*::his-61 is less sensitive to, or can partially bypass the cis-silencing pathway.
Appending the \textit{gfp} ORF approximately triples the open reading frame of the \textit{his-61} gene. We wondered if this added length might overwhelm the ability of \textit{his-61} regulatory sequences to prevent intronless silencing. To explore this possibility, we used CRISPR to replace the \textit{his-61} open reading frame with the 28 amino acid SV40 nuclear localization signal (NLS), shortening the gene by 25\% (Figure 2.5C). This \textit{gfp*::nls} gene driven at the endogenous locus by the \textit{his-61} promoter and UTR remained silent throughout the germline (Figure 2.5C). However, upon the inactivation of \textit{rde-3}, GFP was visible throughout the germline, including within oocyte nuclei (Figure 2.5C middle and bottom panels). Thus, shortening the intronless gene to approximately twice the length of the endogenous histone gene was sufficient to prevent cis-acting silencing throughout the germline but did not prevent small RNA-dependent silencing of \textit{gfp*::nls}. 
Figure 2.5

(A) gfp::his-61

Embryos

100% ON (n=30)

Oocytes

100% ON (n=15)

Distal tip

100% ON (n=15)

rde-3; gfp::his-61

No cis-silencing in distal region

100% ON (n=30)

(B) gfp::his-61

Normalized reads

80

60

40

20

0

rde-3; gfp::his-61

Normalized reads

80

60

40

20

0

(C) gfp::his-61

3’ UTR

CRISPR + HDR

gfp::nls

CRISPR rde-3

rde-3; gfp::nls

gfp::nls

Oocytes

100% ON (n=25)

100% OFF (n=20)

Distal tip

rde-3; gfp::nls

No cis-silencing throughout germline

100% ON (n=20)
Figure 2.5. Regulatory elements of a naturally intronless gene can partially suppress intronless silencing.

(A) Schematic gene structures of intron-containing \textit{gfp::his-61} (left) and intronless \textit{gfp*::his-61} (right), and representative epifluorescence images analyzing GFP expression in embryos (soma), oocytes, or distal tips of dissected gonads (outlined with white dashes in \textit{gfp*::his-61}). CRISPR was used to delete \textit{rde-3(ne4848)} to test the role of RDE-3 in silencing intronless \textit{gfp*::his-61}. In the resulting \textit{rde-3} mutant epifluorescence (upper) and DIC (lower) imaging of a representative gonadal arm shows GFP signal in the distal germline. GFP signal gradually decreases in the pachytene region and is absent in oocytes. Percent GFP+ (ON) or GFP– (OFF) worms and the number of worms analyzed is indicated. Every CRISPR gene editing event was validated by somatic GFP expression (in cases of \textit{gfp} integration), genomic PCR and Sanger sequencing.

(B) Density of antisense small RNAs mapping along \textit{gfp::his-61} (top left) or \textit{gfp*::his-61} (remaining plots). Small RNAs targeting intronless \textit{gfp*::his-61} were examined in wild-type (top right), \textit{rde-3(ne4848)} (bottom left), and \textit{prg-1(ne4766)} (bottom right).

(C) Diagram illustrating replacement of \textit{his-61} with an in-frame SV40 nuclear localization signal (\textit{nls}), and deletion of \textit{rde-3(ne4850)}. Representative epifluorescence images of \textit{gfp*::nls} in \textit{rde-3(ne4850)} or wild-type oocytes (white dashed circles indicate GFP– nuclei). After deleting \textit{rde-3(ne4850)}, GFP signal was visible in the oocytes and in the entire germline (fluorescence and DIC micrographs at the bottom). Percent GFP+ (ON) or GFP– (OFF) worms and the number of worms analyzed is indicated. Every CRISPR gene editing event was validated by genomic PCR and Sanger sequencing.
Intronless silencing requires WAGO Argonautes but is initiated independently of known primary Argonautes

Desilencing of \textit{gfp*:his-61} in the distal germline of \textit{rde-3} mutants suggested its use as an assay to quickly test whether other known small-RNA pathway factors promote intronless silencing. To perform this assay, we used CRISPR editing to either mutate known silencing factors in the \textit{gfp*:his-61} strain or introduce \textit{gfp*} into the \textit{his-61} locus in pre-existing homozygous RNA-silencing mutant strains. The de novo introduction of \textit{gfp*} was not only much faster than crossing, especially when strains bearing multiple mutants were required, but also had the advantage of enabling the assay (in principle) to identify factors required for both the initiation and maintenance of silencing. We validated this assay by inserting intronless \textit{gfp*} into the \textit{his-61} locus of an \textit{rde-3(ne4852)} mutant and confirmed that GFP was visible in the distal gonads of all four independently generated strains (Figure 2.6B). RDE-3 is known to promote the formation of RNA templates used by the partially redundant cellular RdRPs RRF-1 and EGO-1 to amplify silencing signals (Aoki et al., 2007; Chen et al., 2005; Gu et al., 2009; Shukla et al., 2020). \textit{rrf-1} mutants are viable and fertile, but EGO-1 is essential for fertility and embryo viability. We therefore depleted EGO-1 activity using an auxin-inducible degradation system (\textit{ego-1::degron}; see Experimental Procedures) (Zhang et al., 2015). Whereas \textit{gfp*:his-61} was silenced in \textit{rrf-1} and \textit{ego-1::degron} single mutants, it was expressed in the distal germline of the \textit{ego-1::degron, rrf-1} double mutant in the presence of auxin (Figure 2.6B). Mutations in the remaining cellular RdRPs, RRF-2 (of unknown function) and RRF-3 (required for 26G RNA production), did not affect \textit{gfp*:his-61} expression (Figure 2.6B). Consistent with the idea that RRF-
3 and 26G RNAs do not promote intronless silencing, we did not detect 26-nt small RNAs targeting the intronless reporter in our sequencing data (Figure 2.7).

The transitive nature of intronless silencing and the involvement of both RDE-3 and RdRPs suggests the involvement of downstream WAGO Argonautes (Gu et al., 2009; Yigit et al., 2006). To test WAGOs directly, we introduced gfp* into the his-61 locus in WAGO single and multiple mutant strains. The gfp*::his-61 allele was silenced in wago-1 and wago-9/hrde-1 single mutants, but was desilenced in the distal zone in the double mutant and in a previously constructed strain bearing mutations in all 12 wago genes (Gu et al., 2009; Yigit et al., 2006) (Figure 2.6B). These findings suggest that multiple WAGOs, including a predominantly nuclear family member (WAGO-9/HRDE-1) and a predominantly cytoplasmic member (WAGO-1), contribute to intronless silencing (Figure 2.6B) (Ashe et al., 2012; Buckley et al., 2012; Gu et al., 2009; Shirayama et al., 2012).

Consistent with these findings, small RNA sequencing studies on wago-9/hrde-1 and its nuclear RNAi co-factor nrde-4 (Burkhart et al., 2011; Guang et al., 2010; Guang et al., 2008) revealed that many 22G-RNAs targeting gfp were missing in both mutants (for unknown reasons these depleted 22G-RNAs mapped primarily to the second half of gfp, Figure 2.8). Together these findings indicate that intronless silencing engages both the nuclear and cytoplasmic arms of the WAGO pathway.

WAGO-dependent silencing can be initiated by and/or function together with several “primary” Argonautes, including PRG-1, RDE-1, ERGO-1, and redundant ALG-3/ALG-4 (Bagijn et al., 2012; Conine et al., 2010; Correa et al., 2010; Das et al., 2008; Vasale et al., 2010; Yigit et al., 2006). Because PRG-1 is not required for
intronless gene silencing, we were particularly interested to test whether silencing depends on any of the other known primary Argonautes. In addition, we tested the prg-1 homolog prg-2 which is predicted to be a pseudogene. We found that gfp\(^*\)::his-61 was silenced when introduced by CRISPR in all of these Argonaute mutant strains (Figure 2.6B). For each edited strain, we confirmed proper gfp\(^*\) insertion by genomic PCR and sequencing. Thus, none of the primary Argonautes that are known to engage WAGO-dependent silencing are required for intronless gene silencing.
Figure 2.6. **Small RNA-mediated intronless silencing is initiated independently of known primary Argonautes.**

(A) Assay to test if known small RNA factors are required for intronless silencing. Intronless *gfp*:his-61 fusion was generated via CRISPR in small RNA pathway mutants (e.g., rde-3) and animals were examined for germline GFP signal (especially in the distal germline).

(B) List of RdRP, Argonaute, and RNAi co-factor mutants tested in this analysis and percent GFP+ worms in each mutant. Numbers in parentheses indicate the total number of worms examined from the number of independent lines. Every CRISPR gene editing event for *gfp* integration at his-61 locus was validated with phenotypic observation of somatic GFP expression, genomic PCR and Sanger sequencing analyses. Mutants denoted with a double asterisk were generated by CRISPR in the *gfp*:his-61 background, followed by genomic PCR and Sanger sequencing analyses.
Figure 2.7

(A, B) Length and first-nucleotide distribution of small RNA reads mapping to the gfp::his-61 sequence from strains bearing either (A) intron-containing or (B) intronless gfp*:his-61. Horizontal axis indicates length of small RNA reads. Height of vertical bars represents abundance of reads of indicated size and starting nucleotide, expressed as a percent of total (A) 1,125 and (B) 11,695 raw reads depicted. Bars are colored based on the identity of the first 5’ nucleotide.

Figure 2.8

(A-B) Plots showing the density of antisense 22G-RNA species mapping to the intronless gfp*:his-61 in nrde-4(ne4679) and wago-9(ne4872)-mutant background. Broken vertical lines mark the positions corresponding to exon-exon junctions in the cognate intron-containing construct. The height of each bar represents the number of reads that begin at that position per million total reads.

Figure 2.7. Intronless reporters are not targeted by the 26G-RNA pathways.

(A, B) Length and first-nucleotide distribution of small RNA reads mapping to the gfp::his-61 sequence from strains bearing either (A) intron-containing or (B) intronless gfp*:his-61. Horizontal axis indicates length of small RNA reads. Height of vertical bars represents abundance of reads of indicated size and starting nucleotide, expressed as a percent of total (A) 1,125 and (B) 11,695 raw reads depicted. Bars are colored based on the identity of the first 5’ nucleotide.

Figure 2.8. Nuclear RNAi pathway contributes to the small RNA-mediated silencing of intronless reporters.

(A-B) Plots showing the density of antisense 22G-RNA species mapping to the intronless gfp*:his-61 in nrde-4(ne4679) and wago-9(ne4872)-mutant background. Broken vertical lines mark the positions corresponding to exon-exon junctions in the cognate intron-containing construct. The height of each bar represents the number of reads that begin at that position per million total reads.
Trans-acting signals from endogenous genes with intact splicing protect against small RNA silencing

In our oma-1* replacement experiments above, we were intrigued by the RDE-3-dependent accumulation of 22G-RNAs targeting the oma-1 coding sequence (Figure 2.3C). Typically, WAGO-dependent 22G-RNAs only accumulate antisense to foreign sequences of a silenced transgene (Seth et al., 2013; Seth et al., 2018; Shirayama et al., 2012). For example, when we directly inserted an intronless oma-1*:gfp* transgene into chromosome II using MosSCI (Frokjaer-Jensen et al., 2008), 22G-RNA accumulation was limited to the gfp portion of the transgene (Figure 2.9A). In this and all of the previously described MosSCI experiments, the endogenous intron-containing locus was present. We therefore wondered if trans-acting signals emanating from the existing endogenous intron-containing locus prevent the accumulation of 22G-RNAs targeting the cognate sequences in the silenced transgene. To test this possibility, we deleted endogenous oma-1 by CRISPR and then inserted the intronless oma-1*:gfp* into chromosome II using MosSCI (Frokjaer-Jensen et al., 2008) (Figure 2.9B). Strikingly, in the absence of the endogenous locus, antisense 22G-RNAs targeting both the oma-1 and gfp portions of the intronless transgene accumulated to equally high levels (Figure 2.9B). Thus, the endogenous, spliced allele of oma-1 communicates with cognate sequences from oma-1 transgenes, preventing them from templating small RNA-silencing signals.
Figure 2.9. Trans-acting signals from endogenous loci protect against small RNA-mediated silencing.

(A and B) Plots showing the density of antisense small RNAs mapping along intronless *oma-1*:::gfp* transgene in the (A) wild-type or (B) *oma-1* deletion worms. The schematic in (B) outlines the deletion of endogenous *oma-1* (including promoter and 3' UTR) on chromosome IV (LGIV), single-copy insertion of intronless *oma-1*:::gfp* on LGII, and small RNA sequencing. *oma-1* deletion at the endogenous locus was validated by genomic PCR and Sanger sequencing analyses.

(C) Model figure. Introns/splicing protect transcripts from default WAGO-dependent and cis-silencing. (top) Factors associated with splicing (orange ovals) counteract silencing cues (red octagons) deposited on pre-mRNA by default. (bottom) Silencing cues on unspliced transcript recruit RdRP, which makes small RNAs that guide Argonaute-mediated trans-silencing of cognate intron-containing genes (dashed arrow). In addition, possibly in response to the same cues, unspliced transcripts are silenced in cis, for example by disrupting mRNA processing or promoting export to nuage where RNAs are sequestered and used for small-RNA templating.
DISCUSSION

Genetic studies in plants, animals, and fungi have identified mRNA splicing components as factors required for RNA silencing by Argonautes (Akay et al., 2017; Bayne et al., 2008; Czech et al., 2013; Handler et al., 2013; Herr et al., 2006). Moreover, phylogenetic studies suggest that genes encoding spliceosome and RNA silencing components tend to be retained together in eukaryotic genomes (Tabach et al., 2013), consistent with a functional relationship. The essential role of mRNA splicing in gene expression, however, has made it difficult to determine whether splicing factors directly participate in Argonaute-mediated silencing or indirectly promote surveillance by regulating the expression of the silencing machinery—e.g., guide RNAs, Argonautes, or other essential co-factors (Goriaux et al., 2014; Kallgren et al., 2014; Zhang et al., 2014). Removing the introns from reporter genes, rather than impairing splicing by genetic perturbations, has allowed us to circumvent many of these issues and to uncover surprising and robust interactions between intronless mRNAs and Argonaute surveillance pathways.

We have shown that the presence of functional introns—and thus mRNA splicing—prevents the default targeting of *C. elegans* germline transcripts by a piRNA-independent Argonaute pathway. Our findings imply that nascent transcripts are marked in ways that are interpreted by the Argonaute surveillance machinery. For example, positive factors that actively license expression might be deposited on intron-containing transcripts. Alternatively, negative signals that would otherwise trigger default recognition by the silencing machinery might be actively removed in
response to splicing. Of course, some combination of positive and negative cues could also be involved.

Because nearly all germline transcripts template the production of 22G-RNAs in *C. elegans* (Claycomb et al., 2009; Gu et al., 2009; Ruby et al., 2006), RdRPs might be recruited by default to all mRNAs, and cues from splicing might help determine which Argonaute system is loaded: the CSR-1 Argonaute, which appears to protect expression (Claycomb et al., 2009; Conine et al., 2013; Seth et al., 2013; Wedeles et al., 2013), or the WAGO Argonautes that promote silencing (Bagijn et al., 2012; Gu et al., 2009; Shirayama et al., 2012; Vasale et al., 2010). Loading of the CSR-1 Argonaute with 22G-RNAs templated from spliced mRNAs could account for trans-acting signals that promote expression of cognate genes, while the loading of WAGO Argonautes with 22G-RNAs templated from intronless RNAs could promote silencing. Surveillance by the piRNA system could fit into this picture by functioning to overcome either CSR-1 protection or other positive signals put in place by splicing (Model Figure 6C). Indeed, a recent study showed that a piRNA sensor with three introns accumulated fewer WAGO 22G-RNAs than did an otherwise identical sensor with a single intron, strongly suggesting that genes with more introns are less prone to piRNA silencing (Akay et al., 2017). In addition, the type, number, and location of introns have been shown to influence gene expression in the *C. elegans* germline (Aljohani et al., 2020). A previous study suggested that mRNAs with weak splicing signals are retained on the spliceosome and become targets of small RNA silencing (Dumesic et al., 2013; Newman et al., 2018). Perhaps the removal of all introns from a gene causes it to express a nascent transcript where the only remaining splicing signals would, by
definition, be non-optimal. In the absence of strong signals, therefore, weak splicing signals might abound and invariably result in default retention of the RNA in unresolved spliceosomes.

Components of the splicing machinery might also play roles in Argonaute surveillance outside of their functions in mRNA splicing. For example, the exon-junction complex (EJC), which is deposited on exons during splicing, promotes mRNA nuclear export (Luo et al., 2001) and signals the ribosome to promote nonsense-mediated decay of mRNAs with premature stop codons (Kim et al., 2001; Le Hir et al., 2001). Similarly, a recent study has implicated components of the U1 snRNP in a process that prevents the premature cleavage and polyadenylation of transcripts within introns (Kaida et al., 2010). If signals from the abortive recruitment of splicing machinery are responsible for the default silencing of intronless mRNAs observed here, then our findings imply that the presence of functional introns (or their successful removal by the splicing machinery) can counteract these signals. Perhaps the successful resolution of splicing recruits an RNA helicase that then translocates along the mRNA to ensure the removal of stalled spliceosomes or other co-transcriptionally deposited signals (orange oval in the model Figure 6C), liberating the mRNA from influences that would otherwise prevent expression or trigger Argonaute-mediated silencing.

We used RNAi to test whether components of the splicing machinery—including core factors required for spliceosome assembly, activation and catalysis, as well as components of the exon-junction and TREX complexes—promote the silencing of gfp*::his-61. RNAi of these factors failed to desilence gfp*::his-61 even in
the distal region of the gonad, despite clear phenotypic evidence of successful knockdown in many cases (Figure 2.10). In the future, it will be interesting to ask if reporters with some but not all introns removed, or containing weak splice consensus sequences, exhibit increased levels of silencing when splicing is perturbed.

A recent study identified EMB-4, a homolog of the intron-binding helicase Aquarius, as a WAGO-9/HRDE-1-interacting factor that increases the sensitivity of genes with multiple introns to piRNA silencing (Akay et al., 2017). EMB-4 is required for piRNA silencing only when multiple introns are present in the target, a finding consistent with the idea that positive signals from splicing promote resistance to small RNA-mediated silencing. EMB-4 was also enriched in co-IPs with the putative protective Argonaute CSR-1 (Tyc et al., 2017). Clearly further investigation is required to understand whether recruitment of EMB-4 during splicing, or more directly as an Argonaute co-factor, enables it to shape the sensitivity of intron-containing mRNAs to piRNA silencing.

We do not know what mechanisms underlie cis-silencing of intronless mRNAs, nor do we know whether or how these mechanisms trigger downstream recognition by the Argonaute small RNA system. Components of the THO complex, required for mRNA processing and export downstream of splicing (reviewed in Katahira, 2012), suppress the expression of a number of transposon families (Zhang et al., 2021, submitted). Interestingly, these THO factors appear to promote silencing of unspliced mRNAs independently of the piRNA pathway (Gen et al., 2021, co-submitted). Previous studies have linked THO components to the expression of unspliced piRNA precursors (Hur et al., 2016; Zhang et al., 2012a; Zhang et al., 2014), raising the
possibility that components of this complex promote intronless gene silencing directly in cis, and also promote the export of intronless mRNAs as piRNA precursors, suggesting that these factors may lie at or below the divergence of the cis- and trans-silencing pathways.

Many naturally intronless genes are nevertheless abundantly expressed, most of them short, including histone and spermatogenesis genes (Miller et al., 2004; Pettitt et al., 2002). Interestingly, the endogenous retro-element CER1 expresses an ~8-kilobase intronless mRNA in the germline, somehow avoiding both piRNA silencing and the intronless silencing mechanism described here (Dennis et al., 2012). Moreover, our intronless reporters are not silenced in the somatic cells of the worm, and mitotic vs meiotic germ cells exhibit different sensitivities to the cis-arm of intronless silencing. Thus, gene-specific and tissue-specific mechanisms can bypass or counteract intronless silencing.

Histone regulatory sequences appeared to partially counteract the cis-arm of intronless silencing. Indeed, they were sufficient to completely bypass cis-silencing when the intronless reporter was comprised of only a gfp::nls ORF. The histone regulatory sequences, however, did not prevent small RNA-mediated silencing of this transgene. The gfp ORF is approximately twice the size of endogenous his-61, suggesting that regulatory elements, perhaps in the 3’ UTR of his-61, are insufficient to shield the extended ORF from recognition by the intronless small-RNA silencing machinery. The finding that inclusion of introns in gfp protected the fusion mRNA from silencing suggests that splicing and the histone UTR both exert local influences that promote mRNA escape from these surveillance mechanisms.
The finding that histone regulatory regions can bypass cis-silencing but not small RNA-mediated silencing could mean that these mechanisms are initiated independently. Alternatively, it may be that small RNA pathways are simply more sensitive to these cues perhaps due to the amplification of small RNA silencing signals. Regulatory sequences in the histone mRNA 3’ UTR are known to recruit U7 snRNP, which includes core components of the spliceosome (Dominski and Marzluff, 2007), and thus could recruit licensing factors or activities that promote the local removal of default silencing signals in a manner analogous to that envisioned for the successful resolution of splicing (see Model Figure 6B and above discussion).

Our findings provide evidence that mRNA splicing and histone regulatory sequences recruit activities to nascent mRNAs that prevent the default programming of an Argonaute system in the C. elegans germline. Licensing mechanisms of this type could be used as a line of defense against viral gene expression. However, just as eukaryotic cells appear to co-opt an antiviral dsRNA response to regulate endogenous gene expression via miRNAs, it is interesting to contemplate that cells may deploy intronless RNAs in order to program Argonautes as a means to regulate cognate intron-containing mRNAs. This need not involve RdRPs, as many Argonautes engage guide RNAs transcribed by Pol II (Correa et al., 2010; Gu et al., 2012b). For example, by expressing an intronless antisense transcript from a processed pseudogene, the cell might initiate Argonaute programming that acts in trans to regulate the intron-containing gene from which the pseudogene was processed. Such a mechanism could underlie the recently discovered process of transcriptional compensation (El-Brolsy et al., 2019; Serobyan et al., 2020; Watanabe and Lin, 2014).
Conceivably, there are numerous as yet undiscovered cellular mechanisms that program Argonaute silencing, and an equal number of mechanisms that enable transcripts to navigate and evade surveillance by Argonaute systems. Understanding this complexity will require evaluating the response of Argonaute systems to alterations in the myriad RNA binding and modifying activities that engage transcripts throughout the mRNA life cycle.

**Figure 2.10**

<table>
<thead>
<tr>
<th>Target gene</th>
<th>% of <em>gfp</em>::<em>his-61</em> desilencing, # of tested</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>snr-2</em></td>
<td>0%, n=20</td>
</tr>
<tr>
<td><em>sftb-1</em></td>
<td>0%, n=20</td>
</tr>
<tr>
<td><em>yju-2</em></td>
<td>0%, n=20</td>
</tr>
<tr>
<td><em>F33D11.10 + Y65B4A.6</em></td>
<td>0%, n=20</td>
</tr>
<tr>
<td><em>hel-1</em></td>
<td>0%, n=20</td>
</tr>
<tr>
<td><em>mag-1</em></td>
<td>0%, n=20</td>
</tr>
<tr>
<td><em>rnp-4</em></td>
<td>0%, n=20</td>
</tr>
<tr>
<td><em>rnp-2</em></td>
<td>0%, n=20</td>
</tr>
<tr>
<td><em>cyn-13</em></td>
<td>0%, n=20</td>
</tr>
<tr>
<td><em>sel-13</em></td>
<td>0%, n=20</td>
</tr>
<tr>
<td><em>snu-13</em></td>
<td>0%, n=20</td>
</tr>
<tr>
<td><em>rnp-7</em></td>
<td>0%, n=20</td>
</tr>
<tr>
<td><em>mog-1</em></td>
<td>0%, n=20</td>
</tr>
<tr>
<td><em>mut-16</em></td>
<td>70%, n=20</td>
</tr>
</tbody>
</table>

**Figure 2.10. Splicing machinery does not promote default intronless silencing.**

Summary of GFP*::*HIS-61 desilencing in the germline after RNAi of the indicated target genes. L1-staged animals were exposed to RNAi by feeding and scored for GFP expression upon reaching adulthood. Due to larval arrest, to assay *snr-2, sftb-1, yju-2, F33D11.10 + Y65B4A.6*, and *hel-1*, L3-staged individuals were exposed to RNAi instead. Percent of GFP+ worms and the number of animals examined are indicated.
LIMITATIONS OF THE STUDY

There are several limitations of the current study. Foremost, we have not as yet identified a direct link between the splicing machinery and silencing. While the complete absence of introns results in silencing, we have not explored whether the size of exons, or the number of introns are important factors. The genetic mechanism(s) of “cis-silencing” are entirely unknown and may or may not be linked to the same cues that trigger small-RNA silencing. Finally, we used an intronless \textit{gfp}\textsuperscript{*::his-61} reporter to facilitate much of the genetic analysis of small RNA silencing described here. Further studies will be required to determine if findings based on this reporter hold true for other intronless genes.
MATERIALS AND METHODS

C. elegans strains and genetics

All strains in this study were derived from Bristol N2 line and cultured in a 20°C incubator, on normal growth media (NGM) agar plates seeded with E. coli OP50, essentially as described (Brenner, 1974). The majority of the assays were performed on hermaphrodites; males were used solely for crossing to generate additional strains or test transitive silencing as described in corresponding figures. The strains used in this study are listed in Table S1. Details of CRISPR alleles generated in the study, including description of the lesions, guide and donor repair sequences are provided in Table S2. Sanger sequencing trace files of the resulting CRISPR alleles can be found at Mendeley Data repository under https://doi.org/10.17632/c8h2f2prch.2 or using the following link https://data.mendeley.com/datasets/c8h2f2prch/2

Table S1. Strains used in the study

<table>
<thead>
<tr>
<th>Strain name</th>
<th>Genotype</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>N2</td>
<td>wild type (ancestral)</td>
<td></td>
</tr>
<tr>
<td>WM242</td>
<td>neSi12[cdk-1::gfp; Cbr-unc-119(+)] II; unc-119(ed9) III</td>
<td>MosSCI</td>
</tr>
<tr>
<td>WM730</td>
<td>neSi71[cdk-1*::gfp* (intronless); Cbr-unc-119(+)] II; unc-119(ed9) III</td>
<td>MosSCI</td>
</tr>
<tr>
<td>EG4322</td>
<td>ttTi5605 II; unc-119(ed9) III</td>
<td></td>
</tr>
<tr>
<td>WM731</td>
<td>neSi72[oma-1::gfp; Cbr-unc-119(+)] I; unc-119(ed9) III</td>
<td>MosSCI</td>
</tr>
<tr>
<td>WM732</td>
<td>neSi72[oma-1::gfp; Cbr-unc-119(+)] I; neSi71[cdk-1*::gfp* (intronless); Cbr-unc-119(+)] II; unc-119(ed9) III</td>
<td>Cross</td>
</tr>
<tr>
<td>WM733</td>
<td>rde-3(ne3370[null]) II; neSi12[cdk-1::gfp; Cbr-unc-119(+)] II</td>
<td>Cross</td>
</tr>
<tr>
<td>WM734</td>
<td>rde-3(ne3370[null]) II; neSi71[cdk-1*::gfp*(intronless); Cbr-unc-119(+)] II</td>
<td>Cross</td>
</tr>
<tr>
<td>WM735</td>
<td>rde-3(ne4865[null]) II; neSi72[oma-1::gfp; Cbr-unc-119(+)] I; neSi71[cdk-1*::gfp* (intronless); Cbr-unc-119(+)] II; unc-119(ed9) III</td>
<td>CRISPR</td>
</tr>
<tr>
<td>WM736</td>
<td>prg-1(tm872[null]); neSi12[cdk-1::gfp; Cbr-unc-119(+)] II; unc-119(ed9) III</td>
<td>Cross</td>
</tr>
</tbody>
</table>
WM737  prg-1(tm872[null]); neSi78[cdk-1*::gfp* (intronless); Cbr-unc-119(+)] II; unc-119(ed9) III
WM779  oma-1(ne4866[C-ter::3xFLAG]) IV
WM780  oma-1(ne4867[almost full CDS deletion;C-ter::3xFLAG]) IV
WM781  oma-1(ne4868[restored original intron-containing oma-1::3xFLAG]) IV
WM782  oma-1(ne4869[intronless oma-1*::3xFLAG]) IV
WM783  rde-3(ne4870[null]) I; oma-1(ne4868[restored original intron-containing oma-1::3xFLAG]) IV
WM784  rde-3(ne4871[null]) I; oma-1(ne4867[almost full CDS deletion;C-ter::3xFLAG]) IV
WM785  rde-3(ne4870[null]) I; oma-1(ne4868[restored original intron-containing oma-1::3xFLAG]) IV; oma-2(ne4842[null]) V
WM787  prg-1(ne4844[null]) I; oma-1(ne4867[almost full CDS deletion;C-ter::3xFLAG]) IV
WM747  prg-1(ne4845[intronless oma-1*::3xFLAG]) IV
WM748  his-61(ne4846[gfp::his-61]) IV
WM749  his-61(ne4847[intronless gfp*::his-61]) IV
WM750  rde-3(ne4848[null]) I; his-61(ne4847[intronless gfp*::his-61]) IV
WM751  rde-61(ne4849[his-61p::intronless gfp*::nls::his-61 3'UTR]) IV
WM752  rde-3(ne4850[null]) I; his-61(ne4849[his-61p::intronless gfp*::nls::his-61 3'UTR]) IV
WM682  prg-1(ne4766[null]) I
WM753  prg-1(ne4845[intronless oma-1*::3xFLAG]) IV
WM754  rde-3(ne4850[null]) I
WM755  rde-3(ne4852[null]) I; his-61(ne4853[intronless gfp*::his-61]) IV
WM45   rde-1(ne300[null]) V
WM757  rde-1(ne300[null]) V; his-61(ne4854[intronless gfp*::his-61]) IV
WM300  alg-4(ok1041) III; alg-3(tm1155) IV
WM758  alg-4(ok1041) III; alg-3(tm1155) IV; his-61(ne4855[intronless gfp*::his-61]) IV
WM158  ergo-1(tm1860) V
WM759  his-61(ne4856[intronless gfp*::his-61]) IV; ergo-1(tm1860) V
WM760  rrf-3(ne4857[null]) II
WM761  rrf-3(ne4858[null]) II; his-61(ne4858[intronless gfp*::his-61]) IV
WM756  rrf-1(pk1417), rrf-2(pk2040) I; rrf-3(pk1426) II
WM762  rrf-1(pk1417), rrf-2(pk2040) I; rrf-3(pk1426) II; his-61(ne4859[intronless gfp*::his-61]) IV
CA1199  unc-119(ed3) III; ieSi38[sun-1p::tir-1::mRuby::sun-1 3'UTR; Cbr-unc-119(+)] IV
WM763  ego-1(ne4860[degron::ego-1]) I; unc-119(ed3) III; ieSi38[sun-1p::tir-1::mRuby::sun-1 3'UTR; Cbr-unc-119(+)] IV
WM764  ego-1(ne4860[degron::ego-1]) I; unc-119(ed3) III; ieSi38[sun-1p::tir-1::mRuby::sun-1 3'UTR; Cbr-unc-119(+)]; his-61(ne4861[intronless gfp*::his-61]) IV
| WM766 | rrf-1(ne4862[null]); ego-1(ne4860[degron::ego-1]) I; unc-119(ed3) III; ieSi38[sun-1p::tir-1::mRuby::sun-1 3′UTR; Cbr-unc-119(+)]; his-61(ne4861[intronless gfp*::his-61]) IV | CRISPR |
| WM767 | neSi79[oma-1*::gfp* (intronless); Cbr-unc-119(+)] II; unc-119(ed9) III | MosSCI |
| WM768 | ttTi5605 II; unc-119(ed9) III; oma-1(ne4864[full deletion starts upstream of CDS and ends after 3′ UTR]) IV | CRISPR |
| WM769 | neSi80[oma-1*::gfp* (intronless); Cbr-unc-119(+)] II; unc-119(ed9) III; oma-1(ne4864[full deletion starts upstream of CDS and ends after 3′ UTR]) IV | MosSCI |
| WM770 | wago-9(ne4872[null]) III; his-61(ne4847[intronless gfp*::his-61]) IV | CRISPR |
| WM771 | wago-1(ne4873[null]) I; his-61(ne4847[intronless gfp*::his-61]) IV | CRISPR |
| WM772 | nrde-4(ne4679[p.Met12_Glu160del, fsTer21]) IV | T.Ishidate |
| WM773 | nrde-4(ne4679[null]), his-61(ne4874[intronless gfp*::his-61]) IV | CRISPR |
| WM774 | wago-1(ne4873[null]) I; wago-9(ne4875[null]) III; his-61(ne4847[intronless gfp*::his-61]) IV | CRISPR |
| WM191 | mago12: sago-2(tm894) ppw-1(tm914) ppw-2(tm1120) wago-2(tm2686) wago-1(tm1414) I; wago-11(tm1127) wago-5(tm1113) wago-4(tm1019) II; hrde-1(tm1200) sago-1(tm1195) III; wago-10(tm1186) V; nrde-3(tm1116) X. | |
| WM775 | mago12, his-61(ne4876[intronless gfp*::his-61]) IV | CRISPR |
| SX523 | prg-1(n4375); prg-2(n4358) | CRISPR |
| WM776 | prg-1(n4375) I; prg-2(n4358), his-61(ne4877[intronless gfp*::his-61]) IV | CRISPR |
| WM777 | prg-2(tm1094[null]) IV | CRISPR |
| WM778 | prg-2(tm1094[null]), his-61(ne4879[intronless gfp*::his-61]) IV | CRISPR |

**Gonad dissections**

Gonads were dissected in 0.5mM solution of tetramisole (Sigma-Aldrich) in PBSTw (PBS with 0.1% Tween 20) on Rite-On glass slides (Thermo Scientific), fixed with 2.5% paraformaldehyde in Happy buffer (81mM HEPES pH 6.9, 42mM NaCl, 5mM KCl, 2mM MgCl2, 1mM EGTA) (from personal correspondence with James Priess), covered with cover glass (MedSupply Partners), and directly imaged as described below.
**Microscopy**

Worms were mounted using a 0.5mM solution of tetramisole (Sigma-Aldrich) in M9 buffer (Brenner, 1974) on Rite-On glass slides (Thermo Scientific) and cover glass (MedSupply Partners). Epifluorescence and differential interference contrast (DIC) microscopy analysis and images were captured using an Axio Imager M2 Microscope (Zeiss), an ORCA-ER digital camera (Hamamatsu) and Zen (Zeiss) software. Images were processed using Fiji/ImageJ software (Schindelin et al., 2012).

**MosSCI genome editing**

Transgenic lines were generated in EG4322 background following Mos1-mediated single copy insertion (MosSCI) protocol as described (Frokjaer-Jensen et al., 2008).

**CRISPR/Cas9 genome editing**

CRISPR lines were generated by Cas9 ribonucleoprotein (RNP) editing (Dokshin et al., 2018; Ghanta and Mello, 2020). For *gfp* insertion in *gfp*::*his*-61 fusion strains and restored endogenous *oma-1* strains, the annealed PCR products with overhangs served as donors (Dokshin et al., 2018). For deletion mutations and short insertions (such as FLAG and degron), commercially synthesized single stranded DNA oligonucleotides served as repair templates (IDT). Genomic PCR and Sanger sequencing were used for the validation of CRISPR editing. Details of CRISPR alleles generated in the study, including description of the lesions, accession links to Sanger...
sequencing trace files, guide and repair sequences are provided in Table S2 (in the original publication (Makeyeva et al., 2021)).

**Small RNA Cloning and Deep Sequencing**

Synchronized L1 populations were plated, collected and flash frozen in four volumes of TRI Reagent (Sigma) at the young adult stage. On the day of the experiment, after three freeze-thaw cycles, worm pellets were combined with ceramic spheres (Lysing Matrix D, MP Biomedicals) and homogenized using FastPrep system (MP Biomedicals) for three 20-sec cycles at a speed of 6 m/s. Total RNA was extracted using first a phase separation reagent 1-Bromo-3-chloropropane (BCP, Molecular Research Center), then aqueous phase was transferred to Phase Lock Gel tubes (Quanta Bio) for additional separation with phenol-chloroform. Following isopropanol precipitation and 80% ethanol wash, total RNA was resuspended with water and frozen at -80C. Small RNA isolation and cloning were performed as described (Li et al., 2020; Seth et al., 2018). Briefly, small RNAs, size-selected using mirVana miRNA Isolation Kit (Thermo Fisher Scientific), were treated with a recombinant PIR-1 pyrophosphatase (a generous gift from Dr. Weifeng Gu) to remove the 5' γ and β phosphates, and ligated to 3' and 5' adapters using T4 RNA ligase 2 and T4 RNA ligase 1 (New England Biolabs), respectively. cDNA was then generated using SuperScript III Reverse Transcriptase (Thermo Fisher Scientific) and amplified via PCR. After gel size selection, final libraries were sequenced on an Illumina HiSeq
platform at the University of Massachusetts Medical School Deep Sequencing Core Facility.

**Data Analysis**

Following standard removal of adapter sequences using Cutadapt tool (Martin, 2011), reads were mapped using bowtie2 aligner (Langmead and Salzberg, 2012; Langmead et al., 2019), normalized by the number of the reads mapping to the genome (WB272) or transcriptome (WBcel235) and multiplied by 5,000,000. Visualization of antisense 22-nucleotide-long reads starting with a guanine, as well as small RNA length profiles, were generated using ggplot2 (Wickham, 2016). All scripts are available upon request.

**Data availability**

Original and processed datasets are publically available at GEO repository, under the following accession number: GSE178985.

**RNAi**

RNAi was performed by feeding worms *E. coli* strain HT115(DE3) transformed with the control vector L4440 or a gene-targeting construct from the *C. elegans* Ahringer or Vidal RNAi libraries (Kamath and Ahringer, 2003; Rual et al., 2004). Frozen bacterial stocks were streaked on ampicillin (100 μg/ml) and tetracycline (10 μg/ml)-
containing LB agar plates and grown overnight at 37°C. Then individual colonies were inoculated into LB with ampicillin (1:1000) and grown for 16-20 hrs on a shaker at 37°C. NGM plates containing 1 mM isopropyl β-d-thiogalactoside and 100 µg/ml ampicillin were seeded with the liquid culture (80 µl per plate) and incubated at room temperature for one day. L1 or L3 larvae were plated on RNAi plates and kept at 20°C until the phenotypes of the adults were scored.

Bacterial clones targeting the following genes were used: oma-2, snr-2, sftb-1, yju-2, F33D11.10, Y65B4A.6, hel-1, mag-1, rnp-4, rnp-2, cyn-13, sel-13, snu-13, rnp-7, mog-1, mut-16, pos-1.

**Gonad isolation and total RNA extraction**

Ten gonads from N2, gfp::his-61 and gfp*::his-61 young adult worms were dissected in 0.5mM solution of tetramisole (Sigma-Aldrich) in PBSTw on Rite-On glass slides (Thermo Scientific) and incubated first in a PBS-EDTA-ATA buffer (1xPBS, 0.1 mM EDTA, 1mM aurintriicarboxylic acid), followed by total RNA extraction using TRI Reagent (Sigma) and BCP (Molecular Research Center), and isopropanol RNA precipitation, as described in the TRI Reagent manual (Applied Biosystems, Manual 9738M Revision D).
RT-qPCR

For qPCR analysis of *oma-1* samples, total RNA from the whole adult worms was used, as *oma-1* only expresses in the germline. cDNA from both *oma-1* and *gfp::his-61* total RNA sample sets was synthesized with random hexamer primers using a SuperScript IV First-Strand Synthesis System (Invitrogen). qPCR reactions were set up using Fast SYBR Green Master Mix (Applied Biosystems) as per manufacturer’s protocol, and the qPCR assay was run on the QuantStudio Real-Time PCR system. For normalizing *oma-1* transcript level, actin served as an endogenous control. The *gfp::his-61* samples were normalized to a housekeeping *ama-1* gene (RNA polymerase II subunit). Fold change in reporter expression was compared between samples using two-sided Welch’s t-test. P-values are reported in figure legends.

Auxin-inducible depletion of EGO-1

Auxin treatment was performed as described (Zhang et al., 2015). Indole-3-acetic acid (IAA; Fisher Scientific) was dissolved in ethanol with a final stock concentration of 400 mM. NGM plates containing 500 µM IAA (prepared by adding IAA to NGM agar at ~50°C) were seeded with fresh *E.coli* OP50 and incubated at RT in complete darkness for 2 days, then stored at 4°C in a light resistant container. Prior to the experiment, plates were warmed up at RT for 1 hr. Then L1 larvae of *degron::ego-1* strains were plated on NGM plates with or without 500 µM IAA, kept at RT and scored for embryonic lethality and germline GFP expression at the adult stage.
Quantification and Statistical Analysis

Exact sample sizes for every assay are described in figures. RT-qPCR data analysis was performed in R with “stat” package, using two-sided Welch’s t-test. Further statistical analysis details, including sample size and p-values, can be found in the Figure 2.4 legend.

Key Resources Table

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Source</th>
<th>Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Deposited Data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small RNA sequencing data</td>
<td>This paper - (Makeyeva et al., 2021)</td>
<td>GSE178985</td>
</tr>
<tr>
<td>Sanger sequencing trace files</td>
<td>This paper - (Makeyeva et al., 2021)</td>
<td><a href="https://doi.org/10.17632/c8h2f2prch.2">https://doi.org/10.17632/c8h2f2prch.2</a></td>
</tr>
<tr>
<td><strong>Chemicals and Recombinant Proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tetramisole hydrochloride</td>
<td>Millipore Sigma</td>
<td>L9756</td>
</tr>
<tr>
<td>10X PBS buffer PH 7.4</td>
<td>Thermo Fisher Scientific</td>
<td>AM9625</td>
</tr>
<tr>
<td>Rite-On glass slides</td>
<td>Thermo Fisher Scientific</td>
<td>1256820</td>
</tr>
<tr>
<td>Microscope slide cover glass</td>
<td>MedSupply Partners</td>
<td>G07-140410</td>
</tr>
<tr>
<td>TRI Reagent</td>
<td>Millipore Sigma</td>
<td>T9424</td>
</tr>
<tr>
<td>Alt-R® S.p. Cas9 Nuclease V3</td>
<td>Integrated DNA Technologies</td>
<td>1081058</td>
</tr>
<tr>
<td>Lysing Matrix D</td>
<td>MP Biomedicals</td>
<td>116913050-CF</td>
</tr>
<tr>
<td>Molecular Research Center BCP</td>
<td>Fisher Scientific</td>
<td>NC9551474</td>
</tr>
<tr>
<td>Phase Lock Gel tubes (heavy)</td>
<td>Quanta Bio</td>
<td>2302830</td>
</tr>
<tr>
<td>Acid-Phenol:Chloroform</td>
<td>Thermo Fisher Scientific</td>
<td>AM9720</td>
</tr>
<tr>
<td>Recombinant PIR-1 pyrophosphatase</td>
<td>Gift from Dr. Weifeng Gu, University of California, Riverside</td>
<td>N/A</td>
</tr>
<tr>
<td>T4 RNA Ligase 2</td>
<td>New England Biolabs</td>
<td>M0373S</td>
</tr>
<tr>
<td>T4 RNA Ligase 1</td>
<td>New England Biolabs</td>
<td>M0204S</td>
</tr>
<tr>
<td>SuperScript III Reverse Transcriptase</td>
<td>Thermo Fisher Scientific</td>
<td>18080085</td>
</tr>
<tr>
<td>Fast SYBR Green Master Mix</td>
<td>Thermo Fisher Scientific</td>
<td>4385612</td>
</tr>
<tr>
<td>Indole-3-acetic acid</td>
<td>Fisher Scientific</td>
<td>AAA1055614</td>
</tr>
<tr>
<td><strong>Critical Commercial Assays</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mirVANA™ miRNA isolation kit</td>
<td>Thermo Fisher Scientific</td>
<td>AM1560</td>
</tr>
<tr>
<td>Oligonucleotides</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Primer: oma-1 Forward: CCGTTTCACCACCTCGATC AC</td>
<td>Primer: oma-1 Reverse: AAACCTCTGAATCGCGCGA AC</td>
<td>Primer: ama-1 Forward: AACGTGTGGATTCTCTCG CG</td>
</tr>
<tr>
<td>Primer: ama-1 Reverse: ACTTGGCACCAGGATACT GTG</td>
<td>Primer: gfp::his-61 Forward: ACCTGTCCACACAAATCTG CC</td>
<td>Primer: gfp::his-61 Reverse: ACGATGAAGACCGACGTG</td>
</tr>
<tr>
<td>Primer: actin Forward: TGAAGTGCAGCAATTAC GTT</td>
<td>Primer: actin Reverse: CTTGGAGATCCACATCTG TT</td>
<td></td>
</tr>
<tr>
<td>CRISPR donors, see Table S3 in the original publication</td>
<td>This paper - (Makeyeva et al., 2021)</td>
<td>Integrated DNA Technologies</td>
</tr>
<tr>
<td>Recombinant DNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cdk-1p::cdk-1*::gfp*::cdk-1 3'UTR (entirely intronless); Cbr-unc-119(+) for MosSCI insertion into ttTi5605 II</td>
<td>This study - (Makeyeva et al., 2021)</td>
<td>pCCM956</td>
</tr>
<tr>
<td>oma-1p::oma-1*::gfp*::oma-1 3'UTR (entirely intronless); Cbr-unc-119(+) for MosSCI insertion into ttTi5605 II</td>
<td>This study - (Makeyeva et al., 2021)</td>
<td>pCCM957</td>
</tr>
<tr>
<td>Software and algorithms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fiji/ImageJ</td>
<td>(Schindelin et al., 2012)</td>
<td><a href="https://imagej.net/Welcome">https://imagej.net/Welcome</a></td>
</tr>
<tr>
<td>Cutadapt</td>
<td>(Martin, 2011)</td>
<td><a href="https://cutadapt.readthedocs.io">https://cutadapt.readthedocs.io</a></td>
</tr>
<tr>
<td>RStudio</td>
<td>RStudio Team</td>
<td><a href="http://www.rstudio.com/">http://www.rstudio.com/</a></td>
</tr>
<tr>
<td>Ggplot2</td>
<td>(Wickham, 2016)</td>
<td><a href="https://ggplot2.tidyverse.org/">https://ggplot2.tidyverse.org/</a></td>
</tr>
<tr>
<td>BioRender</td>
<td>BioRender Team</td>
<td><a href="https://biorender.com/">https://biorender.com/</a></td>
</tr>
</tbody>
</table>

Experimental Models: Organisms/Strains
| **nrde-4(ne4679[p.Met12_Glu160del, fSTer21]) IV** | Personal correspondence with Dr. Takao Ishidate, University of Massachusetts Medical School, Worcester | WM772 |
CHAPTER III

Intronless silencing – small RNA-independent branch
ABSTRACT

To preserve genome integrity across generations, germ cells employ various mechanisms to detect and halt activity of viruses and parasitic DNA elements. We have shown that in Caenorhabditis elegans, genetic reporters lacking introns are potently silenced in the germline—a phenomenon we termed “intronless silencing.” The small RNA pathway robustly targets intronless reporters, which enables trans-silencing of homologous intron-containing transgenes. Disruption of small RNA-mediated silencing eliminates trans-silencing but fails to restore intronless reporter expression in the germline, indicating that a small RNA-independent pathway (i.e. cis-silencing) also suppresses intronless reporters. RT-qPCR studies show that cis-silencing results in a 50-fold reduction of reporter RNA levels compared to the mRNA levels of a cognate intron-containing transgene. Regulatory elements of a naturally occurring intronless histone gene mitigated cis-silencing of gfp*:his-61 reporter in the germline. Here, using fluorescence in situ hybridization, we show that the cis-branch of intronless silencing can suppress expression of intronless reporters both in the nucleus, likely transcriptionally, and in the cytoplasm, post-transcriptionally. We also show preliminary evidence that cis-silenced intronless reporter transcripts are not polyadenylated. Finally, partially intronless transgene reporters suggest that a long terminal exon triggers small RNA-independent silencing, and that the two modes of intronless silencing can be triggered by different cues.
INTRODUCTION

Safeguarding the integrity of genetic material is paramount to species survival. Organisms employ various mechanisms to surveil intracellular nucleic acid to fend off assaults by viruses and transposable elements. Conserved among metazoans, the PIWI/piRNA pathway presents one of the tools to detect and suppress foreign activity (Ozata et al., 2019). Spearheading the pathway is the PIWI Argonaute loaded with genomically encoded small RNA guides called piRNAs that enable the Argonaute to scan the host transcriptome for target RNA via Watson-Crick base-pairing.

In C. elegans, target recognition engages a silencing cascade, whereby RNA-dependent RNA polymerases (RdRPs) use target mRNA as a template for synthesis of antisense secondary small RNAs (Das et al., 2008; Shirayama et al., 2012; Sijen et al., 2001), termed 22Gs based on their length and 5' guanosine (Gu et al., 2009; Maniar and Fire, 2011). Equipped with target-specific 22G-RNAs, nuclear and cytoplasmic effector worm-specific Argonautes (WAGOs) orchestrate target suppression transcriptionally and post-transcriptionally (Ashe et al., 2012; Bagijn et al., 2012; Das et al., 2008; Gu et al., 2009; Luteijn et al., 2012; Shirayama et al., 2012). RNA interference (RNAi), where double-stranded RNA (dsRNA) – a virus-resembling structure – is processed into silencing agents (Fire et al., 1998), also triggers WAGO 22G-RNA system to amplify the silencing signal and carry out robust target suppression (Sijen et al., 2001). Instead of PIWI, Argonaute RDE-1 recruits RdRPs to RNAi targets (Tabara et al., 1999; Yigit et al., 2006).
In a previous study, we reported that *C. elegans* use yet another mechanism to detect foreign RNA – by monitoring intron content, and, by extension, mRNA splicing (Chapter II or Makeyeva et al., 2021). While the intronless reporter expression was observed in the soma, we failed to detect any signal in the germline (Figure 2.1B and C). Moreover, intronless transgenes transitivity silenced cognate intron-containing constructs (Figure 2.1D and E) – the finding that indicated involvement of small RNAs. Indeed, intronless reporters set off WAGO 22G-mediated silencing independently of PIWI, RDE-1 or other known upstream Argonautes (Figure 2.2; Figure 2.3E; Figure 2.6B), suggesting that transcripts can become RdRP templates by default. In addition, disabling small RNA-mediated silencing abrogated transitive gene suppression, but did not restore expression of the intronless reporter itself (Figure 2.2C), which prompted us to name the observation “cis-silencing.” Therefore, intronless silencing comprises two modes: small RNA-dependent and –independent mechanisms. Notably, a reporter based on a naturally occurring intronless gene – *gfp*::*his*-61 (fusion between GFP and histone) – can partially bypass cis-silencing, as evidenced by a limited germline expression of an intronless *gfp*::*his*-61 reporter (Figure 2.5A).

In addition to the positive role of splicing in multiple mRNA processing steps (phenomenon collectively termed intron-mediated enhancement), studies demonstrate that intron and splicing surveillance is conserved. In the fungal pathogen *C. neoformans*, RNA stalled on spliceosome induces small RNA silencing (Dumesic et al., 2013). In the flowering plant *A. thaliana*, transgenes without introns trigger production of antisense small RNAs, but not cis-silencing (Christie et al.,
Studies in human cell lines reveal that the human silencing hub (HUSH) complex represses actively transcribing intronless genes via chromatin modification (Seczynska et al., 2022). Peculiarly, in HUSH complex mutants, dsRNA-sensing system silences intronless transposons instead (Tunbak et al., 2020).

Here we show that in C. elegans, the two branches of intronless silencing function separately, suggesting that they are triggered by different cues. Molecular biology and FISH approaches suggest that cis-silencing suppresses polyadenylation and likely acts at the chromatin level. In addition, we demonstrate that gfp*::his-61 histone-based intronless reporter experiences a different cis-suppression mechanism that acts post-transcriptionally, in the cytoplasm.

RESULTS

Intronless reporter mRNA is absent from the cytoplasm and likely suppressed transcriptionally

To investigate spatial dynamics of intronless silencing, we performed RNA-fluorescence in situ hybridization (FISH) on a reporter strain bearing an intronless oma-1* gene (asterisk indicates the lack of introns). The wild-type allele of the C. elegans germline gene oma-1 contains five introns, and OMA-1 protein is known to function in oocyte maturation (Detwiler et al., 2001; Shimada et al., 2002). Previously, to create oma-1* allele without introns, we used CRISPR gene editing to remove the majority of the endogenous oma-1 gene and then repaired deletion either with the intronless oma-1 oligo (oma-1*::3xFLAG) or restored the intron-
containing version (restored *oma-1::3xFLAG*; for strain generation schematic see Figure 2.3A) (Makeyeva et al., 2021). RNA FISH assay on the resulting strains showed that restored *oma-1* mRNA is visible in the entire adult female germline, with a prominent signal increase starting in the pachytene region (Figure 3.1A, top panel). However, in worms carrying an intronless *oma-1* allele, *oma-1* RNA is reduced (Figure 3.1A, middle row), in agreement with previous qPCR studies (Makeyeva et al., 2021). The gonads are devoid of cytoplasmic *oma-1* transcript (Figure 3.1B), but we detected a nuclear signal that formed one to three adjacent foci (Figure 3.1B, insets g-i). Upon closer inspection, similar high-intensity nuclear RNA foci were also detected in animals bearing the intron-containing *oma-1* (Figure 3.1B, insets a-c). We reasoned that these foci coincide with the *oma-1* transcription site, but DNA and RNA FISH co-localization studies were not successful. However, unpublished *oma-1* RNA FISH data from our lab show that in animals with two copies of *oma-1*, the number of foci per nucleus doubles, suggesting that these prominent nuclear foci represent nascent transcripts at the gene locus (data not shown). Moreover, in a recent study, the Seydoux group co-stained RNA from two adjacent germline-expressing genes and found that bright nuclear foci from both genes appose each other, evincing that the observed nuclear RNA accumulations occur at transcription sites (Ouyang et al., 2022). Collectively, these data suggest that the observed *oma-1* nuclear foci coincide with *oma-1* transcription start site.

Limited transcript foci restricted to the nucleus observed in *oma-1* allele gonads mirror mRNA pattern during RNA interference (RNAi) (Ouyang et al., 2022). In *C. elegans*, RNAi silences target genes via WAGO 22G-RNAs (Gu et al., 2009; Yigit et
al., 2006): while the cytoplasmic arm of the WAGO pathway silences mRNA post-transcriptionally (Gu et al., 2009; Shirayama et al., 2012; Yang et al., 2012), nuclear RNAi factors deposit H3K9me3 chromatin modifications and suppress transcription elongation (Buckley et al., 2012; Burton et al., 2011; Gu et al., 2012a; Guang et al., 2010) – all of which are in agreement with RNA FISH studies (Ouyang et al., 2022).

Like RNAi, intronless reporters also trigger nuclear and cytoplasmic WAGO 22G-RNA-dependent gene suppression (Figure 2.6B). Therefore, asked if the small RNA-mediated branch of intronless silencing underlies the observed transcript distribution pattern. To test that hypothesis, we disarmed small RNA-mediated silencing by introducing a null rde-3 mutation via CRISPR editing. Ribonucleotidyltransferase RDE-3 is required for WAGO 22G-RNA production (Gu et al., 2009). To our surprise, in rde-3; oma-1* animals, we detected one to three oma-1* mRNA foci in each nucleus, but none in the cytoplasm (Figure 3.1B insets j-l), similar to the transcript distribution in oma-1* worms, which implied transcriptional downregulation (Figure 3.1B bottom right). These observations showed that the small RNA-independent mode of intronless silencing (1) results in mRNA distribution pattern similar to RNAi, (2) leads to a lack of cytoplasmic mRNA, and (3) likely involves a germline-specific splicing-dependent decrease in transcription.
Figure 3.1. Intronless silencing affects *oma-1* mRNA distribution.

(A, B) Single channel and merged confocal photomicrographs of *oma-1* RNA FISH (green) and DNA (blue, stained with DAPI) signal within the adult hermaphrodite germline in worms carrying a restored intron-containing *oma-1::3xFLAG* gene, intronless *oma-1*::*3xFLAG* allele in wild-type (wt) and *rde-3(ne4871)* mutant, or full *oma-1* deletion (as indicated). Gonads are oriented with distal-proximal axis depicted left-to-right. Scale bar: 5 μm.

(A) Each image shows a maximum intensity projection of ~58 z-planes within a single gonad arm including mitotic, transition and pachytene zones (from left to right). The left column shows single channel photomicrographs of *oma-1* RNA signal.
(green). Gonad outline is traced with gray dashes. In the right column, *oma-1* RNA signal (green) is merged with DNA (blue).

(B) Merged confocal photomicrographs of *oma-1* RNA (green) and DNA (blue) signal in pachytene-stage germ cells. Larger images show a maximum intensity projection of ~55 z-planes, while smaller insets depict enlarged single z-planes through individual nuclei (outlined by dashed squares). (a-c, g-l) The brightest *oma-1* RNA signal appearing either as single, double or triple nuclear foci. (d-f) Cytoplasmic, perinuclear *oma-1* RNA.
Cis-mode of intronless silencing can repress cytoplasmic mRNA post-transcriptionally

Histone-based intronless reporter \( gfp^{*}:his-61 \), created by inserting intronless \( gfp^{*} \) at the N-terminus of the endogenous \( his-61 \) locus via CRISPR, is partially resistant to the cis-mode of intronless silencing (Makeyeva et al., 2021). The \( gfp^{*}:his-61 \) triggers WAGO 22G-RNA-dependent silencing, but in \( rde-3 \) mutant worms with disabled small RNA silencing, the \( GFP^{*}:HIS-61 \) is expressed in the mitotic region of the gonad, gradually decreases through pachytene, and is completely absent in oocytes (Makeyeva et al., 2021). This partial immunity to cis-silencing may be imparted by the histone’s naturally occurring regulatory elements. Like other intronless genes, replication-dependent histones contain RNA motifs that promote efficient gene expression despite the lack of splicing (Huang and Carmichael, 1997; Mei et al., 2017; Wang et al., 1996).

To explore how this partial sensitivity to intronless silencing affects mRNA distribution, we performed RNA FISH using \( gfp \) probe. Similar to \( oma-1^{*} \), \( gfp^{*}:his-61 \) RNA was absent from the cytoplasm (Figure 3.2A insets in the middle row), likely reflecting silencing exerted by the cytoplasmic WAGO 22G-RNA pathway. Conversely, in positive-control animals, \( gfp:his-61 \) mRNA was detected in the cytoplasm (Figure 3.2A top row insets). When analyzing nuclear RNA signal, we saw more \( gfp^{*}:his-61 \) RNA punctae than \( oma-1^{*} \) punctae. In addition to the presumptive transcription foci (Figure 3.2B, second column insets i, ii), we observed a weaker signal dispersed in select nuclei throughout the gonad (Figure 3.2A insets in the middle row) – the incidence of these \( gfp^{*}:his-61 \) mRNA-filled nuclei increased in late
pachytene (Figure 3.2B second column). The accumulation of nuclear foci was not specific to the intronless gfp*::his-61, as we observed a similar phenotype in animals with intron-containing gfp::his-61 (Figure 3.2A, top row; Figure 3.2B, first column). Occasional build-up of dispersed nuclear but not cytoplasmic RNA suggests that gfp*::his-61 may be partially resistant to the nuclear arm of WAGO 22G-RNA silencing, unlike oma-1* (Figure 3.1) and other intronless reporters that we have examined (data not shown).

Next we examined how the cis-mode of intronless silencing affects gfp*::his-61 RNA distribution. To our surprise, in rde-3 mutants, cytoplasmic gfp*::his-61 RNA was detected throughout the gonad (Figure 3.2A, bottom row insets; Figure 3.2B third column), even though GFP protein expression is not observed in oocytes and rarely in mid-to-late pachytene region (Makeyeva et al., 2021). This suggests that the cis-branch of intronless silencing suppresses gfp*::his-61 post-transcriptionally, potentially by preventing translation.
Figure 3.2 Spatial distribution of intronless *gfp*:his-61 reporter in the germline.

(A, B) Single channel and merged confocal images of *gfp* RNA FISH (green) and DNA (blue, stained with DAPI) signal within the adult hermaphrodite germline to compare *gfp* RNA distribution between animals with intron-containing *gfp*:his-61 or
intronless \textit{gfp}^{*::his-61}. \textit{Rde-3(4848)} mutants carrying intronless \textit{gfp}^{*::his-61} allele allow to examine cis-branch of intronless silencing. N2 gonad serves as a negative control. Scale bar: 5 \textmu m.

(A) On the left, each gonad image shows a maximum intensity projection of \~33 \textit{z}-planes within a single gonad arm including mitotic, transition and pachytene zones (from left to right). The top gonad in each enclosed horizontal row is a single channel view of \textit{gfp} RNA (green), and the bottom – merge of \textit{gfp} RNA (green) and DNA (blue) signal. Gonads are oriented with distal-proximal axis depicted left-to-right. On the right side of the panel, insets (i) and (ii) show enlarged individual \textit{z}-plane-views of several germline nuclei and shared cytoplasm from distal and proximal gonadal regions, respectively (see corresponding dashed boxes on the left).

(B) Merged photomicrographs of \textit{gfp} RNA (green) and DNA (blue) in late stage pachytene germ cells. Larger images show a maximum intensity projection of \~55 \textit{z}-planes, while smaller insets present zoomed-in single \textit{z}-plane views through individual nuclei. Insets (i) and (ii) focus on the brightest nuclear \textit{gfp} RNA puncta that always appear to overlap with DNA (due to restrictive brightness and contrast settings applied to avoid signal oversaturation, weaker nuclear \textit{gfp} RNA signal is not visible). (iii) Distribution of weaker nuclear \textit{gfp} RNA signal.
mRNA polyadenylation of intronless *oma-1* reporter was not detected

FISH studies revealed that intronless *oma-1* is transcribed. To investigate why *oma-1* RNA fails to be exported and translated, we examined intronless reporter RNA via northern blot analysis using an *oma-1* cDNA probe. While the intron-containing *oma-1* gene produced two transcript isoforms (Figure 3.3A and B), intronless *oma-1* produced only a short isoform (Figure 3.3B). Consistent with qRT-PCR findings, disarming the trans-silencing branch of intronless silencing in *oma-1*; *rde-3* animals slightly increased the abundance of the shorter isoform, but did not restore the longer transcript (Figure 3.3C; (Makeyeva et al., 2021)). Thus the absence of the longer isoform was not caused by small RNA-mediated silencing. The more abundant longer isoform produced by intron-containing *oma-1* could represent an unspliced *oma-1* transcript or a poly-adenylated (poly(A)) mRNA. To test the latter, we performed a polyA tail length assay (Figure 3.3D). Briefly, RNA (in total RNA extracts from young adults) was enzymatically extended with a 3’-polyguanosine/inosine (G/I) tail (reaction specific to RNA species with 3’-terminal adenosines) and reverse transcribed using a polyC primer with an adapter for PCR. The 3’-ends were then amplified by PCR reaction using an *oma-1*-specific primer and a 3’-adapter primer. A prominent 3’-product corresponding to the expected length of the *oma-1* 3’ UTR was detected in the intron-containing *oma-1* sample (Figure 3.3E). Sanger sequencing confirmed that six of eight clones had poly(A) tails ranging from 25 to 90 adenosines, while the remaining two were not polyadenylated but terminated with two adenosines shortly before or after the stop codon prior to reaching polyadenylation signal. By contrast, polyadenylated *oma-1* transcripts
were not detected in extracts from *oma-1* or *rde-3; oma-1* strains (Figure 3.3E), even though we were able to confirm that reverse transcription of *oma-1* mRNA was successful by performing PCR with *oma-1*-specific primers (Figure 3.3F). These findings suggest that the absence of introns prevents polyadenylation of germline transcripts. Additional studies are needed to further characterize *oma-1* mRNA and transcripts of other intronless reporters.
Figure 3.3 Intronless reporter transcript is not polyadenylated.

(A-C) Northern blots showing *oma-1* mRNA in strains carrying either intron-containing or intronless *oma-1*::3xFLAG gene in wild-type (wt) and *rde-3* mutants. Blots with total RNA samples from adult hermaphrodite animals were probed with biotin-labeled-*oma-1* probe, incubated with alkaline phosphatase-conjugated streptavidin and detected via chemiluminescence.

(A) A 2x total RNA dilution series from worms with restored intron-containing *oma-1*::3xFLAG revealing two transcript variants (magenta and green arrows). Decrease in total RNA concentration affects RNA migration (same size bands in adjacent lanes are marked by magenta and green stars). The amount of total RNA per lane: 2 μg, 1 μg, 0.4 μg.

(B) Northern blot showing that *oma-1* mRNA level is 10.3 times higher in animals with a restored intron-containing *oma-1*::3xFLAG allele compared to worms carrying intronless *oma-1*::3xFLAG allele. 10 μg of total RNA/lane. Two transcript isoforms are marked with magenta and green arrows.

(C) Northern blot showing that *rde-3* mutation results in a 1.6x *oma-1* mRNA increase in animals bearing intronless *oma-1*::3xFLAG allele. 10μg total RNA per lane. Two transcript isoforms are marked with magenta and green arrows.

(D) Schematic of the commercially available assay to determine the length of the poly-adenylated (poly(A)) mRNA tail. In the guanosine/inosine (G/I) tailing step, total RNA samples are treated with yeast polyA polymerase that adds a number of G/I residues to polyadenylated mRNA species. cDNA synthesis of selectively G/I-tailed
mRNAs is accomplished with a primer ending with CCCCCCTT at its 3’-end. Finally, to examine the length of *oma-1* mRNA poly(A) tail, the fragment of interest was amplified via PCR using *oma-1*-specific forward primer and the reverse universal primer provided in the kit. Control PCR was performed with the same forward primer and *oma-1*-specific reverse primer upstream of the putative polyadenylation site. Resulting PCR products were resolved on an agarose gel.

(E) Gel showing polyadenylated *oma-1* PCR products in hermaphroditic adults of indicated strains. Expected PCR size is greater than 480 base pairs (bp), depending on the number of terminal adenosine residues. Worms with intronless *oma-1*:*::3xFLAG or full deletion *oma-1* alleles did not exhibit a polyadenylated *oma-1* mRNA product.

(F) Gel showing control PCR fragments using *oma-1*-specific forward and reverse primers in adult hermaphrodites of indicated genotype. Size of the cropped bands is 300 bp. Expected PCR product size is 301 bp. Animals without *oma-1* gene serve as negative control.
Introns in partially intronless reporters prevent trans-silencing but not cis-silencing

Previously, a single copy of *oma-1::gfp* transgene was integrated into worm genome using MosSCI technique (Frokjaer-Jensen et al., 2008; Seth et al., 2013). This transgenic construct with eight introns (Figure 3.4A) yielded a bright GFP signal in worm oocytes (Figure 3.4B). By contrast, a MosSCI-integrated single-copy intronless *oma-1*::*gfp* transgene (Figure 3.4A) failed to express in the germline (Figure 3.4B, (Makeyeva et al., 2021)). We have also generated a strain with an *oma-1::gfp* transgene (Figure 3.4A) – construct containing *oma-1* but not *gfp* introns – and found that the transgene is silent in oocytes (four independent lines), similar to *oma-1*::*gfp* (Figure 3.4B). Next, we set up genetic crosses to test if partially intronless *oma-1::gfp* triggers epigenetic trans-silencing of an expressing, intron-bearing *cdk-1::gfp* transgene (Figure 3.4C). As controls, we crossed *cdk-1::gfp* animals to the worms carrying either the *oma-1::gfp* transgene or its fully intronless version - *oma-1*::*gfp*. As expected, 100% of F1 cross-progeny from the first control cross (n=10) expressed nuclear and cytoplasmic GFP indicating that both *cdk-1::gfp* and *oma-1::gfp* transgenes are expressed (Figure 3.4C). The second control cross showed that intronless *oma-1*::*gfp* trans-silenced *cdk-1::gfp* in the germline, as the resulting offspring (n=20) were negative for the nuclear CDK-1::GFP signal in oocytes but positive in somatic cells (Figure 3.4C). Remarkably, cross-progeny from partially intronless *oma-1::gfp* animals (n=10) expressed nuclear CDK-1::GFP (Figure 3.4C), even though the *oma-1::gfp* transgene remained silent, indicating that, unlike other intronless reporters we had observed (Makeyeva et al., 2021), *oma-1::gfp* does not
transitively silence \textit{cdk-1::gfp} (Makeyeva et al., 2021). Indeed, small RNA sequencing revealed that \textit{oma-1::gfp} did not trigger abundant antisense 22G reads (Figure 3.4D), suggesting that \textit{oma-1::gfp} does not engage the small RNA silencing pathway. In contrast, intronless \textit{oma-1*::gfp} induces robust production of \textit{gfp}-specific antisense small RNAs that mediate trans-silencing of \textit{cdk-1::gfp} described earlier (Figure 3.4C-D). Taken together, these findings suggest that \textit{oma-1} introns in the partially intronless \textit{oma-1::gfp} prevent small RNA-dependent silencing, but are not sufficient to bypass cis-silencing. Thus, the data imply that cis-silencing and trans-silencing are triggered by distinct cues.
Figure 3.4

Figure 3.4 Partially intronless transgenes: removal of 5' and 3'-introns leads to contrasting effects on cis-silencing.

(A) Schematic structures of intron-containing *oma-1::gfp*, intronless *oma-1*::*gfp* *, and partially intronless *oma-1::gfp* transgenes. Purple boxes, *oma-1* coding sequence; green boxes, *gfp* coding sequence; V-shaped lines, introns.

(B) Representative fluorescence (upper) and DIC micrographs (lower) of adult oocytes in *oma-1::gfp*, *oma-1*::*gfp* *, and *oma-1::gfp* *. OMA-1 protein is expected to localize in oocyte cytoplasm as shown in *oma-1::gfp* animals. Scale bar, 20µm. White dashed circles indicate the location of nuclei in GFP– animals.

(C) Genetic crosses testing the ability of worms with fully and partially intronless *oma-1::gfp* constructs to transitively silence *cdk-1::gfp*. Percent of F1 progeny expressing nuclear CDK-1::GFP in oocytes (ON) and the number of analyzed animals (n) is indicated. Unlike *oma-1*::*gfp* *, partially intronless *oma-1::gfp* does not transitively silence *cdk-1::gfp*. White dashed circles indicate the location of nuclei in GFP– animals.

(D) Plots showing the distribution and number of antisense small RNA reads mapping along *oma-1::gfp* cDNA in adult hermaphrodites bearing a transgenic copy of *oma-1::gfp*, *oma-1*::*gfp* *, or *oma-1::gfp* *. The height of each bar represents the number of reads that begin at that gene position per million total reads. In gene cartoons along the x-axis: grey boxes – *oma-1* coding sequence; green boxes – *gfp* coding sequence; positions of exon junctions are indicated by black solid vertical lines (dashed vertical lines mark the corresponding positions in intronless gene regions).

(E) Schematic structures of partially intronless *oma-1*::*gfp* transgene. Purple box – *oma-1* coding sequence; green boxes – *gfp* coding sequence; V-shaped lines – introns.
(F) Representative epifluorescence (top) and DIC (bottom) images of oocytes in animals carrying a partially intronless *oma-1*:::*gfp* (left) or *oma-1*:::*gfp* (right) transgene. Scale bar, 20μm. White dashed circles indicate the location of nuclei in GFP– animals.
“Long” terminal exon can trigger cis-silencing

To explore why *oma-1::gfp* failed to express, we analyzed a strain carrying *oma-1*::*gfp* transgene, which lacks *oma-1* introns but has *gfp* introns (Figure 3.4E). The transgene is expressed in oocytes (Figure 3.4F) and does not accrue antisense WAGO 22G-RNAs (Seth et al., 2018). The presence of *oma-1* or *gfp* introns is sufficient to prevent small RNA-mediated intronless silencing (Figure 3.4D) (Seth et al., 2018), but partially intronless transgenes vary in sensitivity to cis-silencing. Difference in expression states between the two transgenes (Figure 3.4F) cannot be explained by the number of introns or the size of the longest exon: the first exon of active *oma-1*::*gfp* is longer than the terminal exon in silenced *oma-1*::*gfp*.

Therefore, it is likely the 3' position of the unnaturally long exon that makes a transgene prone to cis-silencing. Combined with the finding that intronless reporter mRNA lack polyadenylation (Figure 3.3), these observations suggest that long terminal exons or increased distance between the last intron and polyadenylation site may trigger cis-silencing in the germline.
Figure 3.5. Working model for intronless silencing.

(Top) Splicing (coral and cyan circles mark exon-exon junctions) and the associated factors (orange ovals) enable successful mRNA processing (3’-end formation, nuclear export and expression).

(Bottom) Intronless reporter transcript is devoid of the protective splicing-associated factors (orange ovals) and instead may be marked by silencing cues (red octagons). Either the absence of protective signals or the association with the silencing cues forces default RNA-dependent RNA polymerase recruitment (RdRP) bypassing the initiation step of small RNA silencing pathways – mRNA detection by a primary Argonaute (left arrow labeled WT for wild-type). RdRP-generated antisense 22G-RNAs are loaded on WAGOs (gray rectangle labeled Argonaute) and suppress intronless reporter mRNA and, in some instances, cognate intron-containing gene (dashed arrow labeled “transitive silencing”). In rde-3 mutants with abolished small RNA-mediated silencing, transcript suppression continues in-cis based on FISH, RT-qPCR, and Northern blotting studies. Lack of nuclear mRNA apart from the transcription site suggests intronless reporters may be suppressed transcriptionally. In addition, reporter RNA is not polyadenylated, which may contribute to nuclear export failure (follow arrow labeled “rde-3” pointing to the right) and triggering default RdRP-recruitment. When cis-silencing is partly bypassed using histone gene that
contains regulatory sequences allowing splicing-independent expression, cytoplasmic transcript still undergoes translation failure in oocytes (dashed inhibitory arrow on the right).
DISCUSSION

Cis-silencing acts on transcriptional and post-transcriptional levels

This study revealed that cis-mode of intronless silencing can suppress targets transcriptionally and post-transcriptionally (summarized in model Figure 3.5). FISH in rde-3; oma-1* strain, with defective small RNA silencing, shows that reporter mRNA is present only near the gene’s transcription site (Figure 3.1B), which is reminiscent of the transcriptional silencing pattern during RNAi (Ouyang et al., 2022) and piRNA silencing (data not shown). In both cases, transcriptional silencing is mediated by the nuclear RNAi pathway, which represses the elongation phase of transcription (Guang et al., 2010). Presumably, a minimal level of transcription provides the mRNA template to maintain RdRP-mediated silencing. Why this minimal level of oma-1* transcription persists in rde-3-mutant animals is unknown. Future studies assessing chromatin modification (such as histone H3 lysine 9 trimethylation deposition) and transcription status (RNA pol II ChIP qRT-PCR) of the intronless oma-1* gene locus in rde-3 mutants will test directly if the small RNA-independent branch of intronless silencing suppresses transcription.

Peculiarly, the HUSH complex also requires active transcription to initiate and propagate intronless gene suppression via histone H3K9me3 modification (Seczynska et al., 2022). TASOR, a HUSH complex subunit, is required for complex assembly and concomitant transcriptional repression (Douse et al., 2020). C. elegans TASOR ortholog is expressed in the worm germline and could be implicated in the cis-mode of silencing, which can be tested by assessing intronless reporter mRNA level and distribution in rde-3; tasor double mutant.
FISH analysis of rde-3; gfp*::his-61 revealed the cytoplasmic reporter mRNA in parts of the gonad where GFP is not expressed, indicating localized post-transcriptional suppression. Transcript length could be preventing reporter expression in oocytes, as a shorter gfp*::nls::his-61 3’ utr construct without his-61 coding region expresses in the entire gonad, including oocytes (Figure 2.5C). Alternatively, the histone coding sequence might contribute to translation suppression, though it was not observed with intron-containing gfp::his-61 (Figure 2.5A). To investigate if transcript length or histone coding region prevent gfp*::his-61 expression in oocytes, his-61 coding region can be substituted with a non-histone sequence of a comparable length.

These findings suggest that mitigating transcriptional cis-silencing of non-histone reporters, like oma-1*, may not be sufficient for reporter expression, as they might encounter barriers to nuclear export or ribosomal recruitment, which appear to be more active in the germline than in somatic tissues, based on the successful expression of intronless reporters in the soma (Figures 2.1B and 2.5A).

Lack of polyadenylation as a feature of cis-silencing

Several studies show that splicing of the terminal intron promotes polyadenylation (Chiou et al., 1991; Cooke et al., 1999; Nesic et al., 1993; Niwa et al., 1990). Expression of intronless cdk-1*::gfp* in soma (Figure 2.1B) implies that somatic transcripts are polyadenylated. The absence of polyadenylation in the germline could be a cause or effect of the cis-mode of intronless silencing. This may explain (1) why we did not detect polyadenylation of the intronless oma-1* RNA (Figure 3.3)
and (2) why partially intronless \textit{oma-1}^*::\textit{gfp}, with a 244-nt terminal exon was expressed in the germline, while \textit{oma-1}::\textit{gfp} with a 1152-nt terminal exon was not (Figure 3.4). It would also explain (3) why \textit{gfp}^*::\textit{his}-61 reporter (based on a replication-dependent histone) is partially resistant to cis-silencing (Figure 2.5). Processing of core histones does not require polyadenylation, as they contain a stem loop structure in 3' UTR that recruits alternative 3'-end processing factors (reviewed in Hentschel and Birnstiel, 1981; Marzluff, 1992). Notably, inserting introns into histone genes triggers a canonical splicing-dependent mRNA processing (Pandey et al., 1990).

Further studies will determine whether the polyadenylation status of intronless reporters varies by 3' UTR sequence, and whether polyadenylation failure is a cause or effect of intronless silencing. It has been postulated that aberrant RNA species become RdRP-templates in primer-independent manner (Baulcombe, 2004; Dougherty and Parks, 1995; Wassenegger and Pelissier, 1998). There is a reference to an uncapped mRNA inducing RdRP-dependent gene silencing in \textit{A. thaliana} (Baulcombe, 2004). In \textit{C. neoformans}, mRNA with suboptimal splicing motifs becomes a small RNA silencing target (Dumesic et al., 2013). In their work, Luo ad Chen showed that truncated, non-polyadenylated transgenic \(\beta\)-glucuronidase (GUS) mRNA triggered production of RDR6 RdRP-dependent antisense small RNAs (Luo and Chen, 2007).

Alternatively, it is entirely possible that decreased RNA polymerase II (RNA pol II) retention and transcription elongation caused by the lack of splicing (reviewed in Shaul, 2017) results in premature termination and therefore production of
truncated intronless reporter transcripts, which in turn could serve as RdRP-substrates for silencing initiation. However, northern blotting analysis of intronless \textit{oma-1*} did not reveal a pattern consistent with a heterogeneous population of \textit{oma-1} transcripts (Figure 3.3), though due to the assay’s limited sensitivity, mRNA sequencing studies or comparison of abundance between 5’ and 3’ reporter RNA regions via qRT-PCR present better options for addressing this possibility.

**PIWI-independent trans-silencing and cis-silencing triggered by distinct cues**

Disabling small RNA-mediated silencing (\textit{rde-3}-mutants) did not yield expression of intronless reporters in the germline. Thus we conclude that gene suppression in cis occurs independently of the transitive-arm. That does not, however, exclude the possibility of inverse causality – i.e., that cis-silencing mode could trigger WAGO 22G-RNA-mediated suppression. Using a partially intronless \textit{oma-1::gfp*} construct (Figure 3.4), we have shown that, despite failing to express, the transgene does not become a target of WAGO 22G-RNA silencing, indicating that cis-silencing does not always trigger the transitive mode of suppression. It is possible that small RNA-independent silencing of \textit{oma-1::gfp*} and \textit{oma-1*:gfp*} is mediated by distinct pathways, with only the latter engaging transitive silencing. Future RNA FISH experiments comparing transcript distribution between the two strains will help determine if both transgenes undergo transcriptional suppression.

A different study from the lab also showed that PIWI-independent WAGO 22G-RNA-mediated silencing can be separated from cis-silencing. Partially intronless \textit{oma-1*(recoded)::gfp} construct with recoded \textit{oma-1} cDNA sequence that
yields original OMA-1 protein product is silenced in the germline and exhibits transitive suppression of active cdk-1::gfp, but upon disruption of 22G-RNA biogenesis, oma-1*(recoded)::gfp becomes expressed in oocytes (Seth, 2016). Until we find a factor that initiates default silencing upstream of RdRPs, we cannot definitively tell if both intronless oma-1*::gfp* and partly intronless oma-1*(recoded)::gfp are silenced by the same PIWI-independent rde-3-mediated pathway. However, testing other primary Argonautes for initiation of oma-1*(recoded)::gfp may serve as supporting indirect evidence for default targeting, in which case, the construct would be a good candidate for a screen to identify factors upstream of RdRP recruitment, given its resistance to cis-silencing. To our knowledge, intronless default silencing is the only described instance of RdRP-engagement without a primary Argonaute in C. elegans.
METHODS

C. elegans strains and genetics

All strains in this study were derived from Bristol N2 line and cultured on normal growth media (NGM) agar plates seeded with E. coli OP50 (Brenner, 1974). Strains used in the study:

oma-1::gfp; Cbr-unc-119(+) II; unc-119(ed9) III
oma-1*(intronless)::gfp*(intronless); Cbr-unc-119(+) II; unc-119(ed9) III
oma-1*(intronless)::gfp; Cbr-unc-119(+) II; unc-119(ed9) III
oma-1::gfp*(intronless); Cbr-unc-119(+) II; unc-119(ed9) III
restored oma-1::3xFLAG (endogenous locus) LGIV
oma-1*(intronless)::3xFLAG (endogenous locus) LGIV
rde-3 LGI; oma-1*(intronless)::3xFLAG (endogenous locus) LGIV
oma-1(full deletion) (endogenous locus) LGIV

N2

gfp::his-61 LGIV

gfp*(intronless)::his-61 LGIV

rde-3 LGI; gfp*(intronless)::his-61 LGIV

Epifluorescence and DIC Microscopy

Worms were mounted using a 0.5mM solution of tetramisole (Sigma-Aldrich, L9756) in M9 buffer (Brenner, 1974) on Rite-On glass slides (Thermo Scientific, 1256820) and cover glass (MedSupply Partners, G07-140410). Epifluorescence and differential interference contrast (DIC) microscopy analysis and images were
captured using an Axio Imager M2 Microscope (Zeiss), an ORCA-ER digital camera (Hamamatsu) and Zen (Zeiss) software. Images were processed using Fiji/ImageJ software (Schindelin et al., 2012).

**MosSCI genome editing**

Transgenic lines were generated in EG4322 background following Mos1-mediated single copy insertion (MosSCI) protocol as described (Frokjaer-Jensen et al., 2008).

**Small RNA Cloning and Deep Sequencing**

Synchronized L1 populations were plated, collected and flash frozen in four volumes of TRI Reagent (Sigma, T9424) at the young adult stage. On the day of the experiment, after three freeze-thaw cycles, worm pellets were combined with ceramic spheres (Lysing Matrix D, MP Biomedicals, 116913050-CF) and homogenized using FastPrep system (MP Biomedicals) for three 20-sec cycles at a speed of 6 m/s. Total RNA was extracted using first a phase separation reagent 1-Bromo-3-chloropropane (BCP, Molecular Research Center, NC9551474), then aqueous phase was transferred to Phase Lock Gel tubes (Quanta Bio, 2302830) for additional separation with phenol-chloroform. Following isopropanol precipitation and 80% ethanol wash, total RNA was resuspended with water and frozen at -80C. Small RNA isolation and cloning were performed as described (Li et al., 2020; Seth et al., 2018). Briefly, small RNAs, size-selected using mirVana miRNA Isolation Kit (Thermo Fisher Scientific, AM1560), were treated with a recombinant PIR-1 pyrophosphatase (a generous gift from Dr. Weifeng Gu) to remove the 5’ γ and β
phosphates, and ligated to 3’ and 5’ adapters using T4 RNA ligase 2 (New England Biolabs, M0373S) and T4 RNA ligase 1 (New England Biolabs, M0204S), respectively. cDNA was then generated using SuperScript III Reverse Transcriptase (Thermo Fisher Scientific, 18080085) and amplified via PCR. After gel size selection, final libraries were sequenced on an Illumina HiSeq platform at the University of Massachusetts Medical School Deep Sequencing Core Facility.

**Data Analysis**

Following standard removal of adapter sequences using Cutadapt tool (Martin, 2011), reads were mapped using bowtie2 aligner (Langmead and Salzberg, 2012; Langmead et al., 2019), normalized by the number of the reads mapping to the genome (WB272) or transcriptome (WBcel235) and multiplied by 5,000,000. Visualization of antisense 22-nucleotide-long reads starting with a guanine were generated using ggplot2 (Wickham, 2016). All scripts are available upon request.

**FISH**

*C. elegans* adult hermaphrodite gonads were dissected in 0.5 mM tetramisole (Sigma-Aldrich, L9756) in 0.05% Tween-20 (Thermo Fisher Scientific, BP337500) in 1x phosphate buffered saline (PBS) solution on Rite-On glass slides (Thermo Scientific, 1256820) and transferred with a micropipette to a clean area without worm debris - one of the two 10mm-diameter circles was pre-treated with poly-lysine solution (Sigma-Aldrich, P8920) to aid gonad adhesion to the slide surface. Samples were suspended in Happy Buffer (developed by Jim Priess; 25mM HEPES pH 6.9,
49mM NaCl, 5mM KCl, 2mM MgCl$_2$, 1mM EGTA, H$_2$O) until a sufficient number of gonads from the same sample were pooled (no longer than 20 min). Gonads were then fixed in 5% formaldehyde (Sigma Aldrich, F8775) in PBSTw (0.1% Tween-20 in 1x PBS) for 30 min and permeabilized with 0.1% Triton X-100 (Sigma Aldrich, T8787) in PBSTw for 10 min. Slides were suspended in Coplin jars filled with 70% ethanol at -20°C until the remaining samples were pooled (minimum incubation time 30 min). Next, samples were incubated in wash buffer (10% formamide (Sigma-Aldrich, 4610-OP), 2XSSC [0.3M NaCl, 30mM sodium citrate, pH 7.0] (Thermo Fisher Scientific, BP1325-1)) for 20 min. Finally, wash buffer was replaced with a solution containing 0.01μM oligo probes diluted in hybridization buffer (Stellaris SMF-HB-1) for an overnight incubation at 37°C in the dark. To prevent desiccation, slides were covered with glass coverslips (Globe Scientific, 1404-10) and placed into a light-impermeable slide moisture chamber. Next day, samples were rinsed in wash buffer (pre-warmed to 37°C), followed by two 45 min-long wash buffer incubations at 37°C in the dark (Coplin jars wrapped in foil). After the second incubation, wash buffer solution was replaced with a fresh one, and the samples were mounted with DAPI-containing Vectashield mounting medium (VWR, 101098-044), covered with cover glass (Globe Scientific, 1404-15), sealed with clear nail polish (VWR, 100491-940), and dried overnight. The following day, gonads were imaged using confocal microscopy.
**Northern Blot**

Total RNA was isolated as described in the *Small RNA Cloning and Deep Sequencing* subsection. Total RNA samples were denatured in formamide-containing loading buffer (Thermo Fisher Scientific, R0641) for 13 min at 70°C, resolved on a 1% agarose/1.5% formaldehyde gel using traditional MOPS/sodium acetate conductive medium (Thermo Fisher Scientific, BP2900) and transferred to a nylon membrane by capillary transfer with mild hydrolysis. Membranes were probed with homemade T7-transcribed biotin-labeled oligos at 65°C for 18hrs and visualized via chemiluminescence mediated by alkaline phosphatase linked to streptavidin (SeraCare, 5950-0005) following manufacturer’s recommendations (Detector AP Chemiluminescent Blotting Kit, SeraCare, 5910). For digital blot processing and relative quantity analysis Image Lab software (Bio-Rad) was used.

**mRNA Poly-Adenylated Tail Length Assay**

Total RNA was isolated as described in the *Small RNA Cloning and Deep Sequencing* subsection. 1 µg of total RNA from each sample was processed according to manufacturer’s instructions (Thermo Fisher Scientific, 764551KT). Briefly, a limited number of guanosine and inosine (G/I) residues were added to poly-adenylated mRNA species by yeast poly(A) polymerase. For reverse transcription, a primer, provided in the kit, ending with CCCCCCTTT was used (full primer sequence was not disclosed). Next, *oma-1* fragment containing poly(A) tail was amplified via PCR with a forward *oma-1*-specific primer – CTCTTCTGATGAAATCTGTCGC – and a universal reverse primer (included in the
kit, sequence undisclosed). For control PCR, reverse primer was replaced with an
*oma-1*-specific primer – GGACAAAGAACACGTGAC – mapping upstream of the
gene’s polyA site. PCR products were cloned into Zero Blunt TOPO plasmid vectors
(Thermo Fisher Scientific, 450245), which were then used for One Shot competent
*E. coli* cells (Termo Fisher Scientific, C404003) transformation and sent out for
Sanger sequencing to Genewiz.

**FISH probes**

<table>
<thead>
<tr>
<th></th>
<th><em>gfp</em></th>
<th><em>oma-1</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>aaaagtcttctcttttact</td>
<td>tctgagatgtgctcatcg</td>
</tr>
<tr>
<td>2</td>
<td>caagaatgggcaacaatcga</td>
<td>gtatcgagtgtaaacgg</td>
</tr>
<tr>
<td>3</td>
<td>cccattaacatcaccatcta</td>
<td>caaatctcaatgtggcat</td>
</tr>
<tr>
<td>4</td>
<td>cctctccactgacagaaaaat</td>
<td>agcaactgtttttagatgc</td>
</tr>
<tr>
<td>5</td>
<td>gtaagttttccgtatgttgct</td>
<td>cgcaactgtgcatgattga</td>
</tr>
<tr>
<td>6</td>
<td>gtagttttccagttagtgcag</td>
<td>gaaatggggtgcaactgg</td>
</tr>
<tr>
<td>7</td>
<td>acaagtgtgggccatggaac</td>
<td>ttctgcattcgggctaatc</td>
</tr>
<tr>
<td>8</td>
<td>ggtatatcgagaagcattga</td>
<td>agtttagattcgcagcaag</td>
</tr>
<tr>
<td>9</td>
<td>tcatgccgtttcatatgac</td>
<td>gatgtgacgaaagttcct</td>
</tr>
<tr>
<td>10</td>
<td>gggcatggcacttgtgaaaaa</td>
<td>ttgtcacagtttcggtt</td>
</tr>
<tr>
<td>11</td>
<td>ttctttctgtacataacct</td>
<td>aatggaagggcaccggttcc</td>
</tr>
<tr>
<td>12</td>
<td>gtccccgtcatcttgtgaaaaa</td>
<td>caagaagtttcagcgccga</td>
</tr>
<tr>
<td>13</td>
<td>tgacttcagcagctgttgg</td>
<td>gtccatttgatcagaaatg</td>
</tr>
<tr>
<td></td>
<td>taacaaggggatcacttca</td>
<td>tgctgaatggcagaagcgg</td>
</tr>
<tr>
<td>---</td>
<td>---------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>15</td>
<td>ataccttttaactcgatctct</td>
<td>tggcccaggtgcatttag</td>
</tr>
<tr>
<td>16</td>
<td>gtgtcacaagaatgttttctccat</td>
<td>gagacggtggtaggcatc</td>
</tr>
<tr>
<td>17</td>
<td>gtgaggtttagttgtatcc</td>
<td>atatcacaagctacgtagc</td>
</tr>
<tr>
<td>18</td>
<td>gtctgccaattggtatcatc</td>
<td>ttgcagcgaacccaaac</td>
</tr>
<tr>
<td>19</td>
<td>ctttgaatcctcttttgtt</td>
<td>aaactcctccaggagtattt</td>
</tr>
<tr>
<td>20</td>
<td>ccactctcataatgtgttgtct</td>
<td>actgagatcctgagaagcg</td>
</tr>
<tr>
<td>21</td>
<td>atgtctgtctagttgaacgc</td>
<td>ttgcagattatgcacagcag</td>
</tr>
<tr>
<td>22</td>
<td>gcacaattggagtatttttgtt</td>
<td>agttgcagcagatttcata</td>
</tr>
<tr>
<td>23</td>
<td>gtctggttaaaaggacaggggct</td>
<td>tcctgcaatggagagaatg</td>
</tr>
<tr>
<td>24</td>
<td>aaggccagattgtgtggaca</td>
<td>ccacaattgagatgttgg</td>
</tr>
<tr>
<td>25</td>
<td>tctttcgattggatctttc</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>tcaagaaggaccatgtggtc</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>aatccccagcagctgttacaa</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>tataaggatccatgcatcag</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER IV

Discussion and future work
**Cases of PIWI-independent WAG0-mediated silencing in the C. elegans germline**

Our studies reveal that intronless reporters trigger 22G-RNA-mediated silencing by default, in other words, without the involvement of an upstream primary Argonaute. Is default silencing reserved for genes devoid of splicing? Some evidence from the lab suggests that may not be the case. Two transgenic constructs \textit{cas9::gfp} and \textit{oma-1*(recoded)::gfp} are silenced in the germline by small RNA species in an \textit{rde}-3-dependent manner (Makeyeva et al., 2019; Seth et al., 2018). The data show that 22G-RNA silencing can be induced independently of PIWI (Makeyeva et al., 2019; Seth, 2016), but that does not exclude the possibility that PRG-1 could still serve as a contributing factor. In fact, a study shows that removal of predicted piRNA sites in \textit{cas9} leads to the activation of \textit{cas9} transgene in 9/30 strains, lending support to PIWI involvement (Zhang et al., 2018). While studies of \textit{cas9} and \textit{cas9::gfp} are not entirely comparable for several reasons (i.e. (1) presence of \textit{gfp} fusion and (2) expression assessment: GFP and small RNA sequencing vs. CAS9 nuclease activity – whereby the latter does not show transgene expression pattern), the findings warrant caution in excluding \textit{prg-1} as a factor contributing to the aforementioned \textit{rde}-3-dependent 22G-RNA-mediated silencing.

One common feature between \textit{cas9::gfp} and \textit{oma-1*(recoded)::gfp} is that both contain one or several unusually long exons. Median and average exon size in \textit{C. elegans} is 144 and 201 bases, respectively (Spieth et al., 2014), while the first exon in \textit{oma-1*(recoded)::gfp} is 1517 nt and exons in \textit{cas9::gfp} are 1191, 987, 1063, and 1319 nt. Peculiarly, \textit{oma-1*:gfp} carrying the original \textit{oma-1} coding sequence
contains a 1517-nt first exon as well, but does not trigger silencing (Seth et al., 2018). Either the coding region imparts licensing qualities or the endogenous intron-containing oma-1 gene may exert transitive protection. Additional studies are needed to determine whether cas9::gfp and oma-1*(recoded)::gfp are subjected to default small RNA-mediated silencing and why they trigger PIWI-independent silencing.

**What triggers default small RNA silencing?**

Evidence in *C. elegans* suggests that aberrant cytoplasmic RNA, like the one undergoing nonsense-mediated decay (NMD), does not affect small RNA silencing or protection: piRNA reporters with premature termination codons (PTC) behave as transitive silencing agents, and PTC in RNAa (RNA-induced epigenetic gene activation)-exerting transgenes does not disrupt their RNAa activity (Seth et al., 2018). Peculiarly, in a recently reported transcriptional adaptation phenomenon found in zebrafish, mice, and worms, mutant mRNAs undergoing NMD exert transitive upregulation or activation of genes with sequence-similarity to the mutant mRNA (El-Brolosy et al., 2019; Ma et al., 2019; Rossi et al., 2015; Serobyan et al., 2020). Interestingly, transcriptional adaptation was abolished when translation was blocked with a gene-specific morpholino (Ma et al., 2019). Meanwhile, truncated, unpolyadenylated transcripts (Barry et al., 1993; Luo and Chen, 2007), intronless genes (Makeyeva et al., 2021), and inefficient splicing with subsequent retention on spliceosome (Dumesic et al., 2013) were shown to trigger RdRP-dependent silencing in multiple organisms. In FISH studies with *gfp*::*his-61* reporter, we see
prominent mRNA accumulation in select nuclei (Figure 3.2), which could be the event that recruits RdRPs for intronless reporter silencing. We also see lesser yet clear nuclear accumulation of the expressing intron-containing *gfp::his-61*. As discussed earlier, “determination” to silence appears to be made based on cumulative silencing and protective cues, not one particular feature. While matching piRNAs (Shen et al., 2018; Zhang et al., 2018), homologous WAGO 22G-RNAs (Shirayama et al., 2012) and lack of introns (Makeyeva et al., 2021) serve as silencing cues, PATCs (Frokjaer-Jensen et al., 2016; Zhang et al., 2018), immediate and surrounding chromatin environment (Frokjaer-Jensen et al., 2016; Rechtsteiner et al., 2010), introns (Akay et al., 2017; Makeyeva et al., 2021), coding region (Seth et al., 2018), and RNA-based memory of active expression state (Johnson and Spence, 2011) were shown to impart protection.

It is entirely possible that mRNA exported from the nucleus and/or associated with the ribosome is deemed “safe,” because it passed all of the upstream quality control measures: successful transcription, 3’-end processing and formation, nuclear export, and ribosomal recruitment. On the other hand, if the transcript fails to be efficiently processed, it will likely encounter nuclear export difficulty and may be flagged and treated as “potentially foreign.” Association between nuclear pores and P-granules in *C. elegans* (Pitt et al., 2000; Sheth et al., 2010) allows the latter to assume the role of a checkpoint. P-granules are phase-separated structures (Elbaum-Garfinkle et al., 2015) that contain mRNA processing (Updike and Strome, 2010; Voronina et al., 2012) and small RNA factors (Updike and Strome, 2010), and serve as anchor sites for mutator foci comprised of RdRPs, RDE-3 and other
components required for 22G-biogenesis (Phillips et al., 2012). A unique feature of the *C. elegans* germline is shared cytoplasm (reviewed in Pazdernik and Schedl, 2013) that facilitates and perhaps underlies efficient post-transcriptional silencing by cytoplasmic WAGOs. If some of the nuclei did not detect parasite transcripts, information could still be transmitted via cytoplasm. In Figure 3.2, it is apparent in animals bearing intronless *gfp*::*his-61* that select nuclei accumulate more transgenic mRNA than others, suggesting that target detection and/or silencing process is not uniform and may be partially stochastic. However, cytoplasm is still clearly devoid of the *gfp*::*his-61* mRNA (Figure 3.2 middle row), indicating robustness of post-transcriptional cytoplasmic WAGO silencing.

Another observation in line with the protection of transcripts that reach cytoplasm and/or ribosomes is miRNA target suppression. miRNAs constitute yet another class of small RNA species that are complexed with Argonautes and downregulate gene expression in soma and germline (Bartel, 2009). Studies reveal that miRNA targets are associated with polyribosomes and suggest that instead of actively degrading mRNA, miRNAs halt translation (Olsen and Ambros, 1999). However, some reports show a proportional decrease in mRNA level (Bagga et al., 2005). Evidence supporting both modes of miRNA regulation has been widely discussed (Carthew and Sontheimer, 2009; Eulalio et al., 2008; Filipowicz et al., 2008; Wu and Belasco, 2008). While the aforementioned studies focused on miRNAs expressed during larval *C. elegans* stages, a recent work compared the mechanism of action of the same miRNA species between the germline and somatic tissues (Dallaire et al., 2018). Authors show that translational repression occurs in
both tissues, but target mRNA decay is only observed in somatic tissues. As miRNA targets succeed to recruit ribosomes, perhaps the protective cues placed on germline transcripts license mRNA not only from siRNA-silencing but also prevent miRNA-mediated mRNA decay. It would be curious to examine if mRNA processing failure can enable miRNA-mediated transcript degradation.

Alternatively or, in addition to the nuclear mRNA surveillance, deposited H3K9me3 chromatin marks could trigger WAGO 22G-RNA silencing. In yeast, tethering of RNA-induced transcriptional silencing (RITS) complex triggered RNAi-mediated silencing (Buhler et al., 2006). Similarly, some unpublished data from our lab suggests that tethering NRDE-2/4 chromatin regulators, which enable nuclear RNAi and transgenerational inheritance of silencing memory, leads to RdRP-mediated silencing. Select intronless reporters described here are likely to undergo small RNA-independent transcriptional silencing, as demonstrated by transcript distribution revealed in FISH studies and drastic mRNA reduction assessed by RT-qPCR and Northern assays (cdk-1*::gfp* and oma-1*::3xFLAG). Perhaps, it is the transcriptional suppression that serves as the trigger for unprimed RdRP-mediated WAGO 22G-RNA synthesis. This exciting possibility will be addressed in further studies.

**Screen strategy to identify cis-silencing factors**

An important advantage of *C. elegans* as a model organism is its virtually limitless screening potential. Therefore, to identify factors required for the cis-mode of intronless silencing, I recently designed and created the following strain:
oma-1(te33PTC)::gfp, unc-119cb(+): LGII; unc-119(ed9): LGIII; oma-1::3XFLAG(intronless): LGIV; degron::oma-2, tir-1::mRuby, unc-119cb(+): LGV

To abrogate small RNA-mediated silencing of the intronless oma-1*:3xFLAG without using temperature-sensitive germline mortal mutants like rde-3, I decided to take advantage of the RNAa phenomenon (Seth et al., 2013; Shirayama et al., 2012). Oma-1(te33PTC)::gfp (PTC stands for premature stop codon) that fails to yield protein expression (Detwiler et al., 2001; Seth et al., 2018) can transitively protect piRNA-targeted gfp::cdk-1 (Seth et al., 2018). In addition, intron-containing oma-1 can transitively protect intronless oma-1* sequence from WAGO 22G-RNA accumulation (Figure 2.9A). Based on the aforementioned observations, I reasoned that oma-1(te33PTC)::gfp is likely to prevent WAGO 22G-RNA targeting of the intronless oma-1*:3xFLAG in-trans, which I validated via small RNA sequencing studies (Figure 4.1). While the coding region of oma-1* appears to be protected, 3xFLAG and 3' UTR (untranslated region) portions still accrue antisense small RNAs. According to the previous experiments, abundant but highly localized peaks of antisense small RNAs do not prevent expression of the targeted gene (Seth et al., 2018). Oma-1::gfp targeted with two fully complementary piRNAs exhibited obvious antisense 22G-RNAs near piRNA target sites but OMA-1::GFP expression was still observed. To avoid a potential roadblock in the screen, the FLAG sequence can be removed from oma-1*:3xFLAG via CRISPR-gene editing without affecting the screen, however, altering 3' UTR is not as straightforward given the importance of 3' UTR in germline gene expression (Merritt et al., 2008). Before modifying oma-1*:3XFLAG 3' UTR to avoid small RNA branch of intronless silencing, I tested if
antisense 3’ UTR reads alone can, in fact, prevent gene expression. I used a strain that exhibited a similar occurrence: in animals carrying an intronless *oma-1*::*gfp* transgene, *oma-1* 3’ UTR is also targeted by small RNAs despite the licensing of the *oma-1* coding region (Figure 2.9 A). *Oma-2* depletion via RNAi revealed that intron-containing endogenous *oma-1* gene is expressed, even in the presence of antisense reads mapping to its 3’ UTR, which was indicated by animal fertility following RNAi treatment. Worms deficient for either OMA-1 or OMA-2 are viable and fertile, but the lack of both genes results in failed oocyte maturation (Detwiler et al., 2001; Shimada et al., 2002). While antisense 22G-RNAs targeting 3’ UTR could differentially affect expression of intronless and intron-containing genes, findings of *oma-2* depletion experiment warrant a pilot screen. Moreover, reads in 3’ UTR may indicate the nucleation site of default small RNA silencing, especially considering that intronless *oma-1*::*3xFLAG* mRNA is not polyadenylated (Figure 3.3). Alternatively, observed 3’ UTR reads could represent a signature of RdRP-mediated 22G-RNA synthesis – in the strain carrying *oma-1*::*gfp* and two *oma-1*-specific piRNAs, antisense small RNAs were observed not only near piRNA target sites, but also in the 3’ UTR region (Seth et al., 2018).

Selection of candidates involved in the cis-mode of intronless silencing is facilitated by synthetic sterility of *oma-1*; *oma-2* double mutants. The starting strain does not produce OMA-1 protein, and *degron*::*oma-2* fusion enables conditional depletion of OMA-2 via proteasome-mediated degradation controlled by exposure to auxin. In the preliminary validation, whereby I plated L1-stage screen strain animals on agar plates containing 500μM auxin, 100% of synchronized worms were sterile.
Thus, candidate mutants that disable cis-silencing of *oma-1* and grant its protein expression are expected to yield a fertile strain despite auxin-mediated OMA-2 depletion. If screen candidates disrupt OMA-2 degradation (e.g. proteasome mutants), they will also result in strain survival without disabling cis-silencing. To decrease the probability of the undesired proteasome hits, animals will be subjected to *oma-2* RNAi via feeding along with auxin exposure. Fertile animals will also be examined for GFP expression to exclude candidates with re-activated *oma-1*(te33\textsuperscript{PTC}):gfp allele. Designed strain can be used for a forward or reverse screening strategies.

**Figure 4.1**

Plots showing density of antisense small RNAs mapping along *oma-1::gfp* and *oma-1::3xFLAG* sequences. Assay was performed on *oma-1(te33\textsuperscript{PTC}):gfp, unc-119cb(+)* LGII; *unc-119(ed9) LGIII; oma-1::3XFLAG(intronless) LGIV; degron::oma-2, tir-1::mRuby, unc-119cb(+)* LGV strain. Positions of exon junctions in *oma-1::gfp* and corresponding positions in *oma-1* are indicated by broken vertical lines in gene cartoons along x-axis. The height of each bar represents the number of raw reads that begin at that position.
A word on transitive protection

Transcriptional adaptation is a phenomenon whereby a transcript undergoing NMD transitoriely upregulates homologous genes to compensate for the protein loss caused by the premature termination codon (PTC) (El-Brolosy et al., 2019; Ma et al., 2019; Rossi et al., 2015; Serobyan et al., 2020). Full deletion of the NMD target gene does not exhibit transcriptional adaptation suggesting that the mRNA decay fragments of the NMD target and not the loss of the protein product set the mechanism in motion. Recently, transcriptional adaptation was shown to occur in somatic *C. elegans* tissue (Serobyan et al., 2020). To some extent, transcriptional adaptation is reminiscent of germline RNA-induced epigenetic gene activation (RNAa), except the latter does not require NMD (Seth et al., 2018). While CSR-1 depletion does not interrupt transitive protection in the context of PIWI or intronless reporters (Figure 2.9 and data not shown), we did not test if the factors required for transcriptional adaptation in soma also play role in germline RNAa. Serobyan et al. report that *rrf-3, nrde-3,* and *ergo-1* are required for upregulation of genes homologous in sequence with NMD alleles (Serobyan et al., 2020), though direct association via ERGO-1 and NRDE-3 IP-seq (immunoprecipitation followed by RNA sequencing) remains to be pursued. Gonad mRNA sequencing data reveals that all three express in the germline (Kim et al., 2021). Fellow scientists from the Mello lab, Travers and Shirayama showed that GFP::NRDE-3 is expressed in oocytes (unpublished), and in the Bioxiv study, Claycomb lab shows that GFP::FLAG::ERGO-1 protein is expressed in oocytes and early embryos (Seroussi et al., 2022), consistent with RNA sequencing analyses reporting that *ergo-1* targets
oocyte and early embryo mRNAs (Fischer et al., 2011; Vasale et al., 2010). Peculiarly, RNAa-mediated desilencing of piRNA reporter gfp::cdk-1 in F1 generation appears strongest in oocytes, even though cdk-1::gfp is expressed throughout the entire worm germline (Seth et al., 2013; Shirayama et al., 2012). These findings prompt future studies testing the requirement of rrf-3, nrde-3, and ergo-1 for RNAa. In addition, future bioinformatics analyses could reveal endogenously occurring RNAa and intronless RNAe phenomena.

Parallel with humans

As shown by cell culture studies, HUSH operates in pluripotent and differentiated cells (Seczynska et al., 2022), while piRNA keeps transposable elements at bay in human germ cells (Wang and Lin, 2021). Further studies are required to explore whether intronless gene expression is monitored in human germ cells, and whether the C. elegans mechanisms discussed in this work could be conserved. It is reasonable to postulate that germ cells, with the colossal task of keeping genome intact for the next generation, utilize a multipronged surveillance system to guard the germline from potential genomic invaders.


