Integrator terminates promoter-proximal Pol II to generate *C. elegans* piRNA precursors

Toni Beltran[1,2], Elena Pahita[1,2], Subhanita Ghosh[1,2], Boris Lenhard[1,2] and Peter Sarkies[1,2]*

[1] MRC London Institute of Medical Sciences, Du Cane Road, London W12 0NN, UK

[2] Institute of Clinical Sciences, Imperial College London, Du Cane Road, London, W12 0NN, UK

*correspondence to psarkies@imperial.ac.uk

Abstract

Piwi-interacting RNAs (piRNAs) play key roles in germline development and genome defence in metazoans. In *C. elegans,* piRNAs are transcribed from >15000 discrete genomic loci by RNA polymerase II, resulting in 28 nt short-capped piRNA precursors. Here we investigate transcription termination at piRNA loci. We show that the Integrator complex, which terminates snRNA transcription, is recruited to piRNA loci. We show that the catalytic activity of Integrator cleaves nascent capped piRNA precursors associated with promoter-proximal Pol II, resulting in termination of transcription. Loss of Integrator activity, however, does not result in transcriptional readthrough at the majority of piRNA loci. Our results draw new parallels between snRNA and piRNA biogenesis in nematodes, and provide evidence of a role for the Integrator complex as a terminator of promoter-proximal RNA polymerase II.

Highlights

- Integrator localises to sites of piRNA biogenesis in nematodes

- Integrator cleaves nascent RNAs associated with promoter-proximal Pol II at piRNA loci to release short capped piRNA precursors from chromatin

- Repression of Pol II elongation at the majority of piRNA loci is independent of Integrator

1 **Introduction**

2

3 Piwi-interacting RNAs (piRNAs) are a specific class of small RNAs essential for

4 germline development and transposable element silencing in metazoans (Ozata et

5 al., 2019). piRNA biogenesis is best characterized in the fruit fly *Drosophila*

6 *melanogaster* and in the nematode *Caenorhabditis elegans*. In *D. melanogaster*,

7 piRNAs are transcribed from a small number of genomic clusters to produce >10 kb

8 long piRNA precursors. piRNA precursors are then processed by nuclease activities

9 into 27-35 nucleotide long primary piRNAs. Upon detection of transposable element

10 transcripts, piRNAs initiate a cycle of coupled cleavage reactions involving

11 transposable element mRNAs and piRNA precursor RNAs. This coupling is known

12 as the Ping-pong cycle, which allows amplification of piRNA populations while

13 simultaneously degrading TE mRNAs (Brennecke et al., 2007; Gunawardane et al.,

14 2007).

15

16 *C. elegans* piRNAs present remarkable differences to the *Drosophila* system. Mature

17 piRNAs in *C. elegans* are 21 nucleotides long with a strong bias towards uracil as

18 the first nucleotide (21U-RNAs; (Batista et al., 2008; Das et al., 2008; Ruby et al.,

19 2006)). *C. elegans* piRNA loci are individual transcriptional units. The majority of

20 piRNA loci localize to two ~3 Mb cluster regions on chromosome IV containing more

21 than 15,000 loci, and are demarcated by a conserved GTTTC motif at their promoter

22 regions (Billi et al., 2013; Ruby et al., 2006); these loci are known as motif-

23 dependent piRNAs or type I piRNAs. Another set of low-abundance piRNAs are

24 generated from almost 10,000 loci distributed across the *C. elegans* genome and

25 independently of an upstream motif; these are known as motif-independent piRNAs

26 (Gu et al., 2012) or type II piRNAs. Both types of loci are transcribed as short 27-40

27 nt capped RNA precursors by RNA polymerase II (Pol II) (Gu et al., 2012). These

28 short precursors have been proposed to result from pausing of Pol II at piRNA

29 promoters (Beltran et al., 2019; Gu et al., 2012). We previously described a region of

30 low melting temperature (high AT content) centered 33 nt downstream of piRNA

31 TSSs across multiple nematode species, which may have a role in RNA polymerase

32 II termination (Beltran et al., 2019). However, the exact mechanisms controlling this

33 process are largely unknown. In particular, the role of Pol II-associated factors in

34 promoting pausing and premature termination of Pol II remain unexplored.

35

36 Multiple lines of evidence suggest that the mechanisms of piRNA transcription in

37 nematodes evolved by co-option of factors controlling the transcription of small

38 nuclear RNAs (snRNAs). First, transcription of piRNAs requires the small nuclear

39 RNA activating protein complex SNAPc (Kasper et al., 2014). SNAPc forms a distinct

40 complex in the germline with the nematode-specific pseudokinase PRDE-1 (Kasper

1   et al., 2014; Weick et al., 2014; Weng et al., 2019), which binds to the upstream

2   Ruby motif directly (Weng et al., 2019). In addition, the GTTTC motif is found in

3   snRNA promoters in basal nematodes (Beltran et al., 2019). Transport of piRNA

4   precursors to their processing sites in perinuclear phase-separated P-granules is

5   achieved by a specialised protein complex named PETISCO/PICS (Cordeiro

6   Rodrigues et al., 2019; Zeng et al., 2019). Interestingly, components of this complex

7   are required for the biogenesis of SL1 RNAs, the trans-splicing leaders preceding

8   the majority of *C.elegans* mRNAs, which are also related to snRNAs (Cordeiro

9   Rodrigues et al., 2019). Altogether, these observations suggest that snRNAs and

10   piRNAs share mechanisms of transcription and processing.

11

12   In metazoans, termination of snRNA genes requires the Integrator complex (Baillat

13   et al., 2005). Integrator is a protein complex containing 12-14 subunits, which

14   associates co-transcriptionally with the carboxy terminal domain (CTD) of Pol II. Two

15   of the subunits of the complex, Ints-9 and Ints-11, are homologs of CPSF73, the

16   catalytically active subunit of the cleavage and polyadenylation complex (CPSF)

17   responsible for the formation of 3´ ends of protein-coding gene mRNAs. Ints-11

18   possesses the endonuclease activity responsible for cleavage of nascent snRNAs to

19   generate snRNA 3' ends. This cleavage is coupled to Pol II termination.

20

21   Whilst the best described function for Integrator is at snRNA loci, Integrator may also

22   have functions in transcription of other Pol II-dependent transcripts. Integrator was

23   shown to promote early termination of Pol II at the HIV LTR promoter, attenuating

24   HIV transcription (Stadelmayer et al., 2014). A role for Integrator in conferring

25   processivity to Pol II during the elongation stage of transcription has also been

26   reported (Gardini et al., 2014; Stadelmayer et al., 2014). Additionally, Integrator

27   mediates the biogenesis of enhancer RNAs (eRNAs) by promoting early termination

28   of Pol II (Lai et al., 2015). Finally, Integrator has been proposed to terminate lncRNA

29   transcription (Nojima et al., 2018).

30

31   Here, we discover that the Integrator complex is involved in piRNA biogenesis in

32   nematodes. We show that Integrator activity is required for accumulation of mature

33   piRNAs and their silencing activity. Integrator associates with piRNA clusters in germ

34   cells, and its catalytic activity is essential to produce short capped RNA precursors

35   via co-transcriptional cleavage of nascent piRNA precursors. Our results provide

36   new insights into the mechanisms of piRNA biogenesis in nematodes, and uncover

37   novel functions of the Integrator complex in the regulation of promoter-proximal Pol II

38   pausing and the biogenesis of non-coding RNAs.

39

40

1   **Results**

2

3   *Integrator is required for the biogenesis of short piRNA precursors*

4

5   Prompted by the evolutionary relationship between piRNAs and snRNAs (Beltran et

6   al., 2019), we sought to understand whether Integrator is involved in piRNA

7   biogenesis in *C. elegans*. We performed RNAi against the catalytic subunit of the

8   complex, *ints-11*, and measured changes in piRNA abundance upon knockdown by

9   sequencing small RNAs. We observed a decrease in piRNA abundance in *ints-11*-

10   depleted worms compared to the empty vector controls (Figure 1A, Supp. Fig. 1A-C).

11   To further test this, we assessed the silencing activity of piRNAs using a strain

12   carrying a germline histone H2B::GFP transgene as a sensor of piRNA activity

13   (Bagijn et al., 2012a). Upon knockdown of *ints-11*, we observed an increase in the

14   proportion of GFP-positive animals (Supp. Fig. 1D), suggesting that Integrator

15   functions in piRNA biogenesis.

16

17   To test whether Integrator promotes piRNA transcription, we performed high-

18   throughput sequencing of short (18-36 nt) RNAs with a 5′ cap, allowing us to capture

19   short-capped piRNA precursors. Upon knockdown of *ints-11*, we observed a

20   decrease in piRNA precursors from piRNA loci with the characteristic GTTTC core

21   (Ruby) motif (motif-dependent piRNAs; Figure 1B, Supp. Fig. 1E-F). We also

22   observed a marked increase in motif-dependent piRNA precursor length upon

23   Integrator depletion (Supp. Fig. 1G), suggesting that Integrator plays a role in

24   termination of Pol II at motif-dependent piRNA promoters. To characterize the length

25   distribution of piRNA precursors in greater detail, we performed high-throughput

26   sequencing of capped RNAs up to a length of 75 nt. The previously described peak

27   of ~28 nt (Weick et al., 2014) was accompanied by an additional population of longer

28   precursors up to ~75 nt long (Figure 1C-D). This population was rare in wild type

29   nematodes, but accumulated substantially upon *ints-11* depletion (Figure 1C-D).

30   Altogether, these data strongly suggest a role for Integrator in the termination of

31   short motif-dependent piRNA precursor transcripts. Interestingly, we did not detect a

32   decrease in abundance of motif-independent piRNA precursors, suggesting that their

33   biogenesis does not depend on Integrator (Supp. Fig. 1H).

34

35   *Integrator localizes to sites of piRNA biogenesis in germ cells*

36

37   We investigated whether Integrator is recruited to sites of piRNA biogenesis. We

38   used a strain where an endogenous deletion (*tm1616*) in the *ints-6* Integrator

39   subunit, also known as *dic-1* in *C. elegans*, is rescued by an *ints-6*::3xFLAG::eGFP

40   fusion transgene (Gómez-Orte et. al., 2019). INTS-6::eGFP formed discrete foci in

1   germ nuclei (Figure 2A), which start to form in the mitotic and transition zone regions
2   of the germline, and become most apparent in early-mid pachytene. To test whether
3   these foci correspond to piRNA cluster regions, we crossed this strain with one
4   carrying an mCherry-tagged version of the piRNA biogenesis factor PRDE-1, which
5   is known to bind to piRNA clusters (Weick et al., 2014; Weng et al., 2019). The
6   majority of mCherry::PRDE-1 foci co-localized with INTS-6::eGFP foci across the
7   germline (Figure 2B, quantification in Supp. Fig. 2A). We observed a second
8   population of INTS-6 foci appearing as cells progress into pachytene (Figure 2A),
9   with mid-pachytene cells showing two clear INTS-6 foci, one of which co-localizes
10 with PRDE-1 (Figure 2B, panels 2-3). Towards the end of pachytene, PRDE-1 foci
11 became more diffuse and start to disappear, and INTS-6 signal in PRDE-1 regions
12 was clearly reduced, resulting in no co-localizing foci (Figure 2B, panel 4).
13 Interestingly, however, a single INTS-6 focus remained in late-pachytene cells well
14 after the disappearance of PRDE-1 foci (Figure 2B, panel 5). In order to further
15 confirm these observations, we used CRISPR-mediated genome editing to tag the
16 endogenous INTS-11 protein with an N-terminal FLAG tag, and tested its localization
17 by immunofluorescent staining. FLAG::INTS-11 also co-localized with PRDE-1 foci in
18 germ cells (Supp. Fig. 2B).
19
20 *Chromatin fractionation reveals two distinct populations of nascent piRNA precursors*
21
22 We recently described a sequence signature of increased AT content downstream of
23 motif-dependent 21U-RNAs resulting in a low melting temperature region (hereforth
24 referred to as "termination signal") (Beltran et al., 2019). We showed that the
25 strength of this signal correlates with the length of motif-dependent piRNA
26 precursors, and suggested that it contributes to termination by promoting pausing of
27 early elongating Pol II due to the low stability of the AT-rich nascent RNA-DNA
28 hybrid. We also showed that the elongation factor TFIIS, which is involved in the
29 rescue of backtracked Pol II complexes, is required for efficient motif-dependent
30 piRNA biogenesis (Beltran et al., 2019). On the basis of these observations, we
31 hypothesized that Integrator might function in motif-dependent piRNA biogenesis by
32 processing nascent RNA associated with promoter-proximal Pol II, terminating
33 transcription.
34
35 To test this model, we fractionated purified germ nuclei into chromatin bound and
36 nucleoplasmic fractions. This allowed us to distinguish chromatin-bound nascent
37 piRNA precursors and nucleoplasmic precursors released from chromatin. We
38 generated short capped RNA libraries up to an insert size of 75 bp from both
39 fractions (Figure 3A). We observed longer precursors in the chromatin-bound

1 fraction compared to nucleoplasm, suggesting that this approach indeed captures

2 earlier piRNA precursors (Supp. Fig. 3A).

3

4 In wild type animals, chromatin bound nascent RNA followed a bimodal distribution,

5 with a major peak at ~28 nt and a less abundant peak at ~48 nt. The ~48 nt peak

6 was not present in nucleoplasmic RNA (Figure 3B). Weaker termination signals

7 (lower AT content) downstream of the 21U-RNA locus were associated with a larger

8 proportion of sequences mapping to the 48 nt peak (Figure 3C). Upon *ints-11*

9 knockdown, we observed a decrease in nucleoplasmic and chromatin-bound

10 precursor levels (Figure 3D), and an increase in length in both fractions (Figure 3E).

11 Mutation of *C. elegans* TFIIS (T24H10.1) led to a decrease in nucleoplasmic

12 precursor levels and an increase in their length (Figure 3D,E). However, chromatin-

13 bound precursors had increased levels in *tfiis* mutants, and showed an accumulation

14 of nascent RNAs at the ~48 nt peak, suggesting that TFIIS activity is important for

15 termination of Pol II at the 48 nt peak. Knockdown of *ints-11* in *tfiis* mutants led to a

16 further increase in precursor length in both nucleoplasmic and nascent fractions

17 (Figure 3E). Interestingly, the position of both nascent RNA peaks was shifted in *tfiis*

18 mutants, from 27 to 29-30 nt and from 47-48 to 51 nt. This shift persisted upon *ints-*

19 *11* knockdown in a *tfiis* mutant background, indicating that it is not Integrator-

20 dependent (Figure 3E,F,G). These differences in length are likely to result from

21 reduced efficiency of cleavage of the 3´ end of nascent RNAs by Pol II after pausing

22 and backtracking, suggesting that both peaks indeed correspond to Pol II pause

23 sites on chromatin.

24

25

26 *Integrator cleaves nascent piRNA precursors associated with promoter-proximal Pol*

27 *II*

28

29 On the basis of our preceding observations, we hypothesised that Pol II pauses after

30 it has transcribed ~28 nt, due to increased AT content found downstream of 21U-

31 RNAs (Beltran et al., 2019). The paused polymerase could be resolved either

32 through release from DNA, resulting in release of a ~28 nt piRNA precursor, or by

33 Pol II resuming before arresting again ~48 nt downstream of the TSS. Since ~28 nt

34 nascent RNAs are contained within the RNA exit channel of Pol II, termination of Pol

35 II at the ~28 nt peak would be independent of Integrator. Termination at the ~48 nt

36 peak, in contrast, would result from Integrator cleavage. This model would account

37 for the increased proportion of ~48 nt nascent RNAs upon Integrator knockdown,

38 and the small but consistent 3´ shift in the position of the two peaks in the absence of

39 TFIIS.

40

1   To test this model, we searched for cleavage products with a 5′ monophosphate that
2   might result from Integrator cleavage of nascent RNA (Figure 4A). After removal of
3   piRNA reads, we observed a population of 5′ monophosphate fragments with 5′
4   ends peaking at +38 nt of piRNA TSSs, which we designated as putative cleavage
5   products (Figure 4B). Interestingly, *ints*-11 knockdown resulted in a clear decrease in
6   cleavage products (Figure 4B,C, Supp. Fig. 4A,B), suggesting that these result from
7   Integrator cleavage. These fragments had a median length of 20 nt (Figure 4D), with
8   their 3′ ends centered 58 nt downstream of piRNA TSSs (Supp. Fig. 4C). This data
9   suggests that Integrator processing occurs when Pol II has transcribed
10   approximately 10 nt past the +48 pause site, resulting in the release of 38 nt capped
11   piRNA precursors from chromatin. TFIIS mutation alone also resulted in a reduction
12   in Integrator cleavage products (Figure 4B,C, Supp. Fig. 4A,B). This indicates that
13   TFIIS activity is important for Integrator-mediated processing of nascent RNAs, by
14   promoting Pol II transcription through the +48 pause site. This model is consistent
15   with the accumulation of ~48 nt nascent precursors in *tfiis* mutants (Figure 3D,E).
16   Similarly, *ints-11* knockdown in a *tfiis* background resulted in a reduction in cleavage
17   products (Figure 4B,C, Supp. Fig. 4A,B).
18
19   An important test of our model is that stronger AT-rich signals would result in
20   reduced reliance on Integrator cleavage due to a larger proportion of termination
21   events happening at the 28 nt peak. In support of this, loci with stronger AT-rich
22   signals tended to have reduced abundance of cleavage products (Figure 4E, Supp.
23   Fig. 4D). This data suggests that the AT-rich signal affects pausing and progression
24   of Pol II to the 48 nt peak, and not Integrator processing efficiency. Consistently, the
25   differences in nascent piRNA precursor length between loci with strong and weak
26   termination signals were still present upon knockdown of *ints-11*, suggesting that
27   *ints-11* does not mediate the effect of the termination signal itself (Supp. Fig. 5).
28
29   *Pol II elongation control is independent of Integrator activity at the majority of piRNA*
30   *loci*
31
32   Having established that Integrator cleaves nascent piRNA precursors associated
33   with paused Pol II, we set out to test whether Integrator acts to prevent further Pol II
34   elongation beyond the second pause site at 48 nt downstream of the piRNA TSS. To
35   investigate this, we performed super low-input carrier CAGE (Cvetesic et al., 2018)
36   using chromatin-bound RNA as input material and subjected the libraries to paired-
37   end sequencing. This approach captures 5′ ends of nascent long (>200 nt) capped
38   RNAs, allowing us to identify putative piRNA precursor initiating specifically 2 nt
39   upstream of 21U-RNAs. Paired-end sequencing allowed us to obtain lower-bound

1 estimates of their length from the position of the 3´ end of the fragments generated
2 through random priming.
3
4 Nascent RNA libraries had much higher intron retention rates than a total RNA
5 control (Supp. Fig. 6A,B), validating that chromatin fractionation captures nascent
6 RNAs. Across samples, we observed an enrichment of unique CAGE tags initiating 2
7 nt upstream of motif-dependent 21U-RNAs corresponding to 659 loci in total (3% of
8 Ruby motif-dependent piRNAs, Figure 5A, Supp. Fig 6C). These transcripts were
9 found at low levels in N2 EV controls, as well as in *tfiis* mutants; however, *ints-11*
10 knockdown led to an increase in their abundance (Figure 5B,C), both in an N2
11 background (2.83 median fold increase), and a *tfiis* background (22.62 median fold
12 increase) (Figure 5B). Importantly, this effect was highly specific for CAGE reads
13 initiating 2 nt upstream of piRNAs (Figure 5C), confirming that this signal captures
14 transcription from piRNA promoters. These transcripts peaked at a length of ~400-
15 500 nt (Figure 5D), similarly to the library insert size, suggesting that this is a lower-
16 bound estimate of their length. Thus *ints*-11 knockdown leads to an increase in
17 readthrough transcription at piRNA loci.
18
19 Long capped RNAs initiating 2 nt upstream of motif-independent piRNAs were much
20 more abundant, with 4437 motif-independent piRNA loci producing detectable long
21 capped RNAs (45% of motif-independent loci) (Figure 5E, Supp. Fig. 6D,E). Their
22 abundance, however, was slightly reduced upon Integrator knockdown, in contrast to
23 motif-dependent piRNAs (Figure 5F).
24
25 Despite the clear increase in readthrough transcription at motif-dependent piRNA loci
26 upon *ints-11* knockdown relative to WT, we still only detected such events at a
27 minority of loci (4.65%). This observation is not likely to result from low sensitivity of
28 CAGE since the levels of short piRNA precursors are not significantly different
29 between CAGE-detected and undetected piRNAs (Supp. Fig. 6F); and the number of
30 detected motif-dependent piRNAs is saturated at the sequencing depth we obtained
31 for our CAGE libraries (Supp. Fig. 6G). In order to better understand the factors
32 influencing Pol II promoter escape at piRNA loci, we examined chromatin
33 accessibility (Jänes et al., 2018) and H3K27ac levels (Han et al., 2019) at piRNA
34 promoters. Motif-dependent piRNA promoters showed extremely low accessibility
35 levels below genome-wide average (Supp. Fig. 7A); however, a clear increase in
36 ATAC signal is observed around piRNA TSSs when comparing wild-type adult
37 worms to *glp-1* mutants lacking a germline (1.5-fold mean fold increase) (Supp. Fig.
38 7B). Motif-independent piRNAs, in contrast, showed on average 10 times higher
39 accessibility than genome-wide average, both in wild type and *glp-1* nematodes
40 (Supp. Fig. 7A,B). Both types of piRNA loci also showed clear differences in the

1   levels of H3K27ac: motif-dependent loci showed extremely low levels, while motif-

2   independent loci were highly enriched; this enrichment was especially high in the

3   germline (Supp. Fig. 7C). The subset of motif-dependent piRNA promoters that were

4   detected by CAGE tended to have slightly higher levels of H3K27ac and

5   accessibility, but these are still very low relative to motif-independent loci (Supp. Fig.

6   7D,E). We thus conclude that the differences in readthrough transcription between

7   motif-dependent and motif-independent piRNA promoters correlate with chromatin

8   accessibility and H3K27ac status. However, within motif-dependent loci, accessibility

9   and H3K27ac levels do not explain differences in readthrough.

10

11   This scenario contrasts to snRNA loci, where *ints*-11 knockdown led to a dramatic

12   increase in CAGE signal at all loci (Supp. Fig. 8A,B), reflecting extensive

13   readthrough past snRNA termination sites, which has been described to result in

14   fusion transcripts with downstream protein-coding genes (Figure 6, Gómez-Orte et

15   al., 2019).

16

17   *Integrator attenuates protein-coding gene expression in C. elegans*

18

19   We assessed the role of Integrator in protein-coding gene transcription by examining

20   the changes in short capped RNA and CAGE reads initiating at gene promoters

21   (Chen et al., 2013). We identified a total of 437 genes with changes in short capped

22   RNA signal (FDR<0.1; 371 upregulated, 66 downregulated, Supp. Fig. 8C), as well

23   as 470 genes with changes in CAGE signal (FDR<0.1; 330 upregulated, 140

24   downregulated, Supp. Fig 8D). In addition, changes in short and long capped RNAs

25   tended to correlate well, with very few genes showing changes in the opposite

26   direction (Supp. Fig. 8E-G). The enrichment of upregulated genes suggests that

27   Integrator-mediated termination negatively regulates transcription of a subset of

28   protein-coding genes, as has been recently reported (Elrod et al., 2019). Since motif-

29   independent piRNA loci tend to localize within protein-coding gene promoters

30   genome-wide, we wondered whether they may be generated as a by-product of

31   premature Pol II termination at protein-coding genes by Integrator. Motif-independent

32   piRNA loci, however, were not enriched in the promoters of genes that become

33   upregulated upon Integrator knockdown (Supp. Fig. 8H), consistent with an

34   Integrator-independent biogenesis pathway for motif-independent piRNAs.

35

36   **Discussion**

37

38   In this work, we gain insight on the fundamental mechanism of piRNA biogenesis in

39   *C. elegans*, and explore the role of the Integrator complex in the termination of

40   promoter-proximal Pol II at piRNA loci.

1

2 *Parallels between snRNA and piRNA transcription*

3

4 Multiple lines of evidence suggest that piRNA promoters evolved from snRNA

5 promoters (Beltran et. al., 2019, Domingues et. al., 2019, Weng et. al., 2019). Here,

6 we explore the parallels between the mechanisms of piRNA and snRNA

7 transcription. We show that the Integrator complex is recruited to sites of piRNA

8 biogenesis upon piRNA transcription activation. It is unclear whether Integrator

9 recruitment to piRNA promoters is directly mediated by the USTC complex

10 (SNAPc/PRDE-1/TOFU-5), or whether Integrator is recruited to the initiating Pol II

11 independently of SNAPc. Integrator has been shown to associate with the CTD of

12 Pol II phosphorylated on Ser7 (Egloff et al., 2007, 2010). It will be of interest to

13 explore whether Ser7-P is associated with piRNA biogenesis, and the mechanisms

14 directing this modification. The presence of piRNA-independent Integrator foci may

15 reflect Integrator-dependent regulation of snRNA genes. Indeed, snRNA genes have

16 been shown to spatially interact and coalesce into Cajal bodies (Frey and Matera,

17 2001; Wang et al., 2016). This phenomenon may share some mechanistic aspects

18 with the formation of piRNA foci in the *C. elegans* germline.

19

20 *Two distinct populations of piRNA precursors*

21

22 By isolating chromatin-bound RNA from germ nuclei, we characterize two distinct

23 populations of nascent piRNA precursors; one peaking at ~28 nt, and a second

24 population of ~48 nt long, representing two Pol II pause sites on chromatin. In

25 addition, we provide evidence that Integrator acts to cleave 53-63 nt nascent

26 precursors past the +48 pause site, resulting in the release of 33-43 nt short capped

27 piRNA precursor transcripts (Figure 6). Several studies had set out to address the

28 length of piRNA precursors in the past. Cecere et al., 2012 reported the existence of

29 a ~75 nt long precursor detected by RACE, but high-throughput sequencing-based

30 approaches detected 26-28 nt precursors (Gu et al., 2012; Weick et al., 2014). Our

31 discovery of a second population of 40-60 nt nascent precursors reconciles these

32 previous observations by suggesting that longer precursor transcripts, whilst rare,

33 can be detected at piRNA loci.

34

35 Interestingly, a large fraction of nascent precursors peak at a length of ~28 nt, as

36 observed for nucleoplasmic precursors. This raises the question of whether these

37 are truly chromatin-bound. We previously suggested that increased AT content

38 downstream of 21U-RNAs leads to Pol II pausing ~28 nt downstream of the TSSs

39 (Beltran et al., 2019). Consistent with this model, TFIIS mutation results in a 2-3 nt

40 shift in both +28 and +48 peaks, suggesting pausing and backtracking of Pol II at

1   both sites (Figure 6). In addition, Integrator cleavage fragments are reduced in loci
2   with high AT content downstream of 21U-RNAs, suggesting that Pol II is less likely to
3   reach the +48 peak in these loci. Pol II may be released from the +28 pause site
4   independently of Integrator, and Integrator may act as a failsafe termination
5   mechanism when transcription continues beyond +48 (Figure 6). A recent study also
6   supports the existence of two Pol II pause sites at protein-coding genes in human
7   cell lines (Aoi et al., 2020). Interestingly, accumulation of Pol II at the second pause
8   site is observed upon NELF depletion (Aoi et al., 2020). *C. elegans* does not have
9   NELF (Maxwell et al., 2014) potentially favouring the transition of Pol II to the +48
10 peak.
11
12 *Integrator and Pol II elongation control*
13
14 Our analysis of the role for Integrator at piRNA loci in *C. elegans* fits with a recent
15 study demonstrating that Integrator terminates promoter-proximal Pol II at protein-
16 coding gene promoters to attenuate gene expression (Elrod et al., 2019). The short-
17 capped RNAs produced as a by-product of Integrator cleavage at protein-coding
18 gene promoters are rapidly degraded by the nuclear exosome. At *C. elegans* piRNA
19 loci, instead, the resulting short capped RNAs are handed in to a specialized RNA
20 processing complex resulting in mature piRNA production (Cordeiro Rodrigues et al.,
21 2019; Zeng et al., 2019). How piRNA precursors are recognized and channelled into
22 the downstream processing pathways while avoiding nuclear RNA surveillance will
23 be interesting to explore in the future.
24
25 Our data reveals an interesting interplay between TFIIS and Integrator function in the
26 regulation of promoter-proximal Pol II. TFIIS mutation results in increased nascent
27 precursors, and an accumulation of ~48 nt precursors, but a decrease in degradation
28 fragments, suggesting that TFIIS activity modulates the ability of Integrator to
29 process nascent RNAs. The observation that Integrator processing occurs
30 downstream of the +48 pause site suggests that the role of TFIIS is to promote
31 elongation through the +48 site prior to Integrator termination. Interestingly, *C.*
32 *elegans* mutants of the RPB-9 subunit of Pol II exhibit defects in piRNA biogenesis
33 similar to *tfiis* mutants (Berkyurek et al., cosubmitted). This data is consistent with
34 previous studies indicating that TFIIS binding to Pol II is compromised in Δ*rpb-9*
35 mutants in *Saccharomyces cerevisiae* (Sigurdsson et al., 2010). Indeed, the capacity
36 of TFIIS to promote Pol II transcription through arrest sites *in vitro* depends on RPB-
37 9 (Awrey et al., 1997). Understanding how the regulation of Pol II early elongation
38 influences the ability of Integrator to prematurely terminate transcription will be an
39 interesting avenue for future study.
40

1   At snRNA loci, Integrator is required to terminate transcription and in its

2   absence substantial readthrough transcription is observed (Baillat et al., 2005;

3   Gómez-Orte et al., 2019). In our study, we observed similar evidence of readthrough

4   transcription after Integrator knockdown at snRNA loci. In contrast, Integrator

5   depletion did not result in readthrough transcription at most motif-dependent piRNA

6   loci. This indicates that the cleavage activity of Integrator is not required to prevent

7   elongation at motif-dependent piRNAs. Pol II elongation control is thus likely to

8   operate independently of Integrator via additional mechanisms (Figure 6). One

9   possibility is that elongation repression is chromatin-based. Motif-dependent piRNAs

10   localize to broad domains of H3K27me3 (Beltran et al., 2019), the vast majority of

11   which also have low accessibility and low H3K27 acetylation (Supp. Fig. 8A-D),

12   reminiscent of poised enhancers (Creyghton et al., 2010). In contrast, motif-

13   independent piRNAs, which show extensive readthrough, are in highly accessible

14   H3K27ac-rich genomic regions. The tight control of Pol II elongation at motif-

15   dependent piRNA loci may be a consequence of evolutionary pressure to avoid

16   transcriptional interference between neighbouring piRNA loci given the high density

17   of loci in piRNA cluster regions that are simultaneously activated.

18

19   *Integrator: piRNA functions beyond nematodes?*

20

21   The involvement of Integrator in piRNA biogenesis adds to the existing evidence that

22   snRNA and nematode piRNA transcription are functionally and evolutionarily related.

23   Whether this principle extends to piRNAs in other organisms remains unknown;

24   however, the piRNA biogenesis factor cutoff has been shown to bind to snRNA

25   promoters in *Drosophila melanogaster* (Prytikin et al., 2017). Interestingly, Integrator

26   was identified in genetic screens to be required for piRNA-mediated silencing of a

27   germline reporter in *Drosophila* (Handler et al., 2013)*,* making Integrator a possible

28   candidate to assist in 3´ end formation of piRNAs in *Drosophila* and mouse. Our data

29   adds to the growing evidence indicating co-option of functional modules from a range

30   of existing transcriptional and RNA metabolic processes in metazoan piRNA

31   biogenesis, and expands the repertoire of non-coding RNAs under Integrator control

32   in metazoans.

33

34

35

36

37

38

39

40

Methods

Nematode culture and RNAi

C. elegans nematodes were grown at 20°C in standard nematode growth medium (NGM) agar plates feeding on OP50 *E. coli*. RNAi clones from the Ahringer library (Kamath and Ahringer, 2003) were grown overnight in LB supplemented with 50 µg/ml ampicillin and 25 µg/ml tetracycline. The next morning, cultures were diluted in LB 50 µg/ml ampicillin, grown to an OD of 0.5-0.6, and seeded into 1 week old NGM plates supplemented with 1 mM IPTG and 50 µg/ml ampicillin. Plates were dried for 48 h before plating synchronized L1 worms obtained by hypochlorite treatment followed by 24 h starvation in M9 media.

Live imaging

Live imaging of SX1316 piRNA sensor worms, as well as INTS-6::eGFP::3xFLAG and PRDE-1::mCherry transgenes was carried out in a DeltaVision fluorescence microscope system. Worms were picked into a 2 µl drop of M9 with 0.5 µM levamisole on a slide, and subsequently immobilized in a hydrogel matrix with microbeads as described in Dong et al., 2018. Live imaging was carried out within 15 min of slide preparation.

Immunofluorescent staining

Mouse anti-FLAG M2 monoclonal antibody (F1804, Sigma) preadsorbed with *C. elegans* lysates to remove background was used for immunofluorescent staining. 15-20 worms were picked onto 15 µl M9 Tween 0.1% on a poly-lysine-coated slide, and dissected using a gauge needle. Another 15 µl 2% paraformaldehyde in M9 Tween 0.1% were added, some solution was removed and a coverslip was placed on top to let the germlines fix for 5 minutes. The slides were then frozen in liquid nitrogen, worms were cracked by quickly removing the coverslip, and slides were fixed in -20°C methanol for at least 1 minute. Slides were then washed by immersion in PBST for 5 min, 3 times, and blocked with PBST with 1% BSA for 1 h at room temperature. Slides were then incubated with mouse anti-FLAG M2 antibody diluted 1:500, and rabbit anti-mCherry antibody (GTX128508) diluted 1:500 in blocking solution, in a wet chamber, overnight, at 4°C. The next day, slides were washed in PBST for 10 minutes, 3 times, and then incubated with goat anti-mouse Alexa 488 secondary antibody at 1:200 dilution, and goat anti-rabbit Alexa 594 secondary antibody at 1:200 dilution in PBST for 1-2 h in the dark. Slides were again washed in PBST for 10 minutes, 3 times, and incubated with 100 µl DAPI 1 µg/ml for 5 minutes in the

1   dark. Slides were finally washed in PBST in the dark for 20 minutes. Excess liquid

2   was removed and slides were mounted with Vectashield and sealed with nail polish.

3

4   *RNA extraction*

5

6   For total RNA whole animal samples, young adult synchronized populations of

7   worms were grown and washed with M9 three times to remove bacterial residues.

8   For each 100 µl of worm pellet, 1ml of Trizol was added. 5 cycles of freeze-cracking

9   were carried out, freezing in liquid nitrogen, followed by thawing in a water bath at

10   37ºC. Tubes were vortexed for 30 s on, 30 s off, for 5 minutes, and incubated for

11   5min at room temperature. 200 µl chloroform per ml of Trizol were added, tubes

12   were shaked vigorously, incubated 2-3 min at room temperature, and centrifuged at

13   full speed for 10 min at 4ºC. The top layer was transferred to a new tube, and RNA

14   was precipitated overnight at -20 ºC with 1 µl glycogen and an equal volume of

15   isopropanol.

16

17   *Germ nuclei purification and chromatin fractionation*

18

19   Germ cells were purified similarly to the protocol described in Han et al., 2019.

20   Approximately 50 100 mm standard NGM or RNAi plates of synchronized young

21   adult worms were used per sample. Worms were washed off plates with M9 and

22   collected in a 15 ml conical tube, and washed several times with M9 to clean from

23   bacteria. The final wash was done with ice cold nuclear purification buffer (NPB: 10

24   mM HEPES pH 7.6, 10 mM KCl, 1.5 mM $MgCl_2$, 1 mM EGTA, 0.25 M sucrose,

25   0.025% TritonX-100, 50 mM NaF, 1 mM DTT, 40 mM β-glycerol, 2 mM Na_3VO_4).

26   Worms were resuspended in NPB and transferred to a previously chilled metal

27   grinder (#08-414-20A), and 1-3 strokes were applied, monitoring the extent of

28   homogenization after each stroke, such that no more than 10-15% of worms were

29   broken. The solution was transferred to a prechilled 50 ml conical tube, and vortexed

30   for 30 s at medium speed, then incubated on ice for 5 min, twice. The sample was

31   then filtered through a 100 µm nylon mesh filter (Falcon), followed by three 40 µm

32   (Falcon) and three 20 µm filters (PluriSelect). The filtered solution was recovered

33   and centrifuged at 100 g for 6 min at 4ºC to pellet debris (usually a good germline

34   yield has little or no material pelleted in this step). The supernatant was recovered

35   and centrifuged at 2500 g for 10 min at 4ºC to pellet nuclei, and the supernatant was

36   discarded. Isolated nuclei were then subjected to chromatin fractionation as

37   described in Nojima et al., 2016. Briefly, nuclei were resupended in 67.5 µl of ice-

38   cold NUN1 buffer (20 mM Tris-HCl pH=7.9, 75 mM NaCl, 0.5 mM EDTA, 50%

39   glycerol v/v, 1x cOmplete protease inhibitor cocktail), and mixed gently by pipetting.

40   0.6 ml of ice-cold NUN2 buffer (20 mM HEPES pH=7.6, 300 mM NaCl, 0.2 mM

1 EDTA, 7.5 mM MgCl$_2$, 1% v/v NP-40, 1 M urea, 1x cOmplete protease inhibitor
2 cocktail) were added. The mix was vortexed at maximum speed and incubated on
3 ice for 15 min, vortexing every 3-4 during the incubation to precipitate the chromatin
4 fraction. The mix was then centrifuged at 16000 g for 10 min at 4ºC. The supernatant
5 (nucleoplasm fraction) was collected, split into two aliquots of 300 µl, and 900 µl
6 Trizol LS were added to each aliquot and mixed by vortexing to proceed with RNA
7 extraction. To isolate chromatin-bound RNA, the chromatin pellet was resuspended
8 in 200 µl HSB buffer (10 mM Tris-HCl pH=7.5, 500 mM NaCl, 10 mM MgCl$_2$) with 10
9 µl TURBO DNase (stock 2 U/µl) and 10 µl Superase-In RNase inhibitor (stock 20
10 U/µl), and incubated at 37ºC for 10 min, vortexing to help resuspension and
11 solubilization. 2 µl of proteinase K (stock 20 mg/ml) were then added, and the mix
12 was incubated at 55ºC for 10 min. 600 µl Trizol LS were added to proceed with RNA
13 extraction.

14

15 *CIP-Rpph treatment of RNA and small RNA library preparation*

16

17 To enrich for short-capped RNAs, we first pretreated 2 µg total RNA, or 500 ng of
18 chromatin-bound or nucleoplasmic RNA with 2 μl Quick CIP (NEB) in a total volume
19 of 20 μl for 90 min at 37°C. RNA was phenol-chloroform extracted, and precipitated
20 at -20ºC overnight with 1/10 vol 3 M AcNa, and 3 vol 100% EtOH and 1 μl glycogen.
21 The next day, RNA was resuspended and treated with 7.5 units (1.5 μl) RNA 5′
22 pyrophosphohydrolase (NEB) for 1 h at 37°C in a total volume of 20 μl. RNA was
23 extracted and precipitated overnight as described for CIP treatment. The resulting
24 RNA was used as input for small RNA library preparation using the TruSeq small
25 RNA kit (Illumina) according to the manufacturer's instructions, except for an
26 increase in the number of PCR cycles from 11 to 15. An insert size range of 20-70 nt
27 was gel purified, DNA was eluted from the gel in 0.3 M NaCl, and EtOH precipitated
28 overnight. For direct 5′ P libraries, 1 µg total RNA was used as input, and an insert
29 size range of 20-35 nt was gel purified. Libraries were quantitated using Qubit and
30 Tapestation prior to pooling in groups of 6 to 12 per lane, and sequenced on an
31 Illumina HiSeq2000 (50bp single end libraries), or an Illumina NextSeq instrument
32 (75 bp single end libraries).

33

34 *Small RNA sequencing data processing, mapping and normalization*

35

36 The Illumina universal adapter was trimmed from small RNA reads using cutadapt
37 v1.10 and reads >17nt were mapped to WS252 *C. elegans* genome (ce11) using
38 bowtie v0.12 (Langmead et al., 2009) with parameters –v 0 –m 1. Sam files were
39 converted to bam and bed using samtools v1.2 (Li et al., 2009) and bedtools v2.25.0
40 (Quinlan and Hall, 2010).

1 piRNA and miRNA counts were derived from direct 5′ P libraries by searching
2 collapsed fasta files for exact matches to existing piRNA (Batista et al., 2008; Gu et
3 al., 2012) and miRNA (Kozomara and Griffiths-Jones, 2014) annotations. Direct
4 library sizes were estimated, using total non-structural read counts (1), using total
5 miRNA counts (2), and extracting DEseq2 size factors from a miRNA counts table
6 (3), resulting in similar trends. Log2 fold change in abundance was calculated for loci
7 with more than 5 DEseq2 normalized read counts on average. 1 pseudocount was
8 added to avoid zero counts. Additionally, the change in total mature piRNA counts
9 relative to N2 was estimated using the same size factors.

10 piRNA precursor reads were identified from CIP-Rpph libraries by identifying reads
11 mapping exactly 2 nt upstream of annotated piRNAs. Library sizes were estimated
12 as the total counts of precursor reads mapping to previously annotated TSSs (Chen
13 et al., 2013). The log2 fold change in abundance was calculated for loci with more
14 than 5 reads per million of TSS-mapping scRNA reads on average. 1 pseudocount
15 was added to avoid zero counts. Additionally, the change in total piRNA precursor
16 counts relative to N2 was estimated using the same size factors. To estimate the
17 effects of the termination signal strength on piRNA precursors, we calculated
18 termination signature strength as described in Beltran et al., 2019 considering the
19 entire transcriptional unit from the TSS to +50, and considering the region
20 downstream of 21U-RNAs alone (+25 to +50). We plotted the length distribution of
21 precursors across equal-sized bins of piRNA loci stratified according to pausing
22 signature strength. We additionally plotted the length distributions for the top vs
23 bottom 20% loci.

24

25 *Bootstrap analysis of piRNA precursor peak locations*

26

27 In order to obtain robust estimates of precursor distributions, and sufficient sampling
28 of precursors, we combined precursors identified in the two replicates for each
29 condition. For each condition, we generated 2000 random subsamples of 3000
30 precursor sequences, sampling without replacement and weighting the probability of
31 sampling by the number of reads for each precursor sequence. We fit a smoothing
32 density curve to the distribution, identified the two local maxima of the fit
33 corresponding to first and second peaks, and extracted the positions of the two. The
34 resulting distributions of positions of both peaks for each condition are shown on
35 Figure 4F,G.

36

37 *Degradation fragment analysis*

38 Degradation fragment analysis was carried out from direct 5′ P small RNA libraries
39 generated from RNA purified from germ nuclei, with the data processed as described

1 above but without applying a minimum read length cutoff. Reads >15 nt initiating

2 exactly at annotated 21U-RNA sites were discarded from the libraries, and the read

3 coverage of the remaining 5′ P unique sequences around piRNA promoters was

4 plotted at single nucleotide resolution in the form of an average profile and a

5 heatmap. Additionally, the counts per million and sequences per million of 5′ and 3′

6 ends of 5′ P reads was similarly plotted. We found an enrichment of 21-U RNA

7 sequences initiating within a +-5 nt window from 21U-RNA sites. These likely

8 represent unannotated 21U-RNAs originating from overlapping TSSs from the same

9 piRNA locus (Billi et al., 2013). We additionally removed these 21U-RNA sequences,

10 as well as reads >15 nt initiating at those unannotated 21U-RNA sites. The

11 remaining first nucleotide and length distributions showed an enrichment of unique

12 sequences initiating at the -2 position from annotated 21U-RNAs. Fragments whose

13 5′ ends mapped in between +25 and +50 from the 5′ U of annotated 21U-RNAs,

14 corresponding to the second peak centered at +50 were collected, and their length

15 distribution was plotted. Total fragment counts per locus were calculated and

16 normalized to cpm of total non-structural mapped reads. The total number of loci with

17 detectable fragments, and well as the distribution of fragment cpms were quantified

18 across piRNA loci binned according to the strength of their termination signals.

19

20 *SLIC-CAGE library preparation and sequencing*

21

22 CAGE library preparation from chromatin-bound RNA was performed using the

23 SLiC-CAGE protocol (Cvetesic et al., 2018, 2019). The prepared libraries were

24 pooled and sequenced on an Illumina HiSeq2000 (100bp paired-end libraries). In

25 order to assess the enrichment of potential piRNA precursor reads in chromatin-

26 bound samples, a total RNA sample control was used. Intron retention ratios were

27 estimated using IRFinder (Middleton et al., 2017), which computes a ratio between

28 the abundance of spliced and unspliced introns for all annotated RNAs,confirming

29 the enrichment of unspliced transcripts in nascent RNA samples.

30

31 *Analysis of SLIC-CAGE data*

32

33 CAGE tags of read 1 were mapped to a reference C.elegans genome (ce11) using

34 STAR (Dobin et al., 2013). Uniquely mapped reads were imported into R

35 (http://www.R-project.org/) as bam files using the standard workflow within the

36 CAGEr package (Haberle et al., 2015). Due to the template free activity of the

37 reverse transcriptase enzyme, often a G is added at the 5' end of the read. G's that

38 do not map to the genome at the 5´ end of reads are removed by CAGEr's standard

39 workflow. G's that map to the genome are removed at a rate comparable to the

40 mismatching G's removed. The 5´ ends of reads represent CAGE-supported

1 transcription start sites (CTSSs) and the number of tags for each CTSS reflects
2 expression levels. Raw tags were normalised using a referent power-law distribution
3 and expressed as normalized tags per million (TPMs) (Balwierz et al., 2009).
4 Biological replicates were highly correlated ($r^2 > 0.9$) and were therefore merged ,
5 and the distributions of $\log_2$ fold change in CAGE signal at the -2 position of piRNA
6 loci were calculated. For each pair of samples, piRNA loci with non-zero counts in at
7 least one of the conditions were included, and 0.01 tpm were added to all data points
8 to prevent division by zero. $\log_2$ transformed data for all positions upstream of
9 piRNAs in all conditions was plotted as a heatmap, with the total sum of normalized
10 CAGE tpms by position shown above.

11

12 For fragment length quantification, reads were aligned using using Tophat v2.1.0.
13 (Langmead and Salzberg, 2012) with parameters -X 20000. Alignment files were
14 sorted by read name and converted to bed12 format using pairedBamToBed12.
15 Read pairs initiating at -2 from 21U-RNAs were retrieved and their overall length and
16 abundance normalized to counts per million of mapped fragments was plotted. In
17 addition, library sizes obtained from the total numbers of pair1 reads containing SL1
18 and SL2 spliced leader sequences were used for normalization, resulting in very
19 similar trends to previous normalization strategies. Tophat alignments were
20 converted to bigwig tracks using deeptools v3.1.2 bamCoverage (Ramírez et al.,
21 2016) with parameters -bs 10 -e, tracks from the two replicates for each condition
22 were averaged, and signal average profiles and heatmaps were generated for
23 snRNA and protein-coding gene TSSs.

24

25 CAGE counts were randomly downsampled to multiple total count sizes (from 50,000
26 to 5,000,000 to match library size across libraries), and to multiple percentages of
27 total library size (to examine saturation of piRNA detection with sequencing depth).
28 15 samples were obtained for each size and library, and the total number of piRNA
29 loci with detected signal at -2 nt from the 21U-RNA were recovered. For each of the
30 15 subsamples, the two replicates of each library were averaged. The resulting
31 distributions of total detected piRNAs per condition were plotted as boxplots against
32 the depth of the subsamples.

33

34 *ATAC-seq and ChIP-seq data processing and analysis*

35

36 Processed bigwig ATAC-seq tracks from young adult wild-type nematodes, and day
37 1 *glp-1* mutant adults were obtained from Janes et al., 2018 (GSE114439), and lifted
38 over to ce11. Raw H3K27 acetylation ChIP-seq data from isolated germ nuclei and
39 somatic nuclei was obtained from Han et al., 2019 (SRR9214969 to SRR9214976),
40 mapped to ce11 using bowtie2 (Langmead and Salzberg, 2012) with default

1 parameters, and the resulting bam files were converted to bigwig tracks using
2 deeptools bamCoverage (Ramírez et al., 2016) with parameters -bs 1 -e 150 --
3 normalizeUsing RPGC --effectiveGenomeSize 96945445. Replicates were averaged
4 using bigwigCompare, and heatmaps of H3K27ac and ATAC signal around piRNA
5 loci were generated using deeptools computeMatrix reference-point. The signal data
6 underlying the heatmaps was extracted in order to compare the distributions of
7 ATAC (±100 bp around piRNA TSSs) and H3K27ac signal between groups of loci
8 (±1 kb around piRNA TSSs).
9
10 *Differential expression analysis of gene promoters genome-wide using scRNA and*
11 *CAGE count data*
12
13 scRNA or CAGE TSS annotations were intersected with a set of WormBase protein-
14 coding gene promoters (Chen et al., 2013) lifted over to ce11, in order to generate
15 count matrices of scRNA or CAGE counts for all gene promoters genome-wide.
16 Differential expression analysis comparing N2 empty vector chromatin bound
17 samples to N2 *ints-11* RNAi chromatin bound samples was carried out using
18 DESeq2. Promoters were considered to be differentially expressed using a
19 Benjamini-Hochberg FDR threshold of 0.1. The fold changes of promoters identified
20 as differentially expressed in either of the two experiments were extracted, and the
21 fold change in scRNA signal was plotted against the fold change in CAGE signal. To
22 test for enrichment of motif-independent piRNAs in each of the promoter classes,
23 motif-independent piRNAs were first overlapped with promoters detected in the
24 scRNA and CAGE experiments in order to determine their background enrichment.
25 Enrichment was calculated in each promoter set as an odds ratio relative to its
26 corresponding background; for instance the fraction of CAGE "up" promoters
27 overlapping with motif-independent piRNAs was compared to the fraction of all
28 CAGE detectable promoters overlapping with motif-independent piRNAs. For
29 intersecting promoter sets, genes detected as significant in either experiment, and
30 having a fold change greater than 2 in both experiments were selected. In these
31 cases, background was calculated as the fraction of promoters detected in both
32 experiments overlapping with motif-independent piRNAs.
33
34 **Acknowledgements**
35
38
39
40

1 **References**

2

3 Aoi, Y., Smith, E.R., Shah, A.P., Rendleman, E.J., Marshall, S.A., Woodfin, A.R.,
4 Chen, F.X., Shiekhattar, R., and Shilatifard, A. (2020). NELF Regulates a Promoter-
5 Proximal Step Distinct from RNA Pol II Pause-Release. Mol. Cell.

6

7 Awrey, D.E., Weilbaecher, R., Hemming, S.A., Orlicky, S.M., Kane, C., and Edwards,
8 A.M. (1997). Transcription elongation through DNA arrest sites - A multistep process
9 involving both RNA polymerase II subunit RPB9 and TFIIS. J. Biol. Chem. *272*,
10 14747–14754.

11

12 Bagijn, M.P., Goldstein, L.D., Sapetschnig, A., Weick, E.-M., Bouasker, S.,
13 Lehrbach, N.J., Simard, M.J., Miska, E.A., Das, P.P., Batista, P.J., et al. (2012a).
14 Function, targets, and evolution of Caenorhabditis elegans piRNAs. Science *337*,
15 574–578.

16

17 Bagijn, M.P., Goldstein, L.D., Sapetschnig, A., Weick, E., Bouasker, S., Lehrbach,
18 N.J., Simard, M.J., and Miska, E. a (2012b). Function, Targets and Evolution of
19 Caenorhabditis elegans piRNAs. Science (80-.). *574*.

20

21 Baillat, D., Hakimi, M.-A., Näär, A.M., Shilatifard, A., Cooch, N., and Shiekhattar, R.
22 (2005). Integrator, a Multiprotein Mediator of Small Nuclear RNA Processing,
23 Associates with the C-Terminal Repeat of RNA Polymerase II. Cell *123*, 265–276.

24

25 Balwierz, P.J., Carninci, P., Daub, C.O., Kawai, J., Hayashizaki, Y., Van Belle, W.,
26 Beisel, C., and van Nimwegen, E. (2009). Methods for analyzing deep sequencing
27 expression data: constructing the human and mouse promoterome with deepCAGE
28 data. Genome Biol. *10,* R79.

29

30 Batista, P.J., Ruby, J.G., Claycomb, J.M., Chiang, R., Fahlgren, N., Kasschau, K.D.,
31 Chaves, D.A., Gu, W., Vasale, J.J., Duan, S., et al. (2008). PRG-1 and 21U-RNAs
32 Interact to Form the piRNA Complex Required for Fertility in C. elegans. Mol. Cell
33 *31*, 67–78.

34

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

39

40 Billi, A.C., Freeberg, M.A., Day, A.M., Chun, S.Y., Khivansara, V., and Kim, J.K.
41 (2013). A Conserved Upstream Motif Orchestrates Autonomous, Germline-Enriched
42 Expression of Caenorhabditis elegans piRNAs. PLoS Genet. *9*.

43

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

1

2 Cecere, G., Zheng, G.X.Y., Mansisidor, A.R., Klymko, K.E., and Grishok, A. (2012).
3 Promoters Recognized by Forkhead Proteins Exist for Individual 21U-RNAs. Mol.
4 Cell *47*, 734–745.

5

6 Chen, R.A.-J., Down, T.A., Stempor, P., Chen, Q.B., Egelhofer, T.A., Hillier, L.W.,
7 Jeffers, T.E., and Ahringer, J. (2013). The landscape of RNA polymerase II
8 transcription initiation in C. elegans reveals promoter and enhancer architectures.
9 Genome Res. *23*, 1339–1347.

10

11 Cordeiro Rodrigues, R.J., de Jesus Domingues, A.M., Hellmann, S., Dietz, S., de
12 Albuquerque, B.F.M., Renz, C., Ulrich, H.D., Sarkies, P., Butter, F., and Ketting, R.F.
13 (2019). PETISCO is a novel protein complex required for 21U RNA biogenesis and
14 embryonic viability. Genes Dev. *33*, 857–870.

15

16 Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine,
17 E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone
18 H3K27ac separates active from poised enhancers and predicts developmental state.
19 Proc. Natl. Acad. Sci. *107*, 21931 LP – 21936.

20

21 Cvetesic, N., Leitch, H.G., Borkowska, M., Müller, F., Carninci, P., Hajkova, P., and
22 Lenhard, B. (2018). SLIC-CAGE: high-resolution transcription start site mapping
23 using nanogram-levels of total RNA. Genome Res.

24

25 Cvetesic, N., Pahita, E., and Lenhard, B. (2019). Transcription Start Site Mapping
26 Using Super-low Input Carrier-CAGE. JoVE e59805.

27

28 Das, P.P., Bagijn, M.P., Goldstein, L.D., Woolford, J.R., Lehrbach, N.J.,
29 Sapetschnig, A., Buhecha, H.R., Gilchrist, M.J., Howe, K.L., Stark, R., et al. (2008).
30 Piwi and piRNAs Act Upstream of an Endogenous siRNA Pathway to Suppress Tc3
31 Transposon Mobility in the Caenorhabditis elegans Germline. Mol. Cell *31*, 79–90.

32

33 Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P.,
34 Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner.
35 Bioinformatics *29*, 15–21.

36

37 Dong, L., Cornaglia, M., Krishnamani, G., Zhang, J., Mouchiroud, L., Lehnert, T.,
38 Auwerx, J., and Gijs, M.A.M. (2018). Reversible and long-term immobilization in a
39 hydrogel-microbead matrix for high-resolution imaging of Caenorhabditis elegans
40 and other small organisms. PLoS One *13*, e0193989.

41

42 Egloff, S., O'Reilly, D., Chapman, R.D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D.,
43 and Murphy, S. (2007). Serine-7 of the RNA Polymerase II CTD Is Specifically
44 Required for snRNA Gene Expression. Science (80-.). *318*, 1777–1779.

45

46 Egloff, S., Szczepaniak, S.A., Dienstbier, M., Taylor, A., Knight, S., and Murphy, S.

1 (2010). The integrator complex recognizes a new double mark on the RNA
2 polymerase II carboxyl-terminal domain. J. Biol. Chem. *285*, 20564–20569.
3
4 Elrod, N.D., Henriques, T., Huang, K.-L., Tatomer, D.C., Wilusz, J.E., Wagner, E.J.,
5 and Adelman, K. (2019). The Integrator Complex Attenuates Promoter-Proximal
6 Transcription at Protein-Coding Genes. Mol. Cell *76*, 738-752.e7.
7
8 Frey, M.R., and Matera, A.G. (2001). RNA-mediated interaction of Cajal bodies and
9 U2 snRNA genes. J. Cell Biol. *154*, 499 LP – 510.
10
11 Gardini, A., Baillat, D., Cesaroni, M., Hu, D., Marinis, J.M., Wagner, E.J., Lazar,
12 M.A., Shilatifard, A., and Shiekhattar, R. (2014). Integrator Regulates Transcriptional
13 Initiation and Pause Release following Activation. Mol. Cell *56*, 128–139.
14
15 Gómez-Orte, E., Sáenz-Narciso, B., Zheleva, A., Ezcurra, B., de Toro, M., López, R.,
16 Gastaca, I., Nilsen, H., Sacristán, M.P., Schnabel, R., et al. (2019). Disruption of the
17 Caenorhabditis elegans Integrator complex triggers a non-conventional
18 transcriptional mechanism beyond snRNA genes. PLOS Genet. *15*, e1007981.
19
20 Gu, W., Lee, H.-C., Chaves, D., Youngman, E.M., Pazour, G.J., Conte, D., and
21 Mello, C.C. (2012). CapSeq and CIP-TAP identify Pol II start sites and reveal capped
22 small RNAs as C. elegans piRNA precursors. Cell *151*, 1488–1500.
23
24 Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T.,
25 Siomi, H., and Siomi, M.C. (2007). A Slicer-Mediated Mechanism for Repeat-
26 Associated siRNA 5' End Formation in Drosophila. Science (80-.). *315*, 1587 LP –
27 1590.
28
29 Haberle, V., Forrest, A.R.R., Hayashizaki, Y., Carninci, P., and Lenhard, B. (2015).
30 CAGEr: precise TSS data retrieval and high-resolution promoterome mining for
31 integrative analyses. Nucleic Acids Res. *43*, e51–e51.
32
33 Han, M., Wei, G., McManus, C.E., Hillier, L.W., and Reinke, V. (2019). Isolated C.
34 elegans germ nuclei exhibit distinct genomic profiles of histone modification and
35 gene expression. BMC Genomics *20*, 500.
36
37 Handler, D., Meixner, K., Pizka, M., Lauss, K., Schmied, C., Gruber, F.S., and
38 Brennecke, J. (2013). The genetic makeup of the Drosophila piRNA pathway. Mol.
39 Cell *50*, 762–777.
40
41 Jänes, J., Dong, Y., Schoof, M., Serizay, J., Appert, A., Cerrato, C., Woodbury, C.,
42 Chen, R., Gemma, C., Huang, N., et al. (2018). Chromatin accessibility dynamics
43 across C. elegans development and ageing. Elife *7*, e37344.
44
45 Kamath, R.S., and Ahringer, J. (2003). Genome-wide RNAi screening in
46 Caenorhabditis elegans. Methods *30*, 313–321.

Kasper, D.M., Wang, G., Gardner, K.E., Johnstone, T.G., and Reinke, V. (2014). The C.elegans SNAPc component SNPC-4 coats piRNA domains and is globally required for piRNA abundance. Dev. Cell.

Kozomara, A., and Griffiths-Jones, S. (2014). MiRBase: Annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. *42*, 68–73.

Lai, F., Gardini, A., Zhang, A., and Shiekhattar, R. (2015). Integrator mediates the biogenesis of enhancer RNAs. Nature *525*, 399.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Meth *9*, 357–359.

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup, 1000 Genome Project Data Processing (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Maxwell, C.S., Kruesi, W.S., Core, L.J., Kurhanewicz, N., Waters, C.T., Lewarch, C.L., Antoshechkin, I., Lis, J.T., Meyer, B.J., and Baugh, L.R. (2014). Pol II Docking and Pausing at Growth and Stress Genes in C. elegans. Cell Rep. *6*, 455–466.

Middleton, R., Gao, D., Thomas, A., Singh, B., Au, A., Wong, J.J.-L., Bomane, A., Cosson, B., Eyras, E., Rasko, J.E.J., et al. (2017). IRFinder: assessing the impact of intron retention on mammalian gene expression. Genome Biol. *18*, 51.

Nojima, T., Gomes, T., Carmo-Fonseca, M., and Proudfoot, N.J. (2016). Mammalian NET-seq analysis defines nascent RNA profiles and associated RNA processing genome-wide. Nat. Protoc. *11*, 413–428.

Nojima, T., Tellier, M., Foxwell, J., Ribeiro de Almeida, C., Tan-Wong, S.M., Dhir, S., Dujardin, G., Dhir, A., Murphy, S., and Proudfoot, N.J. (2018). Deregulated Expression of Mammalian lncRNA through Loss of SPT6 Induces R-Loop Formation, Replication Stress, and Cellular Senescence. Mol. Cell *72*, 970-984.e7.

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

Pritykin, Y., Brito, T., Schupbach, T., Singh, M., and Pane, A. (2017). Integrative analysis unveils new functions for the Drosophila Cutoff protein in noncoding RNA biogenesis and gene regulation. RNA *23*, 1097–1109.

1

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

4

5 Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne,
6 S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for
7 deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165.

8

9 Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., and
10 Bartel, D.P. (2006). Large-Scale Sequencing Reveals 21U-RNAs and Additional
11 MicroRNAs and Endogenous siRNAs in C. elegans. Cell *127*, 1193–1207.

12

13 Sigurdsson, S., Dirac-Svejstrup, A.B., and Svejstrup, J.Q. (2010). Evidence that
14 transcript cleavage is essential for RNA polymerase II transcription and cell viability.
15 Mol. Cell *38*, 202–210.

16

17 Stadelmayer, B., Micas, G., Gamot, A., Martin, P., Malirat, N., Koval, S., Raffel, R.,
18 Sobhian, B., Severac, D., Rialle, S., et al. (2014). Integrator complex regulates
19 NELF-mediated RNA polymerase II pause/release and processivity at coding genes.
20 Nat. Commun. *5*, 5531.

21

22 Wang, Q., Sawyer, I.A., Sung, M.-H., Sturgill, D., Shevtsov, S.P., Pegoraro, G.,
23 Hakim, O., Baek, S., Hager, G.L., and Dundr, M. (2016). Cajal bodies are linked to
24 genome conformation. Nat. Commun. *7*, 10966.

25

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

30

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

34

35 Zeng, C., Weng, C., Wang, X., Yan, Y.-H., Li, W.-J., Xu, D., Hong, M., Liao, S.,
36 Dong, M.-Q., Feng, X., et al. (2019). Functional Proteomics Identifies a PICS
37 Complex Required for piRNA Maturation and Chromosome Segregation. Cell Rep.
38 *27*, 3561-3572.e3.

39

40

41

42

43

44

45

Figure legends

Figure 1. Integrator is required for motif-dependent piRNA precursor abundance and their short length.

A. Distributions of $\log_2$ fold change in piRNA abundance in *ints-11* RNAi-treated nematodes compared to empty vector-treated nematodes.

B. Distributions of $\log_2$ fold change in piRNA precursor abundance in *ints-11* RNAi-treated nematodes compared to empty vector-treated nematodes.

C-D. Distributions of piRNA precursor length in *ints-11* RNAi-treated nematodes compared to empty vector-treated nematodes, obtained from deeply sequencing one pair of libraries (inserts sequenced up to 75nt). Distributions of total sequences (C) and total read counts (D) are shown.

Figure 2. Integrator localizes to sites of piRNA biogenesis in germ cells.

A. Live imaging of a rescue transgene expressing an eGFP-tagged version of Integrator subunit 6 (INTS-6) in the germline of *C. elegans.* M=mitotic region, TZ=transition zone.

B. Co-localization between INTS-6::eGFP and mCherry::PRDE-1 in germ cells. PRDE-1 and INTS-6 co-localize in early stages of meiosis (1,2) and up to mid-pachytene (3), where an additional INTS-6 focus appears (3). When PRDE-1 foci are about to disappear, co-localization is lost (4), and the additional INTS-6 focus remains present well after the loss of PRDE-1 foci (5).

Figure 3. Chromatin fractionation identifies nascent piRNA precursors.

A. Experimental design. Germ nuclei isolated from empty vector or *ints-11* RNAi-treated nematodes were fractionated into chromatin-bound and nucleoplasmic fractions. RNA corresponding to each fraction was purified and used as input for short capped RNA libraries.

B. Length distribution of chromatin-bound and nucleoplasmic piRNA precursors in wild type N2 nematodes*.*

C. Length distribution of piRNA precursors generated from loci with weak and strong AT content enrichment downstream of the 21U-RNA.

D. Distributions of log2 fold change in piRNA precursor abundance in nucleoplasmic and chromatin fractions relative to N2 empty vector controls, for N2 *ints*-11 RNAi, *tfiis* empty vector, and *tfiis ints*-11 RNAi nematodes. Significance for each distribution to be different from no change was tested using a two-tailed paired Wilcoxon rank sum test, with all p-values<2.2e-16.

1   E. Length distributions of nucleoplasmic and chromatin-bound piRNA precursors in

2   N2 empty vector, N2 *ints-11* RNAi, *tfiis* empty vector, and *tfiis ints-11* RNAi

3   nematodes. Length shifts in nascent RNA in *tfiis* mutants are indicated with dashed

4   vertical lines and arrows.

5   F-G. Distribution of the position of the ~28 nt (F) and ~48 nt (G) nascent RNA peaks

6   in 2000 bootstrapped subsamples of 3000 precursors. Sampling probability was

7   weighted by precursor abundance.

8

9   **Figure 4. Integrator-mediated cleavage of nascent piRNA precursors generates**

10   **short capped RNAs.**

11

12   A. Model for Integrator cleavage of nascent piRNA precursors associated with

13   promoter-proximal Pol II. Cleavage results in the production of a short capped RNA,

14   and a 3' degradation fragment.

15   B. Signal of 5' P small RNA 5' ends mapping to piRNA loci as a function of the

16   distance to piRNA TSSs, after removal of reads corresponding to mature 21U-RNAs

17   (see Methods). The signal is normalized to counts per million of non-structural

18   mapped reads. A peak of 5' P ends centered at +38 from piRNA TSSs is observed.

19   C. Distributions of counts per million of 5' P ends mapping from +25 to +50 of piRNA

20   TSSs across the different genotypes and conditions. Two independent biological

21   replicates are shown.

22   D. Length distribution of putative 3' degradation fragments with 5' P ends mapping

23   at +28-+58 of the 5' U of annotated 21U-RNAs.

24   E. Number of loci with detected cleavage fragments in N2 empty vector nematodes

25   across percentile bins of increasing termination signal strength (top panel).

26   Distributions of fragment read counts per locus for detected loci across percentile

27   bins of increasing termination signal strength (bottom panel).

28

29   **Figure 5. Integrator knockdown results in transcriptional readthrough in a**

30   **subset of motif-dependent piRNA loci.**

31

32   A. Number of unique CAGE-supported TSSs mapping to a 10 nucleotide region

33   upstream of motif-dependent 21U-RNAs. A clear enrichment of TSS 2 nt upstream of

34   21U-RNAs is observed.

35   B. Distributions of $\log_2$ fold change in CAGE signal corresponding to 2 nt upstream

36   motif-dependent piRNAs upon *ints*-11 RNAi knockdown in an N2 and a *tfiis* mutant

37   background. N2 empty vector was used as a baseline for fold change calculation.

38   C. Heatmap showing changes in CAGE signal upstream of motif-dependent 21U-

39   RNAs upon *ints*-11 knockdown in an N2 and a *tfiis* mutant background. The total

1   cpms corresponding to each nucleotide position upstream to 21U-RNAs are shown

2   as a barplot (top part).

3   D. Length distributions of CAGE fragments initiating 2 nt upstream of motif-

4   dependent 21U-RNAs. The total counts in each length bin are normalized to counts

5   per million of mapped reads.

6   E. Average normalized CAGE signal (transcripts per million) 2 nt upstream of motif-

7   dependent and motif-independent piRNAs (excluding undetected loci).

8   F. Distributions of $\log_2$ fold change in CAGE signal corresponding to 2 nt upstream

9   motif-independent piRNAs upon *ints*-11 RNAi knockdown in an N2 and a *tfiis* mutant

10   background. N2 empty vector was used as a baseline for fold change calculation.

11

12   **Figure 6. Model for the role of Integrator in piRNA transcription termination.**

13

14   At snRNA loci, Integrator generates snRNA 3′ ends leading to Pol II termination.

15   *ints*-11 knockdown results in readthrough past termination sites and polyadenylation

16   of snRNA transcripts.

17   At piRNA loci, Pol II pauses ~28 nt and ~48 nt downstream of piRNA TSSs. ~28 nt

18   short capped RNAs may be generated via direct dissociation of Pol II from chromatin

19   at the ~28 nt pause site. When Pol II transcribes past the ~48 nt pause site,

20   Integrator acts to cleave nascent precursors approximately when Pol II is at +58 from

21   piRNA TSSs, generating ~38 nt short capped RNAs and ~20 nt degradation

22   fragments. *ints*-11 knockdown does not lead to Pol II readthrough, suggesting that

23   additional mechanisms prevent Pol II elongation at piRNA loci.

24

25   **Supplementary Figure legends**

26

27   **Supplementary Figure 1. Quantification of total mature piRNA and piRNA**

28   **precursor levels after RNAi knockdown of *ints*-11 (relates to Figure 1).**

29

30   A-B. Total piRNA read counts normalized using DEseq2 size factors derived from

31   miRNA counts (A) or normalized to the total number of non-structural mapped reads

32   (B). Normalized abundance for each pair of EV and *ints*-11 RNAi samples is shown

33   relative to EV.

34   C. Average mature piRNA abundance relative to EV in EV and *ints*-11 RNAi samples

35   normalized using miRNA-derived size factors. Error bars represent the standard

36   error of the mean.

37   D. Representative examples of a silenced and a desilenced piRNA sensor transgene

38   in the *C. elegans* germline. The proportion of animals desilencing the piRNA sensor

39   upon empty vector and *ints*-11 RNAi treatment is shown in the right panel.

F. Total piRNA precursor counts normalized to the total number of short capped RNA reads mapping to WormBase TSSs (Chen et al., 2013). Normalized abundance for each pair of EV and *ints*-11 RNAi samples is shown relative to EV.

G. Average normalized piRNA precursor abundance relative to EV, in EV and *ints*-11 RNAi samples. Error bars represent the standard error of the mean.

H. Distributions of piRNA precursor length in *ints*-11 RNAi-treated nematodes compared to empty vector-treated nematodes (inserts sequenced up to 50nt).

Supplementary Figure 2. Integrator localizes to sites of piRNA biogenesis in germ cells (relates to Figure 2).

A. Quantification of the number of INTS-6::GFP and mCherry::PRDE-1 foci/cell and their co-localization across the *C. elegans* germline. 1-5 correspond to the germline sections shown in Figure 2B.

B. Co-localization of FLAG::INTS-11 and mCherry::PRDE-1 in the *C. elegans* germline observed by anti-FLAG and anti-mCherry immunofluorescent staining.

Supplementary Figure 3. piRNA precursor length distribution of individual replicates (relates to Figure 3).

A. Boxplots comparing the length of pairs of nucleoplasmic and chromatin-libraries corresponding to the same initial nuclei sample.

B. Length distribution of chromatin-bound and nucleoplasmic piRNA precursors in the two individual replicates of corresponding to each condition.

Supplementary Figure 4. Further analysis of cleavage fragment abundance (relates to Figure 4).

A. Heatmap showing the coverage of unique sequences mapping to piRNA loci after removal of reads corresponding to mature 21U-RNAs (see Methods). The first peak corresponds to 21U-RNA degradation products and 5′ PPP initiation products starting at piRNA TSSs. The second peak corresponds to degradation fragments initiating at +38 on average (Figure 4D), resulting in sequence coverage peaking at +48 due to a median length of fragments of 20 nt (Figure 4D).

B. $\log_2$ fold change distributions of degradation fragment counts per million of mapped reads relative to N2 EV. Fold changes were calculated after addition of 0.01 pseudocpms. Two independent biological replicates are shown.

C. Positions of 3′ ends of 5′ P fragments relative to piRNA TSSs. Signal is normalized to counts per million mapped reads.

1 D. Number of loci with detectable cleavage fragments across bins of loci stratified
2 according to the strength of their termination signals, from weak (GC-rich) to strong
3 (AT-rich). Data from each individual library is shown.
4
5 **Supplementary Figure 5. Integrator and the strength of AT-rich termination**
6 **signals (relates to Figure 4).**
7 Length distribution of chromatin-bound motif-dependent piRNA precursors stratified
8 by the strength of downstream termination signals from weak (GC-rich) to strong
9 (AT-rich), in empty vector and *ints*-11 RNAi-treated nematodes. Two replicates for
10 each condition are shown.
11
12 **Supplementary Figure 6. Nascent SLIC-CAGE detects readthrough**
13 **transcription from piRNA loci (relates to Figure 5).**
14 A. Clustered correlation matrix of promoter-mapping CAGE tpms for the top 1000
15 genes with the highest CAGE signal.
16 B. Intron retention ratio distributions across nascent RNA samples, and in a total
17 RNA control, estimated through IRFinder analysis.
18 C-D. Total raw CAGE tags mapping between -10 and -1 nucleotide positions
19 upstream of motif-dependent piRNAs (C) and motif-independent piRNAs (D).
20 E. Total unique CAGE-detected TSSs mapping between -10 and -1 nucleotide
21 positions upstream of motif-independent piRNAs.
22 F. Mean normalized expression levels of CAGE-detected and non-detected motif-
23 dependent piRNAs at the short-capped RNA level.
24 G. Downsampling analysis of CAGE libraries showing that the detection of piRNA
25 loci is saturated with sequencing depth. Total detected loci in each of the two
26 replicates were averaged for each bootstrapped sample.
27
28 **Supplementary Figure 7. Accessibility and H3K27 acetylation levels at piRNA**
29 **promoters (relates to Figure 5).**
30 A. ATAC-seq read coverage around motif-dependent and motif-independent piRNA
31 promoters, in wild-type young adult wild-type worms, and in day 1 adult germlineless
32 glp-1 mutant worms lacking a germline. 'Motif-dependent CAGE' refers to motif-
33 dependent loci detected by CAGE.
34 B. Log$_2$ ratio in accessibility signal at motif-dependent piRNA loci between young
35 adult wild-type worms and day 1 adult germlineless glp-1 mutants lacking a germline.
36 This reflects a specific increase in accessibility at motif-dependent loci in germ cells.
37 'Motif-dependent CAGE' refers to motif-dependent loci detected by CAGE.
38 C. H3K27ac ChIP signal minus input around motif-dependent and motif-independent
39 piRNA loci in isolated germ nuclei (left panel), and in isolated somatic nuclei (right
40 panel).

D-E. Distributions of H3K27ac (D) and accessibility signal (E) in motif-dependent and motif-independent piRNA loci.

Supplementary Figure 8. Transcriptional changes in snRNAs and protein-coding genes upon *ints*-11 knockdown (relates to Figure 5).

A. Nascent CAGE signal at snRNA genes upon *ints*-11 knockdown, in a wild-type and a *tfiis* mutant background. Extensive readthrough transcription with ~500nt fragments is readily detectable upon *ints*-11 knockdown.

B. Distributions of $\log_2$ fold changes in nascent CAGE signal at snRNA loci upon *ints*-11 knockdown, in a wild-type and a *tfiis* mutant background, using N2 empty vector as a baseline.

C. MAplot showing differentially expressed protein-coding genes at the nascent short capped RNA level, with an excess of upregulated genes.

D. MAplot showing differentially expressed protein-coding genes detected by nascent CAGE, with an excess of upregulated genes.

E. Overlap between upregulated genes in the nascent scRNA and CAGE datasets.

F. Nascent CAGE signal at the set of upregulated protein coding genes upon *ints*-11 knockdown, in a wild-type and a *tfiis* mutant background.

G. Correlation between changes in nascent short-capped RNA and CAGE signal, for genes detected as differentially expressed in at least one of the two assays.

H. $\log_2$ odds ratios of enrichment of motif-independent piRNA loci in protein-coding gene promoters detected as differentially expressed upon *ints*-11 knockdown, at the nascent scRNA level, at the CAGE level, or both (see labels).

Supplementary tables

Table S1. *C. elegans* strains

Strain	Genotype	Description	Reference
SX1316	mjIs144 II	piRNA sensor strain	(Bagijn et al., 2012b)
JCP383	ints-6 (tm1615) IV; jcpSi10 [pJC51 (ints-6p::ints-6::3xFLAG::eGFP::ints-6UTR, unc-119(+))] II	INTS-6::eGFP rescue strain	(Gómez-Orte et al., 2019)

SX2700	mjSi74 I; mjIs144 II; prde-1(mj207)	mCherry::PRDE-1 rescue strain	(Weick et al., 2014)
TEE172	ints-6 (tm1615) IV; jcpSi10 [pJC51 (ints-6p::ints-6::3xFLAG::eGFP::ints-6UTR, unc-119(+))] II; mjSi74 I	Generated by crossing JCP383 and SX2700	This study

1

2 **Table S2. Oligonucleotides**

3

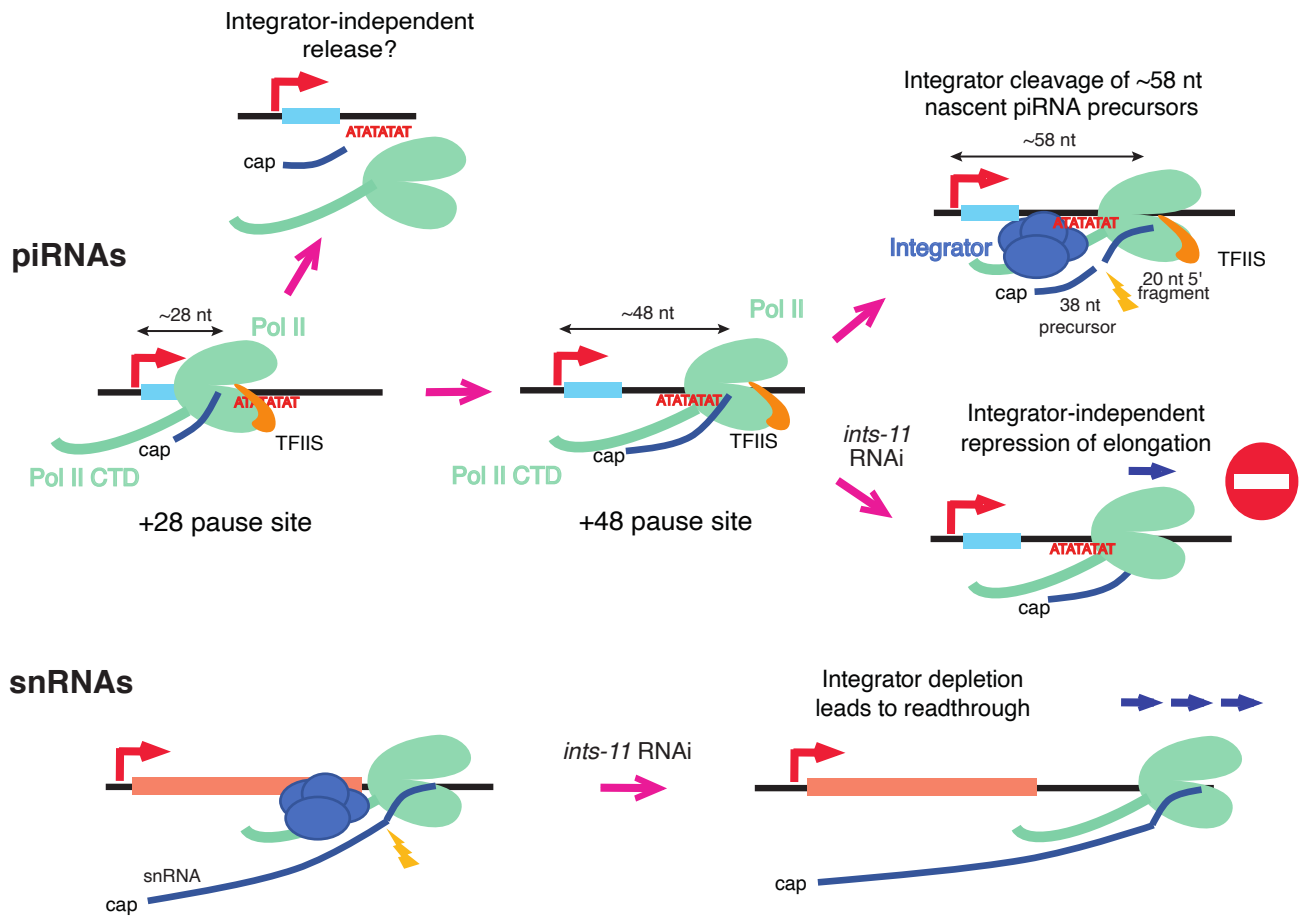| Ints-11_Nterm_F	AAAATCCTATTTCGCGAAGGC	Genotype FLAG::ints-11
Ints-11_Nterm_R	CGACCGTCCAACATCTTGTC	Genotype FLAG::ints-11
Ints-11_Nterm_crRNA	CTTGATTTCAGGAATTTTGT	crRNA
FLAG::Ints-11 repair oligo	GTAGAACTTTATAGCTATATTTCACAATGCCCGACGATTACAAGGATGACGACGATAAGAAGATACCTGAAATCAAGGTGAGTTCTGAAGGACTTTTCTG	repair template
Ints-6_3UTR_F	ACAACTCACTGAACGTCTGC	Genotype jcpSi10
Ints-6_3UTR_R	ACAAAAGTTTTGAAAGAGCACCT	Genotype jcpSi10
Ints-6_tm1616_F	AGGTGGACGAAGCTTCTCAA	Genotype *tm1616*
Ints-6_tm1616_R	GCATTGGGAAATTGAATGGCA	Genotype *tm1616*
Wormcherry_F	GGAACACAAACCGCAAAACT	Genotype mjSi74 I
Wormcherry_R	GCACCGTCTTCAGGGTACAT	Genotype mjSi74 I
MosSCI Chr I F	GAAAAACCCCGAATTTTGGT	Genotype mjSi74 I
MosSCI Chr I R	GGTGGGAACCTTTTCGTTTT	Genotype mjSi74 I

4

Figure 1

A

B
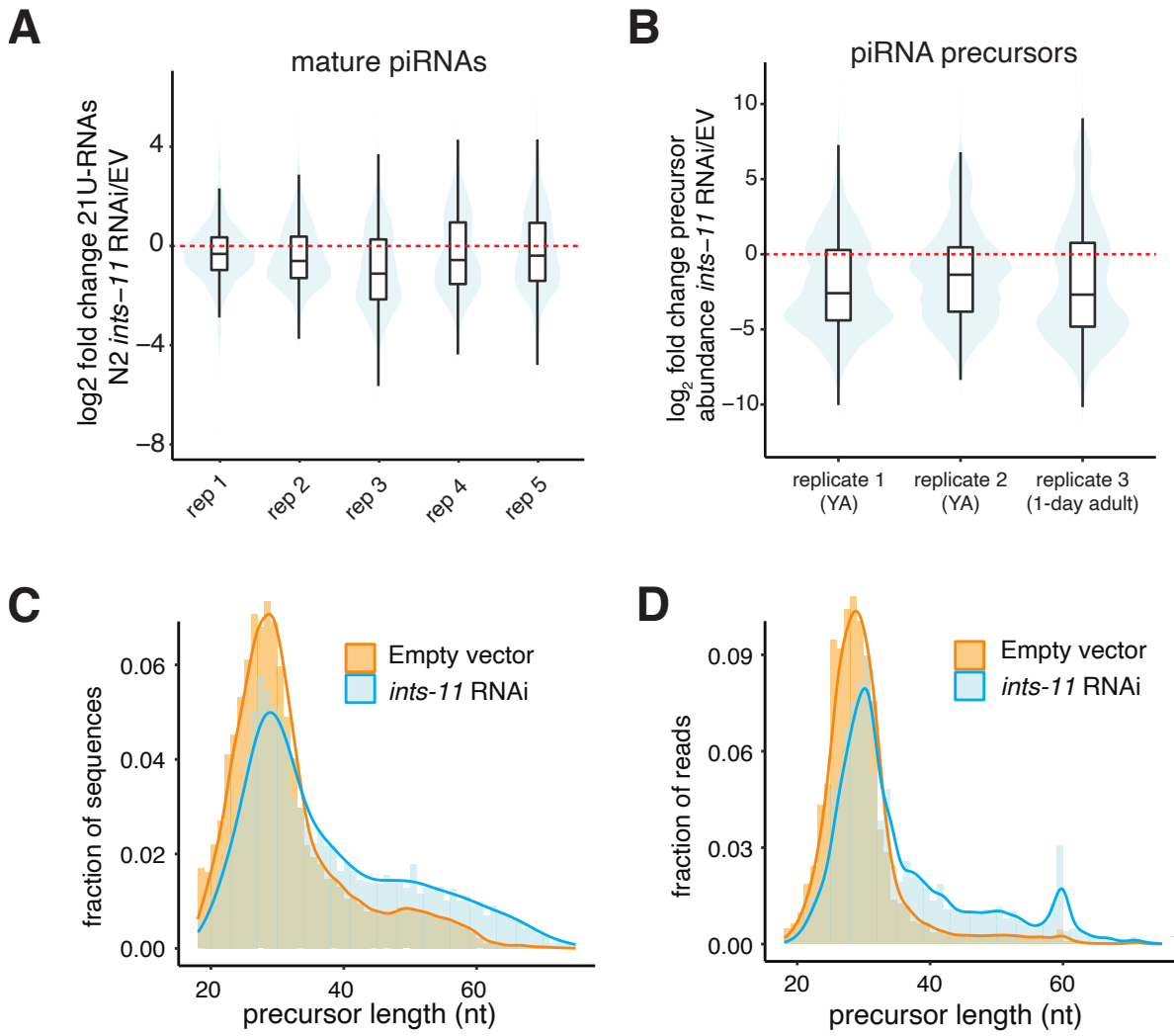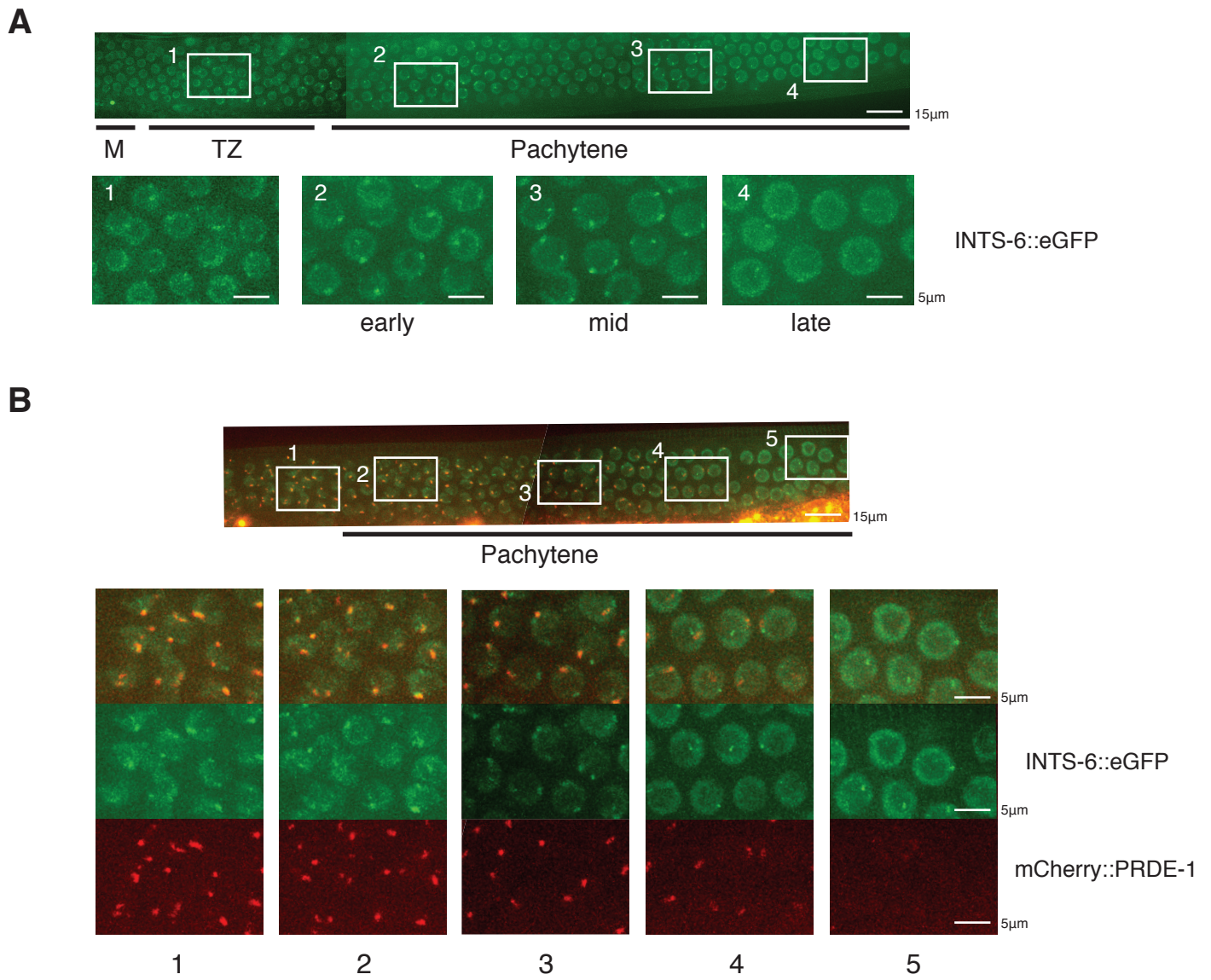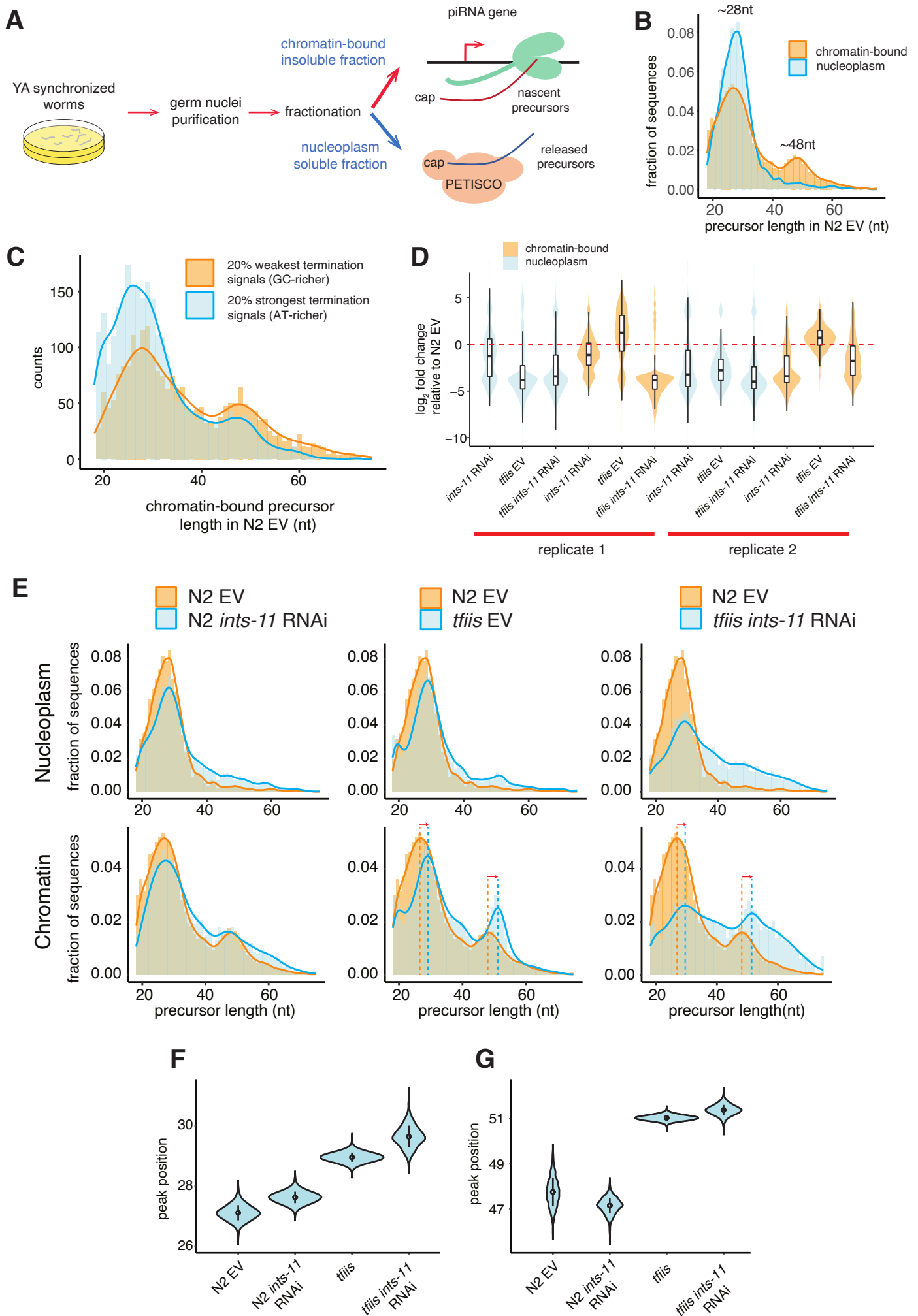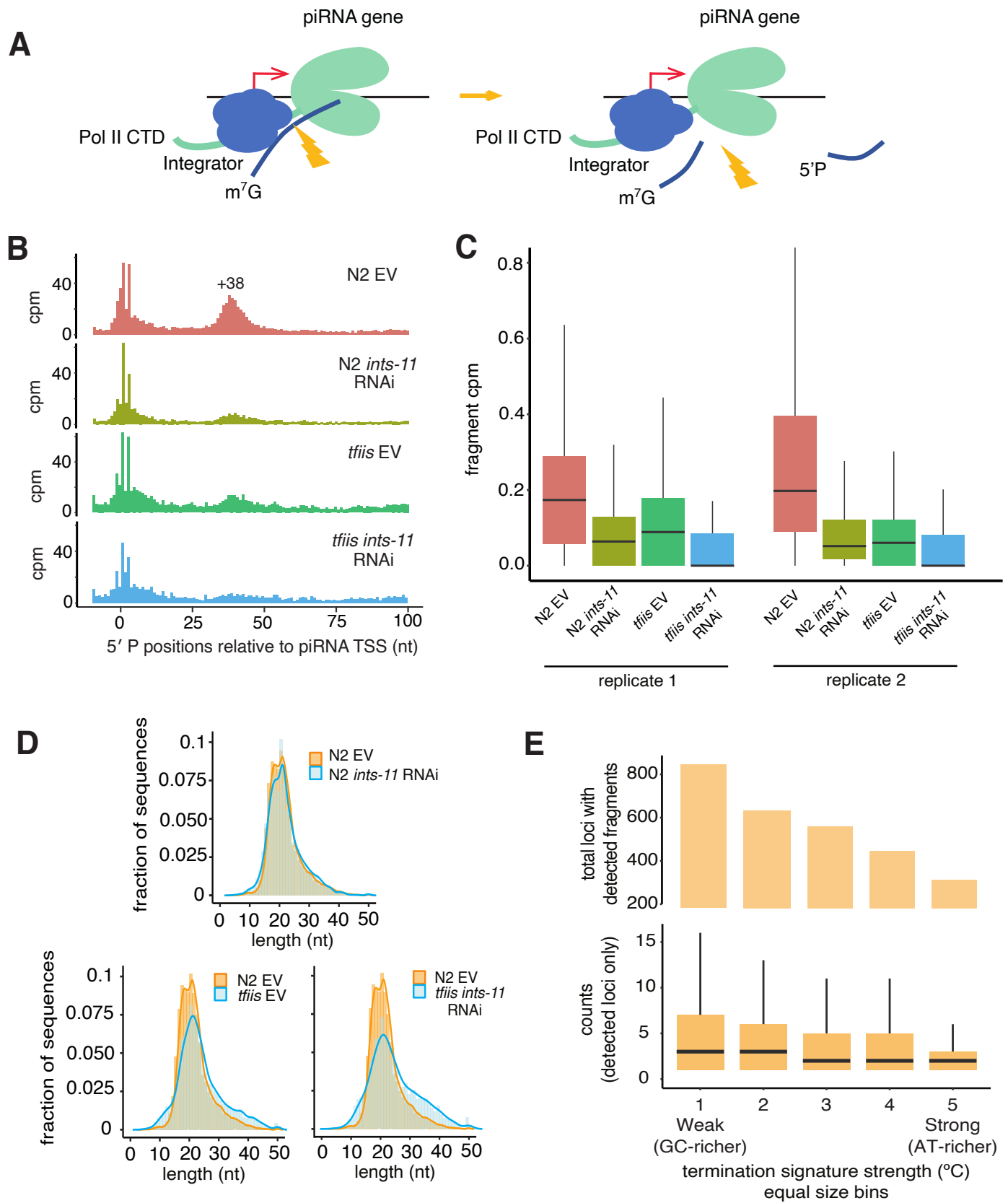
Figure 2

Figure 3

Figure 4

Figure 5

Figure 6