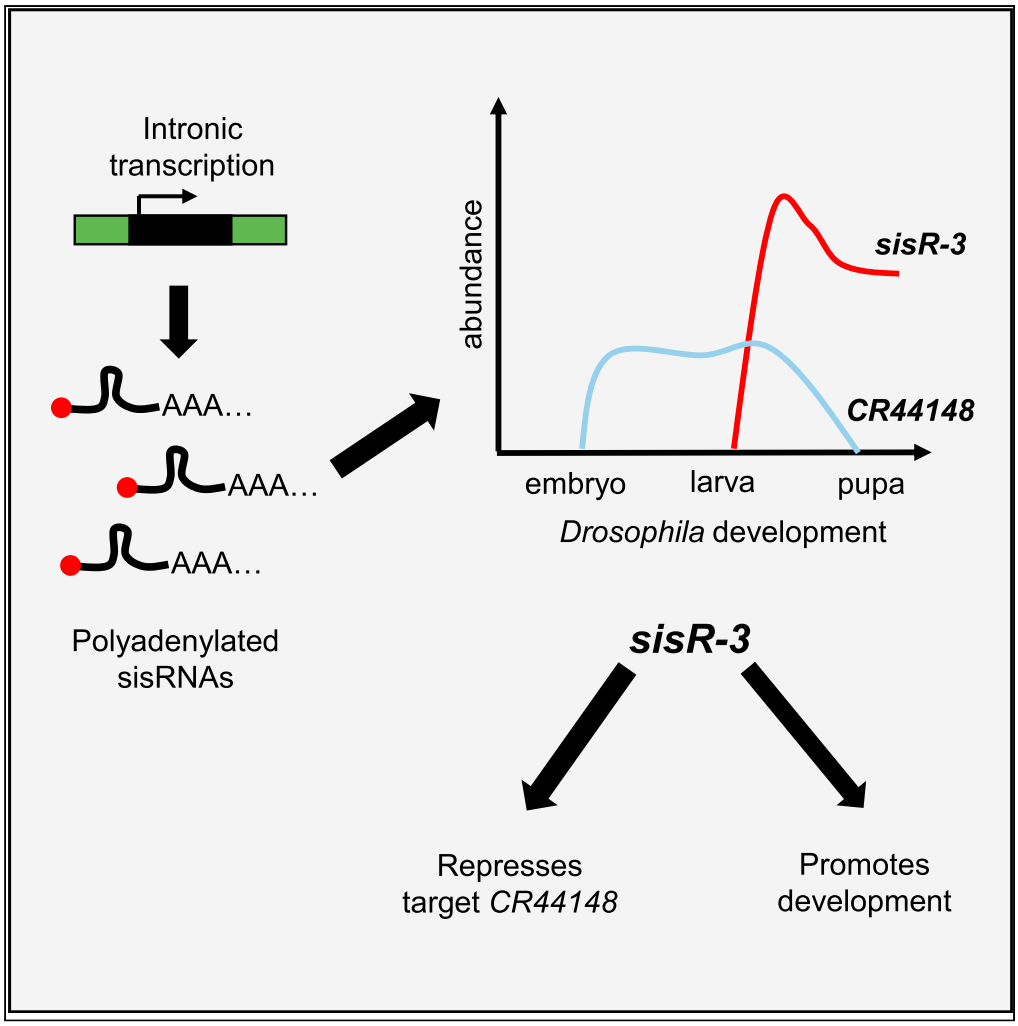
iScience

CellPress

Article

Generation of *Drosophila* sisRNAs by Independent Transcription from Cognate Introns



Sharon Si Jia Ng, Ruther Teo Zheng, Ismail Osman, Jun Wei Pek

junwei@tll.org.sg

HIGHLIGHTS

Identification of polyadenylated sisRNAs

sisRNAs can be produced from independent transcription

sisR-3 regulates a long noncoding RNA

sisR-3 is required during development

Jia Ng et al., iScience 4, 68–75
June 29, 2018
https://doi.org/10.1016/j.isci.2018.05.010

iScience

CellPress

Article

Generation of *Drosophila* sisRNAs by Independent Transcription from Cognate Introns

Sharon Si Jia Ng,[2,5] Ruther Teo Zheng,[3,5] Ismail Osman,[1,4] and Jun Wei Pek[1,6,*]

SUMMARY

Although stable intronic sequence RNAs (sisRNAs) are conserved in plants and animals, their functional significance is still unclear. We identify a pool of polyadenylated maternally deposited sisRNAs in *Drosophila melanogaster*. These sisRNAs can be generated by independent transcription from the cognate introns. The ovary-specific poly(A) polymerase Wispy mediates the polyadenylation of maternal sisRNAs and confers their stability as maternal transcripts. A developmentally regulated sisRNA *sisR-3* represses the expression of a long noncoding RNA *CR44148* and is required during development. Our results expand the pool of sisRNAs and suggest that sisRNAs perform regulatory functions during development in *Drosophila*.

INTRODUCTION

Noncoding RNAs (ncRNAs) have a profound impact on gene expression (Cech and Steitz, 2014; Ghildiyal and Zamore, 2009; Kung et al., 2013; Lee, 2012; Matera et al., 2007; Rinn and Chang, 2012). The identification of stable intronic sequence RNAs (sisRNAs) in the oocyte nucleus of *Xenopus tropicalis* raised the important question on the biological functions of this class of ncRNAs (Gardner et al., 2012). Although sisRNAs had been discovered in humans, mice, *Xenopus*, *Drosophila*, yeasts, and viruses, very little is known about their importance during development (Gardner et al., 2012; Lu et al., 2015; Moss and Steitz, 2013; Osman et al., 2016; Pek, 2018; Pek and Okamura, 2015; Pek et al., 2015; Talhouarne and Gall, 2014; Yin et al., 2012; Zhang et al., 2013; Zheng et al., 2015). *Drosophila sisR-1* has been shown to repress a long ncRNA *ASTR* during embryonic development and regulate stem cell homeostasis (Pek et al., 2015; Wong et al., 2017). In addition, maternally deposited circular *sisR-4* is important for embryonic development by promoting its parental gene transcription in *Drosophila* (Tay and Pek, 2017).

In our continuing effort to identify and characterize more sisRNAs, we performed deep sequencing of polyadenylated and non-polyadenylated maternally deposited sisRNAs in the unfertilized eggs of *Drosophila*. Unexpectedly, we found abundant polyadenylated sisRNAs that are produced by independent transcription from the introns, suggesting an alternative pathway for sisRNA biogenesis. Further analyses on a sisRNA *sisR-3* revealed that *sisR-3* represses a long ncRNA *CR44148* and is essential for proper development.

RESULTS

Deep Sequencing Identifies Polyadenylated sisRNAs

In *Xenopus* and *Drosophila*, maternally deposited sisRNAs are thought to be by-products of splicing (Gardner et al., 2012; Osman et al., 2016; Pek, 2018; Pek et al., 2015; Talhouarne and Gall, 2014; Tay and Pek, 2017). It is unknown how linear sisRNAs are conferred with unusual stability in the oocytes (Gardner et al., 2012; Osman et al., 2016; Pek, 2018; Pek et al., 2015). To identify more sisRNAs, we examined *Drosophila melanogaster* unfertilized eggs, which contain a maternal pool of stable and mature RNAs with no contamination from zygotic transcription (Pek et al., 2015). We performed strand-specific deep sequencing of (1) ribosomal RNA (rRNA)-depleted total RNA, (2) poly(A)+ RNA, and (3) rRNA-depleted poly(A)− RNA (Figure 1A). As a positive control for poly(A)− RNA, we detected the U85 small Cajal body-specific RNA (scaRNA) in the poly(A)− fraction but not in the poly(A)+ fraction (Figure S1A). Conversely, we detected exonic sequences from the messenger RNAs (mRNAs) in the poly(A)+ fraction but not in the poly(A)− sequences (Figures 1B, S1B, and S1C). These observations confirmed the validity of our poly(A)+ and poly(A)− sequencing experiments.

[1]Temasek Life Sciences Laboratory, National University of Singapore, 1 Research Link, Singapore 117604, Singapore

[2]Temasek Polytechnic, 21 Tampines Avenue 1, Singapore 529757, Singapore

[3]Ngee Ann Polytechnic, 535 Clementi Road, Singapore 599489, Singapore

[4]Department of Biological Sciences, National University of Singapore 117543, Singapore, Singapore

[5]These authors contributed equally

[6]Lead Contact

*Correspondence: junwei@tll.org.sg

https://doi.org/10.1016/j.isci.2018.05.010

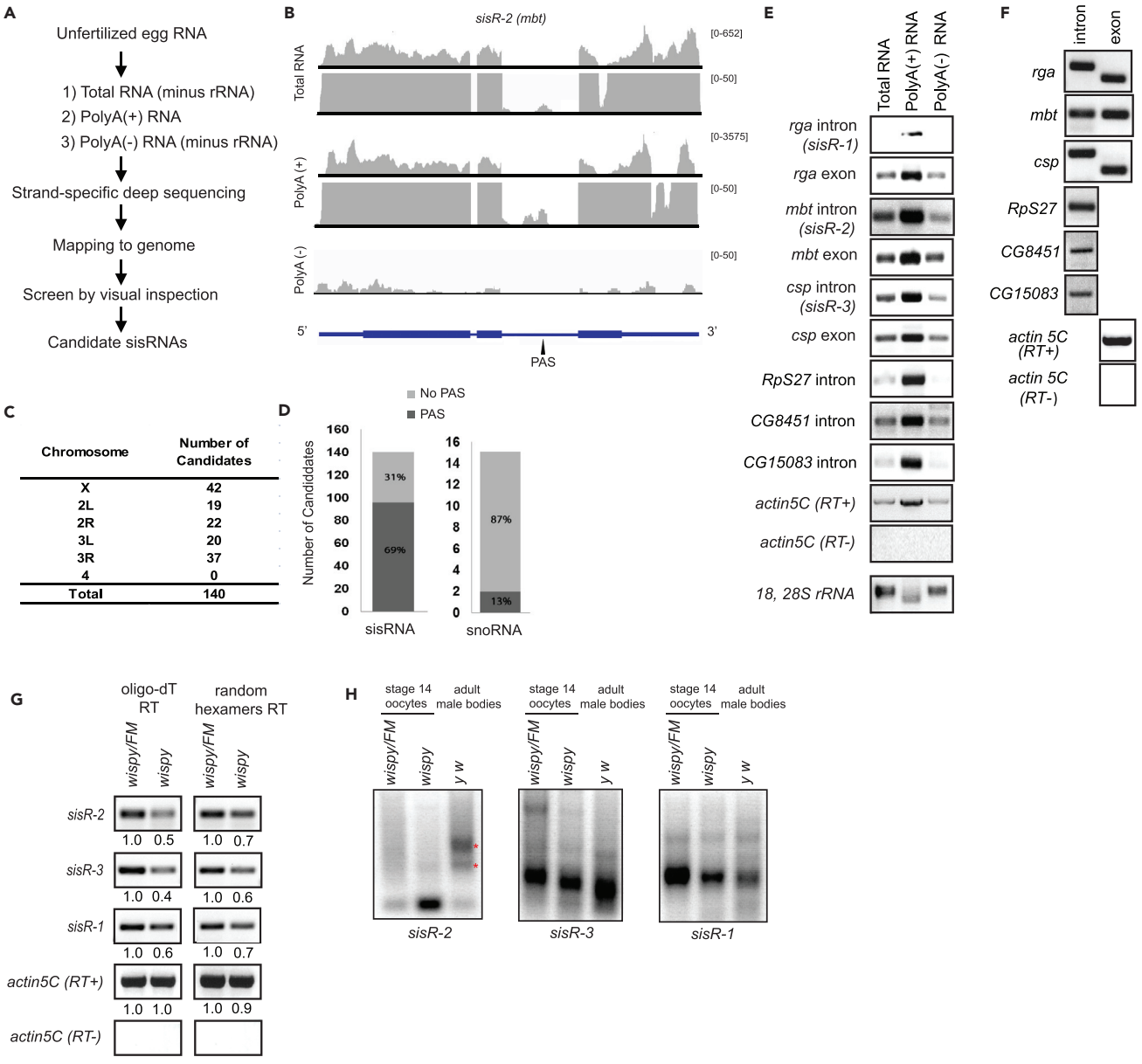


Figure 1. Deep Sequencing Identifies a Pool of Polyadenylated sisRNAs in Unfertilized Eggs

(A) A schematic for identifying candidate sisRNAs in unfertilized eggs.

(B) Genome browser views of *sisR-2 (mbt)* gene locus. RNA sequencing results for total RNA, poly(A)+ RNA, and poly(A)− RNA from unfertilized eggs are shown. PAS, polyadenylation signal near the predicted 3′ ends of the sisRNAs.

(C) Table showing the number of candidate polyadenylated sisRNAs distributed over the chromosomes in *Drosophila*.

(D) Charts showing the numbers and percentages of candidate sisRNAs and snoRNAs with and without PAS.

(E) RT-PCR showing the presence of various intronic and exonic sequences in total, poly(A)+, and poly(A)− RNAs from unfertilized eggs.

(F) RT-PCR showing the presence of intronic and exonic sequences reverse transcribed by oligo-dT in unfertilized eggs.

(G) Semi-quantitative RT-PCR showing the abundance of *sisR-1*, *sisR-2*, and *sisR-3* in stage 14 oocytes of *wispy/FM* versus *wispy* homozygous mutants using oligo-dT or random hexamers during reverse transcription. Numbers below indicate the relative band intensities of respective sisRNAs quantified using ImageJ software.

(H) RACE-PAT assay showing the relative lengths of sisRNA poly(A) tails in *wispy/FM*, *wispy* homozygous mutant stage 14 oocytes, and wild-type adult male body long poly(A) tails of sisR-2, red asterisks.

See also Figure S1 and Table S1.

We previously identified three sisRNAs, *sisR-1*, *sisR-2*, and *sisR-3*, from the *regena* (*rga*), *mushroom body tiny* (*mbt*), and *cysteine string protein* (*csp*) loci, respectively, by northern blotting (Pek et al., 2015). Unexpectedly, we found that these three sisRNAs were present in the poly(A)+ fraction instead of the poly(A)− fraction (Figures 1B, S1B, and S1C), suggesting that some sisRNAs may be polyadenylated. Cleavage and polyadenylation requires the polyadenylation signal (PAS) (AUUAAA, AAUAAA, AUAAAA), which functions to recruit the cleavage and polyadenylation specificity factor (CPSF) to mediate the addition of poly(A) tails (Shi and Manley, 2015). PAS sequences were present upstream of the predicted 3′ ends of these sisRNAs (Figures 1B, S1B, and S1C). By visual inspection of the sequencing data on the genome browser, we identified a total of 140 candidate polyadenylated sisRNAs that mapped to chromosomes 1–3 (Figure 1C and Table S1). Of the 140 candidate sisRNAs, 96 (~69%) have at least a PAS sequence, compared with 13% of the non-polyadenylated small nucleolar RNAs (snoRNAs) (Figure 1D and Table S1).

To verify whether sisRNAs are polyadenylated, we obtained RNA from unfertilized eggs and isolated poly(A)+ and poly(A)− fractions using oligo-dT beads and performed reverse-transcriptase polymerase chain reaction (RT-PCR). Consistent with our deep sequencing data, we detected an enrichment of sisRNAs in the poly(A)+ fraction compared with the poly(A)− fraction (Figure 1E). We finally confirmed the presence of polyadenylated sisRNAs in unfertilized eggs by performing RT-PCR using oligo-dT as the primer for reverse transcription (Figure 1F).

In *Drosophila*, the ovary-specific poly(A) polymerase (PAP) encoded by the *wispy* locus has been shown to polyadenylate maternal mRNAs and adenylate maternal microRNAs (miRNAs) in the cytoplasm during late oogenesis (Benoit et al., 2008; Lee et al., 2014). We asked whether *wispy* is required for polyadenylation and stability of maternal sisRNAs in the oocytes. Females homozygous for $wispy^{12\text{-}3147}$ laid very few eggs, so we examined RNAs from the stage 14 oocytes, which also store mature maternal RNAs. By performing reverse transcription using oligo-dT followed by PCR, we found down-regulation of poly(A)-tail-containing sisRNAs in *wispy* mutants compared with controls (Figure 1G), suggesting that *wispy* is required for the polyadenylation of maternal sisRNAs. We next examined the entire population of sisRNAs by doing reverse transcription using random hexamers followed by PCR. We also observed a decrease in sisRNA abundance in *wispy* mutants (Figure 1G), indicating that polyadenylation promotes the stability of maternal sisRNAs.

We further examined the lengths of poly(A) tails of *sisR-1*, *sisR-2*, and *sisR-3* by performing Rapid amplification of cDNA ends-PCR poly(A) test (RACE-PAT) assay. We examined RNA from stage 14 oocytes in *wispy/FM* and *wispy* mutants. In addition, RNA from adult male bodies (which do not express Wispy) was also included as a comparison for polyadenylation in somatic cells. In control stage 14 oocytes, the poly(A) tails of *sisR-2* were long and heterogeneous, but they were dramatically short in *wispy* mutants (Figure 1H). Interestingly, in adult male bodies, the poly(A) tails of *sisR-2* were long (Figure 1H, red asterisks), suggesting *wispy*-independent polyadenylation in somatic cells. For *sisR-3* and *sisR-1*, the lengths of poly(A) tails were also generally shorter in *wispy* mutants (Figure 1H). Unlike *sisR-2*, the poly(A) tails of these two sisRNAs remained short in adult male bodies (Figure 1H). Taken together, our data suggest that *wispy* is required for the cytoplasmic polyadenylation of maternal sisRNAs to confer their stability. We do not exclude the possibility that sisRNAs are also polyadenylated by a nuclear PAP (Juge et al., 2002).

Production of sisRNAs by Independent Transcription

To examine the nature of these sisRNAs in detail, in addition to previously cloned *sisR-1*, we cloned two previously identified sisRNAs *sisR-2* and *sisR-3* from the *mbt* and *csp* loci, respectively (Pek et al., 2015). By performing 5′ and 3′ RACE analyses, we obtained full-length sequences of *sisR-2* and *sisR-3* in the unfertilized eggs. As predicted from the deep sequencing data, the PAS sequences are near the 3′ ends of all the three sisRNAs (Figures 1B, 2A, S1B, S1C, and S2, data not shown). Furthermore, the 5′ ends of the sisRNAs lie close to the 5′ splice sites of the introns (Figures 2A and S2, data not shown). These observations prompted us to examine the possibility that sisRNAs may be transcribed independently from the cognate introns. We first tested if sisRNAs possess an m^7G cap, a 5′ end modification common to RNA polymerase II transcripts. Using an antibody that recognizes the m^7G cap, we were able to immunoprecipitate sisRNAs from unfertilized eggs (Figure 2B), indicating that sisRNAs have 5′ m^7G caps. As a positive control, *actin5C* mRNA was enriched in the immunoprecipitates but not for U85 intronic scaRNA.

The following observations indicate that the *rga*, *mbt*, and *csp* introns contain sequences that can drive independent transcription of sisRNAs First, transfection of *UAS-dsRed-intron-myc* plasmids containing

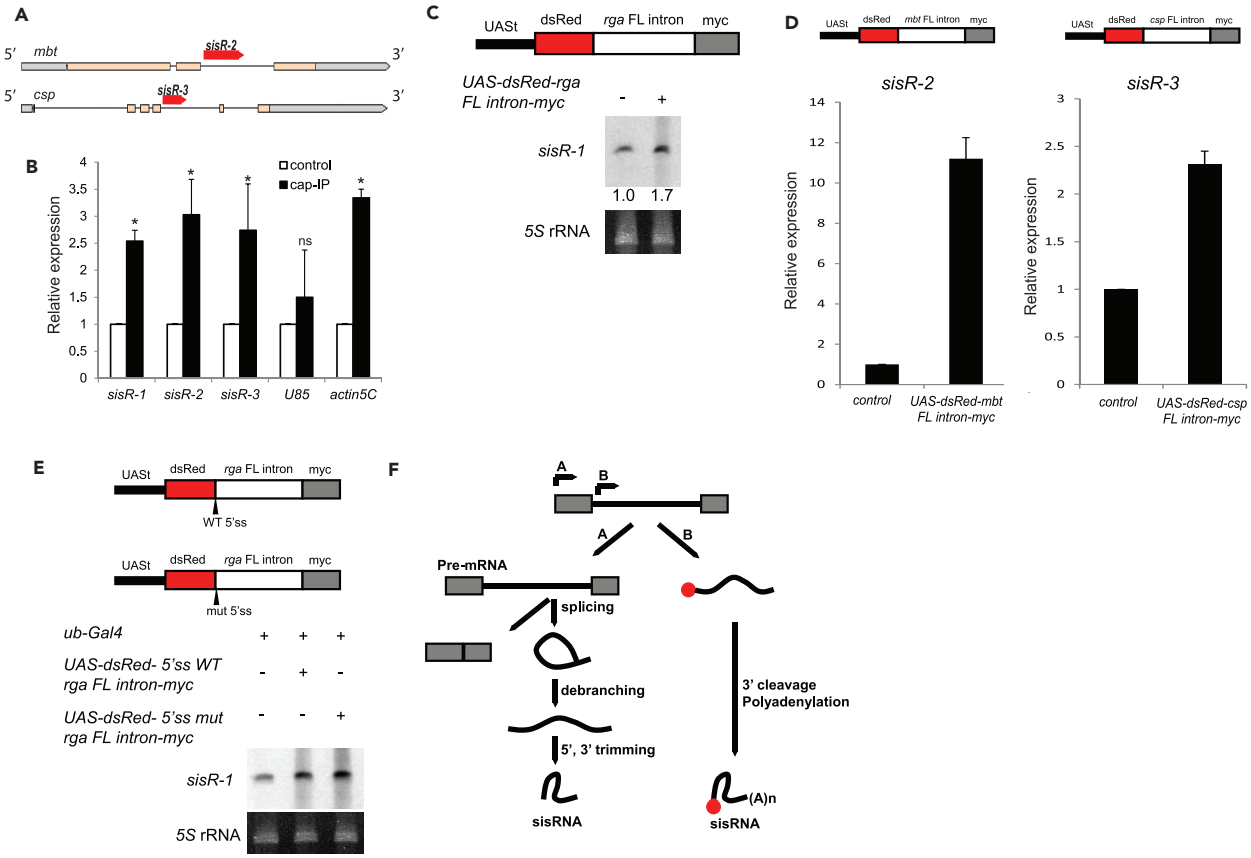


Figure 2. sisRNAs Are Capped, Independently Transcribed, and Polyadenylated by Wispy

(A) The *mbt* and *csp* loci and the locations of *sisR-2* and *sisR-3* in the introns. The *mbt* and *csp* loci are not in the same scale.

(B) qRT-PCT showing enrichment of sisRNAs in m^7G antibody immunoprecipitates in unfertilized eggs. *$p < 0.01$; ns, not significant, $p > 0.05$. N = 3. Two-tailed t test. Data are represented as mean ± SD. SD, standard deviation.

(C) Northern blot showing the expression of *sisR-1* in S2 cells before and after transfection of *UAS-dsRed-rga FL intron-myc*. Numbers below indicate the relative band intensities of *sisR-1* normalized to rRNA quantified using ImageJ software.

(D) qRT-PCR showing the relative expression of *sisR-2* and *sisR-3* in male bodies of *y w* controls (parental strain) and transgenic flies harboring *UAS-dsRed-mbt/csp FL intron-myc* transgenes. *$p < 0.01$, two-tailed t test. N = 3. Data are represented as mean ± SD.

(E) Northern blot showing the expression of *sisR-1* in S2 cells transfected with plasmids containing wild-type or mutated 5′ splice site of *rga* intron.

(F) A proposed model of biogenesis of sisRNAs via splicing-dependent and splicing-independent transcription pathways.

See also Figures S2 and S3.

rga full-length intron (without Gal4 induction) into S2 cells led to an increase in *sisR-1* levels as assayed by northern blotting (Figure 2C). Second, flies harboring extra copies of the *UAS-dsRed-intron-myc* transgenes containing *mbt* or *csp* full-length introns (without Gal4 induction) also expressed higher levels of *sisR-2* or *sisR-3,* respectively, than the parental strains as assayed by quantitative RT-PCR (qRT-PCR) (Figure 2D). The UAS promoter was not leaky as we observed non-significant expression of *dsRed* (Figure S3). Finally, mutation of the 5′ splice site of the *rga* intron in the *UAS-dsRed-intron-myc* plasmid did not perturb the expression of *sisR-1* in S2 cells (Figure 2E), suggesting that sisRNAs can also be processed in a splicing-independent manner. These observations are consistent with an alternative pathway for sisRNA biogenesis via direct transcription from the introns (Figure 2F).

sisR-3 Regulates Long Noncoding RNA *CR44148*

To understand the functional significance of these sisRNAs, we focused on *sisR-3*. Previously, *sisR-1* was shown to repress the expression of *ASTR* ncRNA *in vivo*, possibly via base-pairing of its 3′ tail with the target (Pek et al., 2015). We asked if *sisR-3* also shows similar sisRNA-target relationship properties *in vivo*. We examined the predicted secondary structure of *sisR-3* using the Vienna RNAfold software. Interestingly, *sisR-3* was predicted to form a secondary structure that has an exposed 3′ end (Figure 3A).

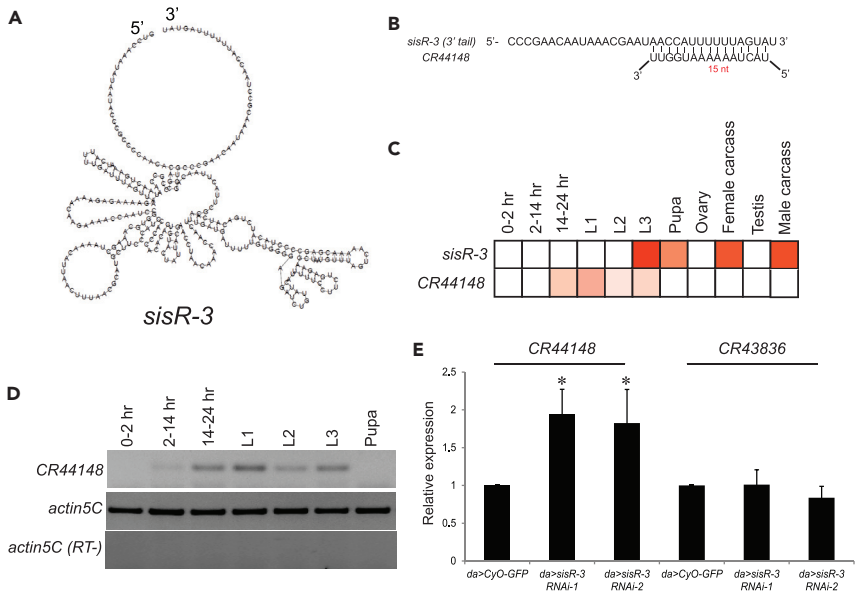


Figure 3. *sisR-3* Represses *CR44148*

(A) Predicted secondary structures of *Drosophila sisR-3*.

(B) Complementary base-pairing between the 3′ end of *sisR-3* and *CR44148*.

(C) Heatmap showing the relative expression of *sisR-3* and its predicted target *CR44148* during development and in adults. Red, high expression; white, low or undetectable expression.

(D) Strand-specific RT-PCR showing the expression of *CR44148* during development. *Actin5C* was used as a loading control.

(E) qPCR showing the relative abundance of *CR44148*, *sisR-2*, and *CR43836* in the indicated genotypes. *Actin5C* was used as a loading control. *$p < 0.05$, two-tailed t test. N = 3. Data are represented as mean ± SD. SD, standard deviation. See also Figure S4.

The characteristic free 3′ end feature of *sisR-3* is reminiscent of the one for *sisR-1* (Pek et al., 2015), suggesting that they belong to a family of structurally related sisRNAs.

We predicted the target of *sisR-3* by performing a BLAST search using the sequences of the exposed 3′ ends. The 3′ end of *sisR-3* is predicted to target a long ncRNA *CR44148* (Figure 3B). We next asked if *sisR-3* and its predicted target *CR44148* have reciprocal expression patterns. The modENCODE temporal and tissue expression data in FlyBase showed that *CR44148* exhibited mutually exclusive temporal and spatial expression patterns to *sisR-3*. *sisR-3* expression is highly expressed in the third-instar larvae, pupae, and adult somatic tissues, whereas *CR44148* is abundantly expressed in the embryos (Figures 3C and 3D). These observations are consistent with a model that *sisR-3* modulates robustness in gene expression by negatively regulating *CR44148*.

To test whether *sisR-3* regulates its predicted target *in vivo*, we knocked down the expression of *sisR-3* by two independent short hairpin RNAs (shRNAs) and then examined if there was any up-regulation of *CR44148* (Figure S4A). The knockdown efficiency was tested in the ovaries by quantitative PCR (qPCR). Expression of *sisR-3*, but not its cognate *csp* mRNA, was down-regulated by driving *sisR-3* shRNAs using *act-Gal4* driver (Figure S4B). Since *sisR-3* and *CR44148* are co-expressed in the third-instar larvae, we examined the effect of *sisR-3* knockdown on target expression during this stage of development. Knockdown of *sisR-3* in the third-instar larvae using *da-Gal4* resulted in up-regulation of *CR44148*, but not another non-targeted ncRNA *CR43836* (Figure 3E), verifying that *sisR-3* specifically represses *CR44184*.

sisR-3 Is Required during Development

To investigate if *sisR-3* is important for development, we ubiquitously knocked down *sisR-3*. Knockdown of *sisR-3* using *act-Gal4* revealed a semi-lethality phenotype. Based on the crossing scheme, we expected the ratio of eclosed *act>CyO:act>sisR-3 RNAi* adults to be 1:1 (Figure 4A). Instead, we observed that the ratio was 1:~0.6 for both RNA interference (RNAi) lines (Figure 4A). This result implied that *sisR-3* knockdown

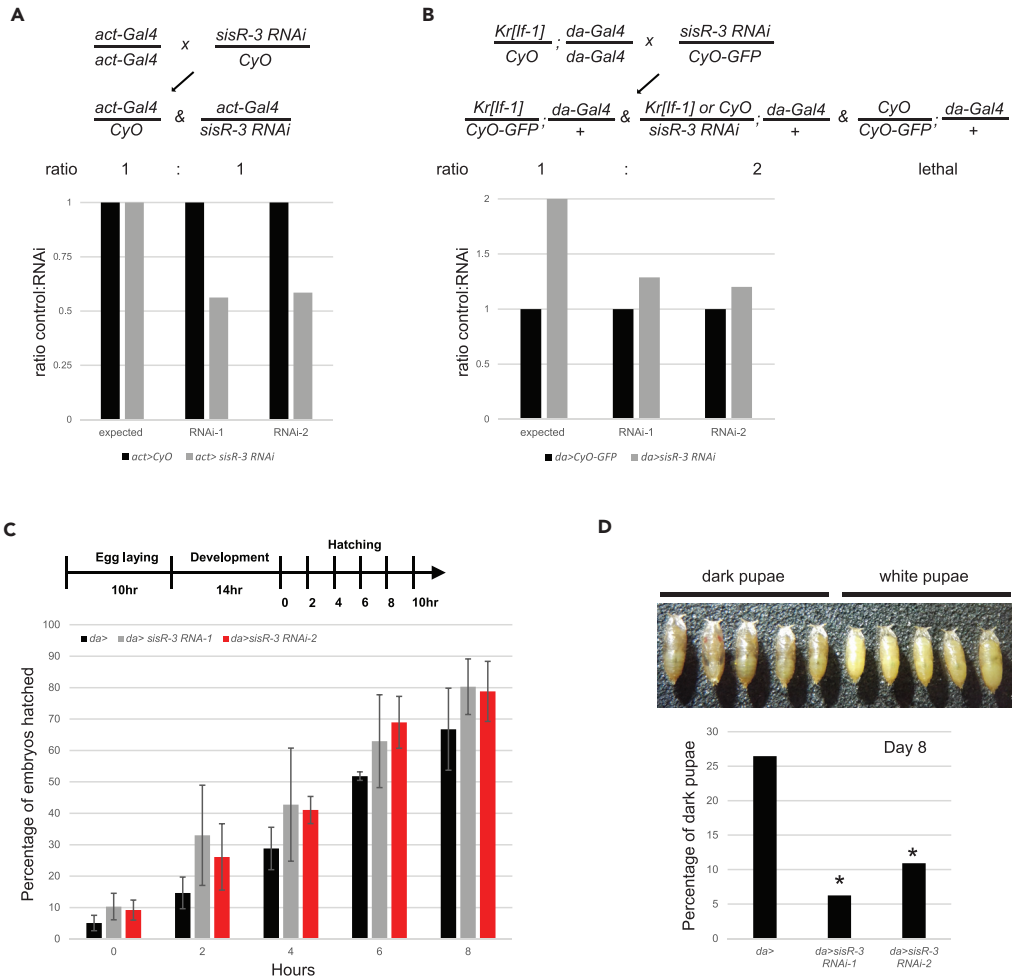


Figure 4. *sisR-3* Is Required for Proper Development

(A and B) Crossing schemes and charts showing the expected and observed ratios of the indicated genotypes.

(C) Chart showing the hatching rates of embryos of the indicated genotypes at different time points.

(D) Chart showing the percentage of pupae that are dark (see photograph) for the indicated genotypes on day 8 of development at 25°C. *p < 0.01, chi-square test. N > 100 pupae in three independent experiments.

may affect development. We used an alternative crossing scheme to check for third-instar larvae development (Figure 4B). In this cross, we expected a ratio of 1:2 for *da>CyO-GFP:da>sisR-3 RNAi*; however, we observed a ratio of 1:~1.2 for both RNAi lines (Figure 4B). This result suggests that *sisR-3* knockdown might have an effect on larvae development. To confirm our results, we next directly examined the developmental effects of *sisR-3* knockdown. Knockdown of *sisR-3* in the embryos did not affect embryogenesis. We observed that *da >sisR-3 RNAi* embryos developed normally and had hatching rates similar to *da-Gal4* controls (Figure 4C). We then compared the development of *da-Gal4* controls and *da >sisR-3 RNAi* pupae. At day 8 of development, whereas ~28% of *da-Gal4* control pupae had begun to mature forming dark pupae, only 8%–11% of *da>sisR-3 RNAi* pupae were mature (Figure 4D, p < 0.01, chi-square test), indicating that *sisR-3* is required for proper development. Taken together, our results demonstrate that *sisR-3* is required for proper larva and pupa development.

DISCUSSION

In this study, we identified an abundant pool of polyadenylated maternal sisRNAs in *Drosophila*. These sisRNAs can be generated by independent transcription from the cognate introns and require Wispy for their stabilization as maternal transcripts. Further characterization of a sisRNA *sisR-3* demonstrates that

sisR-3 represses its target long ncRNA *CR44148* and is required for proper development. Together with our previous study on *sisR-1*, sisRNAs appear to function as regulators of other long ncRNAs, suggesting a robust mechanism to clear off unwanted long ncRNAs (Pek et al., 2015).

Our data suggest an alternative pathway for maternal sisRNA biogenesis (Figure 2F)— sisRNAs are transcribed from intronic promoters and are cleaved and polyadenylated in the nucleus after the PAS sequences in the introns. Maternal sisRNAs are further polyadenylated in the cytoplasm by Wispy to confer stability. Previous work had suggested that splicing and debranching of the intronic transcript is required for *sisR-1* production during development (Pek et al., 2015). Therefore, we propose that there are two non-mutually exclusive pathways generating sisRNAs—a host pre-mRNA splicing-dependent pathway and an independent transcription pathway (Figure 2F), similar to what was reported for snoRNAs (Villa et al., 1998). The relative contribution of each pathway may differ in various tissues or developmental contexts.

Drosophila sisR-1 and *sisR-3* are predicted to form related secondary structures that have exposed 3′ ends. Interestingly, the Epstein-Barr virus (EBV) ebv-sisRNA-1 was also predicted to adopt a similar secondary structure (Moss et al., 2014; Moss and Steitz, 2013). It was proposed that the 3′ tail of the ebv-sisRNA-1 may bind to RNAs by complementary base-pairing and possibly regulate RNA activity or abundance (Moss et al., 2014; Moss and Steitz, 2013). It implies that the 3′ tail is an important element for a sisRNA to function. The 3′ end may provide specificity to the targets by complementary base-pairing, thus providing a new paradigm for sisRNA-mediated gene regulation. This principle of ncRNA-target recognition has been seen in various ncRNAs such as snoRNAs, miRNAs, small nuclear RNAs (snRNAs), and CRISPR RNAs (Cech and Steitz, 2014).

Recent studies have uncovered a role for poly(A) tails in promoting the stability of viral and endogenous long ncRNAs such as MALAT and PAN RNA (Brown et al., 2012; Mitton-Fry et al., 2010; Tycowski et al., 2012, 2016; Wilusz et al., 2012). It is conceivable that the poly(A) tails of sisRNAs may form intra- or intermolecular interactions with other nucleic acids or proteins, to protect them from 3′ exoribonucleases. Polyadenylation of intronic sequences was first described in sea urchin eggs more than 30 years ago (Calzone et al., 1988; Costantini et al., 1980; Ruzdijic and Pederson, 1987). In the *Drosophila* larvae, the *delta* locus was also shown to give rise to multiple polyadenylated intronic sequences more than 20 years ago (Kopczynski and Muskavitch, 1992). More recently, a study identified polyadenylated sisRNAs from the EBV (Cao et al., 2015). The identification of abundant polyadenylated maternal sisRNAs in *Drosophila* suggests that this paradigm may be more widely conserved than previously thought.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Transparent Methods, four figures, and two tables and can be found with this article online at https://doi.org/10.1016/j.isci.2018.05.010.

ACKNOWLEDGMENTS

We thank Joseph Gall and the Bloomington Stock Center for reagents, Steven Ching and Allison Pinder for help in deep sequencing and bioinformatics, Amanda Ng for assistance in qPCR analyses, and members of the Pek laboratory for discussion. The GEO accession number for deep sequencing is GSE77294. The authors are supported by the Temasek Life Sciences Laboratory.

AUTHOR CONTRIBUTIONS

S.S.J.N. and R.T.Z. performed the experiments. I.O. performed the experiments and wrote the paper. J.W.P. conceived the project, performed the experiments, and wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 13, 2018
Revised: April 25, 2018
Accepted: May 14, 2018
Published: June 29, 2018

REFERENCES

Benoit, P., Papin, C., Kwak, J.E., Wickens, M., and Simonelig, M. (2008). PAP- and GLD-2-type poly(A) polymerases are required sequentially in cytoplasmic polyadenylation and oogenesis in *Drosophila*. Development *135*, 1969–1979.

Brown, J.A., Valenstein, M.L., Yario, T.A., Tycowski, K.T., and Steitz, J.A. (2012). Formation of triple-helical structures by the 3'-end sequences of MALAT1 and MENbeta noncoding RNAs. Proc. Natl. Acad. Sci. USA *109*, 19202–19207.

Calzone, F.J., Lee, J.J., Le, N., Britten, R.J., and Davidson, E.H. (1988). A long, nontranslatable poly(A) RNA stored in the egg of the sea urchin *Strongylocentrotus purpuratus*. Genes Dev. *2*, 305–318.

Cao, S., Strong, M.J., Wang, X., Moss, W.N., Concha, M., Lin, Z., O'Grady, T., Baddoo, M., Fewell, C., Renne, R., et al. (2015). High-throughput RNA sequencing-based virome analysis of 50 lymphoma cell lines from the Cancer Cell Line Encyclopedia project. J. Virol. *89*, 713–729.

Cech, T.R., and Steitz, J.A. (2014). The noncoding RNA revolution-trashing old rules to forge new ones. Cell *157*, 77–94.

Costantini, F.D., Britten, R.J., and Davidson, E.H. (1980). Message sequences and short repetitive sequences are interspersed in sea urchin egg poly(A)+ RNAs. Nature *287*, 111–117.

Gardner, E.J., Nizami, Z.F., Talbot, C.C., Jr., and Gall, J.G. (2012). Stable intronic sequence RNA (sisRNA), a new class of noncoding RNA from the oocyte nucleus of *Xenopus tropicalis*. Genes Dev. *26*, 2550–2559.

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

Juge, F., Zaessinger, S., Temme, C., Wahle, E., and Simonelig, M. (2002). Control of poly(A) polymerase level is essential to cytoplasmic polyadenylation and early development in *Drosophila*. EMBO J. *21*, 6603–6613.

Kopczynski, C.C., and Muskavitch, M.A. (1992). Introns excised from the Delta primary transcript are localized near sites of Delta transcription. J. Cell Biol. *119*, 503–512.

Kung, J.T., Colognori, D., and Lee, J.T. (2013). Long noncoding RNAs: past, present, and future. Genetics *193*, 651–669.

Lee, J.T. (2012). Epigenetic regulation by long noncoding RNAs. Science *338*, 1435–1439.

Lee, M., Choi, Y., Kim, K., Jin, H., Lim, J., Nguyen, T.A., Yang, J., Jeong, M., Giraldez, A.J., Yang, H., et al. (2014). Adenylation of maternally inherited microRNAs by Wispy. Mol. Cell *56*, 696–707.

Lu, Z., Filonov, G.S., Noto, J.J., Schmidt, C.A., Hatkevich, T.L., Wen, Y., Jaffrey, S.R., and Matera, A.G. (2015). Metazoan tRNA introns generate stable circular RNAs in vivo. RNA *21*, 1554–1565.

Matera, A.G., Terns, R.M., and Terns, M.P. (2007). Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. Nat. Rev. Mol. Cell Biol. *8*, 209–220.

Mitton-Fry, R.M., DeGregorio, S.J., Wang, J., Steitz, T.A., and Steitz, J.A. (2010). Poly(A) tail recognition by a viral RNA element through assembly of a triple helix. Science *330*, 1244–1247.

Moss, W.N., Lee, N., Pimienta, G., and Steitz, J.A. (2014). RNA families in Epstein-Barr virus. RNA Biol. *11*, 10–17.

Moss, W.N., and Steitz, J.A. (2013). Genome-wide analyses of Epstein-Barr virus reveal conserved RNA structures and a novel stable intronic sequence RNA. BMC Genomics *14*, 543.

Osman, I., Tay, M.L., and Pek, J.W. (2016). Stable intronic sequence RNAs (sisRNAs): a new layer of gene regulation. Cell Mol. Life Sci. *73*, 3507–3519.

Pek, J.W. (2018). Stable intronic sequence RNAs engage in feedback loops. Trends Genet. *34*, 330–332.

Pek, J.W., and Okamura, K. (2015). Regulatory RNAs discovered in unexpected places. Wiley Interdiscip. Rev. RNA *6*, 671–686.

Pek, J.W., Osman, I., Tay, M.L., and Zheng, R.T. (2015). Stable intronic sequence RNAs have possible regulatory roles in *Drosophila melanogaster*. J. Cell Biol. *211*, 243–251.

Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. *81*, 145–166.

Ruzdijic, S., and Pederson, T. (1987). Evidence for an association between U1 RNA and interspersed repeat single-copy RNAs in the cytoplasm of sea urchin eggs. Development *101*, 107–116.

Shi, Y., and Manley, J.L. (2015). The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. Genes Dev. *29*, 889–897.

Talhouarne, G.J., and Gall, J.G. (2014). Lariat intronic RNAs in the cytoplasm of *Xenopus tropicalis* oocytes. RNA *20*, 1476–1487.

Tay, M.L., and Pek, J.W. (2017). Maternally inherited stable intronic sequence RNA triggers a self-reinforcing feedback loop during development. Curr. Biol. *27*, 1062–1067.

Tycowski, K.T., Shu, M.D., Borah, S., Shi, M., and Steitz, J.A. (2012). Conservation of a triple-helix-forming RNA stability element in noncoding and genomic RNAs of diverse viruses. Cell Rep. *2*, 26–32.

Tycowski, K.T., Shu, M.D., and Steitz, J.A. (2016). Myriad triple-helix-forming structures in the transposable element RNAs of plants and fungi. Cell Rep. *15*, 1266–1276.

Villa, T., Ceradini, F., Presutti, C., and Bozzoni, I. (1998). Processing of the intron-encoded U18 small nucleolar RNA in the yeast *Saccharomyces cerevisiae* relies on both exo- and endonucleolytic activities. Mol. Cell. Biol. *18*, 3376–3383.

Wilusz, J.E., JnBaptiste, C.K., Lu, L.Y., Kuhn, C.D., Joshua-Tor, L., and Sharp, P.A. (2012). A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A) tails. Genes Dev. *26*, 2392–2407.

Wong, J.T., Akhbar, F., Ng, A.Y.E., Tay, M.L., Loi, G.J.E., and Pek, J.W. (2017). DIP1 modulates stem cell homeostasis in *Drosophila* through regulation of sisR-1. Nat. Commun. *8*, 759.

Yin, Q.F., Yang, L., Zhang, Y., Xiang, J.F., Wu, Y.W., Carmichael, G.G., and Chen, L.L. (2012). Long noncoding RNAs with snoRNA ends. Mol. Cell *48*, 219–230.

Zhang, Y., Zhang, X.O., Chen, T., Xiang, J.F., Yin, Q.F., Xing, Y.H., Zhu, S., Yang, L., and Chen, L.L. (2013). Circular intronic long noncoding RNAs. Mol. Cell *51*, 792–806.

Zheng, S., Vuong, B.Q., Vaidyanathan, B., Lin, J.Y., Huang, F.T., and Chaudhuri, J. (2015). Non-coding RNA generated following lariat debranching mediates targeting of AID to DNA. Cell *161*, 762–773.

ISCI, Volume *4*

Supplemental Information

Generation of *Drosophila* sisRNAs by Independent Transcription from Cognate Introns

Sharon Si Jia Ng, Ruther Teo Zheng, Ismail Osman, and Jun Wei Pek

Transparent Methods

Fly strains

The *y w* strain and non-shRNA-expressing flies were used as controls unless otherwise stated. The following fly strains were used in this study: *actin-Gal4, Kr^{If-1}/CyO;da-Gal4,* and *$wispy^{12-3147}$*. For collection of eggs, virgin females were fed with wet yeast for several days. Generation of dsRed-intron-myc overexpression flies was done as previously described (Pek et al., 2015). For generation of *csp* sisRNAs shRNA transgenic flies, shRNAs targeting *csp* sisRNAs were designed and cloned into Valium22 plasmid, performed as previously described (Ni et al., 2011). Sequences were chosen to avoid potentially off-target effects. Transgenic flies were generated by Genetic Services and BestGene Inc using phiC31 integrase-mediated insertion into 25C7 landing site (Bischof et al., 2007). Oligo sequences are available in Table S2.

RNA extraction

Tissues were homogenized in 1.5 ml Eppendorf tubes using a plastic pestle and RNA was extracted using the TRIzol extraction protocol (Ambion) or the Direct-zol RNA miniprep kit (Zymo Research). For deep sequencing, rRNA depletion was performed using the Ribo-Zero magnetic kit (Epicentre). RNA was characterized with a Bioanalyzer 2100 (Agilent).

Isolation of polyA(+) and polyA(-) RNAs

Isolation of polyA(+) and polyA(-) RNAs were carried out using the Dynabeads mRNA Purification Kit (Life Technologies). Total RNA was denatured at 65°C and polyA(+) RNAs were

allowed to anneal to the Dynabeads in a microcentrifuge tube at room temperature. The polyA(-) RNA fraction was isolated from the resulting supernatant whereas the remaining polyA(+) RNA-beads was washed and the polyA(+) RNA was subsequently eluted from the beads.

Sequencing and sequence analysis

cDNA library was constructed, sequenced and analyzed as previously described (Gardner et al., 2012; Pek et al., 2015). Reads were aligned to the *Drosophila melanogaster* genome (*dm3* genome release) using TopHat version 1.4.0 and Bowtie version 0.12.9 sequence alignment programs (Langmead et al., 2009; Trapnell et al., 2009).

Identification of candidate polyadenylated sisRNAs

Identification of polyA sisRNAs was done by browsing the polyA(+) deep sequencing data manually on a genome browser. Introns that consist of reads clustering to form a peak were selected. PAS sequence was identified by examining the sequences near the predicted 3' end of the peak.

Isolation of capped RNAs

~75-100 µg of unfertilized egg RNAs were incubated overnight at 4°C with beads conjugated with mouse m^7G antibody (MABE419, Merck Millipore) in NET-2 buffer with continuous rocking. The immunoprecipitates were washed 3 times and RNA was extract using TRIzol. Same volumes of RNA was used for RT-qPCR analyses.

RT-PCR

RT-PCR was performed as previously described (Pek et al., 2012; Pek et al., 2015). For standard RT-PCR, total RNA was reverse transcribed with random hexamers for 1 hour using AMV-RT (New England Biolabs), M-MLV RT (Promega) or Superscript III (Invitrogen). PCR was carried out using the resulting cDNA. For selection of polyA(+) RNAs, RT was carried out with oligo dT primers. For qPCR, SYBR Fast qPCR kit master mix (2X) universal (Kapa Biosystems, USA) was used with addition of ROX reference dye high and carried out on the Applied Biosystems 7900HT Fast Real-Time PCR system. Oligo sequences were reported previously (Pek et al., 2015). RACE-PAT to measure poly(A) lengths was performed as described previously (Salles et al., 1999).

5' RACE and 3' RACE

For 5' RACE, RT was performed using a gene-specific reverse primer. The subsequent cDNA was treated with RNase A (Axygen), T1 (ThermoScientific) and H (New England Biolabs), and tailed with dTTP using terminal deoxynucleotidyl transferase (Promega). Nested PCR was performed twice using gene-specific and polyT-specific primers. For 3' RACE, RT was performed using a polyA-specific reverse primer. Nested PCR was performed twice using gene-specific and polyT-specific primers. For both 5' and 3' RACE, the cDNA were run on an agarose gel and size selected for fragments between a 100 to 500 base pairs before running the nested PCR. PCR products were run on an agarose gel, purified and cloned into pGEM-T-Easy vector (Promega) and sequenced.

Data and software availability

GEO accession number for deep sequencing is GSE77294.

References

Bischof, J., Maeda, R.K., Hediger, M., Karch, F., and Basler, K. (2007). An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. Proceedings of the National Academy of Sciences of the United States of America *104*, 3312-3317.

Gardner, E.J., Nizami, Z.F., Talbot, C.C., Jr., and Gall, J.G. (2012). Stable intronic sequence RNA (sisRNA), a new class of noncoding RNA from the oocyte nucleus of Xenopus tropicalis. Genes & development *26*, 2550-2559.

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

Ni, J.Q., Zhou, R., Czech, B., Liu, L.P., Holderbaum, L., Yang-Zhou, D., Shim, H.S., Tao, R., Handler, D., Karpowicz, P.*, et al.* (2011). A genome-scale shRNA resource for transgenic RNAi in Drosophila. Nature methods *8*, 405-407.

Pek, J.W., Ng, B.F., and Kai, T. (2012). Polo-mediated phosphorylation of Maelstrom regulates oocyte determination during oogenesis in Drosophila. Development *139*, 4505-4513.

Pek, J.W., Osman, I., Tay, M.L., and Zheng, R.T. (2015). Stable intronic sequence RNAs have possible regulatory roles in Drosophila melanogaster. The Journal of cell biology *211*, 243-251.

Salles, F.J., Richards, W.G., and Strickland, S. (1999). Assaying the polyadenylation state of mRNAs. Methods *17*, 38-45.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

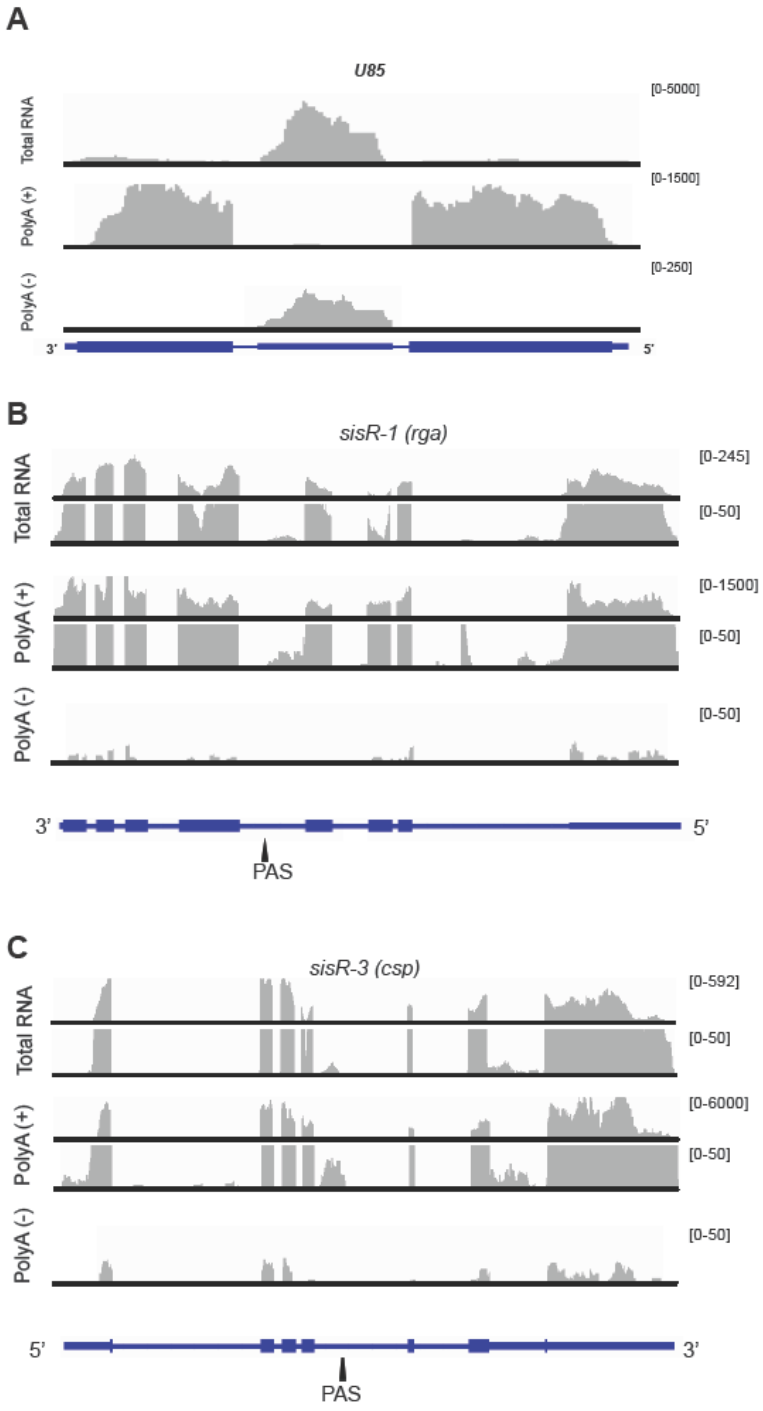


Figure S1. Polyadenylated sisRNAs in unfertilized eggs. Related to Figure 1. (A-C) Genome browser view of *U85*, *rga* (*sisR-1*) and *csp* (*sisR-3*) gene loci. RNA sequencing results for total RNA, polyA(+) RNA and polyA(-) RNA from unfertilized eggs were shown. PAS: polyadenylation signal near the predicted 3' ends of the sisRNAs.

Sequence of *csp* intron 4

GTGAGTGAAGTTACTTGGCTTAGCCAAGTCCAAATATAATACCCGCCCCAACACAGCCCGCCAATAACTGAAAT
CATTTTGATTTAGTTAAGAAAAGAGAAAACAAGAAAACCAATCCGGTGCAAAGCTAAAACATTAACTTTAACGCAT
GCTTCCCATCCACCCACACACTGTACTATTACGAACCTAACAAACCAACTTTGATGATTTTTGTGGGGAGATCTG
TATCATTTTTCCTTCTGAGAAGCTAATGTTTAGTCAAAAACGAGCCCCTACATCTGACATCAACGCTTACTTAACA
GCTGCCCGAACAATAAACGCCTAACCATTTTTTAGTATAATTTTCCTGGGTTATCAGCTATCTTGGCACACCATC
TCCCTCGATTGGTTTTCCCTTACTCACACTACTCCGTACAAACACTACTCTGATTGCACTGCTTTGTAATTAGTTT
ATTTTGGGTAAGTATGAAGTTTGGTTTTGGTTTCCGATTATCAATTCTCCAGACACTGAAGTAATCAGAATACTGA
TTTTTGAGACCAGAGGTAAACACGACTAACCCTAAAACTTTTCCGATCCATCGAACAATTTTTGTATAATTTATTG
TTAGTGTTTGTTTTGTATATATGTACTTCACGAAAATCATCAAATGCTTCCAACGTTTCTCTCTCCATAAAAAAAAT
AATAAATAACGTCAACACATACACGCTCGGATTTCTCGAGACAGACAGTGCAACGATTATTGTTTTTTCCACGAT
TGCTTTTCTTGTTTTTTAAGTGTCCCATTGCGCATACAAATTTATTTTGAAACCACCCAGATTAGAAGTCTTCTTC
CTTTGCTAGTTTGCAAGCTAAGGAACAGAAGCTATAGACTTACGTTTCTATTTTTGGTTAACTTCTTTTTTTTTCTC
AAAAATCTTTGCGATCGACTCATCATATCCACCACCTACTACCACCCACAACCACCACACATTCGCTCTCGACCT
CGCCGTCAG

Red: region of heterogeneity
Green: PAS

Heterogenous 5' end	Number of clones
5' GTCCAAA…3'	3
5' TCCAAA…3'	7
5' CCAAA…3'	7
5' CAAA…3'	1

Heterogenous 3' ends	Number of clones
5'…CATTTTTTAGT 3'	6
5'…CATTTTTTAGTA 3'	8
5'…CATTTTTTAGTAT 3'	3

Figure S2. Sequences of *sisR-3* as determined by 5' and 3' RACE analyses. Related to Figure 2. Sequence of full-length intron is shown. sisRNA sequence is underlined. Red: Region of heterogeneity. Green: PAS.

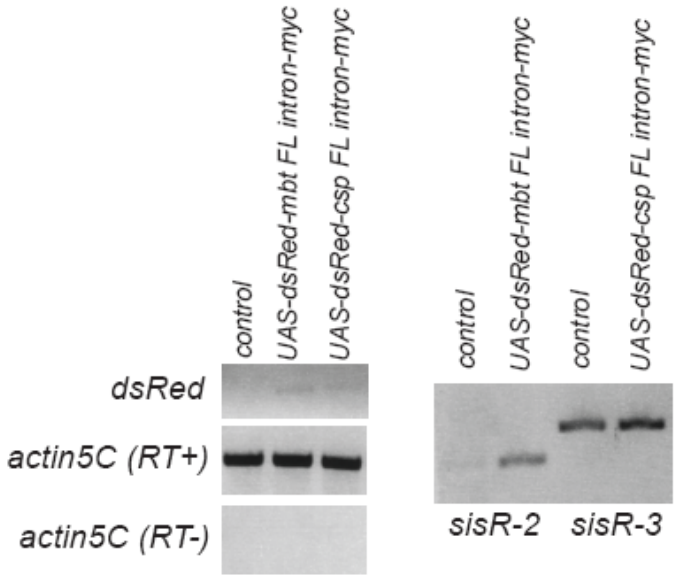


Figure S3. No significant leaky transcription from the UAS promoter. Related to Figure 2. RT-PCR showing the abundance of *dsRed*, *sisR-2* and *sisR-3* in male bodies of the indicated genotypes. *Actin5C* was used as a loading control.

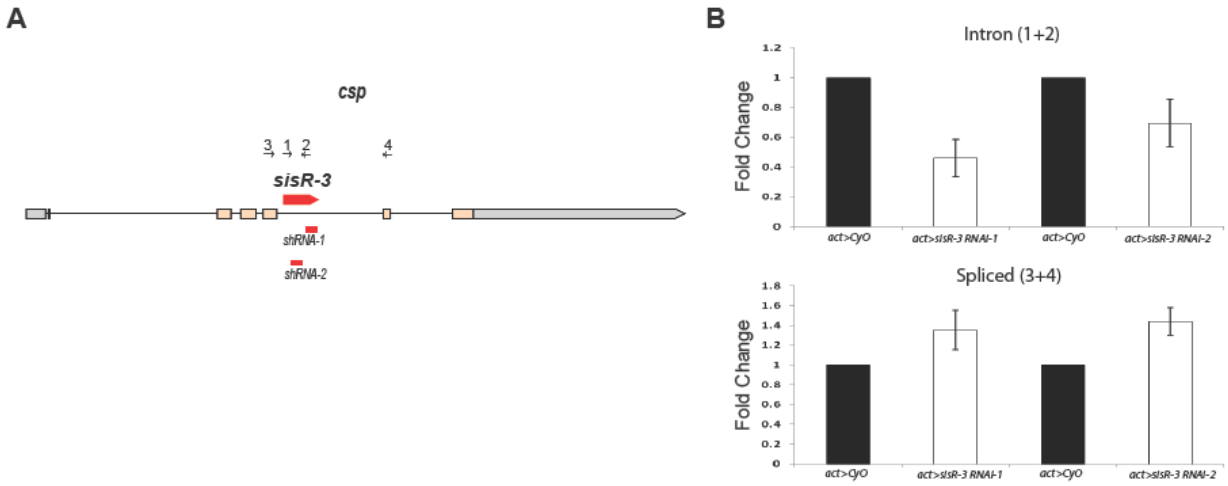


Figure S4. Knockdown of *sisR-3* by shRNA. Related to Figure 3. (A) The *csp* locus showing the regions of *sisR-3*, primers and RNAi constructs. (B) qPCR showing the relative levels of *sisR-3* and *csp* mRNA in ovaries of the indicated genotypes.

Table S2. List of oligos used. Related to Figure 1, 2 and 3.

Oligo names	Sequences
CR44148 Fw	CCACGAATAAATTGTTTGGGCC
CR44148 Rv	ACGACATGGATCTGCGTACT
CR43836 Fw	TTCGCGAATTGTATTCTACGG
CR43836 Rv	ACGGGTATAACGGATGGCTA
csp spliced	TATCTGCTGTGCCGTGATCA
csp spliced Rv	CATATCGTTGCCCTCCCGA
csp intron FW	GGC TTA GCC AAG TCC AAA TA
csp intron Rv	GCT GTT AAG TAA GCG TTG ATG TC
actin5C Fw	TGCCCATCTACGAGGGTTAT
actin5C Rv	AGTACTTGCGCTCTGGCGG
mbt intron *Fw*	CCATCATGCGATTCTTTGTG
mbt intron *Rv*	GATTTCACCTGGCACTTTGG
csp 5' RACE RT	AATGATACAGATCTCCCCACAA
csp 5' RACE 1st Nested PCR Fw	AATGATACAGATCTCCCCACAA
csp 5' RACE 1st Nested PCR Rv	GCGAGCACAGAATTAATACGACTC ACTATAGGAAAAAAAAAAAAVN
csp 5' RACE 2nd Nested PCR Fw	GTTCGTAATAGTACAGTGTGTGGG
csp 5' RACE 2nd Nested PCR Rv	CGCGGATCCGAATTAATACGACTCACTA TAGG
csp 3' RACE RT	GCGAGCACAGAATTAATACGACTCACTA TAGGTTTTTTTTTTTTVN
csp 3' RACE 1st Nested PCR Fw	GGTGCAAAGCTAAAACATTAACTT
csp 3' RACE 1st Nested PCR Rv	GCGAGCACAGAATTAATACGACT

csp 3’ RACE 2nd Nested PCR Fw	CCCACACACTGTACTATTACGAAC
csp 3’ RACE 2nd Nested PCR Fw	CGCGGATCCGAATTAATACGACTCACTA TAGG