Article

SOURCE DATA | TRANSPARENT PROCESS

EMBO reports

Distinct biogenesis pathways may have led to functional divergence of the human and *Drosophila* Arglu1 sisRNA

Seow Neng Chan[1] & Jun Wei Pek[1,2,*]

Abstract

Stable intronic sequence RNAs (sisRNAs) are stable, long noncoding RNAs containing intronic sequences. While sisRNAs have been found across diverse species, their level of conservation remains poorly understood. Here we report that the biogenesis and functions of a sisRNA transcribed from the highly conserved Arglu1 locus are distinct in human and *Drosophila melanogaster*. The Arglu1 genes in both species show similar exon-intron structures where the intron 2 is orthologous and positionally conserved. In humans, Arglu1 sisRNA retains the entire intron 2 and promotes host gene splicing. Mechanistically, Arglu1 sisRNA represses the splicing-inhibitory activity of ARGLU1 protein by binding to ARGLU1 protein and promoting its localization to nuclear speckles, away from the Arglu1 gene locus. In contrast, *Drosophila* dArglu1 sisRNA forms via premature cleavage of intron 2 and represses host gene splicing. This repression occurs through a local accumulation of dARGLU1 protein and inhibition of telescripting by U1 snRNPs at the dArglu1 locus. We propose that distinct biogenesis of positionally conserved Arglu1 sisRNAs in both species may have led to functional divergence.

Keywords Arglu1; autoregulation; orthologous intron; positional conservation; sisRNA

Subject Category RNA Biology

DOI 10.15252/embr.202154350 | Received 17 November 2021 | Revised 28 November 2022 | Accepted 2 December 2022 | Published online 19 December 2022

EMBO Reports (2023) 24: e54350

Introduction

Over the past decade, a number of noncoding RNAs have been discovered and implicated in various biological processes (Cech & Steitz, 2014; Statello *et al*, 2020). One of the latest additions to the list of regulatory long noncoding RNAs (lncRNAs) is the discovery of stable intronic sequence (sis)RNAs. First discovered in the oocyte nucleus of *Xenopus tropicalis*, these introns were—contrary to conventional beliefs—not rapidly degraded but stably maintained (Gardner *et al*, 2012). Since then, subsequent studies have discovered sisRNAs in many different organisms, including viruses, *Arabidopsis*, yeast, fruit fly, zebrafish, mouse, and human cells (Moss & Steitz, 2013; Zhang *et al*, 2013; Tomita *et al*, 2015; Li *et al*, 2016; Osman *et al*, 2016; Talhouarne & Gall, 2018; Wu *et al*, 2018; Morgan *et al*, 2019; Parenteau *et al*, 2019). SisRNAs are found to be involved in the starvation response of yeast and fruit flies (Osman & Pek, 2018, 2021; Morgan *et al*, 2019; Parenteau *et al*, 2019; Voo *et al*, 2021). They have also been found to affect germline stem cell maintenance and embryonic development in *Drosophila* via feedback loops (Pek *et al*, 2015; Tay & Pek, 2017; Ng *et al*, 2018b; Pek, 2018; Koh & Pek, 2022). Accordingly, the definition of sisRNAs has been revised to encompass not just spliced stable intronic sequences but also stable lncRNAs that contain intronic sequences including those produced by premature intronic cleavage, intron retention, and circularization (Chan & Pek, 2019). Therefore, sisRNAs have important regulatory roles in development and homeostasis.

As more studies begin to unravel the biological significance of lncRNAs (including sisRNAs), there is an emerging interest to study their evolution and conservation between different species. Unlike protein coding transcripts, lncRNAs generally have low sequence similarity and are usually rapidly evolving, making them difficult to study in this respect (Ulitsky, 2016). To address this, different criteria have been proposed to interpret conservation of lncRNAs, including splicing patterns or exonic structures, secondary structures, syntenic positions and modes of action (Ulitsky, 2016). Although conserved lncRNAs are generally believed to have similar functions, a recent study by Guo *et al* (2020) reported a pair of positionally conserved lncRNAs in mouse and human, mFAST, and hFAST, which show different functions owing to their differences in RNA processing and subcellular localization. This suggests that there are additional factors that can influence and contribute to the functional evolution of conserved lncRNAs.

Arginine and glutamate-rich protein 1 (Arglu1) is a highly conserved gene found in *Drosophila melanogaster* (known as CG31712) and human. The function of ARGLU1 protein is not well characterized, but it has been reported to interact with Mediator subunit 1 protein (MED1) (Zhang *et al*, 2011) and involved in transcription

1 Temasek Life Sciences Laboratory, National University of Singapore, Singapore, Singapore
2 Department of Biological Sciences, National University of Singapore, Singapore, Singapore
*Corresponding author. Tel: +65 6872 7818; E-mail: junwei@tll.org.sg

and alternative splicing of genes mediated by nuclear receptors like estrogen receptor and glucocorticoid receptor (Zhang *et al*, 2011; Magomedova *et al*, 2019). Interestingly, a previous study by Pirnie *et al* (2017) indicates that the intron 2 of Arglu1 gene is orthologous in vertebrates as it contains an Ultraconserved element (UCE) that shows 95% sequence conservation for 500 nucleotides between human and chicken. Within the UCE region contains regulatory elements affecting alternative splicing (AS) of Arglu1 gene leading to AS-nonsense mediated decay (AS-NMD) and intron retention (Pirnie *et al*, 2017).

Here we report a pair of positionally conserved sisRNAs produced from the Arglu1 gene loci in human and *Drosophila melanogaster*, namely the human Arglu1 sisRNA and *Drosophila* Arglu1 (dArglu1) sisRNA. In both species, the Arglu1 genes have the same exon-intron structures; however, the sequences of the orthologous intron 2 are highly divergent. Arglu1 sisRNA is formed via retention of the orthologous intron, while dArglu1 sisRNA is produced via premature cleavage of the same intron. Both sisRNAs perform autoregulatory functions by affecting the splicing of their host genes, suggesting that sisRNAs help in maintaining homeostatic gene regulation. Interestingly, whereas Arglu1 sisRNA promotes splicing, dArglu1 sisRNA represses splicing and this can be linked to their different biogenesis processes. Taken together, our data suggest that the distinct biogenesis and processing pathways may act as a previously unappreciated factor that contributes to the functional divergence and evolution of these otherwise positionally conserved sisRNAs.

Results

Identification of a sisRNA from the human Arglu1 locus

The human Arglu1 gene locus produces three different splicing isoforms; an mRNA, an unstable NMD transcript and an intron 2-containing transcript (Fig 1A). Notably, the intron 2-containing transcript is stably localized in the nucleus in HEK293, MCF-7 and HeLa cell lines, and it is not regulated by NMD (Pirnie *et al*, 2017). To examine if the intron 2-containing transcript is a sisRNA, we first checked the presence of this transcript using primers spanning the junction of exon 2 and intron 2. We detected the transcript in MCF-7, MDA-MB-231, T47D, and MDA-MB-468 breast cancer cell lines (Fig 1B). Next, we characterized the intron 2-containing transcript (hereafter named Arglu1 sisRNA) in MCF-7 cells. To check the stability of the Arglu1 sisRNA, we subjected the cells to α-amanitin treatment, which inhibits transcription (Fig 1C). Actin and c-Myc mRNA were used as the stable and unstable RNA control respectively (Appendix Fig S1). Indeed, the results show that Arglu1 sisRNA is relatively more stable than the pre-mRNA (Fig 1D).

Consistent with previous observations that Arglu1 sisRNA localizes in the nucleus (Pirnie *et al*, 2017), our nuclear-cytoplasmic fractionation (Fig 1E) and single molecule fluorescent *in situ* hybridization (smFISH) (Fig 1F) experiments showed that Arglu1 sisRNA is localized in the nucleus of MCF-7 cells. We employed smFISH to quantify the transcript copy number of Arglu1 sisRNA (Buxbaum *et al*, 2014). To distinguish Arglu1 sisRNA from Arglu1 pre-mRNA, both of which contain intron 2, we designed specific probes targeting introns 1 and 2 (Fig 1B). Transcripts with signals exclusively from intron 2 probes were considered as sisRNAs, while transcripts with signals from intron 1 or intron 1 and 2 probes were considered as pre-mRNAs. The results showed that Arglu1 sisRNA is approximately 5-fold more abundant than the Arglu1 pre-mRNA, with each MCF-7 cell containing ~ 25 copies of Arglu1 sisRNAs and ~ 5 copies of Arglu1 pre-mRNAs (Fig 1G).

A previous study by Pirnie *et al* (2017) reported that the UCE region is important for unproductive splicing and intron 2 retention, suggesting its role in biogenesis of Arglu1 sisRNA. Manipulation of the UCE region in a reporter minigene and in genomic context led to a decrease in intron retention and Arglu1 sisRNA production (Pirnie *et al*, 2017). To directly examine the role of the UCE in the biogenesis of Arglu1 sisRNA, we employed three antisense morpholino oligonucleotides (AMOs) to block the UCE sequences on Arglu1 pre-mRNA (Fig 1H). The levels of Arglu1 sisRNA were reduced while the levels of spliced Arglu1 mRNA were increased in the AMO-treated cells as compared to the control, indicating a reduction in Arglu1 sisRNA production (Fig 1I). Taken together, these findings suggest that the UCE is important for the biogenesis of Arglu1 sisRNA.

Arglu1 sisRNA is involved in autoregulation of its host gene

Increased levels of ARGLU1 protein have been shown to repress its pre-mRNA splicing suggesting that unproductive splicing of the Arglu1 intron 2 may lead to increased levels of Arglu1 sisRNA (Pirnie *et al*, 2017). We hypothesized that Arglu1 sisRNA may be involved in the autoregulation of its host gene, possibly by modulating the ARGLU1 protein-mediated negative feedback loop. We knocked down Arglu1 sisRNA using two antisense oligonucleotides (ASOs) targeting intron 2 and measured the splicing efficiency of intron 2 (Fig 2A). We found that Arglu1 sisRNA—and not Arglu1 pre-mRNA—levels decreased following Arglu1 sisRNA knockdown (KD), indicating that the ASOs target the sisRNA specifically (Fig 2B and C). Importantly, Arglu1 sisRNA KD led to a significant downregulation of the Arglu1 intron 2 splicing efficiency (Fig 2D), suggesting that Arglu1 sisRNA promotes splicing of Arglu1 intron 2. Similar results were obtained in MDA-MB-231 and HEK293T cells (Fig EV1A–C). We also observed that mature Arglu1 mRNA was downregulated in Arglu1 sisRNA KD cells (Fig 2E and F), concomitant with a drop in ARGLU1 protein levels following a three-day KD (Fig 2G). While Arglu1 sisRNA promotes Arglu1 pre-mRNA splicing, we did not observe an increase in pre-mRNA levels in Arglu1 sisRNA KD cells. This is because the Arglu1 pre-mRNA has been spliced to form the AS-NMD transcript instead, as shown in emetine treated Arglu1 sisRNA KD cells (Fig EV1D).

We are aware that ASOs targeting intronic sequences may target pre-mRNAs and affect their stability (Lai *et al*, 2020; Lee & Mendell, 2020). However, there are also cases that show that ASOs do not target endogenous pre-mRNAs (Yin *et al*, 2012; Zhang *et al*, 2013). Thus, such observations cannot be generalized and the specificity of ASOs has to be determined empirically by optimizing the concentration used. Although the designed ASOs targeting intron 2 can in principle bind to Arglu1 pre-mRNA, we reasoned that at the concentration we were using, the intron 2 ASOs are inclined to specifically target and downregulate the levels of sisRNA due to their higher abundance (Fig 1D, F and G). To examine this, we designed ASOs targeting the flanking introns (introns 1 and 3

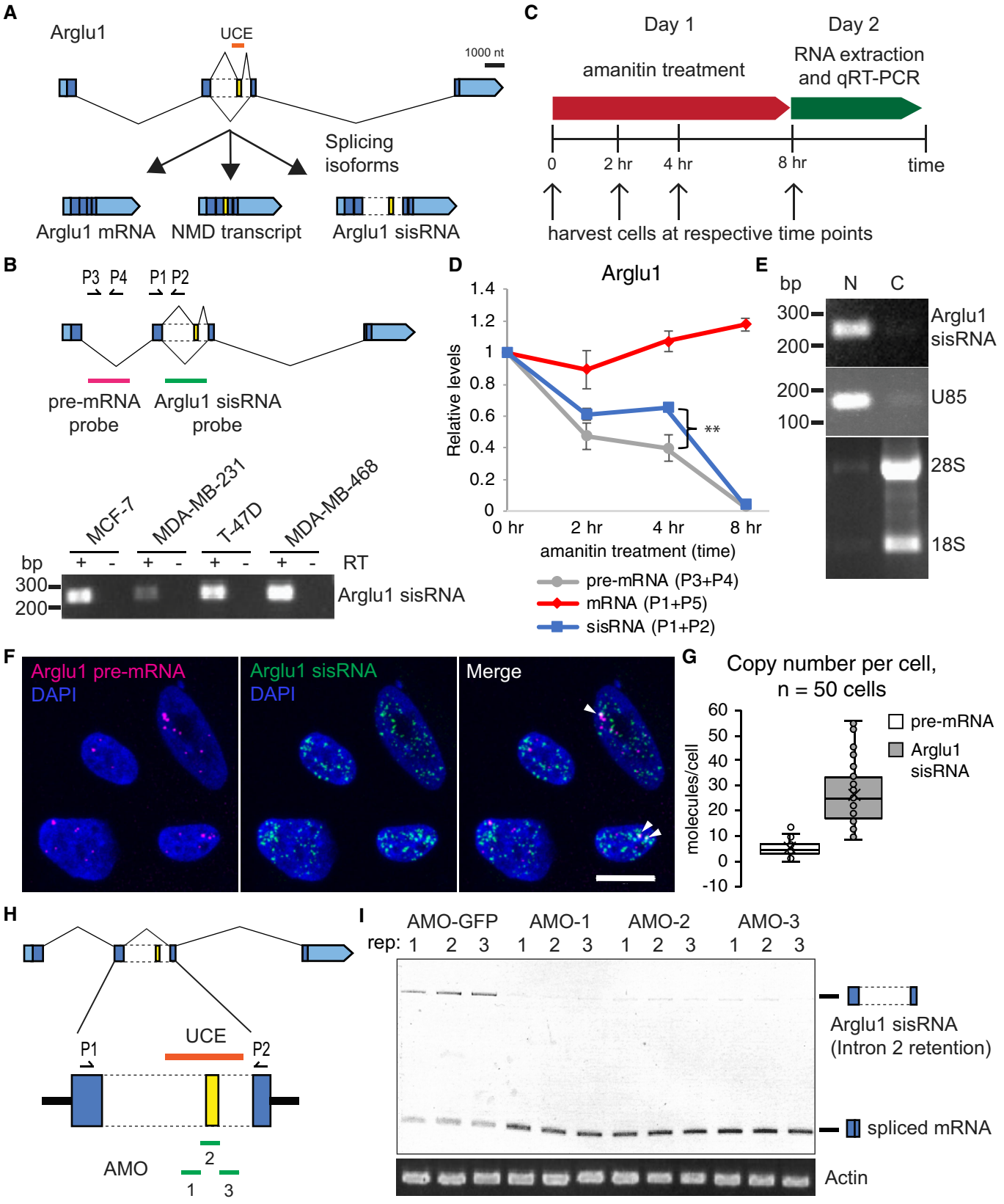
A
Arglu1
UCE
1000 nt
Splicing isoforms
Arglu1 mRNA
NMD transcript
Arglu1 sisRNA
C
Day 1
Day 2
amanitin treatment
RNA extraction and qRT-PCR
0
2 hr
4 hr
8 hr
time
harvest cells at respective time points
B
P3 P4
P1 P2
pre-mRNA probe
Arglu1 sisRNA probe
MCF-7
MDA-MB-231
T-47D
MDA-MB-468
bp
RT
300
200
Arglu1 sisRNA
D
Arglu1
Relative levels
amanitin treatment (time)
pre-mRNA (P3+P4)
mRNA (P1+P5)
sisRNA (P1+P2)
**
E
bp
N
C
Arglu1 sisRNA
U85
28S
18S
F
Arglu1 pre-mRNA
DAPI
Arglu1 sisRNA
DAPI
Merge
G
Copy number per cell, n = 50 cells
molecules/cell
pre-mRNA
Arglu1 sisRNA
H
P1
UCE
P2
AMO
I
AMO-GFP
AMO-1
AMO-2
AMO-3
rep:
Arglu1 sisRNA (Intron 2 retention)
spliced mRNA
Actin

Figure 1.

Figure 1. Characterization and biogenesis of Arglu1 sisRNA.

A Schematics showing the Arglu1 gene model and the different splicing isoforms.
B Top: Arglu1 gene locus indicating the primers' location (black arrow) and smFISH probes designed for Arglu1 sisRNA (green line) and Arglu1 pre-mRNA (magenta line). Bottom: RT-PCR showing the presence of Arglu1 sisRNA (P1 + P2) detected from the different breast cancer cell lines.
C Experimental scheme for transcription inhibition treatment using alpha-amanitin and the indicated time points for cell harvesting.
D qPCR of Arglu1 sisRNA, pre-mRNA and mRNA expression levels normalized to *gapdh* from alpha-amanitin treatment at indicated time points. ($n = 3$ biological replicates). Data are presented as mean ± SEM. $^{**}P \leq 0.01$ (Student's *t*-test). See also Appendix Fig S1.
E RT-PCR indicating presence of Arglu1 sisRNA in the nuclear (N) and but not in the cytoplasmic (C) fraction. U85 gene was used as nuclear marker and 28S and 18S rRNA as cytoplasmic marker.
F smFISH of Arglu1 pre-mRNA (magenta) and Arglu1 sisRNA (green). Merged dots (white, arrowheads) represent Arglu1 pre-mRNA transcript. Scale bar: 20 μm.
G Quantification of Arglu1 pre-mRNA and Arglu1 sisRNA copy number from (F). ($n = 50$ cells) Cross, mean; middle line, median; box, 25^{th}–75^{th} percentiles; whiskers, minimum to maximum.
H Diagram indicating the locations of antisense morpholino oligonucleotides (AMOs) targeting the UCE region of Arglu1 intron 2 and primers used in (I).
I RT-PCR of Arglu1 sisRNA, spliced mRNA levels and *actin* as control after AMOs treatment.

which are not present in the sisRNA) of the lower abundant Arglu1 pre-mRNA and the Arglu1 pre-mRNAs levels were significantly downregulated following transfection of these ASOs at the same concentration (Fig EV2A–C). Thus, at the concentration that we were using, appropriate ASOs can deplete pre-mRNA but not those that were designed to target the sisRNA. To further validate the specificity and effects of the sisRNA ASOs, we checked the levels of Arglu1 sisRNA and pre-mRNA with additional primers at different locations and consistent results were obtained whereby Arglu1 sisRNA levels were downregulated while Arglu1 pre-mRNA remained unchanged (Fig EV2D–F). Moreover, we also quantified the copy number of nascent Arglu1 pre-mRNA via smFISH and the results corroborated the qPCR findings (Fig EV2G). Taken together, these control experiments independently confirm that at the concentration we were using, our ASOs that target Arglu1 sisRNA specifically deplete the sisRNA but not pre-mRNA.

To test whether the UCE in the Arglu1 sisRNA is important for its autoregulatory function, we overexpressed (OE) the full length (WT) Arglu1 intron 2 and Arglu1 intron 2 with UCE-deleted (Δ) (Fig 2H) in MCF-7 cells and checked the effects on endogenous Arglu1 pre-mRNA splicing. OE of Arglu1 intron 2 WT, but not ΔUCE, led to an increase in the splicing index of endogenous Arglu1 (Fig 2I). Concomitantly, the levels of endogenous Arglu1 sisRNA were significantly downregulated in Arglu1 intron 2 WT OE cells as compared to control and ΔUCE OE cells (Fig 2J).

Next, we examined whether Arglu1 sisRNA KD has any effects on the expression of downstream genes regulated by ARGLU1 protein. A previous study by Zhang *et al* (2011) showed that ARGLU1 protein interacts with MED1 and is required for the transcription of estrogen responsive genes such as c-Myc and pS2 (Zhang *et al*, 2011). We knocked down the levels of Arglu1 sisRNA for 3 days and treated the MCF-7 cells using estrogen for 2 h. The expression of c-Myc and pS2 was significantly downregulated after estrogen treatment in Arglu1 sisRNA KD cells (Fig 2K and L). We observed a similar downregulation in cFOS expression (Fig EV1E), another estrogen-responsive gene involved in the growth of breast cancer cells (Lu *et al*, 2005; Bittencourt *et al*, 2008). We also examined the effects of Arglu1 sisRNA on the alternative splicing of AXIN1, a gene involved in breast cancer, since it is also mediated by estrogen receptor (Bhat-Nakshatri *et al*, 2013). Indeed, following Arglu1 sisRNA KD, we saw a downregulation on the splicing of the AXIN1a variant and a reduction in the AXIN1a/AXIN1b splicing ratio (Fig 2M). Taken together, our results suggest that Arglu1 sisRNA is involved in the autoregulation of its host gene by promoting the splicing of intron 2 in a UCE-dependent manner (Fig 2N).

Arglu1 sisRNA binds to ARGLU1 protein and inhibits its interaction with Arglu1 pre-mRNA

We hypothesized that sisRNA-mediated autoregulation of Arglu1 splicing occurs in a two-step process. First, ARGLU1 protein binds to intron 2 of a nascent Arglu1 pre-mRNA that is being actively transcribed at the gene locus and gives rise to Arglu1 sisRNA; Second, Arglu1 sisRNA binds to the ARGLU1 protein and sequesters the protein at a distant location from the gene locus, hence alleviating the inhibitory effect of ARGLU1 protein on Arglu1 pre-mRNA splicing.

We first examined the interaction between ARGLU1 protein with Arglu1 pre-mRNA and sisRNA by performing RNA-immunoprecipitation (RIP). ARGLU1 protein binds to Arglu1 sisRNA with an approximately 5-fold enrichment when compared to the no antibody control, while no enrichment of Arglu1 pre-mRNA was detected (Fig 3A). Since Arglu1 sisRNA is more abundant and stable than the transient Arglu1 pre-mRNA, we speculated that the immunoprecipitated ARGLU1 protein is likely saturated with Arglu1 sisRNA. To test this idea, we knocked down Arglu1 sisRNA using ASO1 for 1 day prior to RIP. Knocking down the Arglu1 sisRNA for 1 day did not affect ARGLU1 protein levels (Fig 3B) in contrast to 3-day KD (Fig 2G). Interestingly, a modest but significant enrichment of Arglu1 pre-mRNA at approximately 1.4-fold could now be observed (Fig 3C). These results suggest that ARGLU1 protein can interact with both Arglu1 sisRNA and Arglu1 pre-mRNA but is bound preferentially to Arglu1 sisRNA due to its abundance. By reducing the levels of Arglu1 sisRNA, more unbound or "free" ARGLU1 protein can then interact with Arglu1 pre-mRNA, resulting in its enrichment.

To visualize the colocalization between ARGLU1 protein, Arglu1 sisRNA, and Arglu1 pre-mRNA, we coupled smFISH of Arglu1 sisRNA and pre-mRNA with ARGLU1 protein immunofluorescence (IF). Z-slices images of the stained MCF-7 cells showed that ARGLU1 protein forms foci and indeed colocalizes with Arglu1 sisRNA and Arglu1 pre-mRNA (Fig 3D and E). Next, to check if levels of Arglu1 sisRNA affect the colocalization of ARGLU1 protein and Arglu1 pre-mRNA, we knocked down Arglu1 sisRNA for 1 day before performing smFISH with ARGLU1 protein IF. The results showed that Arglu1 sisRNAs were clearly reduced after Arglu1 sisRNA KD and the Arglu1 pre-mRNA signals remained unchanged (Fig 3F–H), consistent with our qPCR data (Fig 2B and C). We quantified and

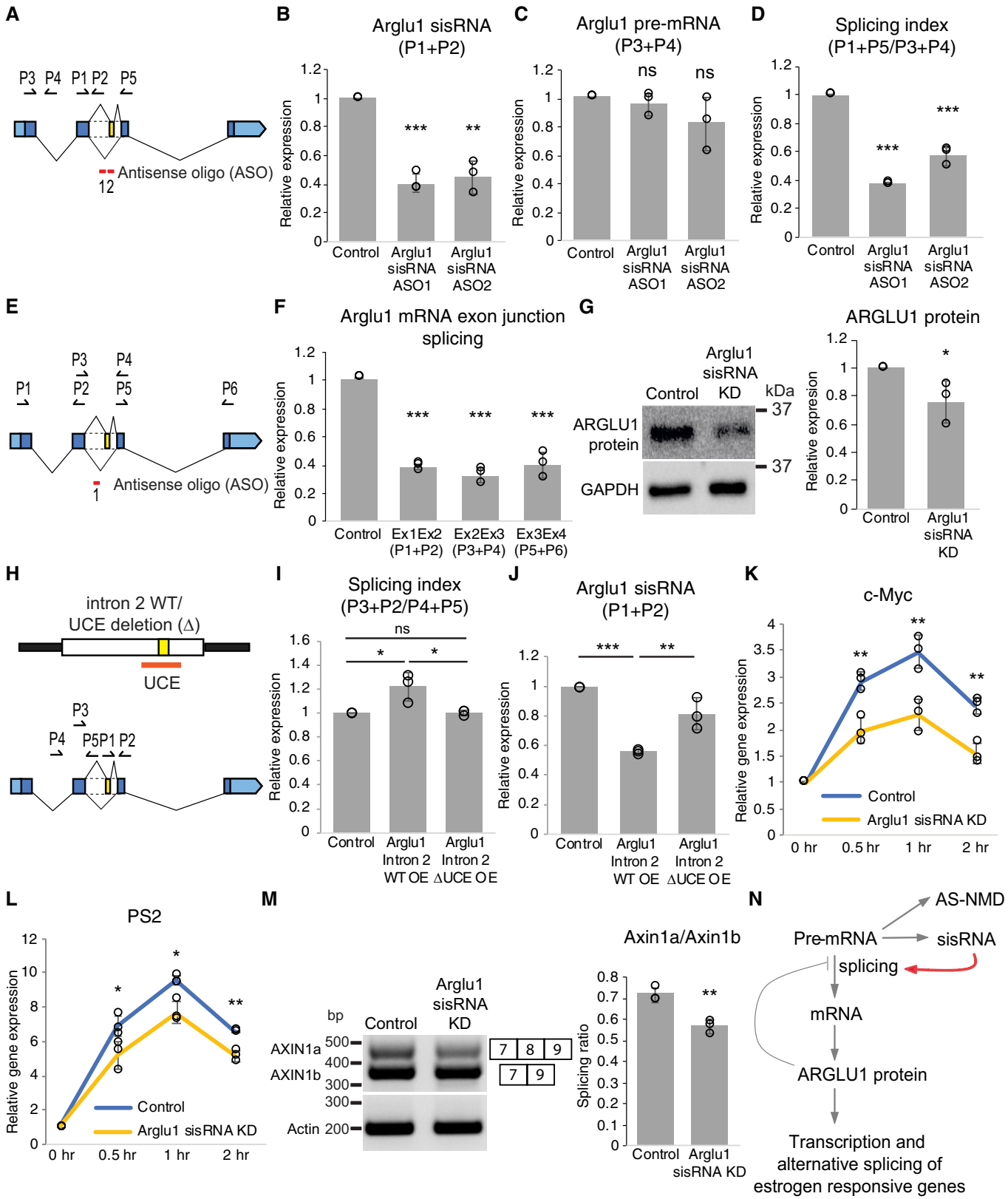
A
P3 P4 P1 P2 P5
Antisense oligo (ASO)
12
B
Arglu1 sisRNA (P1+P2)
Relative expression
Control
Arglu1 sisRNA ASO1
Arglu1 sisRNA ASO2
C
Arglu1 pre-mRNA (P3+P4)
ns
D
Splicing index (P1+P5/P3+P4)
E
P1 P2 P3 P4 P5 P6
Antisense oligo (ASO)
1
F
Arglu1 mRNA exon junction splicing
Ex1Ex2 (P1+P2)
Ex2Ex3 (P3+P4)
Ex3Ex4 (P5+P6)
G
Arglu1 sisRNA KD
kDa
ARGLU1 protein
GAPDH
37
ARGLU1 protein
H
intron 2 WT/ UCE deletion (Δ)
UCE
P3 P4 P5P1 P2
I
Splicing index (P3+P2/P4+P5)
Arglu1 Intron 2 WT OE
Arglu1 Intron 2 ΔUCE OE
J
Arglu1 sisRNA (P1+P2)
K
c-Myc
Relative gene expression
Control
Arglu1 sisRNA KD
0 hr
0.5 hr
1 hr
2 hr
L
PS2
M
Arglu1 sisRNA KD
bp
AXIN1a
AXIN1b
Actin
Axin1a/Axin1b
Splicing ratio
N
AS-NMD
Pre-mRNA
sisRNA
splicing
mRNA
ARGLU1 protein
Transcription and alternative splicing of estrogen responsive genes

Figure 2.

Figure 2. Arglu1 sisRNA is involved in autoregulation of its host gene in a UCE-dependent manner.

A Arglu1 gene locus indicating the locations of antisense oligonucleotides (ASOs) (red lines) designed for Arglu1 sisRNA KD and primers used in (B–D).
B, C qPCR of Arglu1 sisRNA and Arglu1 pre-mRNA expression levels in Arglu1 sisRNA ASO1 and ASO2 cells vs. control cells. ($n = 3$ biological replicates).
D Chart showing the splicing indices (spliced/unspliced ratio) in Arglu1 sisRNA ASO1 and ASO2 cells vs. control cells as measured by qPCR. ($n = 3$ biological replicates).
E Arglu1 gene locus indicating the location of antisense oligonucleotide (ASO) (red line) designed for Arglu1 sisRNA KD and primers used in (F).
F qPCR of Arglu1 mRNA expression levels at different exon junctions in Arglu1 sisRNA KD cells vs. control cells. ($n = 3$ biological replicates).
G Left: Western blot analysis of ARGLU1 protein and GAPDH as loading control in control cells and Arglu1 sisRNA KD cells. Right: Quantification of relative ARGLU1 protein levels normalized to GADPH in Arglu1 sisRNA KD cells vs. control cells from western blot analysis (left). ($n = 3$ biological replicates).
H Top: Schematic showing minigene expressing Arglu1 intron 2 wild type (WT) or intron 2 with UCE deletion (Δ; location of UCE as indicated by orange bar). Bottom: Arglu1 gene locus indicating the primers' location (black arrows) for (I and J).
I Chart showing the splicing indices (spliced/unspliced ratio) in Arglu1 intron 2 WT OE cells and Arglu1 Intron 2 ΔUCE OE cells as measured by qPCR. ($n = 3$ biological replicates).
J qPCR of Arglu1 sisRNA levels in Arglu1 intron 2 WT OE cells and Arglu1 Intron 2 ΔUCE OE cells vs. control cells ($n = 3$ biological replicates).
K, L qPCR of c-Myc and pS2 expression levels in 3 days Arglu1 sisRNA KD cells vs. control cells under 100 nM estrogen treatment at indicated time points. ($n = 3$ biological replicates).
M Left: RT-PCR of AXIN1a and AXIN1b spliced levels and *actin* as loading control in 3 days Arglu1 sisRNA KD cells vs. control cells under 100 nM estrogen treatment for 2 h. Right: Quantification of relative spliced AXIN1a/AXIN1b expression ratio normalized to *actin* from RT-PCR analysis (left). ($n = 3$ biological replicates).
N Schematics of Arglu1 gene autoregulatory loop model indicating involvement of Arglu1 sisRNA in promoting splicing of host gene (red arrow).

Data information: In (B–D, F, G, I–M), data are presented as mean ± SEM. ns, not significant, $*P \leq 0.05$, $**P \leq 0.01$, and $***P \leq 0.001$ (Student's *t*-test). See also Figs EV1 and EV2.

compared the extent of colocalization between Arglu1 pre-mRNA and ARGLU1 protein and our results confirmed that colocalization of ARGLU1 protein with Arglu1 pre-mRNA increased when levels of Arglu1 sisRNA were reduced (Fig 3I; Appendix Fig S2). Taken together, our experiments provide evidence that Arglu1 sisRNA prevents ARGLU1 protein from binding to Arglu1 pre-mRNA.

ARGLU1 protein forms dynamic foci at nuclear speckles

Next, we wanted to see if ARGLU1 protein is indeed sequestered at locations distal to its gene locus in the nucleus. ARGLU1 protein colocalizes with MED1 protein and localizes to nuclear speckles by interacting with SC35 (SRSF2) in the nucleus (Zhang *et al*, 2011; Magomedova *et al*, 2019). Thus, we wondered if ARGLU1 protein may be sequestered in nuclear speckles. We performed IF of ARGLU1 protein against SC35, a marker for nuclear speckle. Our results showed that a majority of the ARGLU1 foci co-localize with SC35 (Appendix Fig S3). Next, we examined the colocalization pattern of ARGLU1 protein with SC35 and MED1 by transiently expressing mCherry-ARGLU1 in MCF-7 cells before performing SC35 and MED1 IF. Our results showed that ARGLU1 protein colocalizes with both MED1 and SC35 in nuclear speckles (Fig 4A).

To examine the dynamics of ARGLU1 protein in the nuclear speckle, we performed fluorescence recovery after photobleaching (FRAP) assay on live cells transiently expressing the mCherry-ARGLU1 protein. As a positive control, we used GFP-SRSF2 (also SC35), which had previously been shown to form phase-separated liquid droplets and is able to recover after photo-bleaching (Greig *et al*, 2020). Interestingly, mCherry-ARGLU1 foci were able to recover to about 75% intensity after photo-bleaching comparable to that of GFP-SRSF2 (Fig 4B and C), suggesting that the ARGLU1 protein is dynamic and exhibits mobility.

Arglu1 sisRNA promotes robust ARGLU1 protein phase separation and localization to nuclear speckles

It has been proposed that lncRNA is able to sequester and drive phase separation of RNA-binding proteins (RBPs) as a mechanism to regulate the activity of the RBPs (Fox *et al*, 2018; Elguindy & Mendell, 2021). We therefore hypothesized that Arglu1 sisRNA sequesters and assists ARGLU1 protein localization to the nuclear speckles by promoting phase separation. To test this hypothesis, we knocked down Arglu1 sisRNA for 1 day prior all experiments (Fig 3B) to exclude effects that may arise from variation in ARGLU1 protein levels and examined if it has any effect on ARGLU1 protein foci formation. ARGLU1 protein IF in Arglu1 sisRNA KD cells showed dispersed ARGLU1 foci formation and reduced co-localization with nuclear speckles (as marked by SC35) (Fig 4D). This result suggests that Arglu1 sisRNA promotes robust localization of ARGLU1 protein to nuclear speckles.

Next, we examined whether Arglu1 sisRNA can promote phase separation of ARGLU1 protein *in vitro*. We generated recombinant ARGLU1 protein tagged with GFP at the C-terminal (ARGLU1-GFP) and tested its ability to phase separate *in vitro*. Droplet assay showed that ARGLU1-GFP protein formed droplets where the size and number of droplets increased in a concentration-dependent manner. GFP protein was used as control and no droplets were formed (Fig EV3A). To directly examine the roles of Arglu1 sisRNA and the UCE region in promoting phase separation of ARGLU1 protein, we added *in vitro* transcribed (IVT) Arglu1 intron 2 RNA with (intron 2 WT) or without UCE region (intron 2 ΔUCE) with ARGLU1-GFP protein and examined the number and size of ARGLU1-GFP protein droplet formed. Addition of intron 2 WT RNA, but not intron 2 ΔUCE RNA, increased the number of droplets formed as compared to ARGLU1-GFP only (Fig 4E and F). Moreover, ARGLU1-GFP protein incorporated with intron 2 WT RNA showed an increase in droplet size as compared to droplets that contain only ARGLU1-GFP protein (Fig 4G). These results suggest that Arglu1 sisRNA promotes phase separation of ARGLU1 protein with the help of the UCE. Taken together, our *in vivo* and *in vitro* experiments suggest that Arglu1 sisRNA promotes ARGLU1 protein phase separation and robust localization to the nuclear speckles.

Formation of ARGLU1 nuclear bodies may assist in transcription of estrogen-responsive genes

We next examined whether the incorporation of ARGLU1 protein into its nuclear bodies is associated with the transcription of estrogen-responsive genes. We first performed 1-day Arglu1 sisRNA

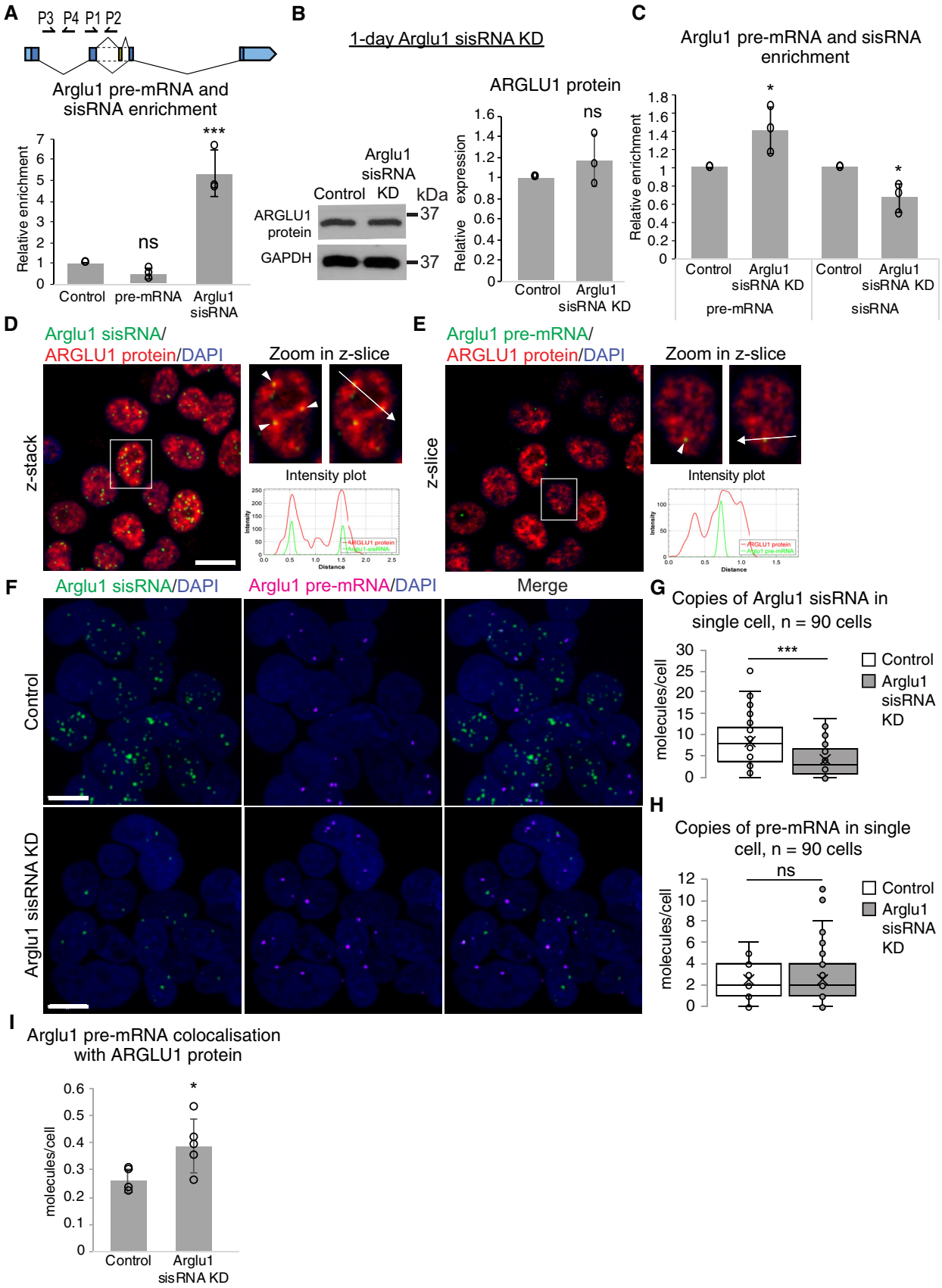
A
P3 P4 P1 P2
Arglu1 pre-mRNA and sisRNA enrichment
Relative enrichment
***
ns
Control pre-mRNA Arglu1 sisRNA
B
1-day Arglu1 sisRNA KD
Arglu1 sisRNA
Control KD kDa
ARGLU1 protein
GAPDH
37
37
ARGLU1 protein
Relative expression
ns
Control Arglu1 sisRNA KD
C
Arglu1 pre-mRNA and sisRNA enrichment
Relative enrichment
Control Arglu1 sisRNA KD
Control Arglu1 sisRNA KD
pre-mRNA
sisRNA
D
Arglu1 sisRNA/
ARGLU1 protein/DAPI
Zoom in z-slice
z-stack
Intensity plot
E
Arglu1 pre-mRNA/
ARGLU1 protein/DAPI
Zoom in z-slice
z-slice
Intensity plot
F
Arglu1 sisRNA/DAPI
Arglu1 pre-mRNA/DAPI
Merge
Control
Arglu1 sisRNA KD
G
Copies of Arglu1 sisRNA in single cell, n = 90 cells
molecules/cell
***
Control
Arglu1 sisRNA KD
H
Copies of pre-mRNA in single cell, n = 90 cells
molecules/cell
ns
Control
Arglu1 sisRNA KD
I
Arglu1 pre-mRNA colocalisation with ARGLU1 protein
molecules/cell
*
Control
Arglu1 sisRNA KD

Figure 3.

Figure 3. Arglu1 sisRNA binds ARGLU1 protein and inhibits its interaction with Arglu1 pre-mRNA.

A Top: Arglu1 gene locus indicating the primers' location (black arrows) for (A–C). Bottom: RIP-qPCR of Arglu1 pre-mRNA (P3 + P4) and Arglu1 sisRNA (P1 + P2) levels from ARGLU1 protein immunoprecipitation. Samples with beads and no antibodies were used as control. ($n = 3$ biological replicates).

B Left: Western blot analysis of ARGLU1 protein and GAPDH as loading control in 1 day Arglu1 sisRNA KD cells vs. control cells. Right: Quantification of relative ARGLU1 protein expression level normalized to GADPH in 1 day Arglu1 sisRNA KD cells vs. control cells from western blot analysis (left). ($n = 3$ biological replicates).

C RIP-qPCR of Arglu1 pre-mRNA and Arglu1 sisRNA levels normalized to *gapdh* in 1 day Arglu1 sisRNA KD cells vs. control cells from ARGLU1 protein immunoprecipitation. ($n = 3$ biological replicates).

D Left: smFISH of Arglu1 sisRNA (green) coupled with immunostaining of ARGLU1 protein (red) shown in z-stack. Right: Zoom in z-slice on the boxed area showing co-localization (arrowheads) of Arglu1 sisRNA with ARGLU1 protein. The intensity plot shows the signal quantification and co-localization of Arglu1 sisRNA and ARGLU1 protein in the direction of the white arrow. Scale bar: 20 μm.

E Left: smFISH of Arglu1 pre-mRNA (green) coupled with ARGLU1 protein (red) immunostaining shown in z-stack. Right: Zoom in z-slice on the boxed area showing co-localization (arrowhead) of Arglu1 pre-mRNA with ARGLU1 protein. The intensity plot shows the signal quantification and co-localization of Arglu1 pre-mRNA and ARGLU1 protein in the direction of the white arrow.

F smFISH of Arglu1 sisRNA (green) and Arglu1 pre-mRNA (magenta) in 1 day Arglu1 sisRNA KD cells vs. control cells. Scale bar: 20 μm.

G, H Quantification of Arglu1 sisRNA and Arglu1 pre-mRNA copy number from (F). ($n = 90$ cells). Cross, mean; middle line, median; box, 25^{th}–75^{th} percentiles; whiskers, minimum to maximum.

I Quantification of Arglu1 pre-mRNA and ARGLU1 protein co-localization from Arglu1 pre-mRNA smFISH coupled with ARGLU1 immunostaining in 1 day Arglu1 sisRNA KD cells vs. control cells. ($n = 5$ biological replicates).

Data information: In (A–C, and G–I), data are presented as mean ± SEM. ns, not significant, $*P \leq 0.05$ and $***P \leq 0.001$ (Student's *t*-test). See also Appendix Fig S2.

KD and IF of ARGLU1 protein at 0, 0.5, and 2 h after estrogen treatment and observed the formation of nuclear bodies (Fig EV3B). To analyze the results quantitatively, we categorized the cells into three groups based on the size of the foci they contain (i.e., big, intermediate, and small; see Materials and Methods; Fig EV3C). Next, we counted the cells belonging to different foci categories at each time point and compared the counts between Arglu1 sisRNA KD and control cells (Fig EV3D). The results showed that there was a significantly lower percentage of cells with big and intermediate foci (about 15–17%) in Arglu1 sisRNA KD cells at 0 and 0.5 h time point as compared to control cells (30–40%). However, at 2 h time point, the percentage of cells containing big and intermediate foci from the Arglu1 sisRNA KD cells reached a similar level to control cells (about 28–30% for both), suggesting a "recovery" in foci formation. Interestingly, the expression of estrogen-responsive genes c-Myc and pS2 correlate with the changes in the ARGLU1 nuclear body formation. Transcript levels of c-Myc under estrogen treatment were significantly reduced during the first 0.5 h but subsequently increased to a level comparable to control cells (Fig EV3E). A similar delay in transcriptional activity was also observed for pS2 (Fig EV3F).

To directly test if ARGLU1 nuclear body formation can regulate c-Myc expression, we overexpressed mCherry-ARGLU1 in MCF-7 cells and observed that ARGLU1 protein formed much bigger nuclear bodies as compared to those in nontransfected cells (Fig EV3G). This was also accompanied by a significant increase in c-Myc expression (Fig EV3H). Our experiments suggest that increase in ARGLU1 protein levels leads to bigger ARGLU1 nuclear body formation and may play a role in promoting the transcription of target genes.

Identification of a sisRNA from the *Drosophila* dArglu1 locus

Having characterized the Arglu1 sisRNA in human cells, we moved on to examine whether sisRNA is produced from the conserved Arglu1 gene locus in *Drosophila*. The CG31712 gene, hereafter called dArglu1, possesses a similar exon-intron structure with the human Arglu1 gene, consisting of 4 exons and 3 introns and the intron 2 is positionally conserved between the conserved exons 2 and 3 (Fig 5A; Appendix Fig S4). Information on FlyBase (CG31712-RB, FlyBase ID: FBtr0331636) and studies on human Arglu1 (Pirnie *et al*, 2017) suggest that the intron 2 of both loci are orthologous (Fig 5A) as they contain regulatory elements that modulate the alternative splicing-NMD pathways. Since human Arglu1 sisRNA biogenesis requires retention of intron 2, we asked whether the orthologous intron 2 in dArglu1 locus can also lead to the production of dArglu1 sisRNA through intron retention.

To identify sisRNAs in *Drosophila*, we used unfertilized eggs (Fig 5B) as they are transcriptionally quiescent and contain a pool of stable RNAs (Lasko, 2012; Pek *et al*, 2015). Based on the previous RNA sequencing data from unfertilized eggs, we found reads that were mapped to the intron 2 of dArglu1 at the poly(A) + fraction, suggesting the presence of transcripts containing intron 2 sequences (Fig 5C) (Ng *et al*, 2018b). To check if the dArglu1 intron 2 is also retained, we designed primers targeting the 5′ and 3′ region of the intron and performed reverse transcription (RT)-PCR using RNAs extracted from unfertilized eggs. However, a band was detected only at the 5′ region but not the 3′ region of intron 2 (Fig 5D). The primer pair targeting the 3′ region was shown to be working by using *Drosophila* genomic DNA (gDNA) as the template (Fig EV4A). Consequently, we searched the FlyBase and found an entry for an unannotated cDNA (FlyBase ID:FBcl0742967, GenBank:BT132987) which covers from exon 2 to a premature termination in intron 2, corresponding to the region where abundant reads were mapped (Fig 5C). Interestingly, two cryptic motifs for premature polyadenylation sites (PASs) are found upstream of the 3′ end of this transcript (Gruber *et al*, 2016; Sanfilippo *et al*, 2017) (Figs 5E and EV4B). These led us to speculate that dArglu1 sisRNA may be produced by premature cleavage of intron 2.

To confirm that, we checked for the presence of the full-length transcript by designing a forward primer from the 5′ end and two reverse primers, before and after the 3′ end respectively (Fig 5C). RT-PCR using RNAs from unfertilized eggs showed that a specific transcript was amplified only with the primer pair before the 3′ end, hereafter named as dArglu1 sisRNA (Fig 5F). Finally, we confirmed the 3′ end sequence of dArglu1 sisRNA using 3′ RACE (Fig EV4B). We also verified the stability of dArglu1 sisRNA in S2 cells using α-amanitin treatment (Fig EV4C). Quantification of dArglu1 sisRNA revealed ~ 160,000–270,000 copies per mature oocyte, which translates to ~ 20–33 copies per somatic cell in the embryo (after 13 cell division with 2^{13} cells). These results show that dArglu1 sisRNA is a stable RNA transcript formed by premature cleavage of intron 2 (Fig 5E).

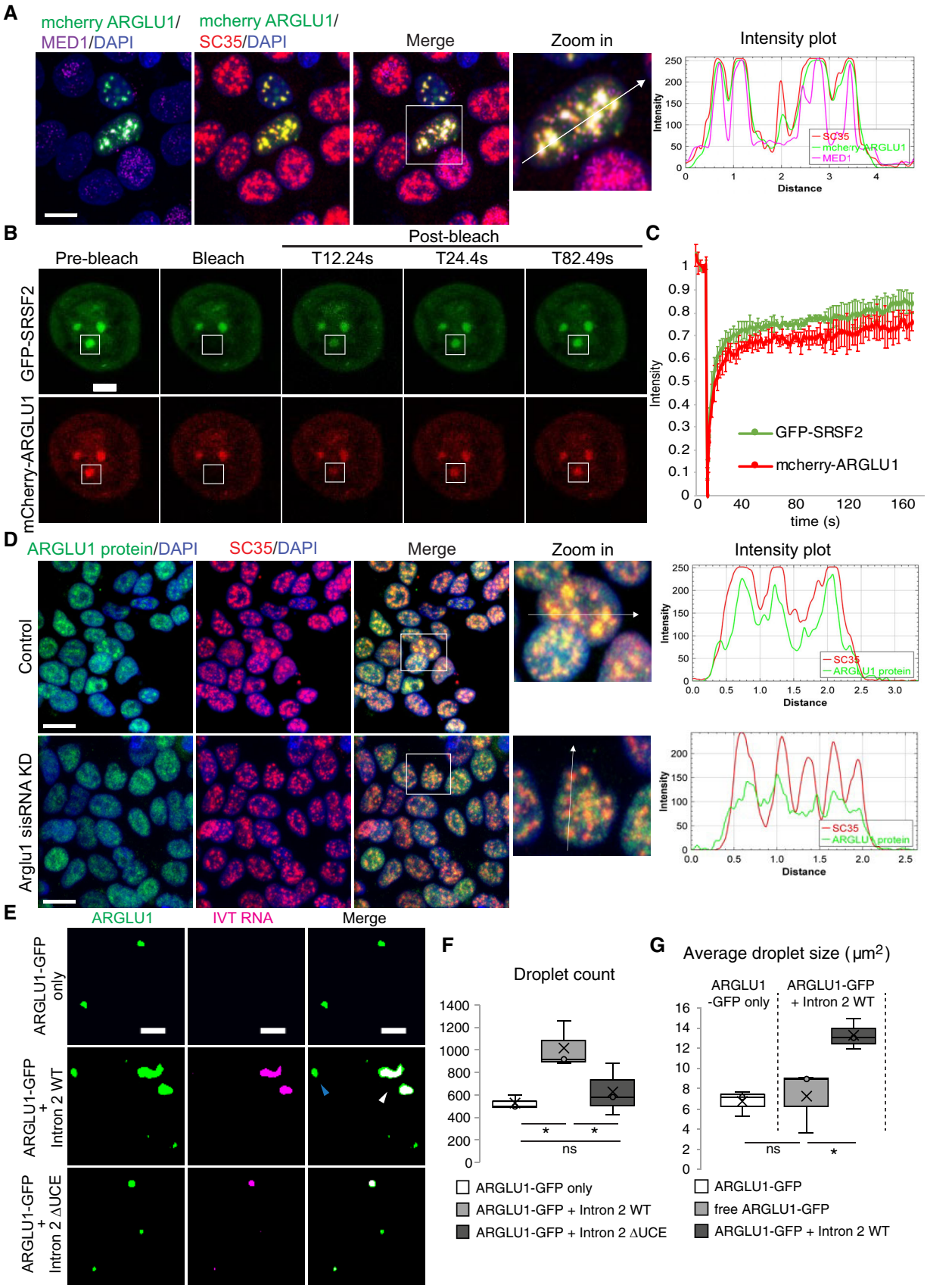

Figure 4.

Figure 4. Arglu1 sisRNA promotes robust ARGLU1 protein phase separation and localization to nuclear speckles.

A Immunostaining of MED1 (magenta) and SC35 (red) with transiently expressed mCherry-ARGLU1 protein (green) in MCF-7 cells. Zoom in of boxed area and the intensity plot shows the co-localization and signal quantification of mCherry-ARGLU1 protein, MED1 and SC35 in the direction of the white arrow. Scale bar: 20 μm.

B Confocal images of GFP-SRSF2 (green, control) and mCherry-ARGLU1 protein (red) at the pre-bleach, bleach, and post-bleach time points in FRAP analysis. The boxed area showing the speckle that was being bleached. Scale bar: 5 μm.

C Fluorescence intensities of the nuclear speckle over time in GFP-SRSF2 (green, control) and mCherry-ARGLU1 protein (red) FRAP analysis. ($n = 3$ separate nuclear speckles).

D Immunostaining of ARGLU1 protein (green) and SC35 (red) in 1 day Arglu1 sisRNA KD cells vs. control cells. Zoom in of boxed area and the intensity plot shows the co-localization and signal quantification of ARGLU1 protein and SC35 in the direction of the white arrow. Scale bar: 20 μm.

E Confocal images of 125 nM ARGLU1-GFP protein (green) only and ARGLU1-GFP protein added with 50 nM IVT RNA (magenta), of either Arglu1 intron 2 with UCE (Intron 2 WT) or IVT Arglu1 intron without UCE (Intron 2 ΔUCE). White arrowhead indicates ARGLU1-GFP droplet incorporated with IVT Intron 2 WT and blue arrowhead indicates free ARGLU1-GFP droplet not incorporated with IVT Intron 2 WT. Scale bar: 10 μm.

F Quantification of total ARGLU1-GFP droplets from (E). ($n = 3$ biological replicates). Cross, mean; Middle line, median; box, 25^{th}–75^{th} percentiles; whiskers, minimum to maximum.

G Quantification of average ARGLU1-GFP droplets size with/without incorporation of IVT intron 2 WT from (E, ARGLU1-GFP + Intron 2 WT). ($n = 3$ biological replicates). Cross, mean; Middle line, median; box, 25^{th}–75^{th} percentiles; whiskers, minimum to maximum.

Data information: In (C), data are presented as mean ± SEM. ns, not significant and $*P \leq 0.05$ (Student's *t*-test). See also Fig EV3 and Appendix Fig S3.

Next, we investigated whether the two cryptic PASs upstream of the 3′ end of dArglu1 sisRNA is important for its biogenesis. We cloned either wild-type (WT) dArglu1 intron 2 or individually PAS-deleted dArglu1 intron 2 (ΔPAS1/ΔPAS2, where PAS1 is nearest to the 3′ end) into the dsred-intron-myc minigene reporter (Fig 5G) and measured their splicing indices after transfection into S2 cells. ΔPAS1 dArglu1 intron 2, but not ΔPAS2, showed an increase in splicing index when compared to the WT dArglu1 intron 2 (Fig 5H). This result suggests that PAS1 suppresses canonical splicing likely by promoting intron cleavage, and hence promotes the production of dArglu1 sisRNA. Together, these observations provide evidence that the production of dArglu1 sisRNA is through premature cleavage of the orthologous intron 2 (Fig 5E).

U1 snRNP, a small nuclear RNA-protein complex involved in RNA splicing, has also been reported to protect pre-mRNA from premature cleavage in introns by lining on the intronic region and suppressing cryptic PAS sites in a process called telescripting (Kaida *et al*, 2010; Berg *et al*, 2012; So *et al*, 2019; Venters *et al*, 2019). Since dArglu1 sisRNA is produced by premature intronic cleavage, we wondered whether U1 snRNP plays a role in suppressing the biogenesis of dArglu1 sisRNA by protecting the pre-mRNA from cleavage. To verify this, we checked the expression of dArglu1 pre-mRNA and sisRNA in the ovaries of heterozygous mutant flies for U1-70K, a component of the U1 snRNP complex. The activity of U1 snRNP is slightly perturbed by reducing one copy of U1-70K gene (Bai *et al*, 2013). U1-70K heterozygous mutant ovaries contained higher levels of dArglu1 sisRNA (Figs 5I and EV4D) and lower levels of dArglu1 pre-mRNA compared to controls (Fig 5J). These results are consistent with the idea that a reduction of U1 snRNP level results in lower protection of the dArglu1 pre-mRNA by telescripting, leading to premature cleavage and higher levels of dArglu1 sisRNA.

Drosophila dArglu1 sisRNA is involved in autoregulation of its host gene

Since Arglu1 sisRNA and ARGLU1 protein are involved in the autoregulatory of the host gene, we wanted to know if the same form of regulation could be observed in dArglu1 as well. We first checked whether dARGLU1 protein has any impact on the expression of dArglu1 mRNA and the production of dArglu1 sisRNA. We overexpressed dArglu1 CDS-FLAG in S2 cells (Fig 5K) and checked the levels of endogenous dArglu1 mRNA by designing primers at the 5′ and 3′ UTR. In addition, we also checked the levels of dArglu1 sisRNA (Fig 5L). The results confirmed that dARGLU1 protein could autoregulate its splicing, by downregulating dArglu1 mRNA and upregulating dArglu1 sisRNA in a manner similar to the human ARGLU1 protein.

Next, we knocked down dArglu1 sisRNA levels by designing two short hairpin RNAs (shRNAs). The two shRNAs specifically reduced dArglu1 sisRNA levels (Fig 5M) and not the dArglu1 pre-mRNA levels (Fig 5N). Surprisingly, downregulation of dArglu1 sisRNA led to an increase in intron 2 splicing efficiency with a greater spliced mRNA to pre-mRNA ratio (Fig 5O). The results suggest that, in contrast to human Arglu1 sisRNA, which promotes splicing, dArglu1 sisRNA represses host gene splicing.

dArglu1 sisRNA suppresses splicing by inhibiting the activity of U1 snRNP

Our previous results suggested that U1 snRNP promotes dArglu1 pre-mRNA splicing by suppressing intronic cleavage through telescripting. We wondered if dArglu1 sisRNA represses splicing by inhibiting the activity of U1 snRNP. We designed two genetic interaction experiments to test this hypothesis. First, we asked if the activity of U1-70K can be enhanced by reducing dArglu1 sisRNA levels. Thus, we attempted to rescue the U1-70K heterozygous mutant phenotype by reducing the levels of dArglu1 sisRNA (Figs 5I and EV4D). Indeed, dArglu1 sisRNA KD restored the levels of dArglu1 pre-mRNA to normal (Fig 5J). The second experiment aimed to test if the increase in splicing in dArglu1 sisRNA KD ovaries was mediated by U1-70K. Consistent with our model, reducing the levels of U1-70K in dArglu1 sisRNA KD ovaries restored the splicing ratio to normal (Figs 5P and EV4E). Together, our results suggest that U1 snRNP promotes splicing and inhibits dArglu1 sisRNA biogenesis via telescripting. In turn, dArglu1 sisRNA suppresses splicing by inhibiting the activity of U1 snRNP (Fig 5Q).

dArglu1 sisRNA acts via promoting dARGLU1 protein phase separation

dArglu1 sisRNA, despite being conserved, possesses a distinct molecular function from its human ortholog. Intrigued, we wanted to investigate the molecular mechanism behind this inhibitory

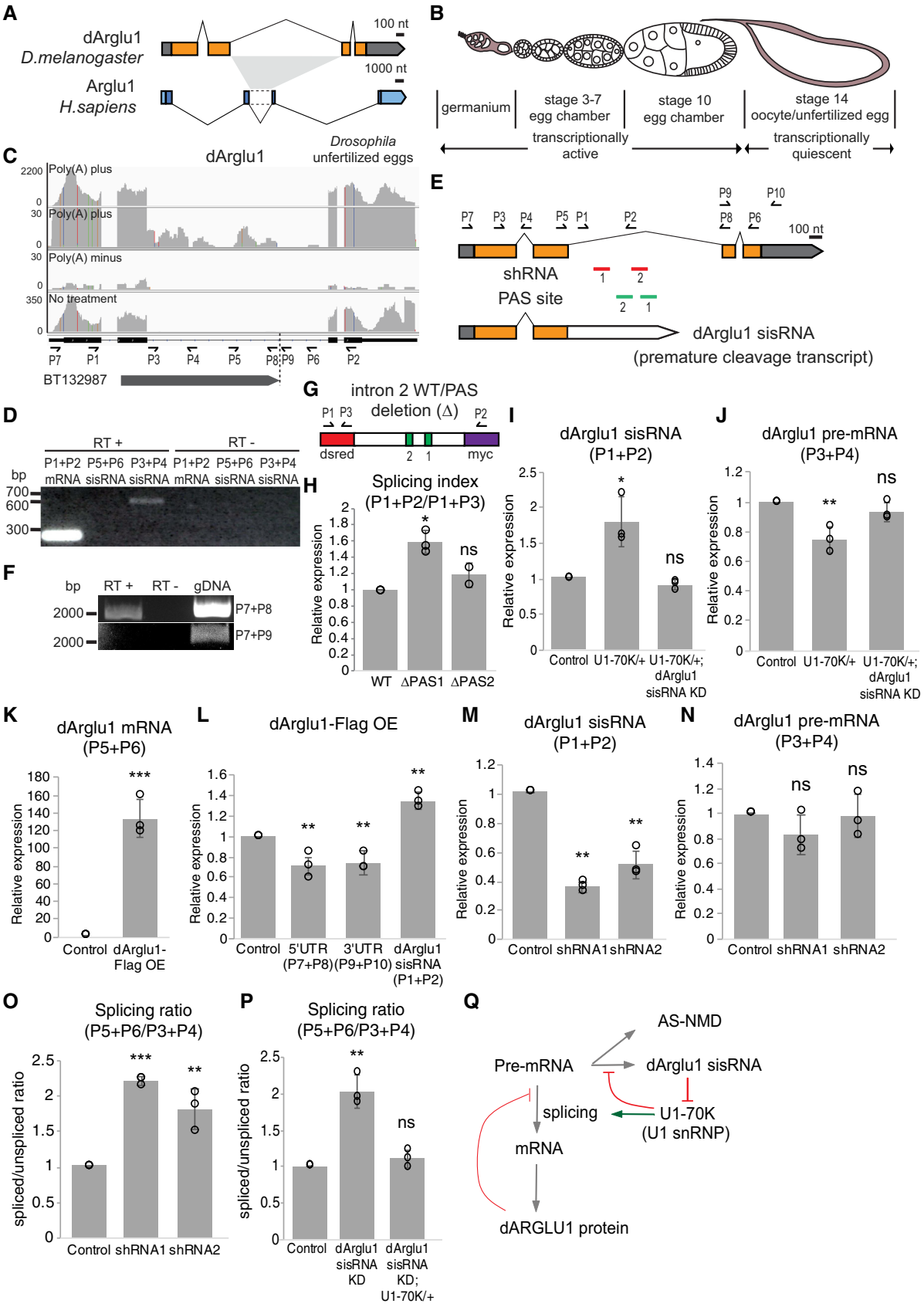
A
dArglu1
D.melanogaster
Arglu1
H.sapiens
100 nt
1000 nt
B
germanium
stage 3-7 egg chamber
stage 10 egg chamber
stage 14 oocyte/unfertilized egg
transcriptionally active
transcriptionally quiescent
C
dArglu1
Drosophila unfertilized eggs
Poly(A) plus
Poly(A) plus
Poly(A) minus
No treatment
BT132987
E
shRNA
PAS site
dArglu1 sisRNA
(premature cleavage transcript)
D
RT +
RT -
P1+P2 mRNA
P5+P6 sisRNA
P3+P4 sisRNA
bp
700
600
300
F
bp
RT +
RT -
gDNA
2000
P7+P8
P7+P9
G
intron 2 WT/PAS deletion (Δ)
dsred
myc
H
Splicing index (P1+P2/P1+P3)
Relative expression
WT
ΔPAS1
ΔPAS2
I
dArglu1 sisRNA (P1+P2)
Control
U1-70K/+
U1-70K/+; dArglu1 sisRNA KD
J
dArglu1 pre-mRNA (P3+P4)
K
dArglu1 mRNA (P5+P6)
dArglu1-Flag OE
L
dArglu1-Flag OE
5'UTR (P7+P8)
3'UTR (P9+P10)
dArglu1 sisRNA (P1+P2)
M
dArglu1 sisRNA (P1+P2)
shRNA1
shRNA2
N
dArglu1 pre-mRNA (P3+P4)
O
Splicing ratio (P5+P6/P3+P4)
spliced/unspliced ratio
P
Splicing ratio (P5+P6/P3+P4)
dArglu1 sisRNA KD
dArglu1 sisRNA KD; U1-70K/+
Q
AS-NMD
Pre-mRNA
dArglu1 sisRNA
splicing
U1-70K (U1 snRNP)
mRNA
dARGLU1 protein

Figure 5.

Figure 5. *Drosophila* dArglu1 sisRNA is involved in autoregulation of its host gene by inhibiting the activity of U1 snRNP.

A dArglu1 and Arglu1 gene structures in *Drosophila* and human, respectively. Orthologous introns 2 between the genes were indicated in light gray.
B Drawing of an ovariole showing germline cells at different stages of oogenesis in *Drosophila* and their transcriptional status.
C Genome browser view showing mapped sequencing reads to dArglu1 gene obtained from RNA-seq of unfertilized eggs of *Drosophila*. Locations of the primers (black arrows) designed for (D and F) were indicated. Bottom: Bar showing the location of an unannotated cDNA entry identified at this locus based on information from FlyBase.
D RT-PCR detecting presence of dArglu1 mRNA (P1 + P2) and dArglu1 sisRNA at the 5′ end (P3 + P4) but not at the 3′ end (P5 + P6) of the orthologous intron 2 from the unfertilized *Drosophila* eggs.
E dArglu1 gene structure indicating locations of shRNAs (red bars), cryptic premature polyadenylation sites (PASs; green bars), locations of designed primers for (H–P) (black arrows), and splicing isoforms from this locus.
F RT-PCR detecting presence of full-length dArglu1 sisRNA using primers before the 3′ end (P7 + P8) and but not after (P7 + P9) as predicted in (C).
G Schematic showing dsred-intron-myc minigene expressing dArglu1 intron 2 wild type (WT) or dArglu1 intron 2 with either single PAS site deletion (Δ) (PAS1/2 sites as indicated by green bar). Location of primers used in (H) was indicated by black arrows.
H Chart showing the splicing indices (spliced/unspliced ratio) of transfected minigenes containing dArglu1 intron 2 WT, dArglu1 intron 2 ΔPAS1/2 cells as measured by qPCR. ($n = 3$ biological replicates).
I, J qPCR of Arglu1 sisRNA and Arglu1 pre-mRNA expression levels in U1-70K heterozygous mutant and U1-70K heterozygous mutant; dArglu1 sisRNA KD vs. control (*y w*). ($n = 3$ biological replicates).
K qPCR of dArglu1 mRNA expression levels in dArglu1 CDS-FLAG overexpression S2 cells vs. control (untransfected S2 cells). ($n = 3$ biological replicates).
L qPCR of dArglu1 pre-mRNA at 5′ untranslated region (UTR), 3′ UTR and dArglu1 sisRNA expression levels in dArglu1 CDS-FLAG overexpression S2 cells vs. control (untransfected S2 cells). ($n = 3$ biological replicates).
M, N qPCR of Arglu1 sisRNA and Arglu1 pre-mRNA expression levels in dArglu1 sisRNA shRNA ovaries vs. control (sibling control, *MTD-GAL4/CyO*). ($n = 3$ biological replicates).
O Chart showing the splicing indices (spliced/unspliced ratio) in dArglu1 sisRNA KD vs. control (sibling control, *MTD-GAL4/CyO*) ovaries as measured by qPCR. ($n = 3$ biological replicates).
P Chart showing the splicing indices (spliced/unspliced ratio) in dArglu1 sisRNA KD and dArglu1 sisRNA KD; U1-70K heterozygous mutant vs. control (sibling control, *MTD-GAL4/CyO*) ovaries as measured by qPCR. ($n = 3$ biological replicates).
Q Schematics of dArglu1 gene autoregulatory loop model involving dARGLU1 protein, U1-70K (a component of U1 snRNP), and dArglu1 sisRNA in repressing splicing of host gene.

Data information: In (H–P), data are presented as mean ± SEM. ns, not significant, $*P \leq 0.05$, $**P \leq 0.01$, and $***P \leq 0.001$ (Student's *t*-test). See also Fig EV4 and Appendix Fig S4.

function of dArglu1 sisRNA. A previous study has shown that dARGLU1 protein interacts with U1-70K (Guruharsha *et al*, 2011). We hypothesized that dARGLU1 protein binds to the dArglu1 pre-mRNA at the gene locus and inhibits the activity of U1 snRNP through interaction with U1-70K, leading to increased production of dArglu1 sisRNAs. Upregulation of dArglu1 sisRNA, which binds to dARGLU1 protein and U1-70K in turn sequesters and accumulates dARGLU1 protein locally. As dARGLU1 protein levels increase at the gene locus, telescripting by U1 snRNP is inhibited. This positive feedback loop results in greater repression of dArglu1 gene splicing.

To test the validity of this model, we first generated an antibody against dARGLU1 protein and verified its specificity for different applications, including IP (Fig 6A), immunoblotting (Fig EV5A), and IF (Fig EV5B). We then examined the interactions between dARGLU1 protein with dArglu1 sisRNA, pre-mRNA, and U1 snRNA (an RNA component of U1 snRNP which also contains U1-70K) using RIP in ovaries samples. The results showed about 1.7–2.3 fold enrichment of the RNAs, but not for U85 snoRNA, in dARGLU1 protein immunoprecipitates compared to IgG controls, confirming their specific interactions with the dARGLU1 protein (Fig 6B).

Next, we wanted to know if dArglu1 sisRNA is required in mediating the interactions between dARGLU1 protein, pre-mRNA, and U1 snRNP. Using ovaries from dArglu1 sisRNA KD flies, we performed RIP for dARGLU1 protein to check its interaction with dArglu1 pre-mRNA and U1 snRNA. As a control, the levels of immunoprecipitated dArglu1 sisRNA were measured (Fig 6C). In dArglu1 sisRNA KD ovaries, we observed a decrease in both dArglu1 pre-mRNA and U1 snRNA enrichment (Fig 6D and E), suggesting that dArglu1 sisRNA is important in maintaining their interactions.

To examine how dArglu1 sisRNA promotes binding of dARGLU1 protein to dArglu1 pre-mRNA and U1 snRNP, we next asked if dArglu1 sisRNA is required in the formation of dARGLU1 nuclear foci. IF of dARGLU1 protein in the germaria showed that dARGLU1 protein localizes as foci in the nucleus, which were absent in dARGLU1 protein KD ovaries (Fig EV5B). Consistent with the higher splicing ratio of dArglu1 mRNA observed in dArglu1 sisRNA KD flies, we also saw an increase in the levels of dARGLU1 protein in dArglu1 sisRNA KD ovaries (Fig 6F). Interestingly, we observed more diffused dARGLU1 protein foci in the nuclei of the germaria from dArglu1 sisRNA KD flies (Fig 6G). We then quantified the percentage of germaria containing dARGLU1 protein foci and saw a significant decrease in numbers when we compared control (90%) against dArglu1 sisRNA KD (30%) (Fig 6H).

Next, we asked if the intron 2 sequence from dArglu1 sisRNA can promote phase separation of dARGLU1 protein *in vitro*. We performed *in vitro* droplet assay using IVT dArglu1 intron 2 (before the dArglu1 sisRNA 3′ end) or dArglu1 pre-mRNA (intron 2 after the 3′ end) added with recombinant dARGLU1-GFP protein (GFP-tagged at C-terminal) and checked the number and size of dARGLU1-GFP protein droplets formed. Like the human counterpart, recombinant dARGLU1-GFP phase separated into droplets *in vitro* where the size and number of droplets increased in a concentration-dependent manner (Fig EV5C). The addition of IVT dArglu1 intron 2, but not the IVT dArglu1 pre-mRNA, led to significantly higher number of droplets (Fig 6I and J). The dARGLU1 droplets incorporated with dArglu1 sisRNA were also bigger as compared to the those without dArglu1 sisRNA (Fig 6K). Taken together, both *in vivo* and *in vitro* results support our hypothesis that dArglu1 sisRNA is important for the formation of dARGLU1

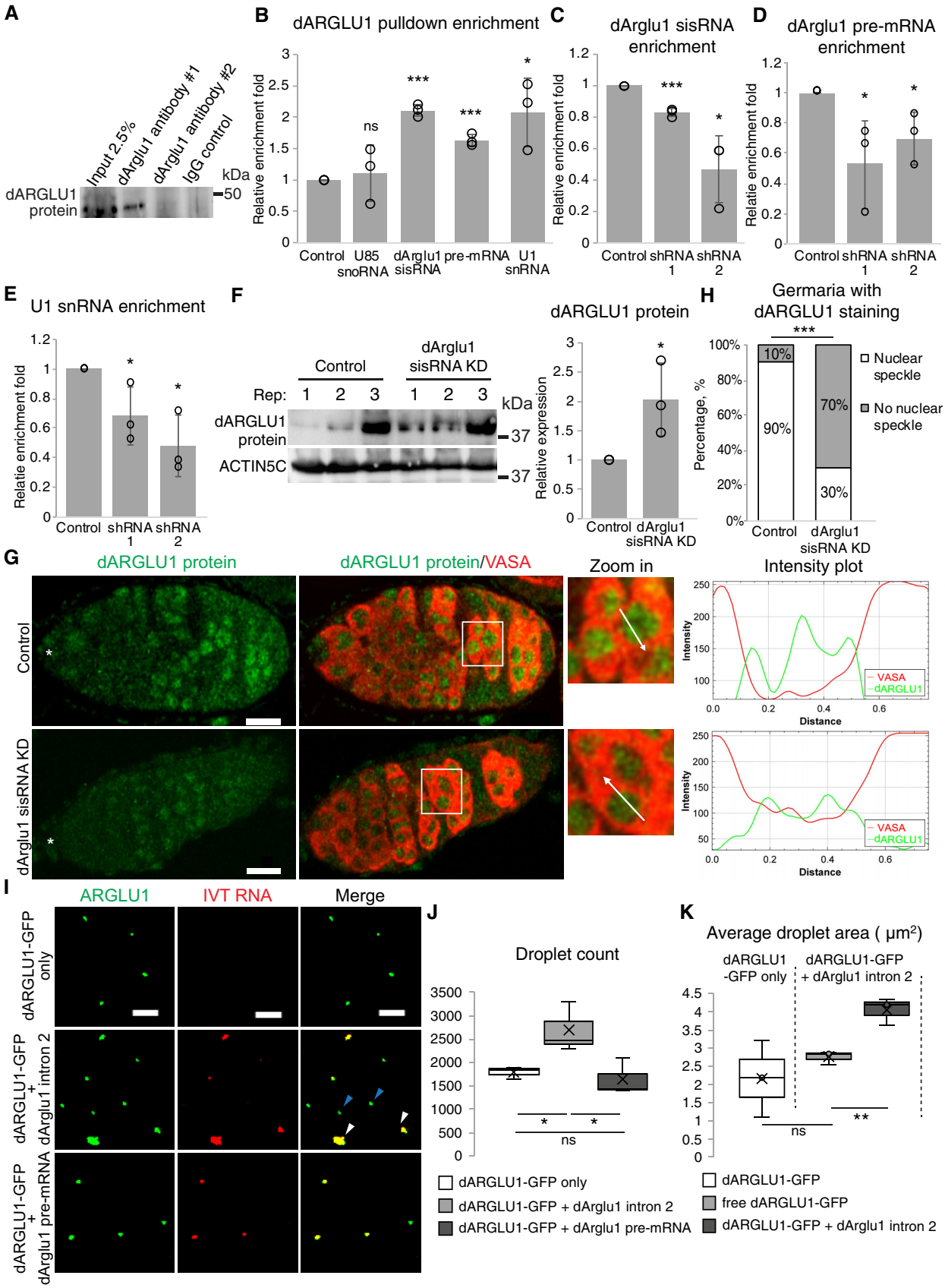
A
Input 2.5%
dArglu1 antibody #1
dArglu1 antibody #2
IgG control
kDa
dARGLU1 protein
50
B
dARGLU1 pulldown enrichment
Relative enrichment fold
Control
U85 snoRNA
dArglu1 sisRNA
pre-mRNA
U1 snRNA
C
dArglu1 sisRNA enrichment
Control
shRNA 1
shRNA 2
D
dArglu1 pre-mRNA enrichment
E
U1 snRNA enrichment
F
Control
dArglu1 sisRNA KD
Rep:
dARGLU1 protein
ACTIN5C
kDa
37
dARGLU1 protein
Relative expression
dArglu1 sisRNA KD
H
Germaria with dARGLU1 staining
Percentage, %
10%
90%
70%
30%
Nuclear speckle
No nuclear speckle
G
dARGLU1 protein
dARGLU1 protein/VASA
Zoom in
Intensity plot
Control
dArglu1 sisRNA KD
Intensity
Distance
VASA
dARGLU1
I
ARGLU1
IVT RNA
Merge
dARGLU1-GFP only
dARGLU1-GFP + dArglu1 intron 2
dARGLU1-GFP + dArglu1 pre-mRNA
J
Droplet count
ns
dARGLU1-GFP only
dARGLU1-GFP + dArglu1 intron 2
dARGLU1-GFP + dArglu1 pre-mRNA
K
Average droplet area (μm²)
dARGLU1 -GFP only
dARGLU1-GFP + dArglu1 intron 2
dARGLU1-GFP
free dARGLU1-GFP
dARGLU1-GFP + dArglu1 intron 2

Figure 6.

Figure 6. dArglu1 sisRNA interacts with U1 snNRP, Arglu1 pre-mRNA, and dARGLU1 protein in forming nuclear bodies and promotes dARGLU1 protein phase separation.

A Western blot analysis of dARGLU1 protein level from dARGLU1 protein immunoprecipitation in S2 cells. Samples with rabbit-IgG antibodies were used as control. ARGLU1 antibody #1, but not #2, worked for IP.

B RIP-qPCR of dArglu1 sisRNA, dArglu1 pre-mRNA, U1 snRNA, and U85 snoRNA levels from dARGLU1 protein immunoprecipitation using ovary lysates. Samples with rabbit-IgG antibodies were used as control. ($n = 3$ biological replicates).

C–E RIP-qPCR of dArglu1 sisRNA, pre-mRNA and U1 snRNA levels normalized to U85 from dARGLU1 protein immunoprecipitation using dArglu1 sisRNA KD 1 and 2 ovaries vs. control (sibling control, *MTD-GAL4/CyO*) ovaries. ($n = 3$ biological replicates).

F Left: Western blot analysis of dARGLU1 protein levels and ACTIN5C as loading control in dArglu1 sisRNA KD ovaries vs. control (sibling control, *MTD-GAL4/CyO*) ovaries. Right: Quantification of relative dARGLU1 protein levels in dArglu1 sisRNA KD ovaries vs. control ovaries from the western blot analysis. ($n = 3$ biological replicates)

G Left: Immunostaining of dARGLU1 protein (green) and VASA (red) in dArglu1 sisRNA KD germaria vs. control (sibling control, *MTD-GAL4/TM3*) germaria. The asterisk (*) marks the anterior of the germarium. Right: Zoom in of the boxed area and the intensity plot shows the signal quantification of dARGLU1 protein and VASA in the direction of the white arrow. Scale bar: 10 μm.

H Quantification of germaria with and without nuclear speckle formation from dARGLU1 protein staining in dArglu1 sisRNA KD germaria vs. control (sibling control, *MTD-GAL4/TM3*) germaria from (G). ($n = 20$ germaria)

I Confocal images of 125 nM dARGLU1-GFP protein (green) only and dARGLU1-GFP protein added with 50 nM IVT RNA (magenta), of either dArglu1 intron 2 until the 3′ end of sisRNA (dArglu1 intron 2) or IVT dArglu1 intron 2 starting after 3′ end of sisRNA (dArglu1 pre-mRNA). White arrowheads indicate dARGLU1-GFP droplets incorporated with IVT dArglu1 intron 2 and blue arrowheads indicate free dARGLU1-GFP droplets not incorporated with IVT dArglu1 intron 2. Scale bar: 10 μm.

J Quantification of total dARGLU1-GFP droplets from (I). ($n = 3$ biological replicates). Cross, mean; middle line, median; box, 25^{th}–75^{th} percentiles; whiskers, minimum to maximum.

K Quantification of average dARGLU1-GFP droplets size with/without incorporation of IVT dArglu1 intron 2 from (I, dARGLU1-GFP + dArglu1 intron 2). ($n = 3$ biological replicates). Cross, mean; middle line, median; box, 25^{th}–75^{th} percentiles; whiskers, minimum to maximum.

Data information: In (B–F), data are presented as mean ± SEM. ns, not significant, $*P \leq 0.05$, $**P \leq 0.01$ and $***P \leq 0.001$ (Student's *t*-test). See also Fig EV5.
Source data are available online for this figure.

protein foci and is likely to localize at the gene locus through the interaction with U1 snRNP.

Cross-species processing and functional analyses of Arglu1 sisRNA suggest species-specificity

We have shown that although in both *Drosophila* and human, Arglu1 intron 2 sequences bind to Arglu1 protein, they are processed into different forms of sisRNAs and have different downstream modes of action in regulating Arglu1 pre-mRNA splicing. Hence, we asked if Arglu1 sisRNA biogenesis is species-specific by performing cross-species *in vitro* droplet assays and mini-gene processing experiments.

We first performed *in vitro* droplet assay by incubating IVT d/hArglu1 intron 2 and d/hARGLU1-GFP protein in different combinations (Fig 7A). Interestingly when IVT Arglu1 intron 2 and ARGLU1-GFP from different species were incubated, the IVT Arglu1 intron 2 was incorporated into ARGLU1-GFP droplets although not as efficiently when the same species were mixed. The number of nonincorporated IVT Arglu1 intron 2 (Fig 7A, arrowheads) is mostly low to none when the same species were mixed, but is significantly higher when different species were incubated, with dARGLU1 + hArglu1 intron 2 showing a significant higher nonincorporation than hARGLU1 + dArglu1 intron 2 (Fig 7B). In addition, the cross-species incorporation of IVT Arglu1 intron 2 into ARGLU1-GFP droplets did not promote more or bigger ARGLU1-GFP droplet formation when compared to those with the same species (Fig 7C–F). The results suggest that IVT Arglu1 intron 2 can interact with ARGLU1 protein cross species, albeit less efficiently, and is unable to promote phase separation of the ARGLU1 protein.

Next, we asked how the orthologous Arglu1 intron 2 is processed when it is being expressed in a different species. First, we cloned the *Drosophila* and human orthologous Arglu1 intron 2 flanked by their cognate exon 2 and exon 3 (Arglu1 E2i2E3) into expression vectors (pcDNA 3.1 or pAW) and transfected them into human MCF-7 or *Drosophila* S2 cells respectively. Interestingly, in MCF-7 cells, dArglu1 E2i2E3 can undergo normal splicing and be processed into a premature cleavage transcript, reminiscent of dArglu1 sisRNA (Fig 7G). Overexpression of dArglu1 E2i2E3 in MCF-7 cells also promotes endogenous hArglu1 splicing with a concomitant drop in hArglu1 sisRNA production (Fig 7H–K). In contrast, overexpression of hArglu1 E2i2E3 in S2 cells did not produce any sisRNA via intron-retention. Instead, it is spliced into canonical mRNA and several alternatively spliced NMD transcripts (Fig 7L). Besides, overexpression of hArglu1 E2i2E3 had no effects on endogenous dArglu1 sisRNA production and splicing in S2 cells (Fig 7M–P).

Expression of dArglu1 E2i2E3 in MCF-7 cells led to production of dArglu1 sisRNA and interfered with endogenous hArglu1 splicing (Fig 7G–K). To examine if dArglu1 sisRNA indeed affects endogenous hArglu1 splicing, we expressed full-length dArglu1 sisRNA in MCF-7 cells and observed similar effects on endogenous hArglu1 splicing (Fig 8A–E). Our results corroborated the idea that dArglu1 sisRNA disrupts the endogenous hARGLU1 protein's autoregulatory function. Since hArglu1 E2i2E3 was not processed into hArglu1 sisRNA in S2 cells, we could not examine the function of hArglu1 sisRNA. To overcome this, we mutated the 5′ splice site of intron 2 and successfully expressed full-length hArglu1 sisRNA in S2 cells (Fig 8F). However, we did not observe any effects on endogenous dArglu1 splicing and sisRNA production (Fig 8G–J). Taken together, our results suggest that while dArglu1 intron 2 can be processed into functional sisRNA in human MCF-7 cells, hArglu1 intron 2 cannot be processed and hArglu1 sisRNA is not functional in *Drosophila* S2 cells.

Discussion

Posttranscriptional alternative splicing events, such as the generation of NMD, intron retention, or premature cleavage transcripts, are known to regulate gene expression. Such regulation is often

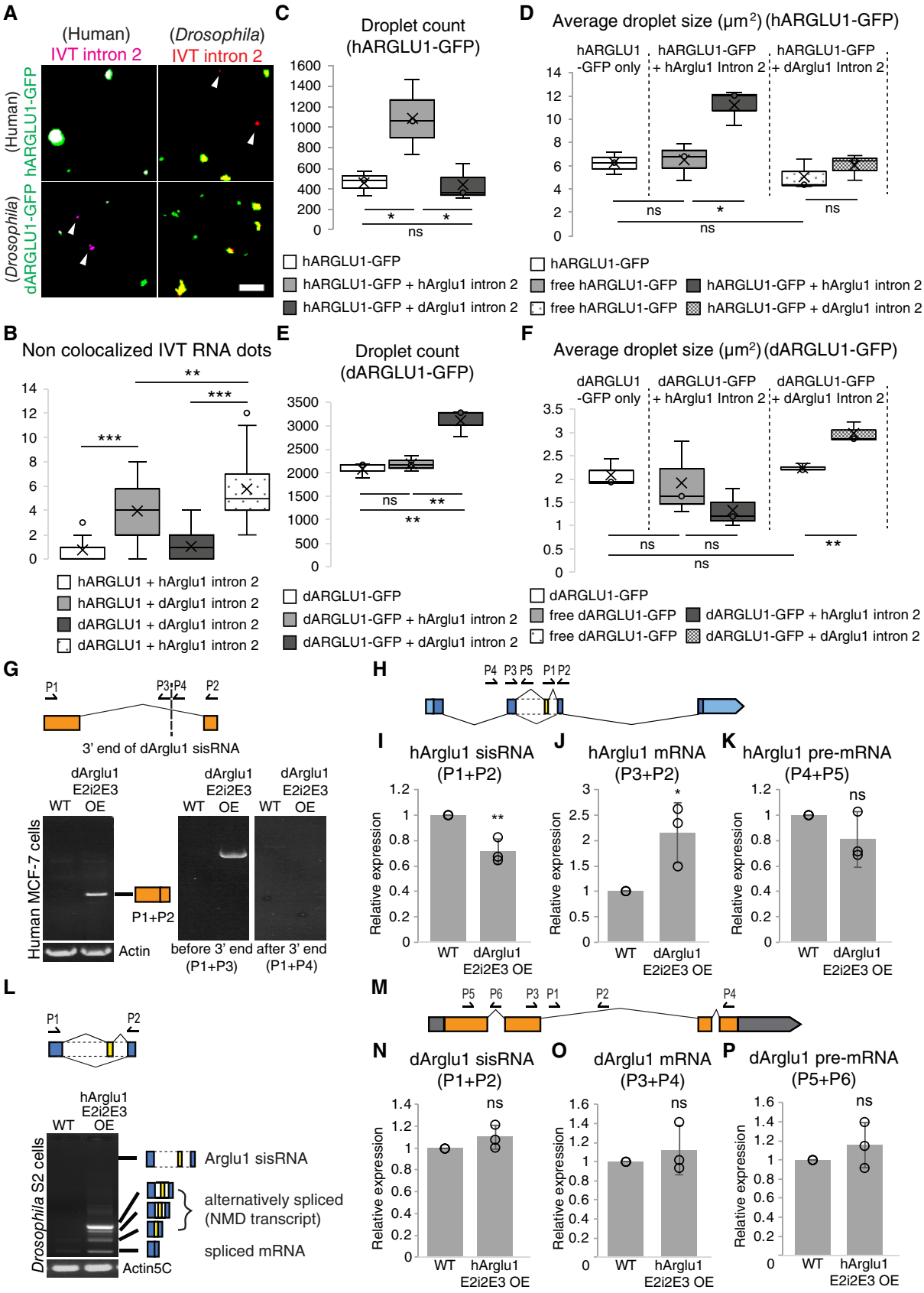
A
(Human) IVT intron 2
(Drosophila) IVT intron 2
(Human) hARGLU1-GFP
(Drosophila) dARGLU1-GFP
C
Droplet count (hARGLU1-GFP)
hARGLU1-GFP
hARGLU1-GFP + hArglu1 intron 2
hARGLU1-GFP + dArglu1 intron 2
D
Average droplet size (μm²) (hARGLU1-GFP)
hARGLU1 -GFP only
hARGLU1-GFP + hArglu1 Intron 2
hARGLU1-GFP + dArglu1 Intron 2
hARGLU1-GFP
free hARGLU1-GFP
hARGLU1-GFP + hArglu1 intron 2
free hARGLU1-GFP
hARGLU1-GFP + dArglu1 intron 2
B
Non colocalized IVT RNA dots
hARGLU1 + hArglu1 intron 2
hARGLU1 + dArglu1 intron 2
dARGLU1 + dArglu1 intron 2
dARGLU1 + hArglu1 intron 2
E
Droplet count (dARGLU1-GFP)
dARGLU1-GFP
dARGLU1-GFP + hArglu1 intron 2
dARGLU1-GFP + dArglu1 intron 2
F
Average droplet size (μm²) (dARGLU1-GFP)
dARGLU1 -GFP only
dARGLU1-GFP + hArglu1 Intron 2
dARGLU1-GFP + dArglu1 Intron 2
dARGLU1-GFP
free dARGLU1-GFP
dARGLU1-GFP + hArglu1 intron 2
free dARGLU1-GFP
dARGLU1-GFP + dArglu1 intron 2
G
P1
P3 P4
P2
3' end of dArglu1 sisRNA
dArglu1 E2i2E3
WT OE
Human MCF-7 cells
P1+P2
Actin
before 3' end (P1+P3)
after 3' end (P1+P4)
H
P4 P3 P5 P1 P2
I
hArglu1 sisRNA (P1+P2)
J
hArglu1 mRNA (P3+P2)
K
hArglu1 pre-mRNA (P4+P5)
Relative expression
WT
dArglu1 E2i2E3 OE
L
P1
P2
hArglu1 E2i2E3
WT OE
Drosophila S2 cells
Arglu1 sisRNA
alternatively spliced (NMD transcript)
spliced mRNA
Actin5C
M
P5 P6 P3 P1 P2 P4
N
dArglu1 sisRNA (P1+P2)
O
dArglu1 mRNA (P3+P4)
P
dArglu1 pre-mRNA (P5+P6)
hArglu1 E2i2E3 OE

Figure 7.

Figure 7. Cross-species binding of Arglu1 intron 2 to ARGLU1 protein and processing into sisRNAs.

A Confocal images of 125 nM d/hARGLU1-GFP protein added with 50 nM IVT RNA of either hArglu1 intron 2 WT (magenta) or dArglu1 intron 2 (red). White arrowheads indicate free IVT RNA dots that are not co-localized with the d/hARGLU1-GFP droplets. Scale bar: 10 μm.

B Quantification of non co-localized RNA dots from (A, white arrowheads). ($n = 16$ biological replicates).

C, D Quantification of total hARGLU1-GFP droplets and the average droplet size with/without incorporation of IVT hArglu1 intron 2 WT or dArglu1 intron 2 from (A). ($n = 3$ biological replicates).

E, F Quantification of total dARGLU1-GFP droplets and the average droplet size with/without incorporation of IVT hArglu1 intron 2 WT or dArglu1 intron 2 from (A). ($n = 3$ biological replicates).

G Top: Schematic showing exon2-intron2-exon3 (E2i2E3) of dArglu1 used for exogenous expression in MCF-7 cells and locations of primers used for RT-PCR (bottom). Bottom: RT-PCR of processed dArglu1 spliced isoforms and *actin* as control in dArglu1 E2i2E3 OE MCF-7 cells vs. untransfected (WT) MCF-7 cells.

H hArglu1 gene locus showing location of primers used in (I–K).

I–K qPCR of hArglu1 sisRNA, mRNA and pre-mRNA levels in dArglu1 E2i2E3 OE MCF-7 cells vs. untransfected (WT) MCF-7 cells. ($n = 3$ biological replicates).

L Top: Schematic showing exon2-intron2-exon3 (E2i2E3) of hArglu1 used for exogenous expression in S2 cells and locations of primers used for RT-PCR (bottom). Bottom: RT-PCR of processed hArglu1 spliced isoforms and *actin5C* as control in hArglu1 E2i2E3 OE S2 cells vs. untransfected (WT) S2 cells.

M dArglu1 gene locus showing location of primers used in (N–P).

N–P qPCR of dArglu1 sisRNA, mRNA and pre-mRNA levels in hArglu1 E2i2E3 OE S2 cells vs. untransfected (WT) S2 cells ($n = 3$ biological replicates).

Data information: In (B–F), cross, mean; middle line, median; box, 25^{th}–75^{th} percentiles; whiskers, minimum to maximum. In (I–K, N–P), data are presented as mean ± SEM. ns, not significant, $*P \leq 0.05$, $**P \leq 0.01$ and $***P \leq 0.001$ (Student's *t*-test).

interpreted from the perspective of protein homeostasis (Colgan & Manley, 1997; Braunschweig *et al*, 2014; Boutz *et al*, 2015). However, it is still not known whether intron retention and premature cleavage transcripts have any noncoding functions as they are often regarded as RNA intermediates or thought to be degraded rapidly. In this study, we identified a pair of orthologous sisRNAs produced from alternative splicing of intron 2 at the highly conserved Arglu1 gene loci in *Drosophila* and human cells. We show that these sisRNAs play a role in regulating Arglu1 splicing by modulating the activity of ARGLU1 protein.

Conservation of Arglu1 sisRNAs in invertebrates and vertebrates

Although the sisRNAs do not show a high degree of sequence conservation, like most conserved lncRNAs, intron 2 exhibits positional conservation by comparing the exon-intron structures of the Arglu1 gene loci (Fig 5A; Appendix Fig S4). Surprisingly, while both sisRNAs regulate the splicing for their cognate genes, they exhibit different modes of action. Previous reports have shown that subcellular localization and RNA processing can affect the functions of conserved lncRNAs (Chen, 2016; Guo *et al*, 2020). Our study provides the first example of how evolutionary divergence in biogenesis and processing can lead to distinct sisRNA functions.

Intronic sequences are known to be rapidly evolving (Ulitsky, 2016). It is possible that the orthologous Arglu1 intron 2 evolved and gave rise to additional posttranscriptional gene regulatory elements, which could affect the biogenesis and modes of action of these sisRNAs in the nucleus. We propose that in *Drosophila*, U1 snRNPs bind to dArglu1 intron 2 and inhibit the activation of cryptic PAS site(s) in a process known as telescripting (Kaida *et al*, 2010; Venters *et al*, 2019). The binding of dARGLU1 protein with U1-70K at intron 2 inhibits the activity of U1 snRNP locally, leading to production of dArglu1 sisRNA via intronic cleavage (Fig 5). dArglu1 sisRNA then causes dARGLU1 protein to be sequestered locally, forming nuclear bodies containing U1-70K at the gene locus (Fig 8K). It has been reported that U1 snRNP plays a role to immobilize and retains lncRNAs at the chromatin (Yin *et al*, 2020). Furthermore, phase separation of polyadenylation factors had been shown to occur at specific poly(A) sites to promote cleavage and polyadenylation (Fang *et al*, 2019). Thus, it is possible that U1 snRNP forms complex with dArglu1 sisRNA and dARGLU1 protein at the gene locus to promote local accumulation of polyadenylation factors via phase separation, which leads to intronic cleavage and repression of pre-mRNA splicing (Fig 8K). At the stoichiometry level, we do not propose that the less abundant dArglu1 sisRNA represses the activity of U1 snRNP globally since U1 snRNP is expressed at a much higher level than dArglu1 sisRNA. Instead, our model proposes that dArglu1 sisRNA acts locally at the dArglu1 gene locus to inhibit local U1 snRNP activity by promoting dARGLU1 phase separation. Such a model will make sense and is consistent with recent work by others that some lncRNAs can act in a sub-stoichiometric level by promoting phase separation of RNA binding proteins (Wu *et al*, 2021; Unfried & Ulitsky, 2022).

In contrast, the intron 2 of human Arglu1 gene contains a stretch of UCE, a highly conserved regulatory sequence that had been reported to be an essential element in the homeostatic splicing of the Arglu1 gene by intron retention as well as the AS-NMD pathway in human cells (Pirnie *et al*, 2017). We propose that ARGLU1 protein binds to the UCE region and promotes sisRNA production via intron retention. The Arglu1 sisRNA then in turn promotes localization of ARGLU1 protein to nuclear speckles, thereby sequestering it away from the gene locus (Figs 4 and 8L). We speculate the human ARGLU1 protein may have evolved to perform additional functions. Human ARGLU1 can interact with SC35 and MED1 (nuclear speckle proteins), but not for fly dARGLU1 (as shown in FlyBase). Hence, it is possible that the interaction with SC35/MED1 or other unknown factors may have driven the localization of ARGLU1 protein to the nuclear speckles.

Our cross-species experiments provide some insights on the evolution of Arglu1 sisRNA biogenesis and function. We observed that human sisRNA cannot be processed in *Drosophila* cells, and the overexpression of full-length hArglu1 sisRNA had no effects in regulating endogenous dArglu1 gene splicing. In contrast, dArglu1 sisRNA can be processed in human cells, and overexpression of full-length dArglu1 sisRNA showed noncoding function in regulating endogenous hArglu1 gene splicing. Hence, the ability to process pre-mRNA to sisRNAs appears to be associated with functionality. We speculate that dArglu1 intron 2 contains the essential sequences required for sisRNA production regardless of binding to *Drosophila*

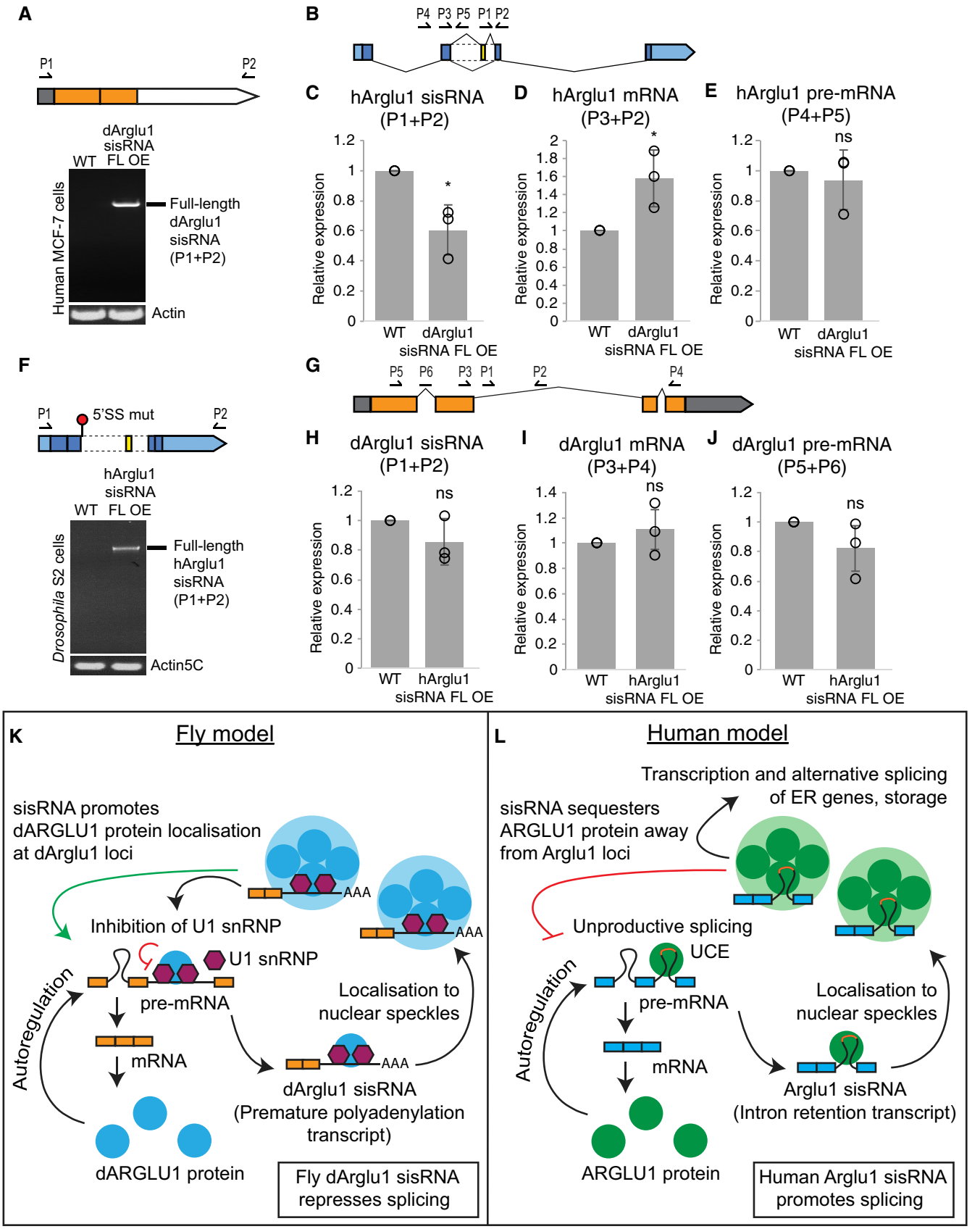
A
P1
P2
dArglu1 sisRNA
WT FL OE
Human MCF-7 cells
Full-length dArglu1 sisRNA (P1+P2)
Actin
B
P4 P3 P5 P1 P2
C
hArglu1 sisRNA (P1+P2)
Relative expression
WT dArglu1 sisRNA FL OE
D
hArglu1 mRNA (P3+P2)
E
hArglu1 pre-mRNA (P4+P5)
ns
F
P1 5'SS mut P2
hArglu1 sisRNA
WT FL OE
Drosophila S2 cells
Full-length hArglu1 sisRNA (P1+P2)
Actin5C
G
P5 P6 P3 P1 P2 P4
H
dArglu1 sisRNA (P1+P2)
WT hArglu1 sisRNA FL OE
I
dArglu1 mRNA (P3+P4)
J
dArglu1 pre-mRNA (P5+P6)
K
Fly model
sisRNA promotes dARGLU1 protein localisation at dArglu1 loci
AAA
Inhibition of U1 snRNP
U1 snRNP
pre-mRNA
mRNA
Autoregulation
Localisation to nuclear speckles
dArglu1 sisRNA (Premature polyadenylation transcript)
dARGLU1 protein
Fly dArglu1 sisRNA represses splicing
L
Human model
Transcription and alternative splicing of ER genes, storage
sisRNA sequesters ARGLU1 protein away from Arglu1 loci
Unproductive splicing
UCE
pre-mRNA
mRNA
Autoregulation
Localisation to nuclear speckles
Arglu1 sisRNA (Intron retention transcript)
ARGLU1 protein
Human Arglu1 sisRNA promotes splicing

Figure 8.

Figure 8. Cross-species functions of full-length sisRNAs and model.

A Top: Schematic showing full-length (FL) dArglu1 sisRNA used for exogenous expression in MCF-7 cells and locations of primers used for RT-PCR (bottom). Bottom: RT-PCR of full-length dArglu1 sisRNA in dArglu1 sisRNA FL OE MCF-7 cells vs. untransfected (WT) MCF-7 cells.

B hArglu1 gene locus showing location of primers used in (C–E).

C–E qPCR of hArglu1 sisRNA, mRNA and pre-mRNA expression levels in dArglu1 sisRNA FL OE MCF-7 cells vs. untransfected (WT) MCF-7 cells. ($n = 3$ biological replicates).

F Top: Schematic showing full-length (FL) hArglu1 sisRNA with 5′ splice site mutated used for exogenous expression in S2 cells and locations of primers used for RT-PCR (bottom). Bottom: RT-PCR of full-length hArglu1 sisRNA in hArglu1 sisRNA FL OE S2 cells vs. untransfected (WT) S2 cells.

G dArglu1 gene locus showing location of primers used in (H–J).

H–J qPCR of dArglu1 sisRNA, mRNA and pre-mRNA expression levels in hArglu1 sisRNA FL OE S2 cells vs. untransfected (WT) S2 cells. ($n = 3$ biological replicates).

K, L Model for Arglu1 sisRNA biogenesis in *Drosophila* and human leading to different localization and autoregulatory splicing effects. In *Drosophila*, dArglu1 sisRNA formed by premature cleavage at intron 2 interacts with dARGLU1 protein and U1 snRNP leading to formation nuclear bodies at gene locus to repress dArglu1 pre-mRNA splicing. In human, hArglu1 sisRNA formed by intron-retention by interacting with hARGLU1 protein and subsequently localizes at the nuclear speckles owing to the additional interacting partners or factors of hARGLU1 protein. UCE in orange.

Data information: In (C–E) and (H–J), data are presented as mean ± SEM. ns, not significant and $*P \leq 0.05$ (Student's *t*-test).

or human ARGLU1 protein. On the other hand, the UCE present in hArglu1 intron 2 appears to be an evolutionary novelty (only found conserved from chicken to human) that can only be processed by hARGLU1 protein.

Arglu1 sisRNA in assisting ARGLU1 nuclear body formation

Arglu1 sisRNA has a relatively higher stability (half-life of more than 4 h) than its host pre-mRNA and NMD isoform (Figs 1D and EV1D). Furthermore, it is also relatively more stable than previously described detained introns (DIs) in polyadenylated transcripts, which were reported to be more abundant and stable than flanking introns (Boutz *et al*, 2015). These DI transcripts are spliced post-transcriptionally, show slower splicing kinetics, and possess a half-life of approximately 1 h (Boutz *et al*, 2015).

In recent years, other examples of intron-retained transcripts showing noncoding RNA functions with RNA-binding proteins have also surfaced. For example, pCharme interacts with MATR3 and PTBP1 on the chromatin for myogenesis (Desideri *et al*, 2020), while the intron-retained transcript produced from SRSF7 locus was found to assist in the formation of nuclear bodies (Königs *et al*, 2020). Our *in vitro* and *in vivo* experiments suggest that although ARGLU1 can phase separate on its own, Arglu1 sisRNA is required to facilitate this process at the nuclear speckles. ARGLU1 protein contains a stretch of arginine serine-rich sequence of low complexity or intrinsically disordered region at the N-terminus (Magomedova *et al*, 2019) and proteins containing such regions are often predicted to form membrane-less condensates through phase separation (Murray *et al*, 2017; Sabari *et al*, 2020). A role for lncRNA in "seeding" nuclear body formation via phase separation, followed by self-sustenance independent of the lncRNA had also been reported for the lncRNAs Xist, SLERT, and NORAD (Xing *et al*, 2017; Pandya-Jones *et al*, 2020; Elguindy & Mendell, 2021).

Nuclear speckle is proposed to be the hub for gene expression where it is enriched with pre-mRNA splicing and transcription-related factors (Hu *et al*, 2008; Spector & Lamond, 2011; Kim *et al*, 2020), such as ARGLU1 protein. Indeed, larger ARGLU1 protein nuclear foci was positively correlated with higher c-Myc transcription (Fig EV3G and H). The role of Arglu1 sisRNA in promoting ARGLU1 protein nuclear speckle localization can account for the delay in transcriptional induction of c-Myc and pS2 after estrogen treatment in 1-day Arglu1 sisRNA ASO cells (Figs 4D and EV3E and F). Although the number and size of ARGLU1 nuclear bodies were low at earlier time point, they recovered back to normal eventually (Fig EV3B–F). Estrogen had been shown to promote phase separation of MED1 with estrogen receptors for gene activation (Boija *et al*, 2018), and MED1 is known to interact and co-localize with ARGLU1 protein in the nuclear speckles (Zhang *et al*, 2011) (Fig 4A). This process can promote the recovery of ARGLU1 nuclear foci over time independent of Arglu1 sisRNA.

Current knowledge on the evolution of intronic sequences and how they affect the biogenesis of conserved lncRNAs are still in their infancy. Many areas remain unexplored, and this hinders researchers from forming a comprehensive understanding of how the processing of intronic sequences is linked to their functions. Here, we propose that the sequence evolution of orthologous introns, as shown in the Arglu1 gene between *Drosophila* and human, can influence the fates of conserved lncRNAs in different ways, including their biogenesis, molecular functions, and modes of action. Flexibility in sequence conservation could be a strategy in nature to confer tenacity in orthologous lncRNAs as they allow functional diversification during evolution from simple organisms to those of higher complexity.

Materials and Methods

Mammalian cell culture and estrogen treatment

MCF-7, MDA-MB-231, T-47D, and MDA-MB-468 breast cancer cell lines were obtained from the American Type Culture Collection (ATCC). The cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM; Gibco, ThermoFisher Scientific) and RPMI 1640 (Gibco, ThermoFisher Scientific), respectively, supplemented with 10% fetal bovine serum (FBS) (Gibco, ThermoFisher Scientific) and incubated in a humidified incubator supplied with 5% of CO_2/air at 37°C. The culture media were changed every 2–3 days or subcultured when the cells reached 80% confluency. For the estrogen treatment, it was performed with slight modification (Bhat-Nakshatri *et al*, 2013). The cells were cultured in phenol-red free DMEM with 5% dextran charcoal treated serum (Gibco, ThermoFisher Scientific) for 3 days before treating with 100 nM estrogen (Sigma-aldrich). The cells were harvested at the indicated time points after treatment and used for RNA extraction.

RT-PCR/qPCR

RNA extraction was performed as described previously (Pek *et al*, 2015), using TRIzol (Ambion) and followed by Direct-zol mini-prep kit (Zymo Research). RT-PCR/qPCR was performed as described previously (Pek *et al*, 2015). RT was performed using random hexamers with M-MLV RT (Promega). PCR products were run on 1% agarose gel to visualize DNA. qPCR was done using SYBR Fast qPCR kit master mix (2×) universal (Kapa Biosystems, USA) and on the Applied Biosystems 7900HT Fast Real-Time PCR system. Oligonucleotides used can be found in Appendix Table S1. Relative gene expression level was normalized to actin and actin5C in human and *Drosophila* samples, respectively, unless specified otherwise.

Alpha-amanitin/emetine treatment

MCF-7 cells were incubated in DMEM with 10% FBS containing 20 μg/ml alpha-amanitin or 100 μg/ml emetine (Sigma) for the indicated time points before RNA was extracted.

3′ RACE

3′ RACE was conducted by using poly-T specific reverse primer for RT, and the cDNA was used for RT-PCR with gene specific primer and poly-A specific reverse primer. The PCR product was run on an agarose gel, excised, purified, and cloned into pGEM-T-Easy vector and sequenced.

Nuclear-cytoplasmic fractionation

Nuclear-cytoplasmic fraction was performed as described previously (Moss & Steitz, 2013) with slight modifications. Cells grown in 10 cm cell culture dishes with 80% confluency were washed twice in cold 1× PBS buffer and harvested in 5 ml cold 1× PBS buffer. The cells were pelleted at 1,000 *g* for 5 min at 4°C and homogenized in 200 μl of cold cell disruption buffer (10 mM KCl, 1.5 mM $MgCl_2$, 20 mM Tris-Cl, 1 mM DTT, 0.1% Triton-X). Lysate was spun at 1,500 *g* for 10 min at 4°C. The supernatant was collected as the cytoplasmic fraction. The crude nuclear pellet was homogenized in cold cell disruption buffer and spun at 1,500 *g* for 10 min at 4°C. The resultant pellet was collected as the nuclear fraction. After nuclear-cytoplasmic fractionation, RNA was extracted and dissolved in equal volumes of water. Equal volumes equivalent of RNA was used for RT-PCR analyses.

RNA single-molecule fluorescence In situ hybridization (smFISH)

The smFISH was performed with reagents and probes designed and purchased from RNAscope (ACDBio) and based on the manufacturer's protocol with modifications. Briefly, MCF-7 cells were grown to 80% confluency overnight on a chamber slide and fixed with 4.0% PFA for 10 min. The slide was washed twice with 1× PBS and a hydrophobic barrier was drawn surrounding the cultured cells on the slide using ImmEdge hydrophobic barrier pen. The cells were permeabilized with RNAscope Protease III in 1:15 dilution with 1× PBS for 10 min and proceeded with smFISH assay using RNAscope Fluorescent Multiplex kit. Probes for Arglu1 sisRNA (C2) and Arglu1 pre-mRNA (C3) were diluted in 1:50 using probe diluent and pipetted to cover the entire area within the hydrophobic barrier on the slide. The probes were allowed to hybridize for 2 h, at 40°C in the HybEZ oven. Then, the slide was washed twice with 1× wash buffer and proceeded with hybridization at 40°C in HybEZ oven using Amp 1-FL for 30 min, Amp 2-FL for 15 min, Amp 3-FL for 30 min, and Amp 4-FL for 15 min with two washes after each Amp hybridization. For experiments that were coupled with immunostaining, the slides were immediately used for blocking in 1× PBX with 5% normal goat serum and continued with immunostaining protocol. If not, the slide was counterstained with DAPI and mounted in Vectashield (Vector laboratories). Images were taken with Spe II upright confocal microscope imaging system (Leica).

sisRNA knockdown by antisense oligonucleotides (ASOs)

Cells were seeded at 2 × 10^5 cells/well in a six-well plate in DMEM supplemented with 10% FBS and allowed to grow overnight to reach about 80% confluency at the point transfection. Transient transfection was performed with ASOs and Lipofectamine 3000 reagent (Invitrogen, ThermoFisher Scientific) in the 2 μl:1 μg ratio per well according to the manufacturer's protocol. Briefly, the ASOs and Lipofectamine 3000 reagent were prepared separately in 150 μl of OPTI-MEM Reduced Serum Medium, mixed, and subsequently incubated at room temperature for 5 min. After the incubation, the cells were washed twice with 1× PBS and ASO-lipid complexes were added to each well containing 1,700 μl of DMEM supplemented with 10% FBS. The plates were placed in a humidified incubator supplied with 5% of CO_2/air, 37°C for 24 h, or 72 h. All ASO transfection assays were conducted in triplicates. Sequences of ASO are ASO Arglu1 sisRNA1 (5′-mA*mA*mU*mG*mA*T*T*T*G*T*A*C*T*G*mU*mU*mA*mG*mC*mU-3′), ASO Arglu1 sisRNA2 (5′-mC*mC*mU*mA*mU*A*C*T*C*C*T*T*T*A*mA*mU*mC*mA*mG*mC-3′), and ASO GFP1 (5′-mC*mU*mG*mC*mC*A*T*C*C*A*G*A*T*C*G*mU*mU*mA*mU*mC-3′).

Blocking UCE by antisense Morpholino oligonucleotides (AMOs)

The synthesis of AMOs and Endo-Porter was acquired from Genetools. MCF-7 cells are grown in a six-well plate to 80% confluency at point of transfection, and the AMOs were delivered into the cells using Endo-Porter according to the manufacturer's protocol. Briefly, 0.6 μM of AMOs were added to 1 ml of the complete culture medium followed by 6 μl of Endo-Porter and mixed well. The plates were returned to the humidified incubator supplied with 5% of CO_2/air, 37°C for 24 h. All AMO transfection assays were conducted in triplicate. Sequences of AMOs are AMO Arglu1 ALE1 (5′-CCATACGCGCCAGCTTCCTCTTTAA-3′), AMO Arglu1 ALE2 (5′-TCCTGCAGAGTGTGCTCCTCGGCTG-3′), and AMO Arglu1 ALE3 (5′-TGAATATTTACTGTCTTGTCCAGTG-3′).

Cloning and plasmid construction

For general cloning, the PCR product of interest was amplified using primers listed in Appendix Table S1 and purified using QIAquick PCR purification kit (Qiagen) according to manufacturer's protocol. The purified PCR product was cloned into pENTR TOPO cloning vector (Invitrogen) and transformed into One Shot chemically competent *Escherichia coli* cells. The inserted sequence of the plasmid

was checked by sequencing before performing LR reactions to swap into the destination vector using Gateway LR Clonase II Enzyme mix (Invitrogen) (Wong *et al*, 2017). As for cloning into dsred-myc minigene, sticky-end cloning using AscI and NotI site was performed (Ng *et al*, 2018b). For preparation of UCE-deleted hArglu1 intron 2, PAS-deleted dArglu1 intron 2 and 5′ splice site mutated hArglu1 intron 2 expression plasmids, we synthesized the respective templates for PCR as long dsDNA sequences cloned in pUC57 plasmid by outsourcing to GenScript. UCE-deleted hArglu1 intron 2 and 5′ splice site mutated hArglu1 intron 2 was cloned into pcDNA3.1/NV5-DEST (ThermoFisher) and pAW (Drosophila Genomics Resource Center), respectively, with Gateway cloning, PAS-deleted dArglu1 intron 2 was cloned into dsred-myc minigene.

Transfection of plasmid into S2 cells

S2 cells were transfected with 2 μg of plasmid of interest using Cellfectin (Invitrogen) in accordance with the manufacturer's protocol and harvested 48 h after transfection. S2 cells that are capable of growing in serum-free conditions were obtained from Steve Cohen's laboratory. They were cultured at 25°C in Schneider's medium (Invitrogen) supplemented with 2 mM glutamine without serum and subcultured every week.

Generation of dARGLU1 antibody

The anti-dARGLU1 antibody generation was outsourced to GenScript, USA. The amino acid sequence, aa213-278 for dArglu1 (UnitProtKB – Q9VL63, FlyBaseID: CG31712), KREELEEILAENNRKIEEAQRKLAEERLAIIEEQRLMDEERQRMRKEQEKRVKEEQKVILGKNNSR, was used as the epitope to generate antibodies in rabbit. The specificity of this antibody was checked for various applications including immunoprecipitation, immunostaining, and western blotting.

Immunoprecipitation

Immunoprecipitation was performed as previously described (Wong *et al*, 2017). MCF-7 cells grown overnight in 1 × 6 wells plate at 80% confluency or ~ 200 fly ovaries were homogenized in protein extraction buffer (50 mM Tris–HCl pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 0.1% NP-40) supplemented with Protease Inhibitor Cocktail (Roche) and blocked using protein A/G agarose beads (Merck Millipore). Eight microliters of rabbit anti-ARGLU1 (Novus Biologicals, NBP1-87921) or rabbit anti-dARGLU1 (this study) were added and incubated overnight at 4°C. As a control, rabbit anti-IgG or beads without antibodies were used as stated in the figure legends. The following day, protein A/G agarose beads were added and incubated for another 4 h before three washes using protein extraction buffer. RNA and protein were then extracted for RT-PCR and western blotting. Western blotting was done using rabbit anti-ARGLU1 (ThermoFisher, PA5-66041) (1:1,000) and rabbit anti-dARGLU1 (1:1,000) as described before (Wong *et al*, 2017).

Western blotting

Western blotting was performed as previously described (Wong *et al*, 2017; Tay & Pek, 2019). For cell lines, the cells were washed twice in cold 1× PBS and harvested. The cell lysates were transferred and homogenized in 2× sample buffer containing β-mercaptoethanol. As for fly samples, the ovaries were dissected in Grace's media and homogenized in 2× sample buffer containing β-mercaptoethanol. Protein lysates were run under denaturing conditions using SDS-PAGE gel and transferred to a PVDF membrane. Antibodies used were anti-Arglu1 (1:1,000; ThermoFisher, PA5-66041), rabbit anti-GAPDH (1:1,000; Abcam, ab9485), rabbit anti-dArglu1 (1:1,000, this study), and mouse anti-Actin5C (1:1,000; Developmental Studies Hybridoma Bank). Western blot detection was done digitally using the ChemiDoc Touch Imaging System (BioRad) and under nonsaturating conditions. The western blot results were analyzed by using ImageJ. The band intensity of the protein of interest was normalized to the loading controls (GAPDH, ACTIN5C), and the ratio was compared to the control sample.

Immunostaining

Immunostaining was carried out as described previously (Pek & Kai, 2011; Wong *et al*, 2017; Osman & Pek, 2018; Ng *et al*, 2018a). For cell culture staining, the cells were grown overnight to 80% confluency on chamber slides and fixed with 4% of paraformaldehyde. The cells on the slides were washed twice with 1× PBX solution (phosphate-buffered saline containing 0.2% Triton X-100) and pre-absorbed in blocking solution (PBX containing 5% normal goat serum) for at least 30 min. Next, primary antibodies were prepared at the respective dilution and incubated for 1 h at room temperature. Antibodies used were rabbit anti-Arglu1 (1:300; ThermoFisher, PA5-66041), mouse anti-SC35 (1:300; kind gift from Joe Gall) and rabbit anti-MED1 (1:100; ThermoFisher, PA5-36114). The slide was washed twice with 1× PBX and continued with secondary antibody incubation for another 1 h at room temperature. After the incubation, the slide was washed twice using 1× PBX before mounting in Vectashield (Vector laboratories).

For fly ovaries, the ovaries were fixed in a solution containing 16% paraformaldehyde and Grace's media at a ratio of 2:1 for 20 min. Fixed samples were rinsed and washed with 1× PBX solution and pre-absorbed in blocking solution for at least 30 min. Primary antibody incubations were done at room temperature overnight, followed by washing with 1× PBX for three times for 20 min each, before secondary antibody incubation at room temperature for 4 h. The samples were then washed three times for 20 min each using 1× PBX before mounting in Vectashield (Vector laboratories). The antibodies used were rabbit polyclonal anti-dARGLU1 (1:2,000; this study), mouse monoclonal anti-alpha Spectrin (3A9, 1:1; Developmental Studies Hybridoma Bank), and guinea pig anti-Vasa (1:500, kind gift from Toshie Kai) (Patil & Kai, 2010). Images were taken using an SPE II confocal microscope (Leica) and processed using ImageJ. The intensity plot of the protein was analyzed using BAR analysis in ImageJ.

Fluorescence recovery after photobleaching (FRAP)

MCF-7 cells were grown in a 35 mm glass bottom dish (Mattek) and transiently transfected with mCherry-hARGLU1 plasmid (kind gift from Carolyn Cummins; Magomedova *et al*, 2019) and pmEGFP-N1 SRSF2 plasmid (kind gift from Gregory Jedd; Greig *et al*, 2020) using Lipofectamine 3000 in the ratio of 2 μl:1 μg for 24 h. The cells were

washed twice with 1x PBS and replaced with fresh DMEM medium supplemented with 10% FBS. The cultured dish was incubated in a live-cell imaging (37°C, 5% CO_2) stage top incubator (Tokai Hit) mounted on an Olympus FV3000 laser scanning confocal microscope. The cells were viewed at 60× magnification lens. Nineteen frames were taken at maximum acquisition speed for one-way scanning before the nuclear speckles were bleached with 100% laser power (at 480 and 561 nM wavelength) at the 20th frame, then a minimum of 380 subsequent frames was taken. FRAP analysis was carried out using cellSens software (Olympus). Normalization was carried out against the background of the imaged cell and non-bleached area of the cells for photobleaching. Experiments and quantification were carried out on three different nuclear speckles.

In vitro RNA transcription

Linear in vitro transcribed (IVT) RNA was generated using MEGAscript T7 Transcription Kit (Thermo Fisher) according to manufacturer's protocol with slight modifications. Fluorescent labeled IVT RNA was generated using either Cyanine 3-UTP or Cyanine 5-UTP (Perkin Elmer). Briefly, a 20 µl reaction was set up using 1 µg of PCR product containing T7-promoter sequence as template. The mixture was incubated at 37°C for 4 h, and the IVT RNA was purified using Illustra Microspin G-50 columns (Cytiva, USA) followed by RNA precipitation with 3 M sodium acetate. The pellet washed with 75% ethanol and resuspended in RNase-free water.

In vitro droplet assay

The generation of purified recombinant GFP-tagged human and fly ARGLU1 proteins were outsourced to GenScript, USA. *In vitro* phase separation assays were performed according to published protocols with slight modification (Elguindy & Mendell, 2021). *In vitro* phase separation assays of increasing concentration of recombinant purified human and fly GFP-tagged ARGLU1 were performed in the presence of 10% polyethylene glycol-3000 (PEG3000). Different concentrations of ARGLU1 proteins were diluted in 50 µl of phase separation buffer with final concentrations of 20 mM Tris HCl pH 7.5, 150 mM NaCl, 1 mM DTT, 10% glycerol, and 10% PEG3000. For IVT RNA-induced phase separation experiments, Cy3-UTP or Cy5-UTP IVT RNA were diluted in water and denatured at 70 °C for 10 min before mixing into 125 nM of human or fly GFP-tagged ARGLU1 protein to a volume of 50 µl modified phase separation buffer, with final concentrations of 50 nM IVT RNA, 20 mM Tris HCl pH 7.5, 150 mM NaCl, 1 mM MgCl2, 1 mM DTT and 10% glycerol. A 20 µl of the protein or protein-RNA mixtures were pipetted on glass base dish (Nunc) and incubated at room temperature for 30–60 min before imaging with Olympus FV3000 laser scanning confocal microscope. The droplets images were analyzed with cellSens software (Olympus) where the droplet number and total area were quantified with the 'Count and Measure' function. Average droplet size (area) was calculated by averaging the total area with the number of droplets.

Scoring of MCF-7 cells based on ARGLU1 protein foci size

For Fig EV3B and C, the cells were categorized into three groups based on the size of ARGLU1 protein foci and intensity plot analysis by ImageJ software. Big foci (size > 1.70 µm): Cells generally contain big ARGLU1 protein foci showing broad peak and high intensity. Intermediate foci (1.20 µm < size < 1.69 µm): Cells generally contain small ARGLU1 protein foci showing narrow peak and high intensity. Small foci (size < 1.20 µm): Cells generally contain small ARGLU1 protein foci with narrow peak and low intensity.

Fly strains

The following fly strains were used in this study: *y w* (used as control in Fig 6K and L), *MTD-Gal4* (Petrella *et al*, 2007), *UASp-dArglu1 KD* (this study), *P{PZ}snRNP-U1-70K^{02107}cn1/CyO; ry506* (Bloomington #11177), *w1118; P{GD10340}v21402/TM3* dARGLU1 protein KD (VDRC #21402). Flies were fed with wet yeast for 3 days at 25°C before dissection unless otherwise stated.

Transgene and plasmid construction

For the generation of *UASp-dArglu1 sisRNA KD*, design of shRNA sequences and cloning were done a previously described (Pek *et al*, 2015; Osman & Pek, 2018). Previously, we found that shRNAs driven by the UAS-GAL4 system were effective in knocking down nuclear localized sisRNAs, but not pre-mRNAs (Pek *et al*, 2015; Osman & Pek, 2018; Ng & Pek, 2021). Sequences used were: dArglu1-shRNA-1 top (CGATGCGCTAGCAGTCACATGATGCAG CAAGCTAAATAGTTATATTCAAGCATATTTAGCTTGCTGCATCAT GTGGCGAATTCATGCTA), dArglu1-shRNA-1 bottom (TAGCATGA ATTCGCCACATGATGCAGCAAGCTAAATATGCTTGAATATAACTA TTTAGCTTGCTGCATCATGTGACTGCTAGCGCATCG), dArglu1-shRNA-2 top (CGATGCGCTAGCAGTCAGACTTATTGCATACTCTT ATAGTTATATTCAAGCATATAAGAGTATGCAATAAGTCTGGCGAA TTCATGCTA), dArglu1-shRNA-2 bottom (TAGCATGAATTCGCCA GACTTATTGCATACTCTTATATGCTTGAATATAACTATAAGAGTA TGCAATAAGTCTGACTGCTAGCGCATCG). Transgenic flies were generated by BestGene using phiC31 integrase-mediated insertion into attP40 landing site.

Calculation of splicing index

Splicing index was calculated as previously described (Teo *et al*, 2018). By using RT-qPCR, the spliced RNA was quantified using primers that flank the intron of interest, while unspliced RNA was quantified using primers that amplify pre-mRNA. Ct values for spliced and unspliced RNA were normalized against housekeeping genes (*actin*, *actin5C*) before the ratio was calculated. Splicing indices for controls were set to 1 for comparison. An increase in splicing index means an increase in spliced/unspliced RNA ratio, hence more efficient splicing, and vice versa.

Calculation of dArglu1 sisRNA copy number

The copy number of dArglu1 sisRNA in unfertilized eggs was quantified using absolute quantification by qPCR method. Briefly, a CT value vs. copy number standard curve for dArglu1 intron 2 and dArglu1 pre-mRNA was plotted using serially diluted cDNA as template. The cDNA template was generated by using 500 µg of IVT (either dArglu1 sisRNA or pre-mRNA consisting of intron 2 sequences after 3′ end of dArglu1 sisRNA), and qPCR was performed using

specific primers targeting dArglu1 sisRNA and pre-mRNA. The copy number of dArglu1 sisRNA and pre-mRNA was acquired based on their CT value amplified from unfertilized egg. As the dArglu1 sisRNA primers also amplify pre-mRNA transcripts, the actual dArglu1 sisRNA per cell was calculated with the following formula (copy number of dArglu1 sisRNA—copy number of dArglu1 pre-mRNA).

Conservation analysis of orthologous intron 2 in Arglu1 gene

Gene annotations and nucleotide sequences of Arglu1 gene for both *Drosophila melanogaster* and human were retrieved from National Center for Biotechnology Information (NCBI) Gene database. The conservation of the orthologous intron 2 was examined using (1) exon-intron structure and (2) positional conservation (Ulitsky, 2016). Exon-intron structures were examined based on the gene annotation. Conservation of the flanking exons sequences (exon 2 and exon 3) of the orthologous intron 2 in Arglu1 gene were pairwisely compared using EMBOSS Needle (Madeira *et al*, 2019).

Statistical analysis

Two-tailed *t*-test for at least 3 biological replicates was performed unless otherwise stated in the figure legends. Chi-square tests were performed for Figs 6H and EV3D. For all analyses, ns: not significant, $*P < 0.05$, $**P < 0.01$, and $***P < 0.001$.

Data availability

This study includes no data deposited in external repositories.

Expanded View for this article is available online.

Acknowledgements
We thank Joe Gall, Carolyn Cummins, Gregory Jedd, Jamie Alan Greig, Toshie Kai, American Type Culture Collection, Bloomington Stock Center, Vienna Drosophila Resource Center, and Developmental Studies Hybridoma Bank for reagents and fly stocks. We thank Jasmine Heng for her assistance in generating the dArglu1 CDS overexpression plasmid. We also thank members of the Pek laboratory for discussion and comments on the manuscript.
This work in the Pek laboratory is supported by the Temasek Life Sciences Laboratory.

Author contributions
Seow Neng Chan: Conceptualization; data curation; formal analysis; investigation; methodology; writing—original draft; writing—review and editing. **Jun Wei Pek:** Conceptualization; data curation; formal analysis; supervision; funding acquisition; investigation; methodology; writing—original draft; project administration; writing—review and editing.

Disclosure and competing interests statement
The authors declare that they have no conflict of interest.

References

Bai B, Hales CM, Chen P-C, Gozal Y, Dammer EB, Fritz JJ, Wang X, Xia Q, Duong DM, Street C *et al* (2013) U1 small nuclear ribonucleoprotein complex and RNA splicing alterations in Alzheimer's disease. *Proc Natl Acad Sci U S A* 110: 16562 – 16567

Berg MG, Singh LN, Younis I, Liu Q, Pinto AM, Kaida D, Zhang Z, Cho S, Sherrill-Mix S, Wan L *et al* (2012) U1 snRNP determines mRNA length and regulates isoform expression. *Cell* 150: 53 – 64

Bhat-Nakshatri P, Song EK, Collins NR, Uversky VN, Dunker AK, O'Malley BW, Geistlinger TR, Carroll JS, Brown M, Nakshatri H (2013) Interplay between estrogen receptor and AKT in estradiol-induced alternative splicing. *BMC Med Genomics* 6: 1 – 18

Bittencourt D, Dutertre M, Sanchez G, Barbier J, Gratadou L, Auboeuf D (2008) Cotranscriptional splicing potentiates the mRNA production from a subset of estradiol-stimulated genes. *Mol Cell Biol* 28: 5811 – 5824

Boija A, Klein IA, Sabari BR, Dall'Agnese A, Coffey EL, Zamudio A v, Li CH, Shrinivas K, Manteiga JC, Hannett NM *et al* (2018) Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* 175: 1842 – 1855.e16

Boutz PL, Bhutkar A, Sharp PA (2015) Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes Dev* 29: 63 – 80

Braunschweig U, Barbosa-Morais NL, Pan Q, Nachman EN, Alipanahi B, Gonatopoulos-Pournatzis T, Frey B, Irimia M, Blencowe BJ (2014) Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res* 24: 1774 – 1786

Buxbaum AR, Wu B, Singer RH (2014) Single β-Actin mRNA detection in neurons reveals a mechanism for regulating its translatability. *Science* 343: 419 – 422

Cech TR, Steitz JA (2014) The noncoding RNA revolution - trashing old rules to forge new ones. *Cell* 157: 77 – 94

Chan SN, Pek JW (2019) Stable Intronic sequence RNAs (sisRNAs): An expanding universe. *Trends Biochem Sci* 44: 258 – 272

Chen LL (2016) Linking long noncoding RNA localization and function. *Trends Biochem Sci* 41: 761 – 772

Colgan DF, Manley JL (1997) Mechanism and regulation of mRNA polyadenylation. *Genes Dev* 11: 2755 – 2766

Desideri F, Cipriano A, Petrezselyova S, Buonaiuto G, Santini T, Kasparek P, Prochazka J, Janson G, Paiardini A, Calicchio A *et al* (2020) Intronic determinants coordinate charme lncRNA nuclear activity through the interaction with MATR3 and PTBP1. *Cell Rep* 33: 108548

Elguindy MM, Mendell JT (2021) NORAD-induced Pumilio phase separation is required for genome stability. *Nature* 595: 303 – 308

Fang X, Wang L, Ishikawa R, Li Y, Fiedler M, Liu F, Calder G, Rowan B, Weigel D, Li P *et al* (2019) Arabidopsis FLL2 promotes liquid–liquid phase separation of polyadenylation complexes. *Nature* 569: 265 – 269

Fox AH, Nakagawa S, Hirose T, Bond CS (2018) Paraspeckles: where long noncoding RNA meets phase separation. *Trends Biochem Sci* 43: 124 – 135

Gardner EJ, Nizami ZF, Conover Talbot J, Gall JG (2012) Stable intronic sequence RNA (sisRNA), a new class of noncoding RNA from the oocyte nucleus of *Xenopus tropicalis*. *Genes Dev* 26: 2550 – 2559

Greig JA, Nguyen TA, Lee M, Holehouse AS, Posey AE, Pappu RV, Jedd G (2020) Arginine-enriched mixed-charge domains provide cohesion for nuclear speckle condensation. *Mol Cell* 77: 1237 – 1250.e4

Gruber AJ, Schmidt R, Gruber AR, Martin G, Ghosh S, Belmadani M, Keller W, Zavolan M (2016) A comprehensive analysis of 3′ end sequencing data sets reveals novel polyadenylation signals and the repressive role of heterogeneous ribonucleoprotein C on cleavage and polyadenylation. *Genome Res* 26: 1145 – 1159

Guo CJ, Ma XK, Xing YH, Zheng CC, Xu YF, Shan L, Zhang J, Wang S, Wang Y, Carmichael GG *et al* (2020) Distinct processing of lncRNAs contributes to non-conserved functions in stem cells. *Cell* 181: 621 – 636.e22

Guruharsha KG, Rual J-F, Zhai B, Mintseris J, Vaidya P, Vaidya N, Beekman C, Wong C, Rhee DY, Cenaj O *et al* (2011) A protein complex network of *Drosophila melanogaster*. *Cell* 147: 690 – 703

Hu Q, Kwon Y-S, Nunez E, Cardamone MD, Hutt KR, Ohgi KA, Garcia-Bassets I, Rose DW, Glass CK, Rosenfeld MG *et al* (2008) Enhancing nuclear receptor-induced transcription requires nuclear motor and LSD1-dependent gene networking in interchromatin granules. *Proc Natl Acad Sci U S A* 105: 19199 – 19204

Kaida D, Berg MG, Younis I, Kasim M, Singh LN, Wan L, Dreyfuss G (2010) U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* 468: 664 – 668

Kim J, Venkata NC, Hernandez Gonzalez GA, Khanna N, Belmont AS (2020) Gene expression amplification by nuclear speckle association. *J Cell Biol* 219: e201904046

Koh XY, Pek JW (2022) Passing down maternal dietary memories through lncRNAs. *Trends Genet*. S0168-9525(22)00198-6

Königs V, de Oliveira Freitas Machado C, Arnold B, Blümel N, Solovyeva A, Löbbert S, Schafranek M, Ruiz De Los Mozos I, Wittig I, McNicoll F *et al* (2020) SRSF7 maintains its homeostasis through the expression of Split-ORFs and nuclear body assembly. *Nat Struct Mol Biol* 27: 260 – 273

Lai F, Damle SS, Ling KK, Rigo F (2020) Directed RNase H cleavage of nascent transcripts causes transcription termination. *Mol Cell* 77: 1032 – 1043.e4

Lasko P (2012) mRNA localization and translational control in *Drosophila* oogenesis. *Cold Spring Harb Perspect Biol* 4: a012294

Lee J-S, Mendell JT (2020) Antisense-mediated transcript knockdown triggers premature transcription termination. *Mol Cell* 77: 1044 – 1054.e3

Li Z, Wang S, Cheng J, Su C, Zhong S, Liu Q, Fang Y, Yu Y, Lv H, Zheng Y *et al* (2016) Intron lariat RNA inhibits MicroRNA biogenesis by sequestering the dicing complex in *Arabidopsis*. *PLoS Genet* 12: e1006422

Lu C, Shen Q, DuPré E, Kim H, Hilsenbeck S, Brown PH (2005) cFos is critical for MCF-7 breast cancer cell growth. *Oncogene* 24: 6516 – 6524

Madeira F, Park Y m, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD *et al* (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* 47: W636 – W641

Magomedova L, Tiefenbach J, Zilberman E, le Billan F, Voisin V, Saikali M, Boivin V, Robitaille M, Gueroussov S, Irimia M *et al* (2019) ARGLU1 is a transcriptional coactivator and splicing regulator important for stress hormone signaling and development. *Nucleic Acids Res* 47: 2856 – 2870

Morgan JT, Fink GR, Bartel DP (2019) Excised linear introns regulate growth in yeast. *Nature* 565: 606 – 611

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

Murray DT, Kato M, Lin Y, Thurber KR, Hung I, McKnight SL, Tycko R (2017) Structure of FUS protein fibrils and its relevance to self-assembly and phase separation of low-complexity domains. *Cell* 171: 615 – 627.e16

Ng AYE, Pek JW (2021) Circular sisRNA identification and characterisation. *Methods* 196: 138 – 146

Ng AYE, Peralta KRG, Pek JW (2018a) Germline stem cell heterogeneity supports homeostasis in *Drosophila*. *Stem Cell Reports* 11: 13 – 21

Ng JSS, Teo ZR, Osman I, Pek JW (2018b) Generation of *Drosophila* sisRNAs by independent transcription from cognate introns. *iScience* 4: 68 – 75

Osman I, Pek JW (2018) A sisRNA/miRNA Axis prevents loss of germline stem cells during starvation in *Drosophila*. *Stem Cell Reports* 11: 4 – 12

Osman I, Pek JW (2021) Maternally inherited intron coordinates primordial germ cell homeostasis during *Drosophila* embryogenesis. *Cell Death Differ* 28: 1208 – 1221

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

Pandya-Jones A, Markaki Y, Serizay J, Chitiashvili T, Mancia Leon WR, Damianov A, Chronis C, Papp B, Chen CK, McKee R *et al* (2020) A protein assembly mediates Xist localization and gene silencing. *Nature* 587: 145 – 151

Parenteau J, Maignon L, Berthoumieux M, Catala M, Gagnon V, Abou Elela S (2019) Introns are mediators of cell response to starvation. *Nature* 565: 612 – 617

Patil VS, Kai T (2010) Repression of Retroelements in *Drosophila* germline via piRNA pathway by the Tudor domain protein Tejas. *Curr Biol* 20: 724 – 730

Pek JW (2018) Stable Intronic sequence RNAs engage in feedback loops. *Trends Genet* 34: 330 – 332

Pek JW, Kai T (2011) DEAD-box RNA helicase belle/DDX3 and the RNA interference pathway promote mitotic chromosome segregation. *Proc Natl Acad Sci U S A* 108: 12007 – 12012

Pek JW, Osman I, Tay MLI, Zheng RT (2015) Stable intronic sequence RNAs have possible regulatory roles in *Drosophila melanogaster*. *J Cell Biol* 211: 243 – 251

Petrella LN, Smith-Leiker T, Cooley L (2007) The Ovhts polyprotein is cleaved to produce fusome and ring canal proteins required for *Drosophila* oogenesis. *Development* 134: 703 – 712

Pirnie SP, Osman A, Zhu Y, Carmichael GG (2017) An ultraconserved element (UCE) controls homeostatic splicing of ARGLU1 mRNA. *Nucleic Acids Res* 45: 3473 – 3486

Sabari BR, Dall'Agnese A, Young RA (2020) Biomolecular condensates in the nucleus. *Trends Biochem Sci* 45: 961 – 977

Sanfilippo P, Wen J, Lai EC (2017) Landscape and evolution of tissue-specific alternative polyadenylation across *Drosophila* species. *Genome Biol* 18: 229

So BR, Di C, Cai Z, Venters CC, Guo J, Oh JM, Arai C, Dreyfuss G (2019) A complex of U1 snRNP with cleavage and polyadenylation factors controls telescripting, regulating mRNA transcription in human cells. *Mol Cell* 76: 590 – 599.e4

Spector DL, Lamond AI (2011) Nuclear speckles. *Cold Spring Harb Perspect Biol* 3: a000646

Statello L, Guo C-J, Chen L-L, Huarte M (2020) Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol* 22: 96 – 118

Talhouarne GJS, Gall JG (2018) Lariat intronic RNAs in the cytoplasm of vertebrate cells. *Proc Natl Acad Sci U S A* 115: E7970 – E7977

Tay MLI, Pek JW (2017) Maternally inherited stable Intronic sequence RNA triggers a self-reinforcing feedback loop during development. *Curr Biol* 27: 1062 – 1067

Tay ML-I, Pek JW (2019) SON protects nascent transcripts from unproductive degradation by counteracting DIP1. *PLoS Genet* 15: e1008498

Teo RYW, Anand A, Sridhar V, Okamura K, Kai T (2018) Heterochromatin protein 1a functions for piRNA biogenesis predominantly from pericentric and telomeric regions in *Drosophila*. *Nat Commun* 9: 1735

Tomita S, Abdalla MOA, Fujiwara S, Matsumori H, Maehara K, Ohkawa Y, Iwase H, Saitoh N, Nakao M (2015) A cluster of noncoding RNAs activates the ESR1 locus during breast cancer adaptation. *Nat Commun* 6: 6966

Ulitsky I (2016) Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nat Rev Genet* 17: 601 – 614

Unfried JP, Ulitsky I (2022) Substoichiometric action of long noncoding RNAs. *Nat Cell Biol* 24: 608 – 615

Venters CC, Oh JM, Di C, So BR, Dreyfuss G (2019) U1 snRNP telescripting: Suppression of premature transcription termination in introns as a new layer of gene regulation. *Cold Spring Harb Perspect Biol* 11: a032235

Voo K, Ching JWH, Lim JWH, Chan SN, Ng AYE, Heng JYY, Lim SS, Pek JW (2021) Maternal starvation primes progeny response to nutritional stress. *PLoS Genet* 17: e1009932

Wong JT, Akhbar F, Ng AYE, Tay MLI, Loi GJE, Pek JW (2017) DIP1 modulates stem cell homeostasis in *Drosophila* through regulation of sisR-1. *Nat Commun* 8: 759

Wu H-W, Deng S, Xu H, Mao H-Z, Liu J, Niu Q-W, Wang H, Chua N-H (2018) A noncoding RNA transcribed from the AGAMOUS (AG) second intron binds to CURLY LEAF and represses AG expression in leaves. *New Phytol* 219: 1480 – 1491

Wu M, Yang LZ, Chen LL (2021) Long noncoding RNA and protein abundance in lncRNPs. *RNA* 27: 1427 – 1440

Xing YH, Yao RW, Zhang Y, Guo CJ, Jiang S, Xu G, Dong R, Yang L, Chen LL (2017) SLERT regulates DDX21 rings associated with pol I transcription. *Cell* 169: 664 – 678.e16

Yin QF, Yang L, Zhang Y, Xiang JF, Wu YW, Carmichael GG, Chen LL (2012) Long noncoding RNAs with snoRNA ends. *Mol Cell* 48: 219 – 230

Yin Y, Lu JY, Zhang X, Shao W, Xu Y, Li P, Hong Y, Cui L, Shan G, Tian B *et al* (2020) U1 snRNP regulates chromatin retention of noncoding RNAs. *Nature* 580: 147 – 150

Zhang D, Jiang P, Xu Q, Zhang X (2011) Arginine and glutamate-rich 1 (ARGLU1) interacts with mediator subunit 1 (MED1) and is required for estrogen receptor-mediated gene transcription and breast cancer cell growth. *J Biol Chem* 286: 17746 – 17754

Zhang Y, Zhang XO, Chen T, Xiang JF, Yin QF, Xing YH, Zhu S, Yang L, Chen LL (2013) Circular Intronic long noncoding RNAs. *Mol Cell* 51: 792 – 806

