Nucleic Acids Research, 2025, **53**, gkaf486
https://doi.org/10.1093/nar/gkaf486
Gene regulation, Chromatin and Epigenetics

RNA-binding protein Ars2 mediates transcriptional silencing of telomeric repeats and transposable elements in the *Drosophila* germline

Valeriya Morgunova[1,†], Anastasiya A. Kobelyatskaya[1,2,†], Maksim Erokhin[3], Olesya Sokolova[1], Tatyana V. Sizova[1], Dmitry A. Kwon[4], Alla Kalmykova[1,*]

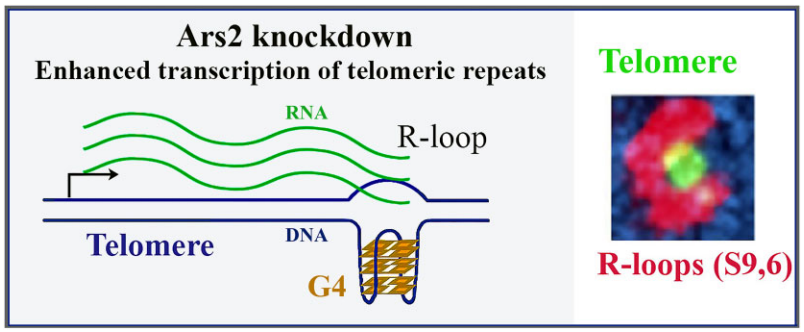
[1]Koltzov Institute of Developmental Biology of Russian Academy of Sciences, Moscow 119334, Russia
[2]Engelhardt Institute of Molecular Biology of Russian Academy of Sciences, Moscow 119991, Russia
[3]Institute of Gene Biology of Russian Academy of Sciences, Moscow 119334, Russia
[4]Syntol LLC Company, Moscow 127434, Russia

[*]To whom correspondence should be addressed. Email: allakalm@idbras.ru
[†]The first two authors should be regarded as Joint First Authors.

Abstract

Telomeres ensure genome stability and the levels of telomeric RNA reflect the integrity of telomeric chromatin. The highly conserved RNA-binding protein Ars2 (Arsenite-resistance protein 2) plays an essential role in the RNA nuclear metabolism and negatively regulates the expression of telomeric transcripts in human cells and in *Drosophila*. We found that germline knockdown of *Drosophila* Ars2 does not affect small RNA abundance but causes overexpression of telomeric repeats and transposable elements (TEs), accompanied by chromatin decompaction of these regions. The expression of a transgene containing the *HeT-A* telomeric retrotransposon was also affected by Ars2 knockdown. The mutation of the G-rich region, which is prone to the formation of G-quadruplex structures, reduces the *HeT-A* transgene's sensitivity to Ars2 depletion. Intriguingly, Ars2-regulated non-telomeric TEs are also enriched by G-quadruplex structures, implying their role in the Ars2 target recognition. Ars2 also prevents the formation of R-loops at telomeres, which are most likely caused by the accumulation of unreleased transcripts. Surprisingly, Ars2 is required for the expression of R1 retrotransposons, which are integrated in rRNA genes and essential for their amplification. Our findings point to a new mechanism for control of expression of telomeric repeats and TEs in the germline involving Ars2.

Graphical abstract



Introduction

Suppression of transposable element (TE) activity and telomere maintenance are the key mechanisms controlling genetic stability. Small RNAs, Piwi-interacting RNAs (piRNAs), mediate transcriptional silencing of transposons in animal gonads [1, 2]. piRNAs are also involved in the regulation of *Drosophila* telomeres, which are maintained by the transpositions of telomeric retrotransposons at the chromosome ends [3, 4]. Despite its high capacity, the piRNA-mediated transcriptional silencing is not the only mechanism involved in TE control. There is a growing body of evidence indicating that the mechanisms of transcriptional repression act in concert with co- and post-transcriptional mechanisms to limit TE activity. In yeast, a coordinated action of transcriptional and co-transcriptional pathways provides silencing of repetitive loci and degradation of the emerging RNAs through the exosome or small RNA pathway [5–10]. In *Drosophila*, the Piwi–piRNA complex is able to promote heterochromatin assembly

Received: June 25, 2024. **Revised:** April 16, 2025. **Editorial Decision:** April 21, 2025. **Accepted:** May 29, 2025

at TEs [11–13], degradation of their transcripts in the cytoplasm accompanied by the amplification of piRNAs [14–17], and co-transcriptional degradation of TE RNAs via recruitment of the nuclear Ccr4–Not deadenylation complex [18].

Our study revealed a new mechanism of transcriptional silencing of TEs and telomeric repeats in *Drosophila* mediated by the RNA-binding protein Ars2. This highly conserved protein with unknown biochemical activity plays an essential role in the nuclear metabolism of many types of RNA in plants, yeasts, and animals [19]. In human cancer cells, ARS2 depletion increased both steady-state and telomere-associated levels of telomeric repeat-containing RNA (TERRA) [20] and regulates telomerase RNA processing [21]. In *Drosophila*, Ars2 also serves as a negative regulator of telomeric retrotransposon expression, suggesting a conserved telomeric function for Ars2 independent of the nature of telomeric repeats [22]. In our study, we were primarily interested in understanding the mechanisms of Ars2 effects on *Drosophila* telomeric repeat transcription.

Telomerase was lost in *Drosophila* species. Telomeres in *Drosophila* are elongated by retrotranspositions of specialized telomeric retroelements *HeT-A*, *TART*, and *TAHRE* to the chromosome ends [23]. *HeT-A*, the basic structural telomeric element, is a non-autonomous retroelement with transcripts that are functionally similar to the telomerase RNA template; *TART* or/and *TAHRE* are thought to provide reverse transcriptase activity. Both species-specific features and conserved mechanisms were identified in studies of the biogenesis of the *Drosophila* telomeres. One of the distinct features of the telomeric chromatin in *Drosophila* is that telomere-specific piRNAs mediate the recruitment of the heterochromatin protein 1 (HP1) to telomeres through the recognition of telomeric retrotransposon transcripts [24]. The nuclear Ccr4–Not deadenylase complex participates in co-transcriptional recognition and degradation of telomeric *HeT-A* transcripts in telomere-associated bodies but does not affect telomeric chromatin state [18]. Our attention was drawn to the fact that Ars2 depletion in the *Drosophila* germline led not only to a hundred-fold increase in steady-state *HeT-A* RNA levels, but also to an increase in nascent *HeT-A* transcript levels, suggesting that RNA-binding protein Ars2 acts at the transcriptional level [22]. Accumulation of *HeT-A* ribonucleoprotein particles following Ars2 depletion in the *Drosophila* germline and syncytial embryos led to centrosome dysfunction, mitotic arrest, and early embryonic lethality, indicating a crucial role of Ars2 in oogenesis and early development [25].

ARS2 and its plant ortholog SERRATE are the critical factors of cell proliferation, early development, and stem-cell maintenance (reviewed in [19]). ARS2 is considered an adaptor protein involved in the formation of higher-order co-transcriptional complexes with diverse compositions, which determine the fate of nascent RNA [26]. ARS2 interacts with the RNA cap-binding complex and regulates the nuclear metabolism of many types of RNAs by affecting 3′-end processing, export, and transcript decay [27–31]. In human cells, ARS2 plays a genome-wide role in transcription termination of small nuclear RNAs (snRNA), enhancer RNAs, promoter upstream transcripts, and histone mRNA (messenger RNA) [27, 32, 33]. Ars2 is required for the biogenesis of histone mRNAs that are normally not polyadenylated and undergo cleavage downstream of a 3′-end stem-loop structure [33]. ARS2 is also required for the processing of a subset of microRNAs in *Drosophila*, mammalian cells [34, 35], and plants [36, 37]. Both histone mRNA and microRNAs contain secondary structures suggesting that Ars2-containing complexes could recognize and process structured RNAs. Despite a broad spectrum of Ars2 functions, its specific biochemical activity remains elusive.

In a few studies, the role of ARS2 in the recruitment of chromatin-modifying enzymes was demonstrated. In the fission yeast, *Schizosaccharomyces pombe*, the Ars2 ortholog was found in the nuclear RNA silencing complex participating in the degradation of meiotic RNAs [38]. *S. pombe* $Pir2^{ARS2}$ was also reported to mediate recruitment of Clr3 histone deacetylase and histone chaperone FACT, which enforces target gene repression [39]. The *Arabidopsis* ARS2 ortholog SERRATE recruits histone methyltransferases to repress TE transcription [40]. In mammals, ARS2 is a transcriptional regulator that is essential for mammalian neurogenesis [41, 42]. ARS2 depletion leads to reduced levels of transcription start site (TSS)-associated RNAPII at protein-coding genes [32]. These examples demonstrate that ARS2 acting co-transcriptionally can be involved in the chromatin-modifying complexes; however, the mechanisms underlying ARS2-mediated chromatin effects are unclear. It also remains an open question how this multifunctional protein is linked to the biogenesis of telomeric transcripts.

Transposon activity, along with telomere length and stability, which are critical for preserving genomic integrity across generations, is most tightly regulated in the germline. The processes governing telomere transcription and the maintenance of telomeric chromatin in the germline appear to differ from those in somatic cells [3, 18, 22, 24, 43, 44]. Here, using a genome-wide chromatin and transcription analysis, we demonstrated that germline knockdown of *Drosophila* Ars2 does not affect small RNA abundance but causes derepression of telomeric repeats and a group of non-telomeric TEs. This effect was accompanied by chromatin changes and telomeric DNA damage. Intriguingly, in contrast to the repressive role of Ars2 in telomeres, it is essential for the transcription of the R1 retrotransposons, which preferentially target rRNA genes and are likely required for their amplification [45]. The biogenesis of the R1 family of domesticated retrotransposons, which are found in many species, is still unknown. The identification of Ars2-mediated regulation of R1 expands the list of significant Ars2 targets.

Overall, our findings demonstrated a novel level of regulation of telomeric repeats and transposable genetic elements in the germline mediated by Ars-dependent chromatin remodeling.

Materials and methods

Drosophila melanogaster strains and genetic constructs

The transgenic strain H111 bearing a construct containing the *HeT-A* 3′ untranslated region (UTR) fragment fused to *lacZ* in the CaSpeR-AUG-β-gal vector was used in the study [46]. X-gal staining of the β-galactosidase in the ovaries was performed according to a previously published protocol [47]. The transgenic strains containing full-length tagged *HeT-A* transgene were described previously [48, 49]. pUAST-HeTA-HA-FLAG expresses *HeT-A* Gag protein tagged with HA and FLAG epitopes. pUAST-HeTA-HA-FLAG_ms2 contains the same HeT-A-HA-FLAG copy tagged by eight MS2 bacteriophage hairpins in the 3′UTR that allows detection of the transgenic *HeT-A* RNA. The strain expressing Ars2-GFP was

#318920 from Vienna Drosophila Research Center (VDRC). The strains bearing the constructs with short hairpin (sh) RNA were as follows: *white (w)*_sh (#30033, VDRC), *Ars2*_sh1 (#106344, VDRC), *Ars2*_sh2 (#22575, VDRC), *Ars2*_sh3 (#35204, Bloomington Drosophila Stock Center, BDSC), *FLASH*_sh (#100082, VDRC), *CPSF160*_sh (#110571, VDRC), *spnE*_sh1 (#103913, VDRC), and *spnE*_sh2 (#34808, BDSC). To induce tissue-specific RNAi knockdowns, we used genetic crosses of strains expressing shRNA with driver strains *P{GAL4-nos.NGT}40* (#25751, BDSC) for the germline knockdown (GLKD) or *P{GAL4-da.G32}UH1* for ubiquitous knockdown. Somatic tissues analyzed were brain and imaginal discs of third instar larvae. The driver strain *P{GAL4-nos.NGT}40* was maintained through "short generations," where offspring are from 3-day-old females. Short generations are expected to lead to depletion of the maternal pool of small RNAs, thereby ensuring stable expression of the short hairpin constructs targeted by endogenous hsp70-directed piRNAs [50, 51].

The construct pUAST-attB-HeTA-HA-FLAG-G4mut was created on the basis of pUAST-attB-HeTA-HA-FLAG. To mutate the nucleotides of the HeT-A_G4_4 (DmHc) fragment, presumably involved in the formation of the G-quadruplex (G4) [52, 53] and located in the *HeT-A* 3′UTR, two fragments were amplified using primers HeTA_3UTR_s/DmHc_1_rev and DmHc_2_fw/DmHc_2_rev on the template of pUAST-attB-HeTA-HA-FLAG (Supplementary Table S1). Then, the two polymerase chain reaction (PCR) fragments were combined and amplified by HeTA_3UTR_s and DmHc_2_rev primers, cut KpnI and XbaI, and cloned into pUASTattB-HeTA-HA-FLAG digested by KpnI–XbaI. The construct was integrated in the attP docking site on chromosome 3 (strain #24862, BDSC) by ϕC31 attB recombination.

Detection of G-quadruplex DNA by native polyacrylamide gel electrophoresis

Oligonucleotides listed in Supplementary Table S1 (3 μl of 5 μM solution) were incubated in 100 mM Tris–HCl (pH 8.0) and 10 mM KCl at 25°C for 20 min, and then run in 20% polyacrylamide gels (acrylamide/bisacrylamide ratio 19:1) in 1× TBE buffer [80 mM Tris–borate, 2 mM ethylenediaminetetraacetic acid (EDTA), pH 8.3]. The gel was stained with a solution of TOR-G4 (Lumiprobe, USA) according to the manufacturer's instructions or with ethidium bromide (0.5 μg/ml). The images were acquired using a ChemiDoc imaging system (Bio-Rad, USA).

Dot-blot analysis of R-loops

XbaI–KpnI fragments of *HeT-A* 3′UTR of pUAST-attB-HeTA-HA-FLAG and pUAST-attB-HeTA-HA-FLAG_G4mut were used to obtain constructs in pBlueScript SK(−) designated HeT-A_G4 and HeT-A_G4mut, respectively. The *HeT-A* open reading frame (ORF) fragment comprising G4_1 and G4_2 motifs was amplified using primers indicated in Supplementary Table S1 on the template of pUAST-attB-HeTA-HA-FLAG and was cloned into pBlueScript SK(−) (HeT-A_G4_1–2). A fragment of the *HeT-A* ORF (nucleotides 4330–4690 of GenBank sequence DMU06920) lacking G4 motifs and cloned in pBlueScript SK(−) was used as a negative control (HeT-A_ORF). The fragments of *gypsy12* (chr2R:4272523–4274243) and *Tirant* (chr2L:21041588–21043676) retrotransposons were amplified using primers indicated in Supplementary Table S1, utilizing the genomic DNA of *iso-1* strain used for the *D. melanogaster* genome sequencing. These PCR fragments were subsequently used to amplify *gypsy12* and *Tirant* fragments comprising several G4 motifs, and the amplicons were cloned in pBlueScript SK(−) to generate the gypsy_G4 and Tirant_G4 constructs. The orientation of fragments in all constructs ensured production of sense RNA from the T7 RNA polymerase promoter, while G4 motifs were located on the template strand. All constructs were verified by Sanger sequencing. The concentrations of linearized templates were measured using a Qubit 4 Fluorometer (Thermo Fisher Scientific, USA). The linearized plasmids (200 ng per sample) were used for *in vitro* transcription with T7 RNA polymerase. Aliquots of all samples were saved for electrophoresis. The sample was then processed with RNase A to remove free RNA and obtain RNA–DNA complexes. Half of the sample was treated with RNase H, which eliminated RNA from an RNA–DNA hybrid, and served as a control for the specificity of R-loop staining. Tris–EDTA buffer (TE), pH 8.0 was used as a negative control. DNA–RNA complexes were dot blotted and mobilized using UV cross-linking onto Hybond N+ membrane (GE Healthcare, USA). The membrane was incubated with the S9.6 antibody that recognizes DNA–RNA hybrids (Millipore, Germany, MABE1095, 1:500), washed, and incubated with anti-mouse IgG alkaline phosphatase conjugated antibody (Jackson ImmunoResearch, USA, 1:7000). The signals were detected by Immun-Star™ AP Substrate (Bio-Rad, USA) and visualized by a ChemiDoc Imaging System (BioRad, USA). Statistical analysis of signal intensity on dot blots ($n = 4$) was performed using ImageJ tools (http://imagej.nih.gov/ij/).

Immunostaining, RNA fluorescence *in situ* hybridization, and immunoprecipitation

Immunostaining and RNA fluorescence in situ hybridization (FISH) combined with immunostaining were carried out according to the published protocols [54]. The previously published protocol [55] was used for the R-loop immunostaining with modifications. After fixation, the specimens were incubated with 50 μg/ml proteinase K solution in 1× phosphate buffered saline for 8 min. RNase III and RNase H treatments were done in blocking buffer (3% bovine serum albumin, 0.2% NP-40, 0.2% Tween 20, 10% dry milk) containing 3 mM $MgCl_2$ with RNase III (New England Biolabs, USA) and RNase H (for RNaseH+ controls) (New England Biolabs, USA) with a 1-h incubation period. The S9.6 antibody (Millipore, Germany, MABE1095) was diluted 1:50 in the blocking buffer. Calculation of S9.6/HOAP foci was done from two biological replicates ($n = 20$, n is the number of nurse cell nuclei quantified). GraphPad Prism 8.0 was used for statistical analysis and graphing. Colocalization analysis, including quantifications of Pearson's correlation coefficient (PCC) and Manders' Colocalization coefficient (MCC), was performed using ImageJ JACoP plugin. In the PCC and MCC plots, each point represents the mean values from 10 slices for each nurse cell nucleus.

A digoxigenin (DIG)-labeled antisense *HeT-A* riboprobe containing a fragment of the ORF (nucleotides 4330–4690 of GenBank sequence DMU06920) was used. The MS2 antisense probe was a 548-bp long PCR fragment containing eight MS2 repeats and a T7 RNA polymerase promoter ampli-

fied using the UAST-HeT-A-HA-FLAG-ms2 construct. Images were captured using a Zeiss LSM 880 confocal microscope and analyzed using ImageJ and Adobe Photoshop. Airyscan detector was used for a super-resolution imaging. The following primary antibodies were used: guinea pig anti-HOAP [25]; rabbit anti-POF (kindly provided by J. Larsson [56]); rat anti-Ars2 (generated in the Immunochemistry Laboratory, Branch of IBC RAS, Pushchino, Russia, using recombinant protein corresponding to the 151–300 amino acids of *D. melanogaster* Ars2); mouse anti-DNA–RNA Hybrid Antibody, clone S9.6 (Millipore, Germany); mouse anti-γH2Av (Developmental Studies Hybridoma Bank, USA); rabbit anti-trimethyl-histone H3 Lys9 (Millipore, Germany); and rabbit anti-HP1a (kindly provided by S. Elgin). Alexa Fluor-conjugated secondary antibodies with minimal cross-reactivity to IgG from non-target species (Jackson ImmunoResearch, USA) were used (dilution 1:500). Immunoprecipitation of Ars2 from ovarian nuclear extract was performed as described [48].

Reverse transcription quantitative PCR

RNA was isolated from 20 pairs of ovaries of 2–3-day-old females. cDNA was synthesized using random hexamers and M-MuLV reverse transcriptase (Biolabmix, Russia). The cDNA samples were analyzed by real-time quantitative PCR (qPCR) on a LightCycler 96 (Roche, Switzerland). Values were averaged and normalized to the expression levels of the ribosomal protein gene *rp49*.

Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) on *Drosophila* ovaries was performed according to a published procedure [57]. ChIP using dual cross-linking with ethylene glycol bis(succinimidyl succinate) and formaldehyde was performed according to the published protocol [58]. Chromatin was immunoprecipitated with the following antibodies: anti-trimethyl-histone H3 Lys9 (Millipore, Germany); anti-dimethyl-histone H3 Lys4 (Millipore, Germany); anti-RNA polymerase II (Millipore, Germany); anti-trimethyl-histone H3 Lys36 (Abcam, UK); and anti-Rhino [46]. The primers used in this study are listed in Supplementary Table S1. qPCR was conducted with a LightCycler 96 (Roche, Switzerland). The obtained values were normalized to input. Standard error of the mean (SEM) of duplicate PCR measurements for four biological replicas was calculated. The libraries for ChIPseq were prepared using MGIEasy FS DNA Library Prep Set (MGI, China) and sequencing was done on an MGI Tech DNBSEQ 400 platform (MGI, China). ChIPseq data are deposited in the GEO database (GSE266038).

Preparation of chromatin-associated and small RNA libraries

The libraries of chromatin-associated RNA were prepared from the ovaries of *Ars2*_GLKD1, *Ars2*_GLKD2, and two independent *w*_GLKD according to a previously published procedure [18]. Small RNAs (19–29 nt) from total ovarian RNA of *Ars2*_GLKD1 and *w*_GLKD were cloned as previously described [59]. Sequencing was performed on the Illumina HiSeq1500 platform (Illumina, USA). ChromRNAseq and small RNAseq data are deposited in the GEO database (GSE266036 and GSE266037).

Sequencing data preprocessing

FastQC (v.0.11.9, Cambridge, UK, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and Trimmomatic (v.0.33, Jülich, Germany) [60] were used for the quality control and trimming of reads, respectively.

ChromRNAseq data processing

The STAR splice-aware aligner (v.2.7.1, Cold Spring Harbor, NY, USA) [61] was used to map the reads to the reference genome (BDGP assembly R6/dm6) and the canonical sequences of TEs (https://github.com/bergmanlab/drosophila-transposons). The Rsamtools package (https://bioconductor.org/packages/Rsamtools) was used to process bam files.

FeatureCounts (RSubread package v.1.6.4, Parkville, VIC, Australia) [62] was used to calculate the read counts for gene (Ensembl) [63] and repeat databases (RepeatMasker) [64] prepared by the biomaRt package (v.2.54.0, Berkeley, CA, USA) [65].

Differential expression analysis was carried out in the R environment (v.3.6.3, Vienna, Austria, https://www.r-project.org/) using the edgeR package (v.3.24.3, NSW, Australia) [66]. To normalize the obtained data, the TMM (trimmed mean of *M*-values) method was applied with the calculation of CPM (counts per million) considering the normalization coefficients.

The quasi-likelihood *F*-test was used to evaluate the significance of the changes. The fold change in the expression level between genotypes (FC, in the $\log_2$ transformation, $\log_2$FC) and the mean expression level in the cohort (CPM, in the $\log_2$ transformation, $\log_2$CPM) were calculated.

Small RNAseq data processing

Small RNA reads that passed quality control and minimal length filter (>18 nt) were mapped (allowing 0 or 0–3 mismatches) to the reference genome (BDGP assembly R6/dm6), the canonical sequences of TEs (https://github.com/bergmanlab/drosophila-transposons), and miRNA sequences of miRbase [67] by bowtie2 [68]. To identify piRNAs, we used reads with a length from 24 to 29 nt, and to identify siRNAs we used 21 nt length reads. The genomic coordinates of piRNA clusters were obtained from [14]. Further data processing (processing bam files, counting) and statistical analysis (differential expression analysis) were similar to ChromRNAseq.

ChIPseq data processing

Bowtie2 aligner [68] was used to map the reads to the reference genome (BDGP assembly R6/dm6) and the canonical sequences of TEs (https://github.com/bergmanlab/drosophila-transposons). The input signal was subtracted from the normalized ChIP signals (CPM). Further data processing (processing bam files, counting) and statistical analysis (differential enrichment analysis) were similar to ChromRNAseq.

Results visualization

To calculate the enrichment per genome regions, computeMatrix (deepTools) [69] was used. For data visualization ggplot2 R package (https://ggplot2.tidyverse.org/) was used.

Results

The expression profile of Ars2 in *Drosophila* ovaries

Ars2 is expressed in a wide range of *Drosophila* tissues, with the highest expression observed in the ovaries (modENCODE Tissue Expression Data). To examine the localization of Ars2 protein in the ovaries, we generated polyclonal rat anti-Ars2 antibodies. Western blot analysis demonstrated that the antibodies specifically recognize Ars2 protein in the ovary lysate of wild-type strain and the transgenic strain expressing Ars2-GFP (Fig. 1A).

The main morphological unit of *Drosophila* ovaries is the ovariole (Fig. 1B), which consists of the germarium and a series of developing egg chambers. The early stages of oogenesis take place in the germarium. The egg chamber consists of a cluster of germ cells (the oocyte and 15 nurse cells) surrounded by somatic follicle cells. Ovary immunostaining detects Ars2 in the nuclei of both somatic follicle and germ cells (Fig. 1B). Ars2 exhibited a broad nuclear staining with an affinity for chromatin in the nurse cells but has a nucleoplasmic localization in the oocyte. Immunoprecipitation of Ars2 from the ovarian nuclear protein extract demonstrated that anti-Ars2 antibodies are efficient in immunoprecipitation (Supplementary Fig. S1A). However, ChIP analysis of Ars2 using standard protocol or dual cross-linking did not reveal the direct binding of Ars2 to telomeric chromatin (Supplementary Fig. S1B).

We then performed co-immunostaining of Ars2 and telomere-specific HOAP protein [70] in wild-type ovaries. Due to the polyploidy of the nurse cell nuclei, HOAP staining reveals numerous signals, many of which appear in the Ars2 staining regions (Fig. 1C). We noticed that Ars2 is excluded from the heterochromatic regions characterized by bright DAPI staining (Fig. 1C). Among them is a large cluster of HOAP signals at nuclear periphery. In the *Drosophila* genome, the heterochromatic fourth chromosome interacts with chromosome-specific painting of fourth (POF) protein [71, 72]. Co-immunostaining of HOAP and POF confirms that the telomeres of the fourth chromosome are localized in this region (Fig. 1D). Thus, Ars2 is excluded from the heterochromatin environment of the fourth chromosome. Moreover, heterochromatin domains and Ars2-occupied regions did not overlap in the nurse and follicle cells of wild-type ovaries, as demonstrated by co-immunostaining and quantitative colocalization analyses of heterochromatin components HP1 and trimethylated at lysine 9 histone 3 (H3K9me3) with Ars2 (Figs. 1E and F, and Supplementary Fig. S2). In contrast, telomeres (indicated by HOAP staining) were localized in both heterochromatic and euchromatic environments (Supplementary Fig. S2). Thus, Ars2 preferentially occupied euchromatic regions, where a fraction of telomeres was also found.

Ars2 depletion causes accumulation of telomeric *HeT-A* transcripts at telomeres and telomere instability in the germline

Since loss of Ars2 is lethal for *Drosophila* [35], we performed RNAi GLKD of Ars2 to study its telomeric role in the germline. Strong overexpression of telomeric retrotransposon *HeT-A* and moderate accumulation of *TART* transcripts were observed in three independent strains expressing Ars2-specific short hairpin RNA in the germline, indicating target-specific effects of knockdowns (Fig. 2A). Ars2 RNAi knockdown in somatic tissues resulted in pupae lethality; therefore, we investigated telomeric retrotransposon expression in the larval brain and imaginal discs. We discovered a very mild effect of Ars2 knockdown on *HeT-A* and *TART* expression in larval somatic tissues (Fig. 2A), suggesting a germline-specific role for Ars2 in telomeric retroelement expression. *Ars2*_GLKD1 was used in most further experiments. Western blot analysis using ovarian protein extracts demonstrated a reduction in Ars2 protein levels following knockdown (Supplementary Fig. S3). However, this result does not accurately reflect the knockdown's efficacy since the ovaries contain both germ and somatic follicle cells, which retain the protein and mask the knockdown effect. We demonstrated the cell-specific impact of Ars2 germline knockdown using Ars2 immunostaining to show that Ars2 was retained in somatic follicle cells but almost lost in the germ nurse cells and the oocyte following *Ars2*_GLKD1 (Supplementary Fig. S3).

Ars2 is a component of multiple alternative complexes acting co-transcriptionally. We performed *HeT-A* RNA FISH on *Ars2*_GLKD1 ovaries to study the distribution of *HeT-A* sense transcripts after Ars2 depletion. We found accumulation of abundant *HeT-A* RNAs in nurse cell nuclei at telomeres stained for telomere-specific HOAP protein (MCC of *HeT-A* RNA with HOAP is 87% ± 7%); however, *HeT-A* RNA was not detected at large heterochromatic telomere clusters (Fig. 2B and D, and Supplementary Fig. S4A).

We next address the role of RNA 3′-end processing factors other than Ars2 in *HeT-A* expression. We chose the CPSF160 subunit of the polyadenylation complex [73] and FLASH, a factor involved in histone mRNA processing in *Drosophila* [73] and previously shown to interact with ARS2 in mammalian cells [26]. Both *FLASH*_GLKD and *CPSF160*_GLKD resulted in the accumulation of *HeT-A* transcripts (Supplementary Fig. S4B). Moreover, *HeT-A* RNAs were detected at telomeres in nurse cell nuclei of *FLASH*_GLKD (Fig. 2C and D, and Supplementary Fig. S4A). These data suggest that the 3′-processing complex regulates the stability of telomeric *HeT-A* transcripts in the nucleus.

Retention of abundant *HeT-A* transcripts at telomeres could result in accumulation of RNA–DNA hybrids—R-loops—and telomere instability. In turn, formation of R-loops can lead to DNA damage caused by induced double-strand (ds) DNA breaks and recombination [74]. We performed immunodetection of R-loops at telomeres after Ars2 knockdown using a S9.6 antibody to recognize RNA–DNA hybrids [75]. It has been demonstrated that RNase III treatment effectively eliminates dsRNAs primarily produced from centromeric repeats, allowing for the genome-wide detection of genuine R-loops utilizing the DNA–RNA immunoprecipitation (DRIP)-seq approach with S9.6 antibody in yeast and human cells [55, 76]. Therefore, *Drosophila* ovaries were pretreated with RNase III to remove dsRNA recognized by S9.6 antibody. While RNase III treatment had no significant influence on telomeric S9.6 staining, a wide genomic region was sensitive to RNase III (Supplementary Fig. S5), demonstrating that RNase III treatment is essential for accurate R-loop detection in tissues and cells using S9.6 immunostaining. The S9.6 immunostaining observed at telomeres of non-heterochromatic chromosomes in *Ars2*_GLKD1 ovaries was RNase H-sensitive suggesting that R-loops accumulate at telomeres upon Ars2

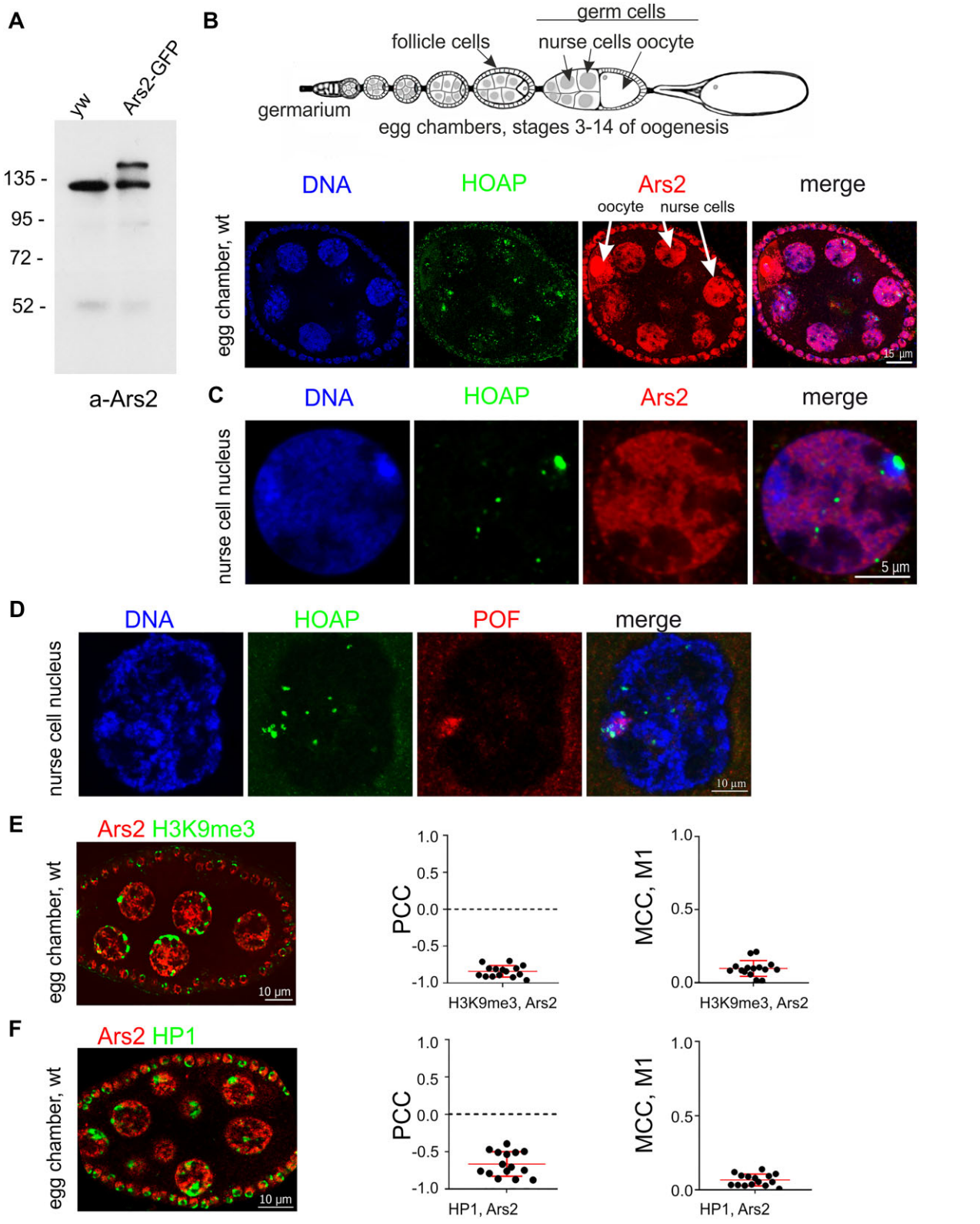


Figure 1. Ars2 subcellular localization in the *Drosophila* ovaries. (**A**) Western blot analysis of ovary lysates prepared from a control *yw* strain and a strain expressing Ars2-GFP probed with anti-Ars2 antibodies. (**B**) The structure of *Drosophila* ovariole is depicted above. The ovariole consists of a germarium and developing egg chambers composed of somatic (follicle) and germ cells (oocyte and nurse cells). Coimmunostaining of HOAP (to detect telomeres, green) and Ars2 (red) was performed on the ovaries of wild-type strain. A stage 6 egg chamber is shown. (**C**) Coimmunostaining of HOAP and Ars2 in the nucleus of a nurse cell of wild-type strain. (**D**) Coimmunostaining of HOAP and POF in the nucleus of a nurse cell of wild-type strain. Co-immunostaining of Ars2 and H3K9me3 (**E**) or HP1 (**F**) was performed on the ovaries of wild-type strain. The representative images of stage 7 egg chambers are shown. The graphs depict PCC and Manders' colocalization coefficients (MCC), with M1 representing the colocalization of H3K9me3 or HP1 signals with Ars2 signals. Error bars indicate SEM, $n = 15$ (n is the number of nurse cell nuclei analyzed). DNA is stained with DAPI.

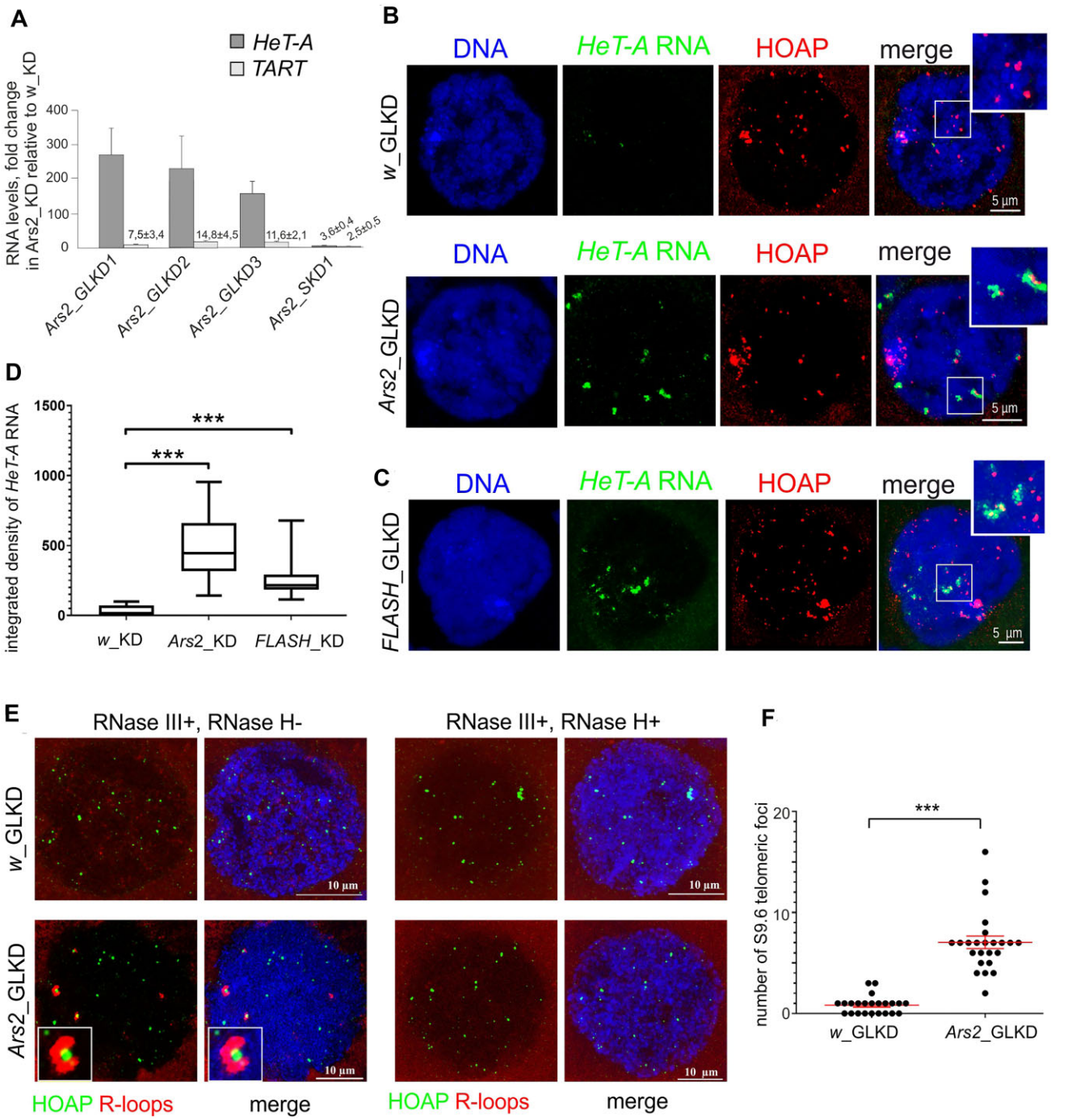


Figure 2. Ars2 and its partners regulate *HeT-A* expression. (**A**) Reverse transcription qPCR (RT-qPCR) analysis of *HeT-A* and *TART* RNA levels in total ovarian RNA from the indicated genotypes was performed relative to *w*_GLKD (normalized to *rp49*). Average levels ($n = 4$) and error bars (SEM) are indicated. (**B**) *HeT-A* RNA FISH (to detect sense RNA) combined with HOAP immunostaining was performed on the ovaries of *w*_GLKD and *Ars2*_GLKD1 strains. The representative images of the nurse cell nuclei and their enlarged regions are shown. (**C**) *HeT-A* RNA FISH combined with HOAP immunostaining was performed on the ovaries of the *FLASH*_GLKD strain. The nurse cell nucleus and its enlarged region are shown. (**D**) The box plots represent the means of integrated density of *HeT-A* RNA FISH signals per nurse cell nucleus in *w*_GLKD, *Ars2*_GLKD, and *FLASH*_GLKD ($n = 15$, three independent experiments). The horizontal line represents the median, box limits represent the first and the third quartiles, and box whiskers span from minimal to maximal values of the data range. ***$P < .001$, unpaired *t*-test. (**E**) Coimmunostaining of HOAP and RNA–DNA hybrids (S9.6 antibody) was performed on ovaries of *w*_GLKD and *Ars2*_GLKD1 strains pretreated with RNase III or a combination of RNase III and RNase H. The nuclei of nurse cells are shown. Enlarged region of the *Ars2*_GLKD nucleus (rectangle) is shown on the left. DNA is stained with DAPI. (**F**) Quantification of the nuclear telomeric S9.6 foci in the nurse cells of *w*_GLKD and *Ars2*_GLKD is shown. ***$P < .001$, unpaired two-tailed *t*-test.

depletion (Fig. 2E). S9.6 foci were revealed in the close proximity of telomeric HOAP staining and partially overlapped it. We counted the S9.6 telomeric foci in *w*_GLKD and *Ars2*_GLKD and found their significantly increased number following *Ars2* knockdown (Fig. 2F). Immunostaining of HOAP and γH2Av, a histone H2 phosphorylated variant associated with dsDNA breaks, combined with *HeT-A* RNA FISH revealed that telomeric regions where *HeT-A* RNA accumulated were associated with DNA lesions in *Ars2*_GLKD1 (Supplementary Fig. S6). In the control, where *HeT-A* expression is low and *HeT-A* RNA does not accumulate in telomeres, co-localization of HOAP and γH2Av was not identified. Thus, *Ars2*_GLKD causes accumulation of telomeric transcripts at telomeres of non-heterochromatic chromosomes accompanied by telomere instability.

Ars2 depletion in the germline causes derepression of telomeric repeats and non-telomeric TEs

Ars2 knockdown leads to the accumulation of *HeT-A* transcripts at telomeres suggesting that these transcripts are recognized as export-incompetent and retained at telomeres. To characterize genome-wide changes in chromatin-bound RNA caused by Ars2 depletion, we performed chromatin-associated RNA sequencing, chromRNAseq [18], from the ovaries of the control *w*_GLKD and two independent *Ars2* GLKDs. This RNA fraction contains nascent RNAs and transcripts that are retained at the sites of transcription [18]. As expected, the chromRNA fraction was enriched with unprocessed RNAs (Supplementary Fig. S7). We performed differential expression analysis of chromRNAseq data for the annotated TEs (*D. melanogaster* genome assembly, dm6) and found derepression of telomeric retrotransposons and certain TE families (Fig. 3A). We generated a list of all annotated TE copies affected by *Ars2* knockdown ($\log_2$ chromRNA fold change $\geq$1.5). Our results show that TEs related to *gypsy*, *Jockey*, *Pao*, and *CR1* superfamilies are the most sensitive to the Ars2 loss among non-telomeric TEs (Fig. 3A and Supplementary Table S2_1). Strongest upregulation in *Ars2*_GLKD was observed for *HeT-A* and *TAHRE* telomeric retrotransposons (Fig. 3B). Surprisingly, *Ars2* knockdown had the opposite effect on the expression of R1 non-LTR retrotransposons (Fig. 3B), which insert specifically into the 28S rRNA genes [77]. Analysis of chromatin-bound transcripts of the piRNA clusters revealed strong upregulation of the telomeric piRNA clusters that consist of *HeT-A*, *TAHRE*, and *TART* (Supplementary Table S2_2). *Ars2*_GLKD showed a small effect on the levels of coding gene transcripts, long noncoding RNAs, and miRNA precursors (Supplementary Fig. S8 and Supplementary Table S2_3).

Next, we estimated how *Ars2*_GLKD affected small RNA abundance in the germline by small RNA sequencing (sRNAseq). We have examined separately three classes of small RNAs, piRNAs, siRNAs, and miRNAs, taking into account the previously reported role of Ars2 in the processing of siRNAs and a subset of miRNAs in *Drosophila* somatic cells [35]. Ars2 is not required and does not affect piRNA production from telomeric regions, transposons, and piRNA clusters (Fig. 3C and Supplementary Table S2_4,5). Although piRNA production is not affected by the Ars2 knockdown, we observed strong accumulation of telomeric TE transcripts and a decrease in R1 RNA levels (Fig. 3E). We did not observe significant changes in the production of siRNAs from TEs and miRNAs (Fig. 3C, Supplementary Fig. S9, and Supplementary Table S2_6,7), suggesting that Ars2 is not generally required for the biogenesis of small RNAs in the germline. Importantly, Ars2 does not compete with the piRNA system for the RNA substrate since Ars2 knockdown has no effect on the abundance of piRNAs derived from the Ars2-regulated TEs.

Although a wide spectrum of TEs is regulated by the piRNA system, Ars2 selectively affects TE expression without interfering with the piRNA pathway, implying a distinct mechanism of action.

Ars2 germline knockdown causes chromatin changes of telomeric and non-telomeric retrotransposons

To address a mechanism of Ars2-mediated repression of telomeric repeats, we performed chromatin analysis of *Ars2*_GLKD ovaries using ChIPseq with antibodies against histone H3 modifications associated with active (H3K4me2) and compact chromatin (H3K9me3). The most pronounced effects were the reduction of H3K9me3 histone modification at telomeric repeats and an increase in H3K9me3 at R1 retrotransposons (Fig. 4A and B). We found enhanced enrichment of H3K4me2 at telomeric retroelements (Fig. 4C). These results are in agreement with the previously observed transcriptional activation of *HeT-A* revealed by nuclear run-on transcription [22]. In contrast, the chromatin of R1 retrotransposons was more compacted in *Ars2*_GLKD, which is in accordance with the reduced levels of R1 expression. *Ars2*_GLKD resulted in the general effect of the increasing of H3K4me2 enrichment at LINE and LTR TEs (Fig. 4D, $\log_2$CPM_ChIPseq > 2). At the same time, *Ars2*_GLKD did not strongly affect the levels of H3K9me3 at these TE groups (Supplementary Fig. S10).

Ars2 depletion also affected chromatin state of DNA class of TEs. For example, *Bari-1*, a transposon related to the Tc1/mariner family, was derepressed in *Ars2*_GLKD and H3K9me3 chromatin marks were lost (Supplementary Fig. S11). RT-qPCR analysis confirmed derepression of *Bari-1* in *Ars2*_GLKD (Supplementary Fig. S11).

HeT-A telomeric element is the most sensitive target of Ars2; therefore, we performed a detailed analysis of *HeT-A* chromatin after Ars2 depletion in the germline by ChIP-qPCR. We analyzed changes in histone modifications associated with transcriptionally active and compact chromatin at *HeT-A* ORF region in the ovaries of *Ars2*_GLKD. A significant increase in active chromatin histone modifications H3K4me2 and H3K36me3 was found at the *HeT-A* ORF region, which is in accordance with ChIPseq data (Fig. 4E). *Ars2*_GLKD did not significantly affect binding of Rhino (Fig. 4E and Supplementary Fig. S12)—a germline-specific ortholog of HP1 required for expression of piRNA precursor transcripts [78, 79]. This observation agrees with the finding that Ars2 depletion did not affect piRNA production from piRNA source loci (Fig. 3C).

Overall, these results indicate that the chromatin of telomeric and non-telomeric TEs has a mixed pattern of active and repressive chromatin marks and that the erasure of active histone modifications from the many TEs is to a large extent dictated by Ars2. In contrast, Ars2 is required for the expression of R1 retrotransposons in the germline, suggesting that Ars2 can mediate different mechanisms of TE regulation.

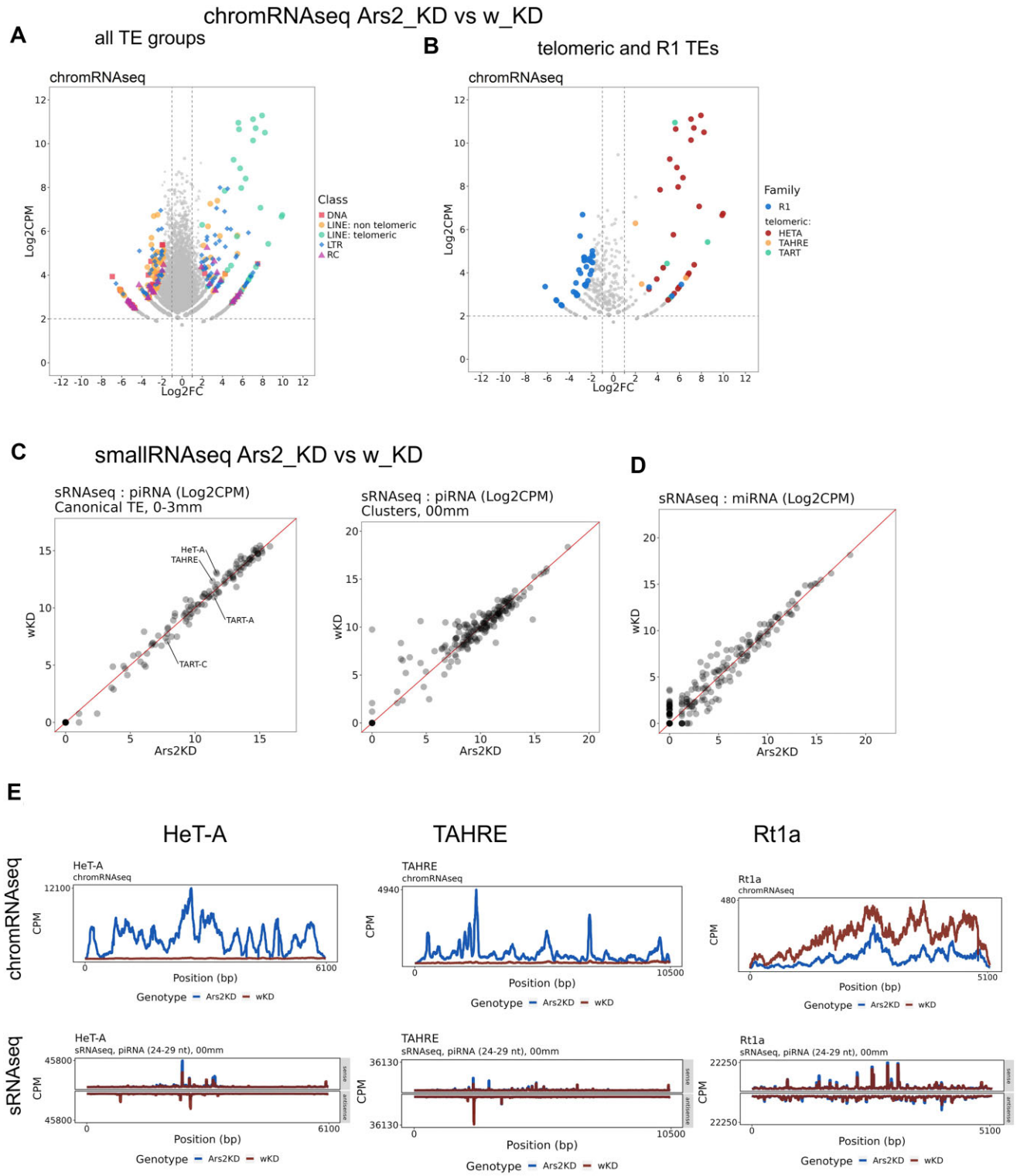


Figure 3. Ars2 depletion in the germline causes overexpression of telomeric retrotransposons and distinct non-telomeric TEs. (**A**) Differential analysis of TE chromatin-bound RNA abundance using ovarian chromRNAseq of *w*_GLKD and *Ars2*_GLKD. (**B**) Differential analysis of chromatin bound RNA abundance of telomeric and R1 retrotransposons using ovarian chromRNAseq of *w*_GLKD and *Ars2*_GLKD. (**C**) Scatter plots compare expression of piRNAs derived from TEs and piRNA clusters between *w*_GLKD and *Ars2*_GLKD ovaries. (**D**) Scatter plot demonstrating similar expression of miRNAs in *w*_GLKD and *Ars2*_GLKD ovaries. (**E**) Densities of the RNAseq reads (top plots) and small RNAseq reads (24–29 nt) (bottom plots) for *HeT-A*, *TAHRE*, and *Rt1a* in *w*_KD and *Ars2*_KD. CPM, counts per million.

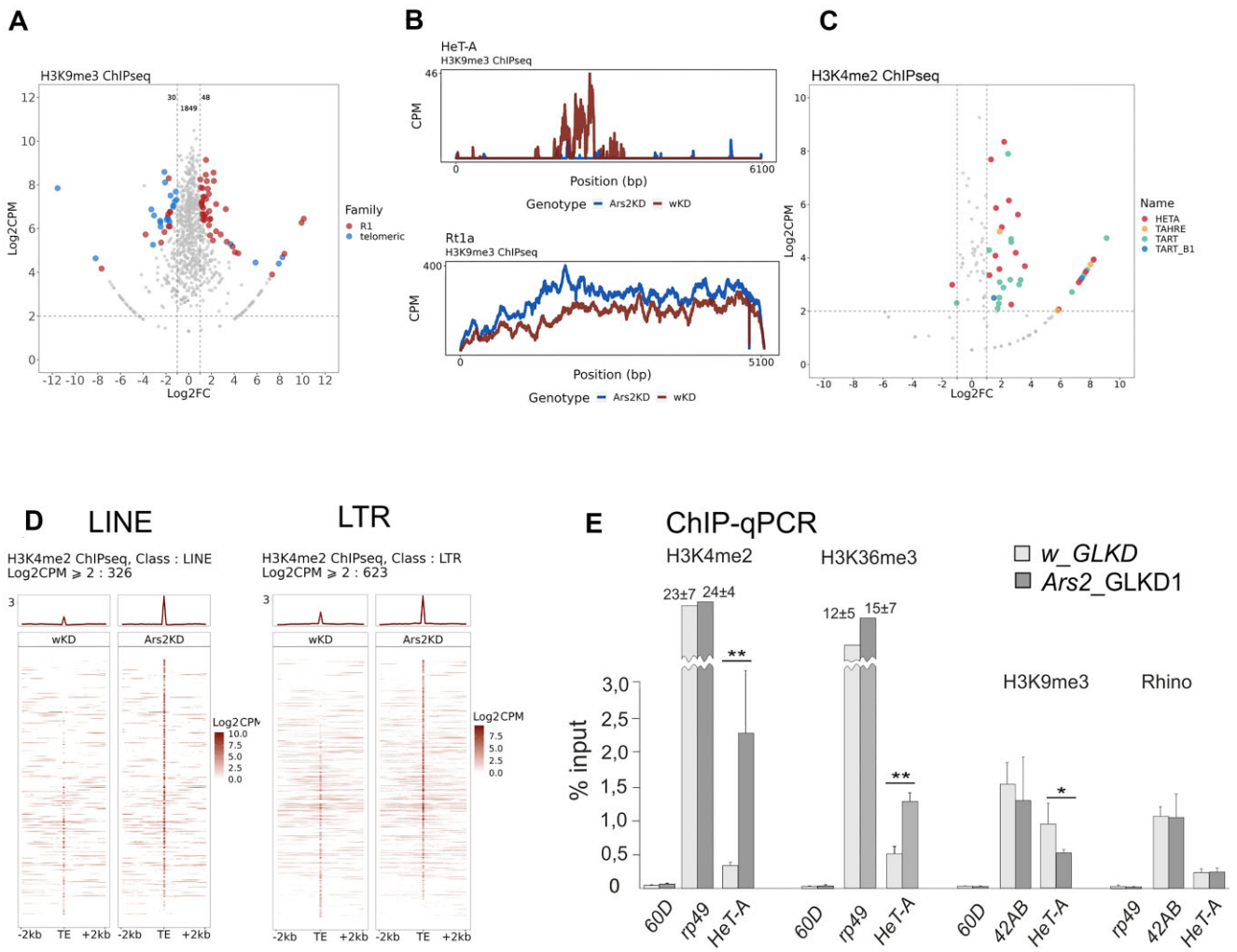


Figure 4. Ars2 regulates the chromatin state of telomeric and non-telomeric retrotransposons in the germline. (**A**) Volcano plot showing differential H3K9me3 ChIPseq enrichments at telomeric and R1 retrotransposons. (**B**) H3K9me3 ChIPseq reads plotted to the canonical sequence of *HeT-A* and *Rt1a* in *w*_KD and *Ars2*_KD. (C) Differential analysis of H3K4me2 ChIPseq enrichments for telomeric retrotransposons in the ovaries of *w*_GLKD and *Ars2*_GLKD. (**D**) Heatmaps show H3K4me2 ChIPseq signals at LINE and LTR TEs ($\log_2$CPM $\geq$ 2) and 2-kb genomic regions upstream and downstream of TE insertions in *w*_GLKD and *Ars2*_GLKD. (**E**) H3K4me2, H3K36me3, H3K9me3, and Rhino binding to *HeT-A* in the ovaries of *w*_GLKD and *Ars2*_GLKD were estimated by ChIP–qPCR. Average levels ($n = 4$) and error bars (SEM) are indicated. Asterisks indicate statistically significant differences (*$P < .05$ to .01, **$P < .01$ to .001, unpaired *t*-test).

A G-rich sequence in the 3′UTR of *HeT-A* is required for Ars2-mediated regulation

The strong effect of Ars2 on endogenous telomeric repeats could be mediated by the telomere-specific components or by telomere-specific genomic context. The *HeT-A* promoter has an unusual organization, as it is positioned within the 3′UTR and can drive transcription of the downstream element in the telomeric retrotransposon arrays [80]. Furthermore, the *HeT-A* promoter is bidirectional and can drive divergent transcription [46].

To address the potential role of genomic context in the Ars2-mediated regulation, we investigated the expression of *HeT-A* transgenes in a euchromatic context. Previously, we demonstrated that the expression of HeT-A-lacZ construct in which *HeT-A* promoter drives the expression of *lacZ* reporter is silenced by the piRNA pathway [46, 47]. To evaluate the role of Ars2 in the regulation of *HeT-A* promoter activity, we analyzed the expression of HeT-A-lacZ upon *Ars2*_GLKD. The expression of this construct was significantly upregulated after knockdown of *spnE*, a piRNA pathway component, as expected, and was only weakly activated in *Ars2*_GLKD (Supplementary Fig. S13). This result demonstrated that activity of the transgenic *HeT-A* promoter is not strongly affected by Ars2.

In our previous experiments, we used the transgenic strain HeTA-HA-FLAG, which carried euchromatic full-length *HeT-A* fused with the hemagglutinin (HA) and FLAG tags and was driven by the UAST promoter (Fig. 5A), and we observed an accumulation of both endogenous and transgenic *HeT-A* transcripts after piRNA pathway depletion [48, 49]. Here, we genetically combined the HeTA-HA-FLAG transgene and *Ars2*_sh1 to examine transgenic *HeT-A* expression. RT-qPCR using transgene-specific primers demonstrated strong upregulation of transgenic *HeT-A* after the knockdowns of both *Ars2* and *spnE*, the piRNA pathway gene (Fig. 5B). Using the HeTA-HA-FLAG-ms2 transgenic strain carrying euchromatic full-length *HeT-A* tagged with MS2 bacteriophage hairpins, we revealed nuclear foci of *HeT-A_ms2* transcripts that did

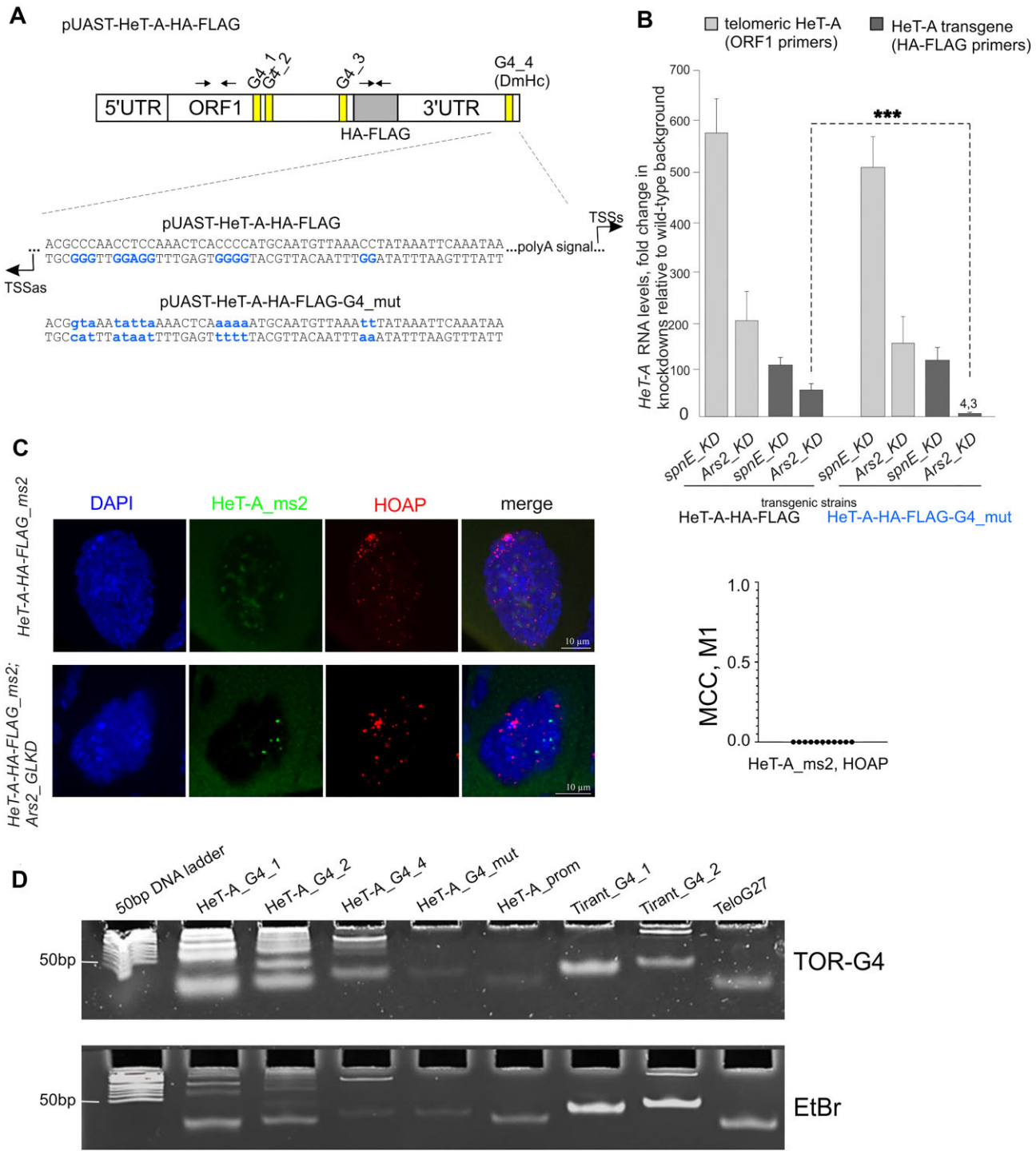


Figure 5. A G-quadruplex motif in the 3′UTR of telomeric retrotransposon *HeT-A* is required for its Ars2-mediated regulation. (**A**) The scheme of the pUAST-attB-HeTA-HA-FLAG construct is shown. Transgene-specific primers corresponding to the tag sequences and ORF1-specific primers are indicated by the arrows. The G4 motifs are indicated by yellow. pUAST-attB-HeTA-HA-FLAG-G4mut was created on the basis of pUAST-attB-HeTA-HA-FLAG by cloning of the mutated G4_4 (DmHc) motif into the *HeT-A* 3′UTR. The sequence of G4_4 motif in the *HeT-A* 3′UTR is shown. The original and mutated nucleotides are colored blue. Sense and antisense TSSs are indicated by the broken arrows. (**B**) RT-qPCR analysis of the relative RNA levels of endogenous telomeric *HeT-A* and transgenic HeT-A-HA-FLAG (to the left) or HeT-A-HA-FLAG-G4mut (to the right) was performed on total ovarian RNA from indicated genotypes (normalized to *rp49*). $n = 4$; error bars indicate SEM. Asterisks indicate statistically significant differences (***$P < .001$, unpaired *t*-test). (**C**) *HeT-A_ms2* RNA FISH combined with HOAP immunostaining was performed on the ovaries of HeT-A-HA-FLAG_ms2 transgenic strain in wild-type background and in *Ars2*_GLKD. The representative images of the nurse cell nuclei are shown. The graph depicts MCC, with M1 representing the colocalization of *HeT-A_ms2* RNA FISH signals with telomeric HOAP signals. $n = 11$ (n is the number of nurse cell nuclei analyzed). DNA is stained with DAPI. (**D**) The detection of G-quadruplexes in the oligonucleotides was performed by polyacrylamide gel electrophoresis (PAGE). The representative images of the gels stained with TOR-G4 (upper panel) or with ethidium bromide (lower panel) are shown. Bright bands are revealed by TOR-G4 staining for HeTA G4_1, HeTA G4_2, HeTA G4_4, Tirant G4_1, Tirant G4_2, and TeloG oligonucleotides that comprise G4 motifs.

not overlap with telomeric protein HOAP staining in the nurse cells in *Ars2*_GLKD (Fig. 5C). MCC analysis revealed no colocalization of *HeT-A_ms2* RNA to telomeres (i.e. HOAP staining). Transgenic transcripts appear to be retained at the sites of their transcription similar to endogenous *HeT-A* RNAs. The finding that transgenic full-length *HeT-A* driven by UAST promoter behaves similarly to endogenous *HeT-A* implies that Ars2-mediated regulation of *HeT-A* expression is independent of genomic context and does not require the *HeT-A* promoter.

Most likely, Ars2 recognizes sequence specificity or structural features of *HeT-A* transcripts. It was previously found that *HeT-A* comprises sequences capable of forming G-quadruplex DNA [52, 53]. In most organisms, telomeres are composed of short G-rich repeats that form double-stranded regions with a single-strand overhang of the G-rich sequences. G-quadruplexes in the telomeric single-strand overhang are thought to protect telomeres from degradation and regulate telomere end accessibility. In contrast, G-quadruplex formation in the double-strand telomere region during transcription and replication can cause telomere damage and instability [81, 82]. G-quadruplex motifs have been discovered in *HeT-A*, *TART*, and *TAHRE Drosophila* telomeric retroelements [53], which differ significantly from the short telomeric repeats generated by the telomerase, demonstrating that G-rich motifs at telomeres are preserved regardless of the structure of telomeric repeats.

Four G-quadruplex motifs were identified bioinformatically in the *HeT-A* coding region within the *gag* gene and in the 3′UTR [53] (Fig. 5A). The DmHc region, forming G-quadruplex and corresponding to the HeT-A_G4_4 motif [53], was independently identified by electrophoretic mobility assay [52] on the template strand between the sense and antisense TSSs in the *HeT-A* 3′UTR [83] (Fig. 5A). This region is strongly enriched by the H3K4me2 histone modification (Supplementary Fig. S14), which is typical for the open chromatin regions. G4 DNA is implicated in maintaining the open structure of promoters and can cause arrest of RNA polymerase II progression when located on the template strand [84–86]. It has been proposed that Ars2 is capable of identifying transcripts in regions where G-quadruplexes alter DNA structure and transcriptional activity. We mutated the HeT-A_G4_4 (DmHc) fragment (Fig. 5A) and generated transgenic flies bearing the construct pUASTattB-HeTA-HA-FLAG-G4mut at the same genomic site as the original construct HeTA-HA-FLAG. Then, we examined the expression of this transgene using transgene-specific primers in *Ars2*_GLKD ovaries. RT-qPCR revealed strong upregulation of endogenous *HeT-A*, whereas transgenic HeTA-HA-FLAG-G4mut was less affected in contrast to original transgene HeTA-HA-FLAG (Fig. 5B). At the same time, *spnE* knockdown still caused strong upregulation of transgenic HeTA-HA-FLAG-G4mut expression (Fig. 5B). This result suggests that Ars2-mediated regulation of *HeT-A* expression depends on G-quadruplex DNA structures. Other G4 motifs found in the *HeT-A* coding region may potentially have an impact on transgene expression, as the DmHc mutation in the *HeT-A* 3′UTR diminished but did not completely eliminate the Ars2-mediated effect on transgene expression.

Computational analysis predicted G4-forming sequences in a subset of TEs, including the *Jockey* superfamily of non-LTR retrotransposon retroelements, LTR retrotransposons *Gypsy* and *Pao*, and *Helitron* DNA transposon [53]. Non-telomeric TEs upregulated in *Ars2*_GLKD were mostly represented by those TE families that are enriched by the sequences capable to form G4 structures (Table 1). The number of predicted G4 within TE families is strongly positively associated with the number of TE copies derepressed in the *Ars2*_GLKD ovaries (Pearson's $r = 0.867$, P-value $= 1.531e{-}06$, adj.$R^2 = 0.737$). It is noteworthy that the number of G4 motifs is not proportional to the TE ratio in the genome. Moreover, it was reported that G4 motifs were predominantly found at the template strand of TEs [53]. This suggests that Ars2 primarily controls the expression of TEs that have the ability to form G-quadruplexes, which likely contribute to TE transcript recognition.

Table 1. TEs enriched by the G-quadruplex motifs are overexpressed in the Ars2 knockdown

TE family	The Number of derepressed TE copies	Proportion of TEs from a specific family relative to all TEs	The Number of predicted G4 [53]	TE class
Gypsy	63	0.381	623	LTR
telomeric	37	0.023	141	LINE
Pao	22	0.100	213	LTR
Jockey	16	0.094	350	LINE
Helitron	15	0.155	110	RC
CR1	14	0.041	77	LINE
Copia	6	0.028	2	LTR
R1	3	0.038	184	LINE
TcMar-Tc1	3	0.029	3	DNA
P	2	0.063	4	DNA
CMC-Transib	1	0.016	11	DNA
I	1	0.009	35	LINE
LOA	1	0.011	51	LINE
hAT-hobo	0	0.005	0	DNA
MULE-NOF	0	0.001	0	DNA
PiggyBac	0	0.001	0	DNA
R2	0	0.001	5	LINE
TcMar-Mariner	0	0.000	0	DNA
TcMar-Pogo	0	0.002	0	DNA

To validate the presence of G4 in the retrotransposons regulated by Ars2, we utilized a thiazole orange derivative that demonstrated binding selectivity for G-quadruplex DNA in the gel electrophoresis and *in vivo* studies [87, 88]. To this end, we synthesized the oligonucleotides comprising G4 motifs from *HeT-A* and *Tirant* LTR retrotransposon, which are upregulated in *Ars2*_GLKD, along with TeloG, mimicking mammalian telomere sequences. These oligonucleotides, as well as the HeT-A_G4mut and HeT-A_prom oligonucleotides lacking G4, were separated by PAGE and stained with either TOR-G4 dye, which binds selectively to structured DNA, or ethidium bromide, which more efficiently stains dsDNA. HeT-A_G4_1, HeT-A_G4_2, HeT-A_G4_4, Tirant_G4_1, Tirant_G4_2, and TeloG oligonucleotides were consistently stained more brightly by TOR-G4 compared to HeT-A_G4mut and HeT-A_prom oligonucleotides (Fig. 5D). HeT-A G4_1, HeT-A_G4_2, and HeT-A_G4_4 were highly structured, as evidenced by TOR-G4 staining, and these structured DNA forms were less visible on an ethidium bromide-stained gel. These findings suggest the presence of G-quadruplex DNA structures in the *HeT-A* and *Tirant* retrotransposons, both regulated by Ars2.

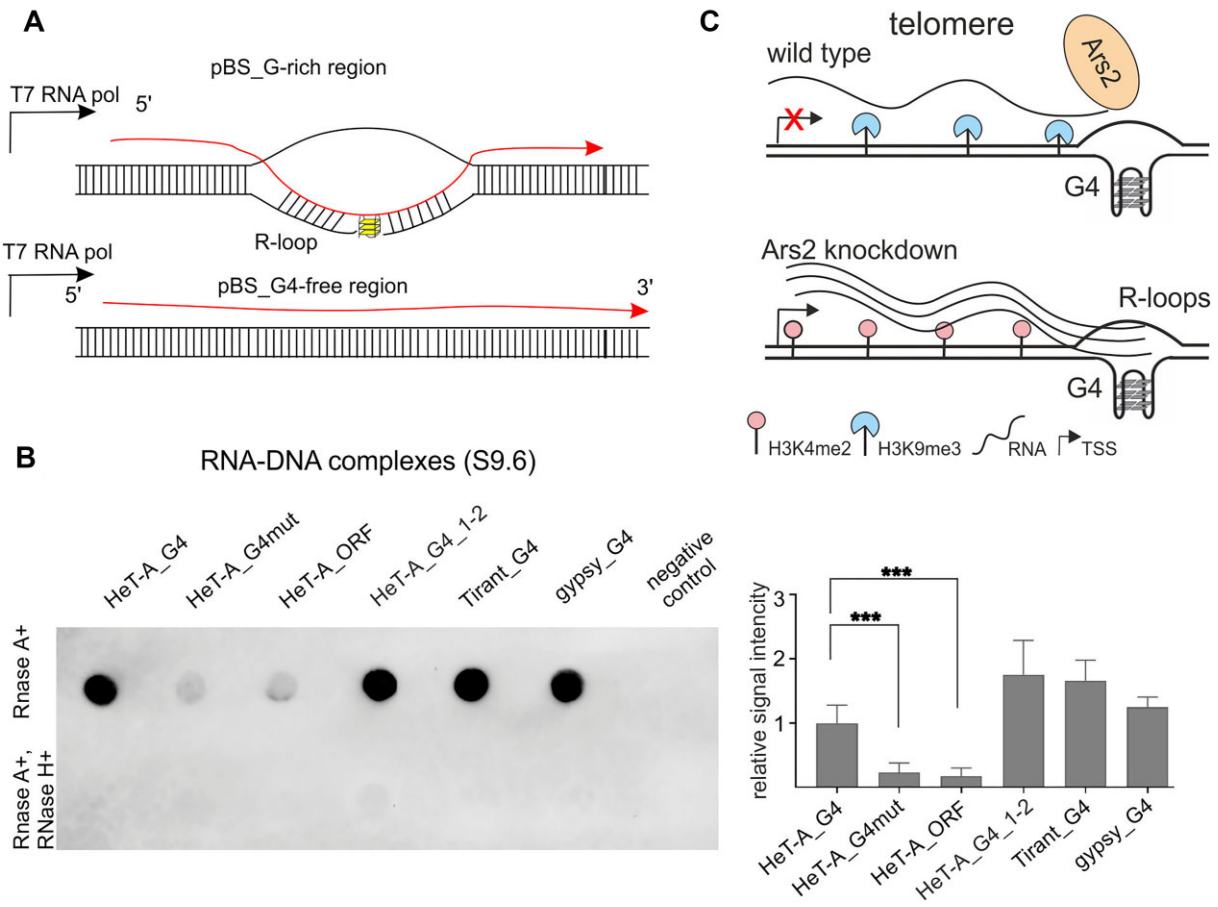


Figure 6. The G-rich *HeT-A* 3′UTR region is prone to formation of R-loops. (**A**) Scheme showing the strategy for *in vitro* detection of R-loops. Two different types of pBS-based constructs with TE fragments were used in T7 RNA polymerase *in vitro* transcription: those with G4 (yellow) on the template strand and those without G4 motifs. It was proposed that the G4 structure, which maintains the open DNA state, would encourage the generation of R-loops. RNA is shown by a red arrow. (**B**) Immunodetection of R-loops using S9.6 antibody by dot-blot analysis of purified DNA–RNA complexes treated with RNaseA or RNaseA and RNAseH. Negative control is TE buffer. Representative image is shown. Quantification of R-loop staining measured from four independent dot-blot analyses is shown on the right. ***$P < .001$, unpaired two-tailed *t*-test. (**C**) Schematic model depicting chromatin and transcription dynamics of *HeT-A* telomeric retrotransposon after Ars2 depletion. In germ cells, Ars2 identifies telomeric RNAs that form R-loops in G-rich regions, facilitating heterochromatinization and transcriptional silencing. A deficiency in Ars2 leads to the accumulation of nascent telomeric RNAs and the development of R-loops.

Retrotransposon regions with G4 motifs are prone to producing R-loops (RNA–DNA hybrids) *in vitro*

We proposed that the 3′UTR of *HeT-A* is prone to the R-loop formation due to G4 structure that maintains the open DNA state. If this hypothesis is correct, mutation of the *HeT-A* G-rich motif should resolve R-loops at this region. We tested this hypothesis by an *in vitro* experiment. Fragments of *HeT-A* 3′UTR containing either the original G4_4 region or the mutated one (Fig. 5A) were cloned into the pBlueScript SK(−) and *in vitro* transcribed by T7 RNA polymerase (Fig. 6A). Free RNA was removed by RNaseA. Plasmid concentrations and *in vitro* transcription efficiency were found to be similar by measuring template DNA concentrations and using agarose gel electrophoresis (Supplementary Fig. S15). Purified RNA–DNA complexes were used for the dot-blot immunodetection of R-loops. R-loop staining was significantly lower in *HeT-A* G4-mutated construct as compared to the original (Fig. 6B). Next, we tested the generation of R-loops in the *HeT-A* ORF region containing G4_1 and G4_2 motifs (HeT-A_G4_1–2 construct), as well as in the G-rich regions of the *Tirant* (Tirant_G4) and *gypsy12* (gypsy_G4) retrotransposons, both of which are targets of Ars2 (Supplementary Fig. S16). The construct containing *HeT-A* ORF region lacking the G-rich motifs was also evaluated (HeT-A_ORF). Dot-blot analysis of R-loops revealed that HeT-A_G4_1–2, Tirant_G4, and gypsy_G4 constructs but not HeT-A_ORF were positive for staining of the R-loop. Treatment with RNase H, which removes RNA in RNA–DNA hybrids, was used as a control for specificity for the R-loop staining. These results demonstrated that the G-rich regions of *HeT-A*, *gypsy12*, and *Tirant* retrotransposons promoted the formation of R-loops *in vitro*. Ars2 prevents R-loop development at telomeres and retrotransposons *in vivo* most likely by eliminating excessive nascent RNAs and chromatin compaction (Fig. 6C).

Discussion

We show that Ars2 mediates transcriptional silencing of telomeric retroelements and a subset of non-telomeric TEs in the *Drosophila* germline. Ars2-mediated TE silencing is a novel anti-transposon defense mechanism acting alongside the piRNA pathway in the germline. Depletion of *Drosophila*

Ars2 caused telomeric accumulation of chromatin factors associated with active transcription, namely enrichment of H3K4me2 and H3K36me3 histone modifications. Our findings imply that Ars2-mediated silencing may be triggered by changes in DNA structure and transcriptional activity at G-quadruplex-forming regions. Here, we showed that Ars2 inhibited the formation of R-loops in telomeres, strengthening the argument for Ars2 function in preserving functional telomeres and genome stability. Because Ars2 has no effect on piRNA production, yet regulates a subset of TEs, we believe that the piRNA pathway and Ars2 work in tandem to silence TEs. However, it cannot be excluded that both pathways share common chromatin-modifying partners.

RNA-dependent mechanisms play an important role in the assembly of inactive chromatin at specific genomic loci. Physical and functional interaction between nuclear RNA decay and chromatin remodeling pathways is required for the elimination of prematurely terminated TE transcripts in mouse embryonic stem cells [89]. Meiotic mRNAs in *S. pombe* contain specific hexanucleotide sequences recognized by RNA degradation machinery [90], which in turn promotes heterochromatin assembly at these loci [6]. Differential splicing of cryptic introns is recognized by RNA processing machinery that triggers heterochromatin assembly at retrotransposons and promotes telomerase RNA maturation in yeast [5]. However, the mechanisms for recognition of nascent RNA and the triggering of chromatin changes at specific genomic loci in higher eukaryotes are far from understood. It is believed that nuclear RNA biogenesis is directed by both RNA-specific signals and context-dependent factors, and in its turn, nascent RNA features could also affect the recruitment of regulatory proteins to chromatin [91]. A well-known example is the recruitment of chromatin remodeling or DNA methylation factors to TEs mediated by the complementarity of small piRNAs to the nascent TE transcripts [92, 93]. Our study has revealed a new example of RNA-dependent chromatin modulation: Ars2, a component of the nuclear RNA surveillance, has been shown to mediate the repression of TE transcription through chromatin modifications.

Although Ars2 was found in a variety of RNA processing complexes, its biochemical activity remains unclear. Ars2 is required for the regulation of a special subset of transcripts, including promoter and enhancer transcripts, some pre-miRNA, histone mRNA, and telomeric RNAs, suggesting that Ars2 uses very specific targeting mechanisms. A secondary stem-loop structure at the 3′ end of transcripts was proposed to be required for Ars2 binding to its RNA targets [33]. The list of Ars2 targets suggests that CG-rich sequences or G-quadruplex structures could be responsible for Ars2-mediated transcriptional control. Indeed, promoter/enhancer regions, transcription termination sites, or telomeric repeats are G-rich sequences prone to the formation of G-quadruplexes [81, 94, 95]. Within transcribed regions, G-rich sequences of particular orientation and composition can cause blockage and stalling of RNA polymerase II activity [86]. The 3′UTR of the *D. melanogaster* telomeric element *HeT-A* contains sequences capable of forming G-quadruplexes on the template DNA strand [52]. G4-forming motifs were identified within *HeT-A*, *TART*, and *TAHRE Drosophila* telomeric elements [53], suggesting a similar mechanism for their Ars2-mediated repression. Interestingly, telomeric DNA in species that use non-telomerase telomere maintenance mechanisms also reveals a G/C strand bias, with G4 DNA enriched on the noncoding strand [53]. It is believed that formation of G-quadruplexes at telomeric 3′ ssDNA overhangs may have a telomere-capping function. Moreover, emerging evidence implicates G4 DNA and proteins interacting with this structure in the regulation of R-loop formation and telomere stability [81]. G4-forming sequences are a hallmark of young lineages of LINE1 retrotransposons contributing to their mobility [96]. Viral G-quadruplex structures and G4-binding proteins are also implicated in the regulation of viral reverse transcription and replication [97]. Given the role of G4 structures in the transposon- and virus-associated pathogenesis, the recognition and regulation of their state is essential for genome stability.

Formation of stable G4 requires open chromatin, as found at active promoters and enhancers in human and mouse genomes [94, 95]. It was suggested that G4 in the 3′UTR may be involved in transcriptional termination [98]. R-loops associated with transcriptional termination at G-rich terminator elements can on their own induce assembly of the repressive chromatin [99]. The G-rich motif found in the 3′UTR of *HeT-A* is located on the template strand between sense and antisense TSSs of the promoter region [46, 83]. In *Drosophila* telomeres, telomeric retrotransposons are organized in head-to-tail tandem repeats so that the upstream copy can drive transcription of the downstream one. Localization of a G-quadruplex 30 bp upstream of the *HeT-A* polyadenylation signal suggests a scenario in which G4 could cause premature termination of *HeT-A* transcription and subsequent activation of the Ars2-mediated silencing. The expression of *HeT-A* may be similarly impacted by G4 DNA structures present in the *HeT-A* coding region. Overall, this mechanism adds another layer of regulation to the levels of telomeric RNAs, which are critical for cell fate and development. Computational analysis revealed G4 motifs at the template strand of many TEs [53]; noteworthy, those TEs that contain more G4 motifs are derepressed after the Ars2 knockdown. Given the Ars2 activity in the RNA quality control, it seems unlikely that Ars2 is directly involved in the recognition of G-rich sequences. Instead, Ars2 may induce chromatin modifications by targeting RNAs that have been prematurely terminated or are engaged in R-loop structures. Ars2 participates in several pathways implicated in the activation of RNA decay and RNA degradation [19, 27]. Ars2 most likely stimulates both degradation of *HeT-A* transcripts and transcriptional silencing of the locus. FLASH is involved in histone pre-mRNA 3′-end processing in *Drosophila*; however, Ars2 is dispensable for this regulation in contrast to mammalian cells [73]. We found that *HeT-A* expression is regulated by both Ars2 and FLASH, suggesting the cooperation of these factors in telomere protection. Future research investigating Ars2 interactors in the germline will shed light on the RNA-guided pathway for the telomeric epigenetic silencing.

Intriguingly, Ars2 regulates telomeric and rDNA-specific retrotransposons, both of which are considered symbiotic and their integrations are beneficial for the host genome [100]. R1/R2 insertions into 28S rRNA genes are proposed to generate dsDNA breaks that initiate homologous recombination resulting in the amplification of rDNA copies [101]. R1 retroelements have sequence-specific integration patterns and are found in 28S rDNA as well as in non-rDNA regions [102, 103]. R1/R2 retrotransposons are transcribed by RNA polymerase I as a part of read-through transcripts from the 28S gene [104, 105]. Since Ars2 participates in the biogenesis of the RNA polymerase II-transcribed RNA, it is un-

likely that it regulates metabolism of RNA PolI-derived 28S rRNA–R1 fusion transcripts. Most likely, Ars2 targets R1 transcripts derived from R1 insertions into the non-rDNA regions. Silkworm R1 transcripts are not polyadenylated; addition of poly(A) at the 3′ end of *Bombix mori* R1 strongly suppresses its retrotransposition activity, which is mediated by the base pairing between R1 3′UTR and 28S rDNA [103]. Given that Ars2 promotes 3′-end processing of non-adenylated histone mRNAs, snRNAs, and snoRNAs [29, 33], it is reasonable to suggest that Ars2 also promotes 3′-end formation of R1 transcripts. Ars2-mediated stabilization of R1 RNA described here is a novel mechanism that can modulate rDNA-specific retrotransposition activity of R1.

Acknowledgements

We thank Pavel A. Komarov for useful advice on data processing and critical comments on the manuscript. We thank Jan Larsson for the anti-POF antibodies. We thank A.S. Seleznev and A.A. Schukina for their contributions to G4 detection and RT-PCR analysis, respectively. We thank BDSC and VDRC, and FlyORF for fly strains. We thank the User Centre of the Institute of Developmental Biology RAS supported from the IDB RAS Government basic research program 0088-2024-0010 for the opportunity to use microscope and PCR equipment. Fly transgenesis was performed using the facilities of the Center for Precision Genome Editing and Genetic Technologies for Biomedicine, IGB RAS.

Author contributions: Valeriya Morgunova (Investigation [lead], Methodology [equal], Visualization [lead]), Anastasiya Kobelyatskaya (Data curation [lead], Formal analysis [lead], Validation [equal], Visualization [equal]), Maksim Erokhin (Investigation [equal], Methodology [equal], Resources [equal]), Olesya Sokolova (Investigation [equal]), Tatyana V. Sizova (Investigation [equal]), Dmitry A. Kwon (Investigation [equal]), and Alla Kalmykova (Conceptualization [lead], Supervision [lead], Visualization [equal], Writing—original draft [lead], Writing—review & editing [lead]).

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

None declared.

Funding

This work was supported by the Russian Science Foundation (24-14-00043 to A.K.). Funding to pay the Open Access publication charges for this article was provided by Russian Science Foundation (24-14-00043).

Data availability

ChIPseq data are deposited in the Gene Expression Omnibus (GEO) database under the accession number GSE266038 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE266038). ChromRNAseq data are deposited in the GEO database under the accession number GSE266036 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE266036). Small RNAseq data are deposited in the GEO database under the accession number GSE266037 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE266037). All other data are available in the article and in its online supplementary material.

References

1. Czech B, Munafo M, Ciabrelli F *et al.* piRNA-guided genome defense: from biogenesis to silencing. *Annu Rev Genet* 2018;**52**:131–57. https://doi.org/10.1146/annurev-genet-120417-031441
2. Ozata DM, Gainetdinov I, Zoch A *et al.* PIWI-interacting RNAs: small RNAs with big functions. *Nat Rev Genet* 2019;**20**:89–108. https://doi.org/10.1038/s41576-018-0073-3
3. Savitsky M, Kwon D, Georgiev P *et al.* Telomere elongation is under the control of the RNAi-based mechanism in the *Drosophila* germline. *Genes Dev* 2006;**20**:345–54. https://doi.org/10.1101/gad.370206
4. Kalmykova AI, Sokolova OA. Retrotransposons and telomeres. *Biochemistry (Mosc)* 2023;**88**:1739–53. https://doi.org/10.1134/S0006297923110068
5. Lee NN, Chalamcharla VR, Reyes-Turcu F *et al.* Mtr4-like protein coordinates nuclear RNA processing for heterochromatin assembly and for telomere maintenance. *Cell* 2013;**155**:1061–74. https://doi.org/10.1016/j.cell.2013.10.027
6. Zofall M, Yamanaka S, Reyes-Turcu FE *et al.* RNA elimination machinery targeting meiotic mRNAs promotes facultative heterochromatin formation. *Science* 2012;**335**:96–100. https://doi.org/10.1126/science.1211651
7. Yamanaka S, Mehta S, Reyes-Turcu FE *et al.* RNAi triggered by specialized machinery silences developmental genes and retrotransposons. *Nature* 2013;**493**:557–60. https://doi.org/10.1038/nature11716
8. Sugiyama T, Thillainadesan G, Chalamcharla VR *et al.* Enhancer of rudimentary cooperates with conserved RNA-processing factors to promote meiotic mRNA decay and facultative heterochromatin assembly. *Mol Cell* 2016;**61**:747–59. https://doi.org/10.1016/j.molcel.2016.01.029
9. Buhler M, Haas W, Gygi SP *et al.* RNAi-dependent and -independent RNA turnover mechanisms contribute to heterochromatic gene silencing. *Cell* 2007;**129**:707–21. https://doi.org/10.1016/j.cell.2007.03.038
10. Buhler M, Spies N, Bartel DP *et al.* TRAMP-mediated RNA surveillance prevents spurious entry of RNAs into the *Schizosaccharomyces pombe* siRNA pathway. *Nat Struct Mol Biol* 2008;**15**:1015–23. https://doi.org/10.1038/nsmb.1481
11. Le Thomas A, Rogers AK, Webster A *et al.* Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev* 2013;**27**:390–9. https://doi.org/10.1101/gad.209841.112
12. Rozhkov NV, Hammell M, Hannon GJ. Multiple roles for Piwi in silencing *Drosophila* transposons. *Genes Dev* 2013;**27**:400–12. https://doi.org/10.1101/gad.209767.112
13. Sienski G, Donertas D, Brennecke J. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* 2012;**151**:964–80. https://doi.org/10.1016/j.cell.2012.10.040
14. Brennecke J, Aravin AA, Stark A *et al.* Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 2007;**128**:1089–103. https://doi.org/10.1016/j.cell.2007.01.043
15. Gunawardane LS, Saito K, Nishida KM *et al.* A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in *Drosophila*. *Science* 2007;**315**:1587–90. https://doi.org/10.1126/science.1140494
16. Han BW, Wang W, Li C *et al.* Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA

production. *Science* 2015;**348**:817–21. https://doi.org/10.1126/science.aaa1264
17. Mohn F, Handler D, Brennecke J. Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. *Science* 2015;**348**:812–7. https://doi.org/10.1126/science.aaa1039
18. Kordyukova M, Sokolova O, Morgunova V *et al.* Nuclear Ccr4-Not mediates the degradation of telomeric and transposon transcripts at chromatin in the *Drosophila* germline. *Nucleic Acids Res* 2020;**48**:141–56.
19. Lykke-Andersen S, Rouviere JO, Jensen TH. ARS2/SRRT: at the nexus of RNA polymerase II transcription, transcript maturation and quality control. *Biochem Soc Trans* 2021;**49**:1325–36. https://doi.org/10.1042/BST20201008
20. Scheibe M, Arnoult N, Kappei D *et al.* Quantitative interaction screen of telomeric repeat-containing RNA reveals novel TERRA regulators. *Genome Res* 2013;**23**:2149–57. https://doi.org/10.1101/gr.151878.112
21. Pakhomova T, Moshareva M, Vasilkova D *et al.* Role of RNA biogenesis factors in the processing and transport of human telomerase RNA. *Biomedicines* 2022;**10**:1275. https://doi.org/10.3390/biomedicines10061275
22. Morgunova V, Akulenko N, Radion E *et al.* Telomeric repeat silencing in germ cells is essential for early development in *Drosophila*. *Nucleic Acids Res* 2015;**43**:8762–73. https://doi.org/10.1093/nar/gkv775
23. Casacuberta E. *Drosophila*: retrotransposons making up telomeres. *Viruses* 2017;**9**:192. https://doi.org/10.3390/v9070192
24. Radion E, Morgunova V, Ryazansky S *et al.* Key role of piRNAs in telomeric chromatin maintenance and telomere nuclear positioning in *Drosophila* germline. *Epigenetics Chromatin* 2018;**11**:40. https://doi.org/10.1186/s13072-018-0210-4
25. Morgunova V, Kordyukova M, Mikhaleva EA *et al.* Loss of telomere silencing is accompanied by dysfunction of Polo kinase and centrosomes during *Drosophila* oogenesis and early development. *PLoS One* 2021;**16**:e0258156. https://doi.org/10.1371/journal.pone.0258156
26. Schulze WM, Stein F, Rettel M *et al.* Structural analysis of human ARS2 as a platform for co-transcriptional RNA sorting. *Nat Commun* 2018;**9**:1701. https://doi.org/10.1038/s41467-018-04142-7
27. Andersen PR, Domanski M, Kristiansen MS *et al.* The human cap-binding complex is functionally connected to the nuclear RNA exosome. *Nat Struct Mol Biol* 2013;**20**:1367–76. https://doi.org/10.1038/nsmb.2703
28. Porrua O, Libri D. Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat Rev Mol Cell Biol* 2015;**16**:190–202. https://doi.org/10.1038/nrm3943
29. Hallais M, Pontvianne F, Andersen PR *et al.* CBC-ARS2 stimulates 3′-end maturation of multiple RNA families and favors cap-proximal processing. *Nat Struct Mol Biol* 2013;**20**:1358–66. https://doi.org/10.1038/nsmb.2720
30. Fan J, Kuai B, Wu G *et al.* Exosome cofactor hMTR4 competes with export adaptor ALYREF to ensure balanced nuclear RNA pools for degradation and export. *EMBO J* 2017;**36**:2870–86. https://doi.org/10.15252/embj.201696139
31. Bajczyk M, Lange H, Bielewicz D *et al.* SERRATE interacts with the nuclear exosome targeting (NEXT) complex to degrade primary miRNA precursors in *Arabidopsis*. *Nucleic Acids Res* 2020;**48**:6839–54. https://doi.org/10.1093/nar/gkaa373
32. Iasillo C, Schmid M, Yahia Y *et al.* ARS2 is a general suppressor of pervasive transcription. *Nucleic Acids Res* 2017;**45**:10229–41. https://doi.org/10.1093/nar/gkx647
33. Gruber JJ, Olejniczak SH, Yong J *et al.* Ars2 promotes proper replication-dependent histone mRNA 3′ end formation. *Mol Cell* 2012;**45**:87–98. https://doi.org/10.1016/j.molcel.2011.12.020
34. Gruber JJ, Zatechka DS, Sabin LR *et al.* Ars2 links the nuclear cap-binding complex to RNA interference and cell proliferation. *Cell* 2009;**138**:328–39. https://doi.org/10.1016/j.cell.2009.04.046
35. Sabin LR, Zhou R, Gruber JJ *et al.* Ars2 regulates both miRNA- and siRNA-dependent silencing and suppresses RNA virus infection in *Drosophila*. *Cell* 2009;**138**:340–51. https://doi.org/10.1016/j.cell.2009.04.045
36. Lobbes D, Rallapalli G, Schmidt DD *et al.* SERRATE: a new player on the plant microRNA scene. *EMBO Rep* 2006;**7**:1052–8. https://doi.org/10.1038/sj.embor.7400806
37. Yang L, Liu Z, Lu F *et al.* SERRATE is a novel nuclear regulator in primary microRNA processing in *Arabidopsis*. *Plant J* 2006;**47**:841–50. https://doi.org/10.1111/j.1365-313X.2006.02835.x
38. Egan ED, Braun CR, Gygi SP *et al.* Post-transcriptional regulation of meiotic genes by a nuclear RNA silencing complex. *RNA* 2014;**20**:867–81. https://doi.org/10.1261/rna.044479.114
39. Thillainadesan G, Xiao H, Holla S *et al.* Conserved protein Pir2^{ARS2} mediates gene repression through cryptic introns in lncRNAs. *Nat Commun* 2020;**11**:2412. https://doi.org/10.1038/s41467-020-16280-y
40. Ma Z, Castillo-Gonzalez C, Wang Z *et al.* *Arabidopsis* serrate coordinates histone methyltransferases ATXR5/6 and RNA processing factor RDR6 to regulate transposon expression. *Dev Cell* 2018;**45**:769–84. https://doi.org/10.1016/j.devcel.2018.05.023
41. Yu Y, Andreu-Agullo C, Liu BF *et al.* Regulation of embryonic and adult neurogenesis by Ars2. *Development* 2020;**147**:dev180018. https://doi.org/10.1242/dev.180018
42. Andreu-Agullo C, Maurin T, Thompson CB *et al.* Ars2 maintains neural stem-cell identity through direct transcriptional activation of Sox2. *Nature* 2012;**481**:195–8. https://doi.org/10.1038/nature10712
43. Cacchione S, Cenci G, Raffa GD. Silence at the end: how *Drosophila* regulates expression and transposition of telomeric retroelements. *J Mol Biol* 2020;**432**:4305–21. https://doi.org/10.1016/j.jmb.2020.06.004
44. Khurana JS, Xu J, Weng Z *et al.* Distinct functions for the *Drosophila* piRNA pathway in genome maintenance and telomere protection. *PLoS Genet* 2010;**6**:e1001246. https://doi.org/10.1371/journal.pgen.1001246
45. Ye J, Eickbush TH. Chromatin structure and transcription of the R1- and R2-inserted rRNA genes of *Drosophilamelanogaster*. *Mol Cell Biol* 2006;**26**:8781–90. https://doi.org/10.1128/MCB.01409-06
46. Radion E, Ryazansky S, Akulenko N *et al.* Telomeric retrotransposon HeT-A contains a bidirectional promoter that initiates divergent transcription of piRNA precursors in *Drosophila* germline. *J Mol Biol* 2017;**429**:3280–9. https://doi.org/10.1016/j.jmb.2016.12.002
47. Shpiz S, Olovnikov I, Sergeeva A *et al.* Mechanism of the piRNA-mediated silencing of *Drosophila* telomeric retrotransposons. *Nucleic Acids Res* 2011;**39**:8703–11. https://doi.org/10.1093/nar/gkr552
48. Kordyukova M, Morgunova V, Olovnikov I *et al.* Subcellular localization and Egl-mediated transport of telomeric retrotransposon HeT-A ribonucleoprotein particles in the *Drosophila* germline and early embryogenesis. *PLoS One* 2018;**13**:e0201787. https://doi.org/10.1371/journal.pone.0201787
49. Olovnikov IA, Morgunova VV, Mironova AA *et al.* Interaction of telomeric retroelement HeT-A transcripts and their protein product Gag in early embryogenesis of *Drosophila*. *Biochemistry (Mosc)* 2016;**81**:1023–30. https://doi.org/10.1134/S000629791609011X
50. DeLuca SZ, Spradling AC. Efficient expression of genes in the *Drosophila* germline using a UAS promoter free of interference by Hsp70 piRNAs. *Genetics* 2018;**209**:381–7. https://doi.org/10.1534/genetics.118.300874

51. Grentzinger T, Armenise C, Brun C *et al.* piRNA-mediated transgenerational inheritance of an acquired trait. *Genome Res* 2012;**22**:1877–88. https://doi.org/10.1101/gr.136614.111
52. Abad JP, Villasante A. The 3′ non-coding region of the *Drosophila melanogaster* HeT-A telomeric retrotransposon contains sequences with propensity to form G-quadruplex DNA. *FEBS Lett* 1999;**453**:59–62. https://doi.org/10.1016/S0014-5793(99)00695-X
53. Jedlicka P, Tokan V, Kejnovska I *et al.* Telomeric retrotransposons show propensity to form G-quadruplexes in various eukaryotic species. *Mobile DNA* 2023;**14**:3. https://doi.org/10.1186/s13100-023-00291-9
54. Morgunova V, Sukhova MM, Kalmykova A. Whole-mount RNA FISH combined with immunofluorescence for the analysis of the telomeric ribonucleoproteins in the *Drosophila* germline. *Methods Mol Biol* 2022;**2509**:157–69.
55. Smolka JA, Sanz LA, Hartono SR *et al.* Recognition of RNA by the S9.6 antibody creates pervasive artifacts when imaging RNA:DNA hybrids. *J Cell Biol* 2021;**220**:e202004079. https://doi.org/10.1083/jcb.202004079
56. Larsson J, Svensson MJ, Stenberg P *et al.* Painting of fourth in genus *Drosophila* suggests autosome-specific gene regulation. *Proc Natl Acad Sci USA* 2004;**101**:9728–33. https://doi.org/10.1073/pnas.0400978101
57. Akulenko N, Ryazansky S, Morgunova V *et al.* Transcriptional and chromatin changes accompanying *de novo* formation of transgenic piRNA clusters. *RNA* 2018;**24**:574–84. https://doi.org/10.1261/rna.062851.117
58. Zeng PY, Vakoc CR, Chen ZC *et al.* *In vivo* dual cross-linking for identification of indirect DNA-associated proteins by chromatin immunoprecipitation. *BioTechniques* 2006;**41**:694. https://doi.org/10.2144/000112297
59. Olovnikov I, Ryazansky S, Shpiz S *et al.* *De novo* piRNA cluster formation in the *Drosophila* germ line triggered by transgenes containing a transcribed transposon fragment. *Nucleic Acids Res* 2013;**41**:5757–68. https://doi.org/10.1093/nar/gkt310
60. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;**30**:2114–20. https://doi.org/10.1093/bioinformatics/btu170
61. Dobin A, Davis CA, Schlesinger F *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;**29**:15–21. https://doi.org/10.1093/bioinformatics/bts635
62. Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res* 2019;**47**:e47. https://doi.org/10.1093/nar/gkz114
63. Cunningham F, Allen JE, Allen J *et al.* Ensembl 2022. *Nucleic Acids Res* 2022;**50**:D988–95. https://doi.org/10.1093/nar/gkab1049
64. Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics* 2009;**Chapter 4**:4.10.1–14.
65. Durinck S, Spellman PT, Birney E *et al.* Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* 2009;**4**:1184–91. https://doi.org/10.1038/nprot.2009.97
66. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;**26**:139–40. https://doi.org/10.1093/bioinformatics/btp616
67. Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Res* 2019;**47**:D155–62. https://doi.org/10.1093/nar/gky1141
68. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;**9**:357–9. https://doi.org/10.1038/nmeth.1923
69. Ramírez F, Dündar F, Diehl S *et al.* deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 2014;**42**:W187–91. https://doi.org/10.1093/nar/gku365
70. Cenci G, Siriaco G, Raffa GD *et al.* The *Drosophila* HOAP protein is required for telomere capping. *Nat Cell Biol* 2003;**5**:82–4. https://doi.org/10.1038/ncb902
71. Johansson AM, Stenberg P, Bernhardsson C *et al.* Painting of fourth and chromosome-wide regulation of the 4th chromosome in *Drosophilamelanogaster*. *EMBO J* 2007;**26**:2307–16. https://doi.org/10.1038/sj.emboj.7601604
72. Tzeng TY, Lee CH, Chan LW *et al.* Epigenetic regulation of the *Drosophila* chromosome 4 by the histone H3K9 methyltransferase dSETDB1. *Proc Natl Acad Sci USA* 2007;**104**:12691–6. https://doi.org/10.1073/pnas.0705534104
73. Sabath I, Skrajna A, Yang XC *et al.* 3'-End processing of histone pre-mRNAs in *Drosophila*: U7 snRNP is associated with FLASH and polyadenylation factors. *RNA* 2013;**19**:1726–44. https://doi.org/10.1261/rna.040360.113
74. Skourti-Stathaki K, Proudfoot NJ. A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev* 2014;**28**:1384–96. https://doi.org/10.1101/gad.242990.114
75. Boguslawski SJ, Smith DE, Michalak MA *et al.* Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. *J Immunol Methods* 1986;**89**:123–30. https://doi.org/10.1016/0022-1759(86)90040-2
76. Hartono SR, Malapert A, Legros P *et al.* The affinity of the S9.6 antibody for double-stranded RNAs impacts the accurate mapping of R-loops in fission yeast. *J Mol Biol* 2018;**430**:272–84. https://doi.org/10.1016/j.jmb.2017.12.016
77. Stage DE, Eickbush TH. Origin of nascent lineages and the mechanisms used to prime second-strand DNA synthesis in the R1 and R2 retrotransposons of *Drosophila*. *Genome Biol* 2009;**10**:R49. https://doi.org/10.1186/gb-2009-10-5-r49
78. Mohn F, Sienski G, Handler D *et al.* The rhino–deadlock–cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in *Drosophila*. *Cell* 2014;**157**:1364–79. https://doi.org/10.1016/j.cell.2014.04.031
79. Volpe AM, Horowitz H, Grafer CM *et al.* *Drosophila* rhino encodes a female-specific chromo-domain protein that affects chromosome structure and egg polarity. *Genetics* 2001;**159**:1117–34. https://doi.org/10.1093/genetics/159.3.1117
80. Danilevskaya ON, Arkhipova IR, Traverse KL *et al.* Promoting in tandem: the promoter for telomere transposon HeT-A and implications for the evolution of retroviral LTRs. *Cell* 1997;**88**:647–55. https://doi.org/10.1016/S0092-8674(00)81907-8
81. Bryan TM. G-quadruplexes at telomeres: friend or foe? *Molecules* 2020;**25**:3686. https://doi.org/10.3390/molecules25163686
82. Shiekh S, Kodikara SG, Balci H. Structure, topology, and stability of multiple G-quadruplexes in long telomeric overhangs. *J Mol Biol* 2024;**436**:168205. https://doi.org/10.1016/j.jmb.2023.168205
83. Shpiz S, Kwon D, Rozovsky Y *et al.* rasiRNA pathway controls antisense expression of *Drosophila* telomeric retrotransposons in the nucleus. *Nucleic Acids Res* 2009;**37**:268–78. https://doi.org/10.1093/nar/gkn960
84. Agarwal T, Roy S, Kumar S *et al.* In the sense of transcription regulation by G-quadruplexes: asymmetric effects in sense and antisense strands. *Biochemistry* 2014;**53**:3711–8. https://doi.org/10.1021/bi401451q
85. Du Z, Zhao Y, Li N. Genome-wide analysis reveals regulatory role of G4 DNA in gene transcription. *Genome Res* 2008;**18**:233–41. https://doi.org/10.1101/gr.6905408
86. Belotserkovskii BP, Soo Shin JH, Hanawalt PC. Strong transcription blockage mediated by R-loop formation within a G-rich homopurine-homopyrimidine sequence localized in the vicinity of the promoter. *Nucleic Acids Res* 2017;**45**:6589–99. https://doi.org/10.1093/nar/gkx403

87. Lubitz I, Zikich D, Kotlyar A. Specific high-affinity binding of thiazole orange to triplex and G-quadruplex DNA. *Biochemistry* 2010;**49**:3567–74. https://doi.org/10.1021/bi1000849
88. Long W, Lu Y-J, Zhang K *et al.* Boosting the turn-on fluorescent signaling ability of thiazole orange dyes: the effectiveness of structural modification site and its unusual interaction behavior with nucleic acids. *Dyes Pigm* 2018;**159**:449–56. https://doi.org/10.1016/j.dyepig.2018.07.008
89. Garland W, Muller I, Wu M *et al.* Chromatin modifier HUSH co-operates with RNA decay factor NEXT to restrict transposable element expression. *Mol Cell* 2022;**82**:1691–707. https://doi.org/10.1016/j.molcel.2022.03.004
90. Yamashita A, Shichino Y, Tanaka H *et al.* Hexanucleotide motifs mediate recruitment of the RNA elimination machinery to silent meiotic genes. *Open Biol* 2012;**2**:120014. https://doi.org/10.1098/rsob.120014
91. Skalska L, Begley V, Beltran M *et al.* Nascent RNA antagonizes the interaction of a set of regulatory proteins with chromatin. *Mol Cell* 2021;**81**:2944–59. https://doi.org/10.1016/j.molcel.2021.05.026
92. Aravin AA, Sachidanandam R, Bourc'his D *et al.* A piRNA pathway primed by individual transposons is linked to *de novo* DNA methylation in mice. *Mol Cell* 2008;**31**:785–99. https://doi.org/10.1016/j.molcel.2008.09.003
93. Sienski G, Batki J, Senti KA *et al.* Silencio/CG9754 connects the Piwi–piRNA complex to the cellular heterochromatin machinery. *Genes Dev* 2015;**29**:2258–71. https://doi.org/10.1101/gad.271908.115
94. Hansel-Hertsch R, Beraldi D, Lensing SV *et al.* G-quadruplex structures mark human regulatory chromatin. *Nat Genet* 2016;**48**:1267–72. https://doi.org/10.1038/ng.3662
95. Lyu J, Shao R, Kwong Yung PY *et al.* Genome-wide mapping of G-quadruplex structures with CUT&Tag. *Nucleic Acids Res* 2022;**50**:e13.
96. Sahakyan AB, Murat P, Mayer C *et al.* G-quadruplex structures within the 3′ UTR of LINE-1 elements stimulate retrotransposition. *Nat Struct Mol Biol* 2017;**24**:243–7. https://doi.org/10.1038/nsmb.3367
97. Abiri A, Lavigne M, Rezaei M *et al.* Unlocking G-quadruplexes as antiviral targets. *Pharmacol Rev* 2021;**73**:897–923. https://doi.org/10.1124/pharmrev.120.000230
98. Huppert JL, Bugaut A, Kumari S *et al.* G-quadruplexes: the beginning and end of UTRs. *Nucleic Acids Res* 2008;**36**:6260–8. https://doi.org/10.1093/nar/gkn511
99. Skourti-Stathaki K, Kamieniarz-Gdula K, Proudfoot NJ. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature* 2014;**516**:436–9. https://doi.org/10.1038/nature13787
100. Fujiwara H. Site-specific non-LTR retrotransposons. *Microbiol Spectr* 2015;**3**:MDNA3-0001-2014. https://doi.org/10.1128/microbiolspec.MDNA3-0001-2014
101. Nelson JO, Slicko A, Yamashita YM. The retrotransposon R2 maintains *Drosophila* ribosomal DNA repeats. *Proc Natl Acad Sci USA* 2023;**120**:e2221613120. https://doi.org/10.1073/pnas.2221613120
102. Browne MJ, Read CA, Roiha H *et al.* Site specific insertion of a type I rDNA element into a unique sequence in the *Drosophila melanogaster* genome. *Nucleic Acids Res* 1984;**12**:9111–22. https://doi.org/10.1093/nar/12.23.9111
103. Anzai T, Osanai M, Hamada M *et al.* Functional roles of 3′-terminal structures of template RNA during *in vivo* retrotransposition of non-LTR retrotransposon, R1Bm. *Nucleic Acids Res* 2005;**33**:1993–2002. https://doi.org/10.1093/nar/gki347
104. Fefelova EA, Pleshakova IM, Mikhaleva EA *et al.* Impaired function of rDNA transcription initiation machinery leads to derepression of ribosomal genes with insertions of R2 retrotransposon. *Nucleic Acids Res* 2022;**50**:867–84. https://doi.org/10.1093/nar/gkab1276
105. Raje HS, Lieux ME, DiMario PJ. R1 retrotransposons in the nucleolar organizers of *Drosophila melanogaster* are transcribed by RNA polymerase I upon heat shock. *Transcription* 2018;**9**:273–85. https://doi.org/10.1080/21541264.2018.1506682

Received: June 25, 2024. **Revised:** April 16, 2025. **Editorial Decision:** April 21, 2025. **Accepted:** May 29, 2025