

# Mod(mdg4) variants repress telomeric retrotransposon *HeT-A* by blocking subtelomeric enhancers

Chikara Takeuchi<sup>1</sup>, Moe Yokoshi<sup>2</sup>, Shu Kondo<sup>3</sup>, Aoi Shibuya<sup>1</sup>, Kuniaki Saito<sup>3</sup>, Takashi Fukaya<sup>2,4</sup>, Haruhiko Siomi<sup>1,\*</sup> and Yuka W. Iwasaki<sup>1,5,6,\*</sup>

<sup>1</sup>Department of Molecular Biology, Keio University School of Medicine, Tokyo 160-8582, Japan, <sup>2</sup>Laboratory of Transcription Dynamics, Research Center for Biological Visualization, Institute for Quantitative Biosciences, The University of Tokyo, Tokyo 113-0032, Japan, <sup>3</sup>Department of Chromosome Science, National Institute of Genetics, Research Organization of Information and Systems (ROIS), Shizuoka 411-8540, Japan, <sup>4</sup>Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, Tokyo 113-0032, Japan, <sup>5</sup>Laboratory for Functional Non-coding Genomics, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan and <sup>6</sup>Japan Science and Technology Agency (JST), Precursory Research for Embryonic Science and Technology (PRESTO), Saitama 332-0012, Japan

Received September 16, 2022; Revised October 17, 2022; Editorial Decision October 18, 2022; Accepted October 22, 2022

## ABSTRACT

**Telomeres in *Drosophila* are composed of sequential non-LTR retrotransposons *HeT-A*, *TART* and *TAHRE*. Although they are repressed by the PIWI-piRNA pathway or heterochromatin in the germline, the regulation of these retrotransposons in somatic cells is poorly understood. In this study, we demonstrated that specific splice variants of Mod(mdg4) repress *HeT-A* by blocking subtelomeric enhancers in ovarian somatic cells. Among the variants, we found that the Mod(mdg4)-N variant represses *HeT-A* expression the most efficiently. Subtelomeric sequences bound by Mod(mdg4)-N block enhancer activity within subtelomeric TAS-R repeats. This enhancer-blocking activity is increased by the tandem association of Mod(mdg4)-N to repetitive subtelomeric sequences. In addition, the association of Mod(mdg4)-N couples with the recruitment of RNA polymerase II to the subtelomeres, which reinforces its enhancer-blocking function. Our findings provide novel insights into how telomeric retrotransposons are regulated by the specific variants of insulator proteins associated with subtelomeric sequences.**

## INTRODUCTION

Telomeres are specialized structures that protect the ends of chromosomes, and their loss of function can lead to various defects, including the fusion of chromosome ends (1). In mammals, telomere sequences are generally composed of short tandem repeats, such as TTAGGG, which are elongated by telomerase (2). Telomere elongation and terminal protection are controlled by multiple mechanisms. For example, the heterochromatin machinery (3), end-capping proteins (4–9) or telomeric repeating-containing RNA (TERRA) (10) play a critical role in telomeric regulation. Another important regulatory component is the subtelomere, which is a sequence neighboring telomeres. Subtelomeres are vital for telomere regulation, e.g. serving as transcription start sites of non-coding RNA, which are essential for telomere maintenance (10). Subtelomeres can also act as chromatin insulators, i.e. sequences that possess enhancer-blocking activity and act as epigenetic modification boundaries involved in the regulation of gene expression (11). In mammals, CTCF binds to subtelomeres and regulates the transcription of TERRA, which is essential for telomeric protection (12–14). However, it is unknown whether insulator-based telomeric control exists in diverse species.

In contrast with these telomere maintenance mechanisms, some invertebrate species lack telomerase activity, and telomeres are replaced by retrotransposons: *TRAS* and *SART* in *Bombyx mori* (15,16) and *HeT-A*, *TART*, and *TAHRE* in *Drosophila melanogaster* (17–20). Telomeres of

\*To whom correspondence should be addressed. Tel: +81 3 5363 3529; Fax: +81 3 5363 3266; Email: [iwasaki@keio.jp](mailto:iwasaki@keio.jp)

Correspondence may also be addressed to Haruhiko Siomi. Email: [awa403@keio.jp](mailto:awa403@keio.jp)

Present address: Shu Kondo, Department of Biological Science and Technology, Faculty of Advanced Engineering, Tokyo University of Science, Tokyo 125-8585, Japan.

*D. melanogaster* lack tandem repeat sequences completely, and it has been proposed that telomere length depends on the transpositions of these retrotransposons (20–22) (Figure 1A). *HeT-A*, *TART* and *TAHRE* have notable features: (i) they exist only at the telomeres, (ii) they line up unidirectionally and (iii) *HeT-A* is the most abundant among these three telomeric retrotransposons. *HeT-A* contains Gag protein but not Pol protein, which includes reverse transcriptase; therefore, *HeT-A* uses Pol from the other telomeric transposons upon its transposition (21,22). These processes are regulated by multiple fast-evolving telomere-capping proteins (23–28). The subtelomere sequences of *D. melanogaster* also have unique features. These subtelomere sequences are composed of TAS-L or TAS-R repeats (29,30) (Figure 1A). Although the TAS-L monomer is composed of a non-coding sequence, TAS-R is highly similar to the Invader4 long terminal repeat (LTR) element (31,32), suggesting an additional relationship between retrotransposons and telomere/subtelomere evolution in this organism.

These unique features of telomere maintenance by retrotransposons in *D. melanogaster* resemble a symbiosis between the host and retrotransposon (20). However, de-repression of *HeT-A* retrotransposon in germ cells causes severe embryo abnormalities (33,34), and a recent genomic analysis of other *Drosophila* species has suggested that these retrotransposons evolve rapidly and replicate their copy number selfishly rather than symbiotically (21,26,35). Therefore, the host must regulate the telomeric retrotransposon activity. Multiple mechanisms repress telomeric retrotransposons (36). For example, the PIWI-piRNA pathway is a small RNA-mediated RNA-silencing mechanism (37) known to repress transposons, including telomeric retrotransposons in fly germline cells. The dysfunction of this pathway leads to strong *HeT-A* expression (38–41). The heterochromatin machinery also plays a critical role in regulating telomeric retrotransposons. The mutation of heterochromatin protein 1 (HP1; also known as Su(var)205) causes telomere fusion and increases the expression of telomeric retrotransposons (42–45). Additionally, some transcription factors, such as *Woc*, *Stwl*, *Hmr* or *Lhr*, and poly-A tail regulation machinery contribute to *HeT-A* repression in the ovaries (33,43–46). Most of these repressive machineries function in a germline-specific manner (47), but the expression of telomeric transposons is also regulated within somatic cells. For example, the expression of *HeT-A* is limited in the proliferating brain cells of larvae in a cell cycle-specific manner (48,49), suggesting that the dynamic expression pattern of *HeT-A* also has functions in somatic cells. However, the regulatory mechanism of telomeric transposons in somatic cells remains to be elucidated.

In this study, we demonstrated that specific splice variants of Mod(mdg4), an insulator protein in *Drosophila*, act as repressors of telomeric *HeT-A* retrotransposons by blocking subtelomeric enhancer activity in ovarian somatic cells (OSCs). Using RNAi knockdown (KD) screening, we demonstrated that the Mod(mdg4) variants N, V and AF (hereafter, Mod(mdg4)-N, V, AF, respectively), which are variants of Mod(mdg4) insulators with unknown functions, are responsible for *HeT-A* repression. Among these variants, Mod(mdg4)-N homozygous mutant flies display *HeT-*

*A* de-repression in the ovary and a female sterility phenotype. Further analysis revealed that each Mod(mdg4) variant has binding specificity, and Mod(mdg4)-N is bound to both subtelomeric and telomeric loci, in addition to genome-wide binding sites. A live-imaging analysis demonstrated strong enhancer-blocking activity of the subtelomeric TAS-R sequences and weak enhancer-blocking activity of Mod(mdg4)-N-binding *HeT-A* sequences. These results indicate that Mod(mdg4)-N represses *HeT-A* expression by blocking subtelomeric enhancers. Further investigation into how Mod(mdg4) blocks the enhancer activities at TAS-R revealed that the important factors for inducing a strong enhancer blocking effect, specifically at the TAS-R regions, are the Pol II association site within TAS-R, capable of recruiting Pol II upon Mod(mdg4)-N association, and the highly repetitive nature of TAS-R sequences. Our findings provide novel insights into how insulator proteins mediate the link between subtelomeres and telomeres.

## MATERIALS AND METHODS

### List of *Het-A* binding proteins

The list of *HeT-A* binding proteins (the names of the ChIP antigens) was collected from the ChIP-Atlas (50,51). Using the dm6 datasets, we downloaded datasets with the following parameters: ‘ChIP: TFs and others’ for Experiment type, ‘All cell type’ for Cell type Class and ‘500’ for the threshold for significance. With this dataset, the peaks that overlapped with full-length *HeT-A* (chr2R:25261552–25268656) were extracted using bedtools (<https://github.com/arq5x/bedtools2>).

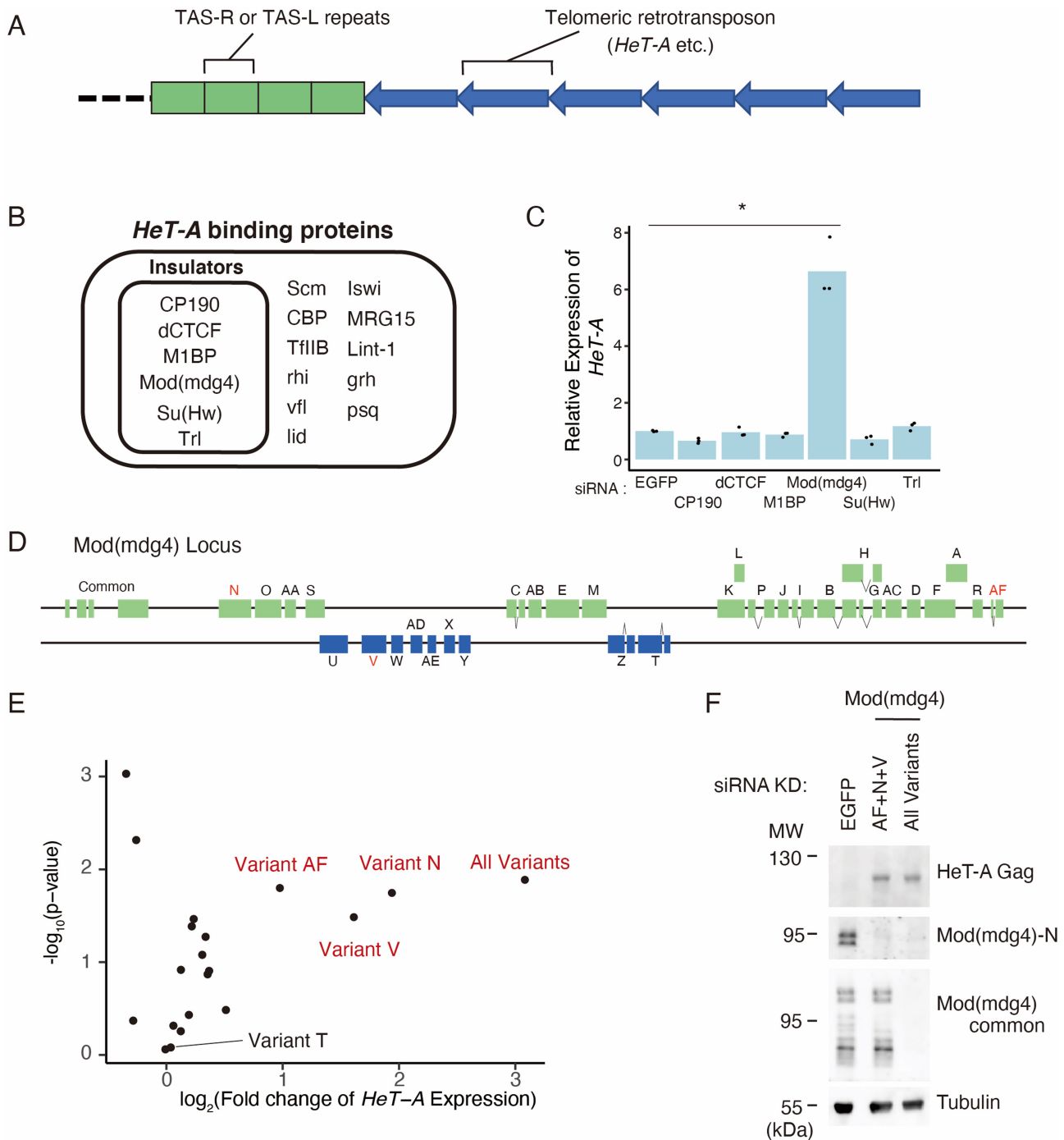
### Cell lines and culture condition

Ovarian somatic cells (OSC) were cultured as described (52,53). Briefly, OSC were cultured in Shield and Sang M3 Insect Medium (Sigma) supplemented with 10% fly extract, 10% fetal bovine serum, 0.6 mg/ml glutathione, and 10 µg/ml insulin. OSC were passed every second day.

### Fly strains

New alleles of mod(mdg4) were generated using the transgenic Cas9 system as previously described (54). The 20-bp sequence of the gRNA targeting the RN variant-specific exon of mod(mdg4) was as follows: GTAGTTGCGGAA-CACCAGCT. Multiple candidate mutant lines were established from individual male offspring of parents carrying both the nos-Cas9 and U6-gRNA transgenes. PCR was performed on their genomic DNA to amplify a region surrounding the gRNA target and their sequences were determined to search for indel mutations. Two lines, RN1-2 and RN1-4, were found to carry distinct frameshift mutations and were subjected to further functional analysis.

In all live-imaging experiments, *D. melanogaster* embryos at nuclear cycle 14 were analyzed. The following fly lines were used in this study: *nanos> MCP-GFP*, *His2Av-mRFP/CyO* (55), *DSCP<sub>WT</sub>-MS2-yellow-sna shadow enhancer* (55), *DSCP<sub>WT</sub>-MS2-yellow-gypsy-sna shadow enhancer* (56), *DSCP<sub>WT</sub>-MS2-yellow-HeT-A-sna shadow enhancer* (this study), *DSCP<sub>WT</sub>-MS2-yellow-HETRP-sna*



**Figure 1.** Mod(mdg4) restricts *HeT-A* expression in a variant-specific manner. (A) Schematic representation of a typical *D. melanogaster* telomere/subtelomere structure. On the right arms of chromosomes 2 and 3, the subtelomeres contain several TAS-R sequences and on the left arms of chromosomes 2 and 3, the subtelomeres contain several TAS-L sequences. Next to these subtelomeric TAS-R/L repeats, telomeric retrotransposons, such as *HeT-A* are connected unidirectionally. (B) Protein list with significant ChIP-seq peaks on *HeT-A* as revealed by the ChIP-Atlas (Oki *et al.*, 2018). Insulator proteins are highlighted by an inner rectangle. (C) siRNA knockdown (KD) screening for insulator genes in OSC, followed by qRT-PCR of *HeT-A* transcript expression levels. Values on the y-axis were normalized to *RP49* ( $n = 3$ ). Each dot indicates the value obtained from different experiments. Asterisks indicate  $P < 0.05$  in the two-sided *t*-test. (D) Schematic diagram of Mod(mdg4) locus. Mod(mdg4) has four common exons and 31 variant-specific exons. The names of the variants are indicated on the exons. Variants of interest are highlighted in red. (E) Volcano plot showing *HeT-A* expression level measured by qRT-PCR and the corresponding significance upon KD using siRNA targeting 19 Mod(mdg4) variants and all variants. Three variants (variant N, V, and AF) that upregulate *HeT-A* upon KD are labeled. Each dot represents the KD of the different variants ( $n = 3$ ). The x-axis is the  $\log_2$  ratio, and the y-axis is the  $\log_{10}$  ratio. (F) Western blotting showing protein levels of tubulin (loading control), total Mod(mdg4), Mod(mdg4)-N, and *HeT-A* Gag upon the KD of EGFP (control), Mod(mdg4)-AF + N + V, or all Mod(mdg4) variants. The molecular weight (MW) is indicated on the left side of each image.

*shadow enhancer* (this study), *DSCP<sub>WT</sub>-MS2-yellow-HeT-A-sna shadow enhancer-HeT-A* (this study), *DSCP<sub>WT</sub>-MS2-yellow-HETRP-sna shadow enhancer-HETRP* (this study).

### Cell transfection

For transfection of small interfering RNA (siRNA), 200 pmol siRNA duplex was transfected using the Cell Line 96-well Nucleofector Kit SF (Lonza) and program DG150 of the 96-well Shuttle Device (Lonza). For transfection of expression vectors, expression vectors were transfected using Xfect Transfection Reagent (TaKaRa Clontech), in accordance with the manufacturer's instructions. All siRNA sequences used in this study are listed in Supplementary Table S1.

### Generation of antibodies

The 1469–1901 bp region (for Mod(mdg4)-N antibody) or the 543–1364 bp region (for Mod(mdg4) common region antibody) of Mod(mdg4) variant N mRNA (Flybase: FBtr0084073) and the 3984–4511 bp region of HeT-A 23Zn-1 (GenBank: U06920.2) were subcloned into pMAL-c2G and pGEX-5X-1 from OSC or ovary cDNA. These vectors were expressed in Rosseta-gami B(DE3) (Novagen) with 1 mM IPTG at 18°C overnight. MBP-tagged proteins and GST-tagged protein were purified with Amylose Resin (New England Biolabs) and Glutathione Sepharose 4B (Cytiva). These purified proteins were used as antigens, and hybridomas with SP2/O were generated and screened according to standard protocol (57).

### Quantitative reverse-transcription PCR

For OSC, RNA extraction and cDNA preparation were performed with SuperPrep II Cell Lysis & RT kit for qPCR from  $1.0 \times 10^5$  OSC. For fly ovaries, total RNA extraction was performed with ISOGEN II (Nippon Gene: 311-07361) and cDNA was synthesized with Transcription First Strand cDNA Synthesis Kit (Roche 04379012001). TB Green Premix Ex Taq II (Clontech) was used for quantitative PCR with indicated qPCR primers (Supplementary Table S1).

### Western blot

Samples were lysed to the concentration of  $5.0 \times 10^5$  OSC/1 ovary per 10  $\mu$ l 1 $\times$  Laemmli buffer and heated at 95°C for 5 min. Proteins were separated by SDS–polyacrylamide gel electrophoresis (PAGE) and transferred to a 0.45  $\mu$ m nitrocellulose membrane (Cytiva). Membrane was washed by PBS and blocked with 3% skim milk in PBS-T, then incubated with primary antibody for 1 h at room temperature. For primary antibody, anti-beta tubulin (DSHB, E7) (1:5000), anti-c-Myc (DSHB, 9E10) (1:1000), anti-Ty1 antibody (Diagenode, C15200054) (1:2000), anti-HeT-A Gag supernatant (this study) (1:1), anti-Mod(mdg4) common region supernatant (this study) (1:4) and anti-Mod(mdg4) variant N supernatant (this study) (1:1) were used with indicated dilutions. After three washes with PBS-T, membrane was incubated with HRP-conjugated secondary antibody (MP Bioscience, 0855558) (1:5000) for 30 min, followed

by three PBS-T washes. The membranes were incubated at room temperature for 1 h. The membrane was incubated with ECL Prime Western Blotting Detection Regent (Cytiva) and was exposed to Amersham Hyperfilm ECL (Cytiva). Exposed film was developed by X-ray film developer (KONICAMINOLTA TCX-101).

### Immunofluorescence

Ovaries were fixed with 4% paraformaldehyde for 20 min and washed once by PBS-T (0.2% Tween20 in PBS), followed by blocking by PBS-BT (0.2% Tween20 and 10 mg/ml BSA in PBS). Primary antibodies were diluted with PBS-BT and incubated with samples overnight at 4°C. For primary antibodies, anti-Orb (DSHB, 4H8) (1:200) was used with indicated dilutions. Following three washes with PBS-T, secondary fluorophore-conjugated antibody (ThermoFisher, A32727) was diluted at 1:1000 in PBS-BT and incubated with samples for 2 h at room temperature. The stained samples were washed three times with PBS-T and mounted by DAPI-containing medium. Samples were imaged with FV3000 (Olympus) confocal microscope.

### RNA fluorescence in situ hybridization (FISH)

*HeT-A* RNA fluorescence in situ hybridization (FISH) was performed on the ovaries as described previously (58) and according to the manufacturer's instructions using Quasar-labeled 670 Stellaris oligo probes (Supplemental Table 4). Briefly, the ovaries were subjected to FISH with 4% PFA and dissected into ovarioles. The ovarioles were incubated overnight with 100% methanol for penetration. The methanol was then gradually replaced with PBS-T (0.2% Tween20 in PBS). The ovarioles were fixed again with 4% PFA followed by 1  $\mu$ g/ml proteinase K for 13 min. The proteinase K was then quenched twice with 2 mg/ml glycine. The ovarioles were washed twice with 1 ml of PBS-T and hybridized with FISH probes for 16 h at 37°C. The ovarioles were washed twice with wash buffer A, PBS and 1  $\mu$ g/ml DAPI-containing PBS. The samples were imaged using an FV3000 confocal microscope (Olympus).

### Isolation of Ty1-Mod(mdg4) variants stable line OSC

Stable cell lines were established as described previously (59). In short, Mod(mdg4) variants were cloned into pPB-2  $\times$  Ty1-Tjen-EGFP-P2A-BlastR. Stable clones were selected by medium containing 50  $\mu$ g/ml blasticidin.

### Plasmid construction

*HeT-A/HETRP sequence cloning.* *HeT-A* or HETRP sequences were cloned from the OSC genome. These sequences were amplified by PCR using specific primers (rep\_HeT-A\_F/R or rep\_HETRP\_F/R), and cloned to pBlueScript SK(+) EcoRI/NotI treated fragment by using NEBuilder HiFi DNA Assembly Master Mix (NEB). These plasmids are named pBS-HeT-A or pBS-HETRP respectively. Primer sequences for cloning are shown in Supplementary Table S1, and whole sequences of fragments are shown in Supplementary Table S2.



*DSCP<sub>WT</sub>-MS2-yellow-HeT-A-sna shadow enhancer*: A DNA fragment containing *HeT-A* sequence was amplified from pBS-HeT-A using primers (5'-ACA TGA AGC TTC TCC GTT CTA CCT CAA T-3') and (5'-ACG GGA AGC TTC ACA GGG TGC CGC AAA AAT TG-3') and digested with HindIII. The resulting fragment was inserted into the unique HindIII site of pbphi-DSCP-MS2-yellow-sna shadow enhancer (55).

*DSCP<sub>WT</sub>-MS2-yellow-HETRP-sna shadow enhancer*: A DNA fragment containing HETRP sequence was amplified from pBS-HETRP using primers (5'-ACG GGA AGC TTG TGT GTC ATC CAT TTC GTT T-3') and (5'-ACA TGA AGC TTC GAC GCG TAC ACA TAT TTC G-3') and digested with HindIII. The resulting fragment was inserted into the unique HindIII site of pbphi-DSCP-MS2-yellow-sna shadow enhancer (55).

*DSCP<sub>WT</sub>-MS2-yellow-HeT-A-sna shadow enhancer-HeT-A*: A DNA fragment containing *HeT-A* sequence was amplified from pBS-HeT-A using primers (5'-ACA TGG CTA GCC TTC TCC GTT CTA CCT CAA TAT ATC-3') and (5'-TTA AAG CTA GCC ACA GGG TGC CGC AAA AAT TG-3') and digested with NheI. The resulting fragment was inserted into the unique NheI site of *DSCP<sub>WT</sub>-MS2-yellow-HeT-A-sna shadow enhancer*.

*DSCP<sub>WT</sub>-MS2-yellow-HETRP-sna shadow enhancer-HETRP*: A DNA fragment containing *HeT-A* sequence was amplified from pBS-HETRP using primers (5'-ACA TGG CTA GCG TGT GTC ATC CAT TTC GTT TAT TC-3') and (5'-TTA AAG CTA GCC GAC GCG TAC ACA TAT TTC GC-3') and digested with NheI. The resulting fragment was inserted into the unique NheI site of *DSCP<sub>WT</sub>-MS2-yellow-HETRP-sna shadow enhancer*.

### Site-specific transgenesis by phiC31 system

All reporter plasmids were integrated into a unique landing site on the third chromosome using the VK00033 strain (60). PhiC31 was maternally provided using the *vas-phiC31* strain (61). Microinjection was performed as previously described (62). In brief, 0–1 h embryos were collected and dechorionated with bleach. Aligned embryos were dried with silica gel for ~7 min and covered with FL-100-1000CS silicone oil (Shin-Etsu Silicone). Subsequently, microinjection was performed using FemtoJet (Eppendorf) and DM IL LED inverted microscope (Leica) equipped with M-152 Micromanipulator (Narishige). Injection mixture typically contains ~500 ng/μl plasmid DNA, 5 mM KCl, 0.1 mM phosphate buffer, pH 6.8. The mini-white marker was used for screening.

### MS2 live-imaging

Virgin females of *nanos> MCP-GFP, His2Av-mRFP/CyO* (55) were mated with males carrying the MS2 allele. The resulting embryos were dechorionated and mounted between a polyethylene membrane (Ube Film) and a coverslip (18 mm x 18 mm), and embedded in FL-100-450CS

(Shin-Etsu Silicone). Embryos were imaged using LSM900 (Zeiss). Temperature was kept between 23.5 and 24.5°C during imaging. Plan-Apochromat 40×/1.4 N.A. oil immersion objective was used. A stack of 26 images separated by 0.5 μm was acquired at each time point, and the final time resolution was 16.8 sec/frame. Images were captured in 16-bit. Images were typically taken from the end of nuclear cycle 13 to the onset of gastrulation at nuclear cycle 14. During imaging, data acquisition was occasionally stopped for a few seconds to correct z-position, and data were concatenated afterwards. For each cross, three biological replicates were taken. The same laser power and microscope setting were used for each set of experiments. The laser power was measured using X-Cite XR2100 power meter (Lumen Dynamics).

### mRNA-seq

Total RNAs were isolated with the Isogen II reagent (Nippon Gene) according to the manufacturer's protocol. Poly(A)+ RNAs were obtained using Oligo-dT beads. Libraries were prepared with Illumina TruSeq stranded mRNA kit according to the manufacturer's protocol.

### Chromatin immunoprecipitation (ChIP)

ChIP experiments of Ty1-tagged Mod(mdg4) variants were performed using the truChIP Chromatin Shearing Kit according to the manufacturer's instructions with minor modifications.  $2 \times 10^7$  OSC were fixed with 1% formaldehyde for 10 min and quenched and lysed with buffers from the truChIP Chromatin Shearing Kit. Fixed chromatin was sheared with Bioruptor II (BMBio BR2012A) for 20 cycles of 30 s ON/30 s OFF, high settings. Sheared chromatin was immunoprecipitated by antibody-conjugated Dynabeads protein G for 2 h. Immunoprecipitated chromatin was incubated with Proteinase K and RNase, and de-crosslinked at 65°C overnight.

For ChIP experiments of RNA polymerase II,  $2 \times 10^7$  OSC were fixed with 1% formaldehyde for 10 min and quenched with final 0.1 M glycine for 5 min. After washing twice with PBS, nuclei were isolated with 1 ml of swelling buffer (25 mM HEPES-KOH (pH 7.5), 1.5 mM MgCl<sub>2</sub>, 10 mM KCl, 0.1% NP-40, 1 mM DTT and 1x Protease Inhibitor). Isolated nuclei were lysed in 400 μl of sonication buffer (50 mM HEPES-KOH (pH 7.4), 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS and 1 × protease inhibitor) and were sheared with Bioruptor II (BMBio BR2012A) for 20 cycles of 30 s ON/30 s OFF, high settings. 3 μg of Pol II antibody were incubated with chromatin overnight (63). The next day, 20 μl of Dynabeads M-280 Sheep anti-Mouse IgG was added to the sample and incubated for 1 h. Beads were washed and treated with Proteinase K and RNase as described (64). DNA was purified with isopropanol precipitation using Pellet Paint NF Co-precipitant (Merck: 70748). Fragments from the ChIP experiment were sheared to ~200 bases using Covaris S220. These were used for library preparation with the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB) following the manufacturer's protocol.

### Dual luciferase assay

For the plasmid construction, the Nluc-PEST sequence was cloned downstream of the *Tj* enhancer and *Drosophila* synthetic core promoter, and the TAS-R sequences were cloned at both sides of the *Tj* enhancers using standard cloning techniques (Supplemental Table 2).

For the transfection, 200 pmol siRNAs were transfected into  $2.0 \times 10^6$  OSCs as described above, and the cells were passaged into four wells of a 24-well plate. On the day after the siRNA transfection, 5 ng of each indicated plasmid was transfected into the OSCs using Xfect. After 24 h of plasmid transfection, dual-luciferase assays were performed according to the manufacturer's instructions (Promega). The luciferase activity in each well was measured using Cytation 5 (BioTek) with a 1-s integration time.

### Enhancer blocking reporter assay in OSC

Reporters were constructed from OSC.Reporter.UAS.traffic\_jam.BoxB.d2eGFP.t2a.Blast (65) and pAc5.1 vector in which Flag-mCherry:P2A:PuroR gene was cloned using the indicated primer to add an additional restriction site (Supplemental Table 1). Wild-type and mutant *HETRP* sequences were inserted into reporter with NruI or SalI restriction enzyme and NEBuilder HiFi DNA assembly (NEB). These reporters were transfected with pHsp70-Myc-PBase plasmid as described (59). Cells were selected with puromycin and fluorescence was measured with SH800Z (Sony). Normalized EGFP intensity is the original EGFP signal intensity divided by mCherry intensity, which allows normalization of the effect of the insertion copy number.

### Fertility assay

7–10 virgin female flies were collected 3–5 days before mating. Virgin female flies were mated with the indicated male and kept at 25°C overnight. The next day, mated flies were transferred onto grape juice plate and kept at 25°C without light. After 3 h, eggs were counted and kept at 25°C again. After 24 h, hatching eggs were counted.

### Image analysis

All the image processing methods and analysis were implemented in MATLAB (R2020a, MathWorks).

### Segmentation of nuclei

For each time point, maximum projection was obtained for all z-sections per image. His2Av-mRFP was used to segment nuclei.  $512 \times 512$  maximum projection images were initially cropped to  $300 \times 430$  to remove nuclei at the edge, and used for subsequent analysis. For nuclei segmentation, His2Av images were first blurred with Gaussian filter to generate smooth images. Pixels expressing intensity higher than 5% of the global maxima in the histogram of His2Av channel were removed. Processed images were converted into binary images using a custom threshold-adaptive segmentation algorithm. Threshold values were determined at each time frame by taking account of (i) histogram distribution

of His2Av channel and (ii) the number and the size of resulting connected components. Boundaries of components were then modified to locate MS2 transcription dots inside of nearest nuclei. In brief, pixels with intensity twice larger than mean intensity of MS2 channel were considered as transcription dots, and new binary image was created for each time frame. The Euclidean distances between the centroid of binarized transcription dot and all boundaries of segmented nuclei were calculated. Boundary of the nucleus with the smallest Euclidean distance was modified in order to capture transcription dot within a nucleus. Centroids of connected components in nuclei segmentation channel were used to compute the Voronoi cells of the image. Resulting binary images were manually corrected using Fiji (<https://fiji.sc>).

### Tracking of nuclei

Nuclei tracking was done by finding the object with minimal movement across the frames of interest. For each nucleus in a given frame, Euclidean distances between the centroids of the nucleus in the current time frame and the nuclei in the previous time frame were determined. Nucleus with the minimum Euclidean distance was considered as a same lineage.

### Recording of MS2 signals

3D raw images with all z-sections of MCP-GFP channel were used to record MS2 fluorescence signals. Using segmented regions from max projected images of His2Av-mRFP channel, fluorescence intensities within each nucleus were extracted. 3D fluorescence values were assigned to the nearest segmented regions of projected images. Signals of MS2 transcription dots were determined by calculating an integral of fluorescence intensities around the brightest pixel within each nucleus using a 2D Gaussian fitting method as described below. (i) The xyz position of transcription site was determined as the brightest pixel in each nucleus. (ii) A 2D Gaussian fitting was performed in a  $11 \times 11$  pixels region with a single z-plane centering the transcription site to estimate a fluorescent dot intensity and a local background. Fitting was performed with the following formula

$$I(x, y) = \alpha + I_0 \exp\left(-\left(\frac{(x-x_0)^2}{2\sigma_x^2} + \frac{(y-y_0)^2}{2\sigma_y^2}\right)\right)$$

where  $\alpha$  is the local background intensity,  $I_0$  is the amplitude of the peak fluorescence intensity,  $x_0$  and  $y_0$  are the center of the peak,  $\sigma_x$  and  $\sigma_y$  are the spreads of the fluorescent dot. (iii) The intensity of MS2 transcription dot was calculated as  $2\pi\sigma_x\sigma_y I_0$  from fitting parameters as an estimated integral value after subtracting the local background (66). Subsequently, minimum MS2 intensities were determined for individual trajectories and subtracted to make the baseline zero.

### Detection of transcriptional bursting

A transcriptional burst was defined as a local change in fluorescence intensity. First, signal trajectories were smoothed

by averaging within a window of 5 timeframes. When a nucleus had above-threshold transcription activity, burst was considered to be started. Burst was considered to be ended when the intensity dropped below 55% of the local peak value. When the burst duration was less than 5 timeframes, it was considered as a false-positive derived from detection noise. When the signal trace exhibited continuous decreasing at the beginning of burst detection, it was also not considered as a burst. Location of defined burst was then moved two timeframes afterwards to better capture the center of individual bursting event. The same method and threshold value were used for each set of experiments.

### Description of bursting properties

From each trajectory, number of bursts, amplitude and duration of each burst, and total integrated signal (output) produced by each nucleus were measured. To determine amplitude, the peak value during the burst was measured using trajectories after smoothing by averaging within a window of 5 timeframes. Duration was determined by measuring the length of each burst. Total RNA production was measured by taking the area under the raw trajectory. Amplitude and duration for each nucleus were determined by taking the average of all analyzed bursts in a single nucleus.

### ChIP-seq data analysis—mapping and peak call

Adaptors added by NEBNext Ultra II were cut by cutadapt (67) using the following parameters (cutadapt -j 12 -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-A AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT -m 20). For mapping to the genome, subsequent reads were mapped to the dm6 by bowtie2 using the following parameters (bowtie2 -p 16 -N 1). Peak call was performed by MACS using the following parameter: for Mod(mdg4) variants (macs2 callpeak -f BAM -q 1e-15 -g dm) and for PolII (macs2 callpeak -f BAM -g dm -outdir ../peakcall). To note, peak call for Mod(mdg4) was performed with strict cutoff values to rule out ambiguous peaks and we confirmed almost no peak was detected in Ty1-tag ChIP for wildtype OSC dataset. Binding motif was found with MEME-ChIP (68) with standard options. Coverage tracks were visualized by pyGenomeTracks (69) with standard options.

### ChIP-seq data analysis—handling of Pol II ChIP-seq data

Pausing index was calculated as previously described (70). Pol II signal of both promoter region (TSS-250~TSS + 250) and gene body (TSS + 500~TES-500) was calculated from a public dataset (71). Genes whose length is shorter than 1500 bp were removed from downstream analysis. To obtain the pausing index for each gene, we divided reads of promoter region by reads of gene body. Cumulative distribution plot was illustrated by Python. For identifying Pol II loss genes, maximum coverage was calculated for each Pol II peak. Fold change of this maximum value was calculated between EGFP KD and Mod(mdg4) variant N KD.

### mRNA-seq data analysis

The mRNA reads were mapped to dm6 by RSEM and STAR using the following parameters (rsem-calculate-expression -paired-end-p 8 -star -star-path ~/anaconda3/bin -gzipped-read-file). The differentially expressed genes (DEGs) were detected using limma and edgeR. A gene ontology analysis was performed using DAVID (72). All the gene lists were used as the background.

For the transposon mapping, the paired-end reads were mapped with STAR with the-outFilterMultimapNmax 100 option, and first reads were extracted using samtools. Then, only the sense reads were counted with featureCounts with -M -fraction -s 2 options to exclude anti-sense transcripts. The statistical tests were performed with edgeR using the counts per million reads.

### Reanalysis of single-cell RNA-seq data

The data were obtained from Jevitt *et al.* (73). We used Cellranger (version 7.0.0) (74) to map the transcriptome. As a reference, we used the dm6 data with a custom annotation (GTF format) file. The GTF file was downloaded from the ensemble database and modified to distinguish between each Mod(mdg4) transcript. This GTF file was deposited in GitHub. After mapping, each UMI was annotated into cell-type clusters as reported in the original paper (73), and the transcripts were counted within each cluster using the original Python script. The clusters were combined into the following categories: germ cells (1. Germline cluster 1, 2. Germline cluster 2) and early follicle cells (7. The mitotic follicle cells (stg. 1–5), 8. Postmitotic follicle cells (stg. 6–8), 7. Vitellogenic MBFCs (g. 8) 1, 8. Vitellogenic MBFCs (g. 8) 2, 11. Vitellogenic MBFCs (g. 8) 3, 12. Vitellogenic MBFCs (g. 8) 4, 13. Vitellogenic MBFCs (g. 9–10A) 1, 14. Vitellogenic MBFCs (g. 9–10A) 2, 15. Vitellogenic MBFCs (g. 9–10A) 3, 16. The choriogenic MBFCs (g. 10B)), and late follicle cells (17. dorsal appendage-forming follicle cells, 18. The choriogenic MBFCs (g. 12), 19. The choriogenic MBFCs (g. 14) 1, 20. The choriogenic MBFCs (g. 14) and 2), and the RPM was calculated within each category.

## RESULTS

### Mod(mdg4) variants N, V and AF repress *HeT-A* expression independently of heterochromatin formation

*HeT-A*, the most abundant of the telomeric transposons in *Drosophila* (Figure 1A) (35), is repressed by the PIWI-piRNA pathway in *Drosophila* germline cells. This is due to the expression of piRNAs against *HeT-A*, processed from the Rhino-dependent transcription of piRNA precursor transcripts (75); therefore, *HeT-A* is dramatically upregulated in the ovaries of *piwi* mutant flies. In contrast, piRNAs against *HeT-A* are not present in the ovarian follicle cells and the cell line derived from follicle cells, named OSC (52,53), where only Rhino-independent PIWI-piRNA regulation occurs (53). A reanalysis of previously published RNA-seq data revealed that *HeT-A* upregulation was not observed upon *Piwi* knockdown (KD) in the OSC (Supplementary Figure S1A) (76). The KD of



linker histone H1, which is indispensable for heterochromatin maintenance, also did not cause the upregulation of *HeT-A* (Supplementary Figure S1A) (76–78) and the levels of the heterochromatin histone mark H3K9me3 on *HeT-A* were relatively low compared with heterochromatic transposons, including *mdg1*, in the OSC (Supplementary Figure S1B). These results suggested the existence of a piRNA- and heterochromatin-independent *HeT-A* repression mechanism in the OSC. To gain insight into the transcriptional regulation of the retrotransposons at the telomeres, we searched the ChIP-Atlas database, which contains previously published ChIP-seq data sets, for proteins associated with *HeT-A* sequences (50,51). We found that several insulator proteins were potentially associated with the *HeT-A* regions (Figure 1B, Materials and Methods). To test the possibility that *HeT-A* is regulated by these insulator proteins, we performed a KD screen using siRNAs. This revealed that Mod(mdg4) KD led to the de-repression of *HeT-A* retrotransposons in the OSC (Figure 1C).

The *Mod(mdg4)* locus produces as many as 31 variants owing to trans-splicing (Figure 1D) (79). Among the 31 variants, only variants H and T have been reported (80–84). Almost all the variants were expressed in the OSC at various levels (Supplementary Figure S1C). To determine which variants were responsible for the *HeT-A* repression, we performed a KD screen using variant-specific siRNAs. The KD of variants N, V, and AF [Mod(mdg4)-N, V, AF], all of which are variants with unknown functions, resulted in the upregulation of *HeT-A* expression levels (Figure 1E). We further performed western blotting using specific antibodies generated against Mod(mdg4)-N (Supplementary Figure S1D), a common exon of Mod(mdg4) (Supplementary Figure S1E) and HeT-A Gag (Supplementary Figure S1F). This revealed that, consistent with the RNA levels, the *HeT-A* Gag protein became detectable upon the KD of these three variants (N + V + AF) to a level close to that observed upon the KD of all the variants (Figure 1F, S2A–C). These results indicate that Mod(mdg4)-N, V, and AF repress *HeT-A* expression in a piRNA- and heterochromatin-independent manner.

### Mod(mdg4)-N mutant flies show elevated *HeT-A* expression

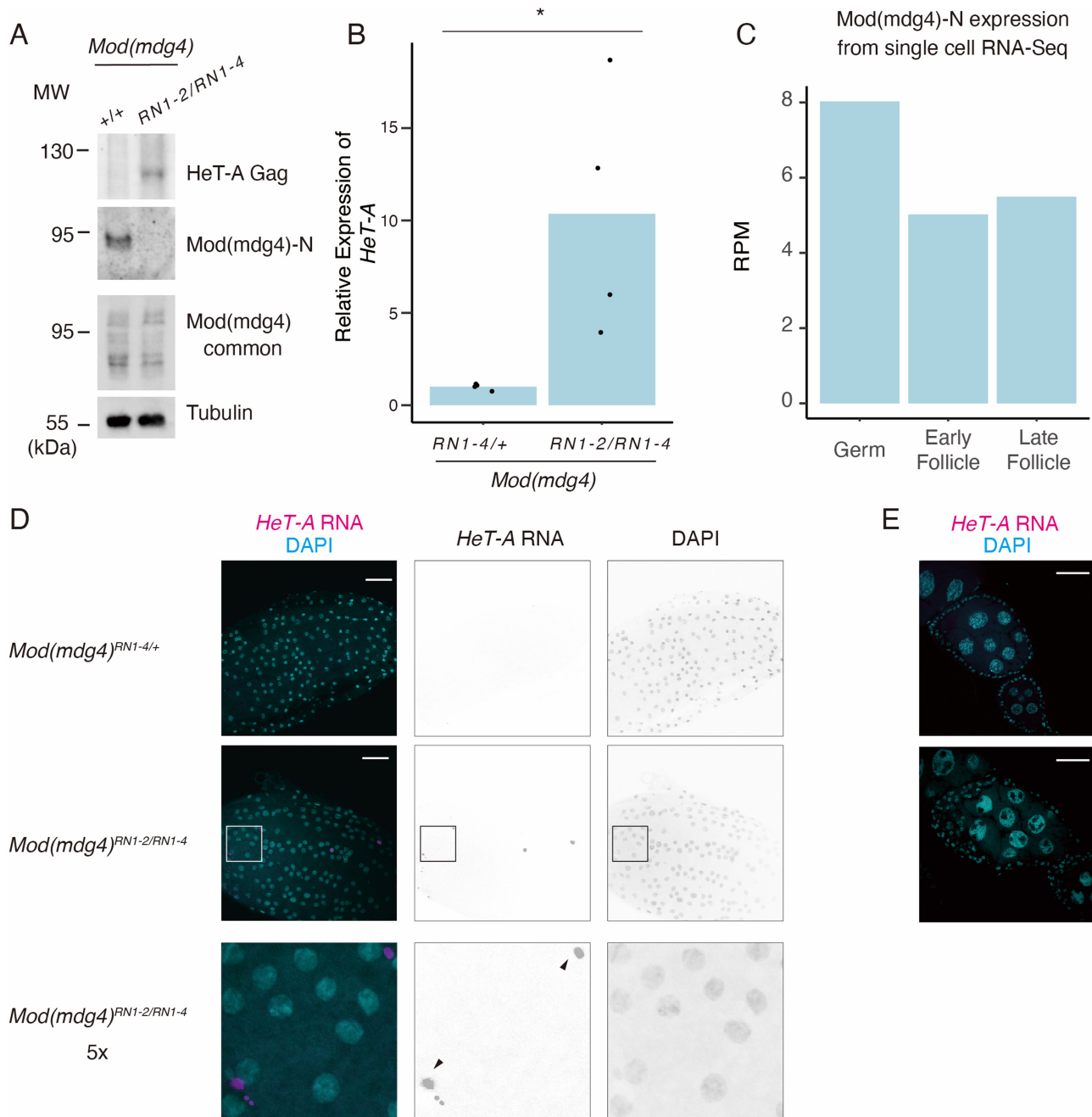
The Mod(mdg4)-N KD resulted in the greatest degree of de-silencing of *HeT-A* expression among all the tested variants in the OSC (Figure 1E). To investigate whether *HeT-A* upregulation could be observed *in vivo* upon Mod(mdg4)-N loss and the physiological role of this specific variant, we generated two lines of Mod(mdg4)-N mutant flies (*RNI-2* and *RNI-4*) that harbor nucleotide deletions at the variant-specific exon using CRISPR-Cas9 (54) (Supplementary Figure S3A). Both deletions caused a frameshift in the coding sequence of the Mod(mdg4)-N-specific exon, and the Mod(mdg4)-N protein was eliminated completely in the homozygous or trans-heterozygous mutant flies (Figure 2A, S2D–F, S3B). In these homozygous or trans-heterozygous mutant ovaries, we observed increased expression of both *HeT-A* transcripts and *HeT-A* Gag proteins (Figure 2A–B, S3B–C). This result indicates that Mod(mdg4)-N represses *HeT-A* expression *in vivo* as well as in the OSC.

To understand which cells in the ovary possess this Mod(mdg4)-N-mediated *HeT-A* regulation, we reanalyzed previously published single-cell RNA-seq (scRNA-seq) data from *Drosophila* ovaries (73). We selected cell clusters of three categories as defined in a previous publication: germ cells, early follicle cells, and late follicle cells. We searched for reads mapped specifically to the Mod(mdg4)-N variant. This revealed that Mod(mdg4)-N was expressed in both germ and follicle cells (Figure 2C). Despite the ubiquitous expression of Mod(mdg4)-N in the ovary, RNA-FISH of *HeT-A* transcripts revealed that *HeT-A* was de-repressed specifically in the late-stage follicle cells in the Mod(mdg4)-N mutant ovaries (Figure 2D, E). These results indicate that, although Mod(mdg4)-N is expressed in both germ and somatic cells, it is especially functional for *HeT-A* repression in somatic cells. In the germ cells, the PIWI-piRNA pathway is active; therefore, *HeT-A* is likely to be still silenced without regulation by Mod(mdg4)-N.

Interestingly, the flies that were homozygous for each mutant allele and the trans-heterozygous flies were viable yet exhibited female sterility (Supplementary Figure S3D). Although the mutants of previously described Mod(mdg4) variants (variants T or H) did not show such a female sterility phenotype, this phenotype resembled previously reported hypomorphic mutants in common regions of Mod(mdg4) (85). The expression levels of variants other than Mod(mdg4)-N were not affected in this mutant (Figure 2A, Supplementary Figure S3B), suggesting that Mod(mdg4)-N was responsible for the female sterility phenotype. Macroscopically, the ovaries of the trans-heterozygous mutants were not atrophic (Supplementary Figure S3E), indicating that ovary development was not drastically impaired. Histologically, although the tissue was partly distorted, Orb (oocyte marker)-positive cells were present in the ovarioles of the trans-heterozygous Mod(mdg4)-N mutants (Supplementary Figure S3F). However, we observed eggs in the ovaries of trans-heterozygous Mod(mdg4)-N mutants and follicle cells remaining in the egg chambers (Supplementary Figure S3G), suggesting ovulation defects in the Mod(mdg4)-N mutant flies.

To understand the basis of the ovulation defect in the Mod(mdg4)-N mutant, we performed an RNA-seq analysis of the mutant ovaries and detected 505 upregulated and 911 downregulated genes (Supplementary Figure S3H, Supplemental Table 3). We performed a gene ontology analysis with both upregulated and downregulated genes (72). This revealed that in the Mod(mdg4) mutant, the GO term ‘cytoplasmic translation’ was enriched for the downregulated genes (Supplementary Figure S3I), and a series of cytoplasmic ribosomal proteins were included in the list of downregulated genes (Supplemental Table 3). The loss of ribosomal proteins can cause defects in follicle cells (86,87); therefore, the misregulation of genes encoding ribosomal proteins caused by the loss of Mod(mdg4)-N may induce follicle cell rupture defects. Whether the misregulation of genes encoding ribosomal proteins is linked to *HeT-A* expression remains unclear. Altogether, these results indicate that Mod(mdg4)-N plays a vital role in *HeT-A* repression *in vivo* and female fertility in the ovary.





**Figure 2.** *Mod(mdg4)*-N plays a critical role in *HeT-A* repression in the follicle cells of the ovaries. (A) Western blots showing protein levels of *HeT-A* Gag, *Mod(mdg4)*-N, total *Mod(mdg4)*, and tubulin (loading control) in ovaries of wild-type (+/+; OregonR) or trans-heterozygous *Mod(mdg4)*<sup>RN1-2/RN1-4</sup> mutants. The molecular weight (MW) is indicated on the left side of each image. (B) Bar plot showing *HeT-A* expression as measured by RT-qPCR. Values on the y-axis were normalized to *RP49* (n = 4). Each dot indicates the value obtained from different experiments. Asterisks indicate p < 0.05 in the two-sided t-test. (C) Bar plot showing *Mod(mdg4)*-N expression in germ cells, early follicles, and late follicle cells obtained from single-cell RNA-seq data from Jevitt *et al.* (73). The y-axis indicates the RPM for these cell types. (D) RNA-FISH images of stage 13–14 egg chambers. Images show DAPI staining (right), *HeT-A* RNA FISH (middle), and merging of the two (left). The bottom images are enlarged images of the *Mod(mdg4)*-N<sup>RN1-2/RN1-4</sup> mutants. Black arrows indicate *HeT-A* foci. (E) RNA-FISH images of stage 5–7 egg chambers. The egg chambers were stained with DAPI (cyan) and *HeT-A* RNA FISH (magenta).

### Mod(mdg4)-N associates with telomeric and subtelomeric repeats

To determine how Mod(mdg4)-N represses *HeT-A* expression, we performed a ChIP-seq analysis by generating stable cell lines expressing Ty1-tagged Mod(mdg4) variants (Supplementary Figure S4A). The monoclonal antibody generated for Mod(mdg4)-N was unsuitable for ChIP experiments, so we performed ChIP-seq using the tagged Mod(mdg4) variants. The Ty1-tag was used because it lacks lysine, which is the primary target of formaldehyde, and has been successfully used for ChIP analyses (88). Mod(mdg4)-T, which forms protein complexes with Su(Hw) and CP190 as gypsy insulators and binds to the same location as Su(Hw) (89), was used as a control. To avoid artifacts caused by phantoms or pseudo-peaks (90), wild-type (WT) cells were also used to generate ChIP-seq libraries. Of the 31 Mod(mdg4) variants, 27 have FLYWCH domains in variant-specific regions (91). FLYWCH domains are C2H2-type zinc finger domains conserved in eukaryotes and are necessary for DNA binding or recruitment to chromatin (92,93), suggesting that the genomic binding patterns of the Mod(mdg4) variants may be different (85). Consistent with this view, the Mod(mdg4)-N and -T ChIP-seq peaks in the genome showed variant-specific chromatin localization patterns and did not overlap (Figure 3A–C), and the peaks were associated with genomic regions with different annotations (Figure 3D). Furthermore, a motif analysis using MEME revealed the sequence specificity of each Mod(mdg4) variant (Figure 3E, Supplementary Figure S4B). Consistent with a previous report (89), the Mod(mdg4)-T motif was highly similar to the previously identified Su(Hw) motif (Supplementary Figure S4C) (94), confirming that the results of the Ty1 tag-based ChIP-seq reflect the endogenous distribution of each Mod(mdg4) variant.

Next, we investigated how Mod(mdg4)-N is associated with telomere/subtelomere sequences. As expected from the KD screen (Figure 1E), mapping to the consensus sequence of *HeT-A* and reference genome revealed that Mod(mdg4)-N bound to *HeT-A*, whereas Mod(mdg4)-T did not (Figure 3F). Additionally, Mod(mdg4)-N was strongly enriched in the TAS-R repeats (Figure 3G), but not in the TAS-L repeats (Supplementary Figure S4D). Taken together, the accumulation of Mod(mdg4)-N in both the telomeric *HeT-A* and subtelomeric TAS-R repeats suggests that this variant regulates the expression of *HeT-A* by associating at these loci.

### The regulation of expression by Mod(mdg4)-N occurs specifically at *HeT-A*

Mod(mdg4)-N is associated with various genomic loci in OSCs (Figure 3). Therefore, we aimed to determine the extent to which Mod(mdg4)-N affects gene expression. To answer this question, we performed RNA-seq on Mod(mdg4)-N KD cells, which specifically depleted the -N isoform of Mod(mdg4) (Figure 4A). Unexpectedly, the expression of most protein-coding genes remained unchanged (Figure 4B). Only two genes were identified as differentially expressed genes (DEGs), despite the association of Mod(mdg4)-N at numerous gene bodies (331 peaks) and

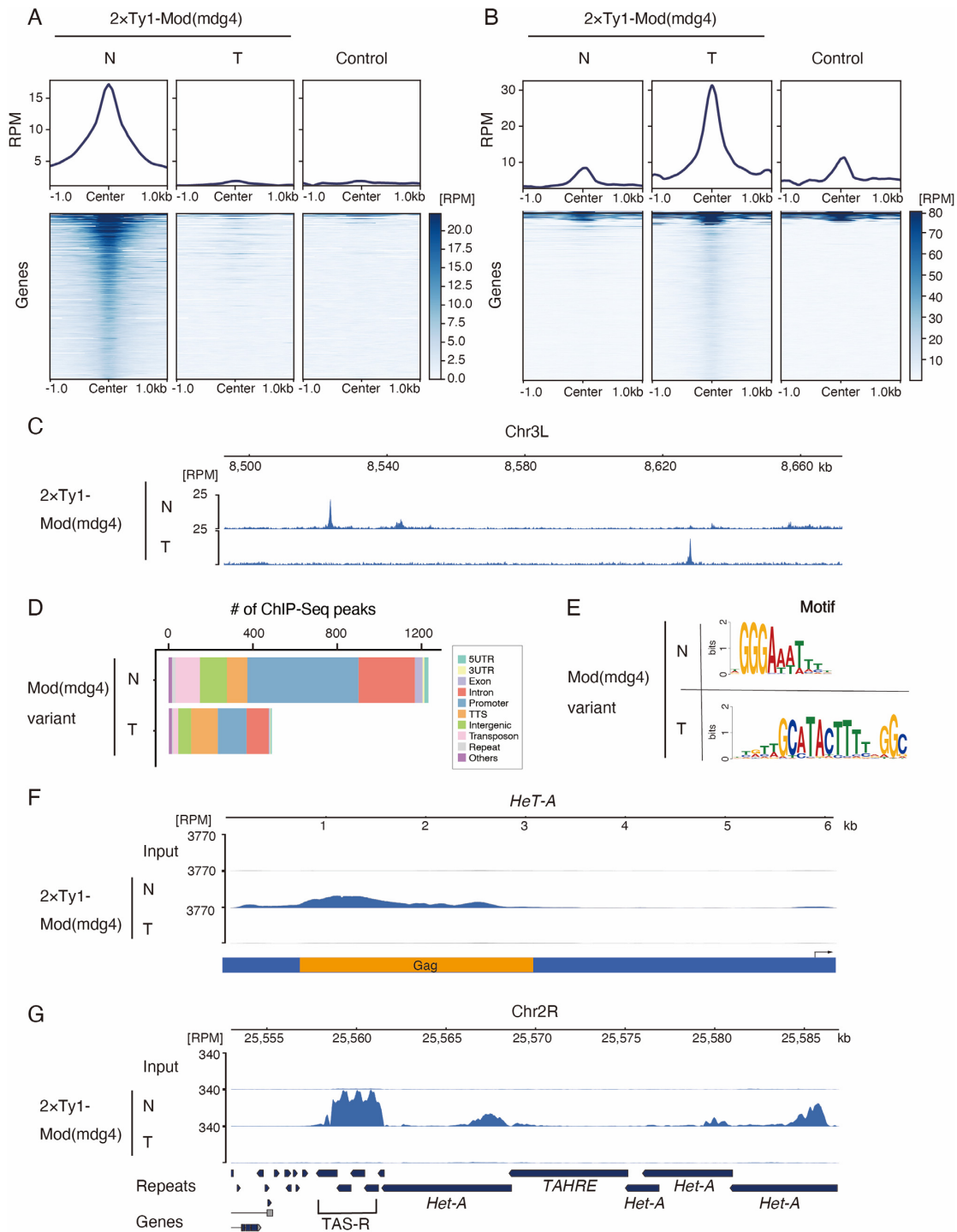
promoters (527 peaks) (Figure 3D). These two genes were *CG9628* and *5-HT2A*, whose biological functions are unrelated to telomere/subtelomere or transposon expression. In addition, when we analyzed the expression levels of transposons in the Mod(mdg4)-N KD cells, we found that the effect was limited to *HeT-A* (Figure 4C). These results show that Mod(mdg4)-N specifically regulates expression of *HeT-A*, despite its genome-wide association. Additionally, these results suggest that Mod(mdg4)-N does not directly regulate the expression of associated genes, but rather regulates expression in a context-specific manner.

### Subtelomeric TAS-R regions harbor enhancer activity

Since Mod(mdg4)-N cannot directly regulate most of its associated genes, we hypothesized that Mod(mdg4)-N has a specific function at telomeric regions. Previous reports have suggested the enhancer-blocking activity of Mod(mdg4)-T as a component of gypsy insulators (82,95). Given the strong binding of Mod(mdg4)-N to the telomeric and subtelomeric repeats, Mod(mdg4)-N may block enhancers located near telomeres, thereby repressing *HeT-A* expression. To search for enhancers near the telomeres, we reanalyzed STARR-seq [self-transcribing active regulatory region (STARR) sequencing] data obtained using OSCs (96). STARR-seq is a method for the genome-wide detection of enhancer activity by inserting randomly fragmented genomic sequences into reporter plasmids. By examining the regulation levels of the reporter gene, the enhancer activity of the randomly inserted fragments can be analyzed. The inserted fragments are deep-sequenced, and the genomic regions with accumulated reads indicate the regions with strong enhancer activity (96). Using this dataset, we analyzed the genomic regions near the telomeres and observed that TAS-R had strong enhancer activity in the OSC, whereas enhancer activity was not observed for *HeT-A* or TAS-L (Figure 4D, Supplementary Figure S5A, B). This result, together with the finding that Mod(mdg4)-N does not bind to TAS-L (Supplementary Figure S4D), indicates that different modes of regulation occur for the *HeT-A* associated with TAS-R and TAS-L. TAS-L does not possess enhancer activity, and therefore, the lack of Mod(mdg4)-N at TAS-L does not result in the de-repression of *HeT-A*. It is plausible that the TAS-R repeats possess enhancer activity because they are derived from the LTR sequences of the *Invader4* retrotransposon (31,32) and because LTR sequences generally have both enhancer and promoter activity (97). With these observations on the Mod(mdg4)-N-binding features and cis-regulatory element landscape, we hypothesized that Mod(mdg4)-N blocks subtelomeric enhancers at TAS-R to suppress *HeT-A* expression.

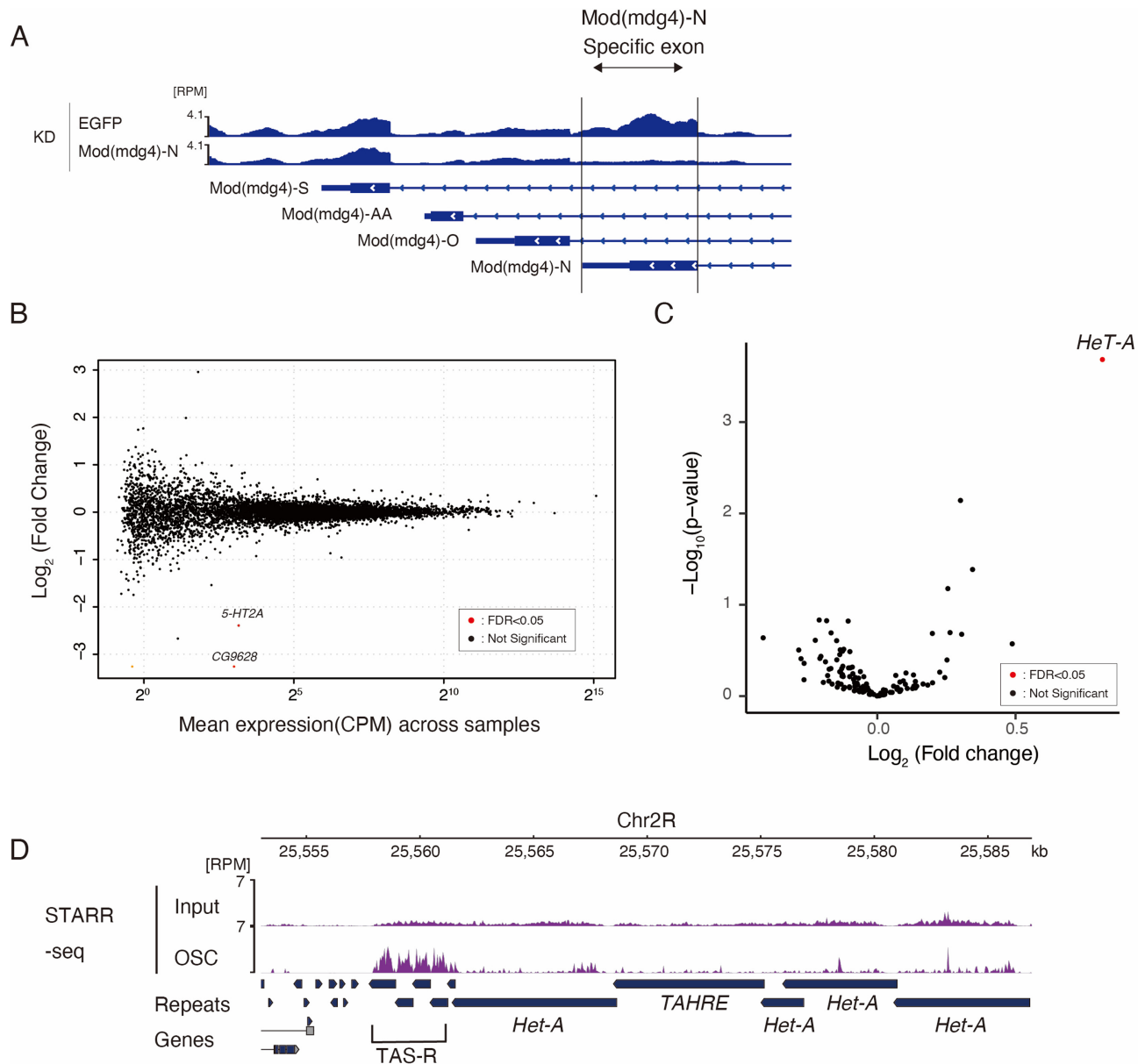
### Subtelomeric TAS-R repeats possess enhancer-blocking activity

We performed a live-imaging analysis of the enhancer-blocking activities of the Mod(mdg4)-N-binding sequences in *HeT-A* and TAS-R in living embryos, where Mod(mdg4) is highly expressed (85). In this imaging system, the *sna* shadow enhancer causes a transcriptional burst from the *Drosophila* synthetic core promoter (DSCP), a collection of



**Figure 3.** Mod(mdg4)-N associates with telomeric/subtelomeric regions in a variant-specific manner. **(A, B)** Metaplot and heatmap indicating Mod(mdg4)-N, -T, and control (anti-Ty1 ChIP against wild-type OSC) signal level within 1.0 kb of Mod(mdg4)-N (A) or -T (B) peaks. The heatmap is sorted by the intensity of the Mod(mdg4)-N (A) or -T (B) signals. The y-axes of the metaplots are normalized using the RPM. Note that some sites of the Mod(mdg4)-T peaks (B) display high background signals because Mod(mdg4)-T strongly accumulates on gypsy retrotransposons, which have many copies within the genome. **(C)** Distribution of Mod(mdg4) variant. Coverage panel showing ChIP-seq results for 2 × Ty1-tagged Mod(mdg4) variants (N, T) in the dm6 chr3L 8.50 Mb-8.66 Mb region. **(D)** Functional annotation of the peaks of Mod(mdg4)-N and -T identified using HOMER. The y-axis represents the number of peaks for each Mod(mdg4) variant. We used only highly confident (FDR < 1e-15) ChIP-seq peaks for the analysis to avoid ambiguous peaks. **(E)** Sequence logos showing clear enrichment around the ChIP-seq peaks of Mod(mdg4)-N and -T. These motifs were identified via de novo motif discovery analysis using MEME (Multiple Em for Motif Elicitation). **(F)** Coverage panel showing ChIP-seq of Mod(mdg4)-N and -T on the consensus sequence of *HeT-A* from rebase. Coverage of each ChIP-seq experiment is shown with blue. The y-axis is normalized with RPM. Below coverage panel of *HeT-A*, the reading frame of HeT-A Gag protein is shown in orange, and the black arrow indicates the *HeT-A* promoter. **(G)** Coverage panel showing ChIP-seq of Mod(mdg4)-N and -T near the subtelomere (TAS-R)/telomere region in dm6. The coverage of each ChIP-seq experiment is indicated in blue. The y-axis is normalized to RPM.

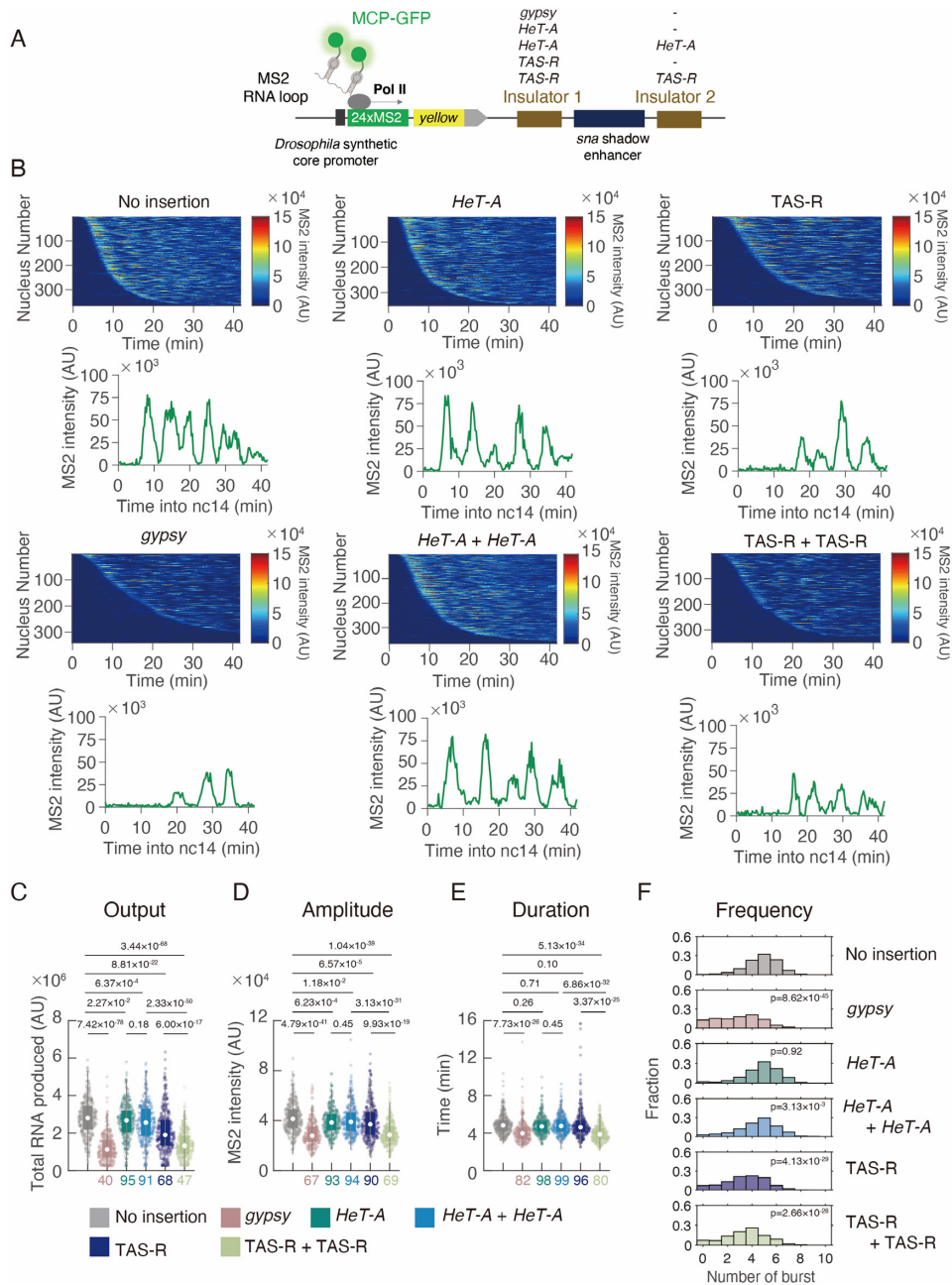




**Figure 4.** Specific regulation of *HeT-A* expression by *Mod(mdg4)-N*. (A) Coverage tracks of RNA-seq on EGFP KD or *Mod(mdg4)-N* KD at (*mdg4*)-locus (dm6 chr3R:21,370,172–21,373,552). The y-axis is normalized to RPM. The *Mod(mdg4)-N*-specific exon is highlighted with an arrow. Below the coverage, annotations of each *Mod(mdg4)* variant are indicated. (B) MA plot of TPM values ( $\log_{10}$  scale) for mRNA in the *Mod(mdg4)-N* KD versus EGFP KD OSC ( $n = 2$  for each condition). Differentially expressed genes (DEGs) are denoted by red dots. Note that *Mod(mdg4)* was not included in the DEGs because only *Mod(mdg4)* variant N was knocked down. (C) Volcano plot showing the differential expression of retrotransposons in *Mod(mdg4)-N* KD versus EGFP KD OSC. The x-axis is the  $\log_2$  ratio, and the y-axis is the  $\log_{10}$  ratio. Differentially expressed genes (DEGs) are denoted by red dots. (D) Coverage panel showing STARR-seq signal (purple) for OSC and its input near the subtelomere (TAS-R)/telomere region from dm6. The y-axis is normalized to RPM. These data are from the reanalysis of the STARR-seq results of Arnold *et al.* (2013).

core promoter motifs. The 24xMS2-yellow synthetic gene is transcribed from the DSCP under the control of the *sna* shadow enhancer. The MS2 coat protein (MCP) is a single-stranded RNA phage capsid protein that binds to the MS2 19-nucleotide RNA stem loop with high affinity (98), and the MCP-GFP fusion protein recognizes the MS2 repeats in this synthetic gene (Figure 5A, Supplemental Video 1–2). With this imaging system, the transcriptional dynamics and enhancer function can be precisely visualized (99–101). For the enhancer-blocking assay, we

tested *Mod(mdg4)-N*-binding TAS-R (520 bp) and *HeT-A* (999 bp) sequences (Supplementary Figure S6, Supplemental Table 2). We compared the TAS-R and *HeT-A* sequences with the gypsy insulator sequence, which strongly inhibits enhancer-promoter interactions via the gypsy insulator proteins (101–103). Both *HeT-A* and TAS-R sequences decreased the total output of transcription although the effect of the *HeT-A* sequence was significantly weaker than that of the TAS-R sequence (Figure 5B, C: decreased to 95% and 68% compared with control). With the *HeT-A* sequence



**Figure 5.** Subtelomeric TAS-R repeats possess enhancer-blocking activity. **(A)** Schematic representation of the yellow reporter gene containing the 155-bp *Drosophila* synthetic core promoter (DSCP), the 1.5-kb *sna* shadow enhancer, and 24  $\times$  MS2 RNA stem loops within the 5' UTR. Insulator candidate sequences were inserted on both sides of the *sna* shadow enhancer. The expression of this reporter gene was visualized using MCP-GFP proteins binding to MS2 RNA stem loops. **(B)** (Upper) MS2 trajectories for all analyzed nuclei. Each row represents the MS2 trajectory for a single nucleus. A total of 365 (no insertion), 340 (*gypsy*), 352 (*HeT-A*), 365 (*HeT-A + HeT-A*), 365 (*TAS-R*), and 349 (*TAS-R + TAS-R*) ventral-most nuclei, respectively, were analyzed from three independent embryos for the reporter genes with no insertion of insulators, *gypsy* insulator, single insertion of *HeT-A*, double insertion of *HeT-A* (*HeT-A + HeT-A*), single insertion of *TAS-R*, or double insertion of *TAS-R* (*TAS-R + TAS-R*). (Lower) Representative trajectory of transcriptional activity of the MS2 reporter gene with no insertion of insulators, *gypsy* insulator, single insertion of *HeT-A*, double insertion of *HeT-A* (*HeT-A + HeT-A*), single insertion of *TAS-R*, or double insertion of *TAS-R* (*TAS-R + TAS-R*). **(C–E)** Boxplots showing the distribution of the total output (C), burst amplitude (D), and burst duration (E). The box indicates the lower (25%) and upper (75%) quantiles, and the open circle indicates the median. Whiskers extend to the most extreme non-outlier data points. A total of 365 (no insertion), 340 (*gypsy*), 352 (*HeT-A*), 365 (*HeT-A + HeT-A*), 365 (*TAS-R*), and 349 (*TAS-R + TAS-R*) ventral-most nuclei, respectively, were analyzed from three independent embryos for reporter genes with no insertion of insulators, *gypsy* insulator, single insertion of *HeT-A*, double insertion of *HeT-A* (*HeT-A + HeT-A*), single insertion of *TAS-R*, or double insertion of *TAS-R* (*TAS-R + TAS-R*) from left to right. The median values relative to the control reporter are shown at the bottom. The p-values of the two-sided Wilcoxon rank-sum test are shown at the top. AU; arbitrary unit. **(F)** Histograms showing the distribution of the burst frequency during the nuclear cycle 14 stage. A total of 365 (no insertion), 340 (*gypsy*), 352 (*HeT-A*), 365 (*HeT-A + HeT-A*), 365 (*TAS-R*), and 349 (*TAS-R + TAS-R*) ventral-most nuclei, respectively, were analyzed from three independent embryos for reporter genes with no insertion of insulators, *gypsy* insulator, single insertion of *HeT-A*, double insertion of *HeT-A* (*HeT-A + HeT-A*), single insertion of *TAS-R*, or double insertion of *TAS-R* (*TAS-R + TAS-R*) from top to bottom. The p-values of the two-sided Wilcoxon rank-sum test are shown within the windows.

insertion, only amplitude per burst slightly decreased (Figure 5D–F). On the other hand, single insertion of TAS-R decreased both amplitude and frequency of the transcriptional burst (Figure 5D–F).

Considering that the Mod(mdg4)-N mutants displayed a 5–10-fold higher *HeT-A* expression in the ovary (Figure 2B), this enhancer-blocking activity of a single TAS-R or *HeT-A* was weaker than expected. Both TAS-R and *HeT-A* exist as tandem repeats in the *D. melanogaster* genome (20,29), and the pairing of insulators generally facilitates enhancer-blocking function (104). Therefore, we hypothesized that mimicking the *in vivo* tandem repeats of TAS-R and *HeT-A* may be important for assessing the enhancer-blocking activity of these sequences. To test this, we inserted TAS-R or *HeT-A* sequences on both sides of the *sna* shadow enhancer. With two copies of TAS-R (TAS-R + TAS-R), total output decreased considerably compared with wild-type and single insertion of TAS-R (Figure 5C), accompanied with significant decrease of both amplitude and duration of the transcriptional burst (Figure 5D, E). This result suggests that multiple repeats in the genome additively facilitate the enhancer-blocking activity of the TAS-R sequences. In contrast, two copies of *HeT-A* sequences bound by Mod(mdg4)-N (*HeT-A* + *HeT-A*) did not result in a significant difference in either the total output of transcription, amplitude, or duration of transcriptional burst, when compared to single *HeT-A* insertion (Figure 5C–E). These results indicate that Mod(mdg4)-N-binding TAS-R and *HeT-A* sequences are *bona fide* insulators while the effect of multiple repeats of insulator is specific to TAS-R sequences. Since enhancer-blocking function of *HeT-A* is weaker than that of TAS-R, and multiple repeats of TAS-R can result in stronger enhancer blocking activity, repetitive TAS-R has a major function in enhancer blocking activity, while *HeT-A* may assist that function.

To confirm the impact of the Mod(mdg4)-N-mediated enhancer-blocking function on TAS-R sequences, we performed luciferase assays upon loss of Mod(mdg4). Because the Mod(mdg4)-N mutant could not lay eggs and the Mod(mdg4) transcripts are maternally deposited (85), we used the OSC for further analysis. In this experiment, the TAS-R sequences were inserted on both sides of the *Tj* enhancer, and the downstream NanoLuc gene was transcribed from the DSCP. Firefly luciferase plasmids were used as internal controls (Supplementary Figure S7A). Using this reporter, we were able to determine that the insertion of the TAS-R repeats could regulate the expression of the NanoLuc reporter, consistent with live-imaging analysis (Supplementary Figure S7B). Furthermore, the KD of Mod(mdg4)-N resulted in de-silencing of the reporter activity (Supplementary Figure S7C, D), confirming that Mod(mdg4)-N was involved in the enhancer-blocking activity of TAS-R.

### RNA polymerase II is recruited to Mod(mdg4)-N binding sites

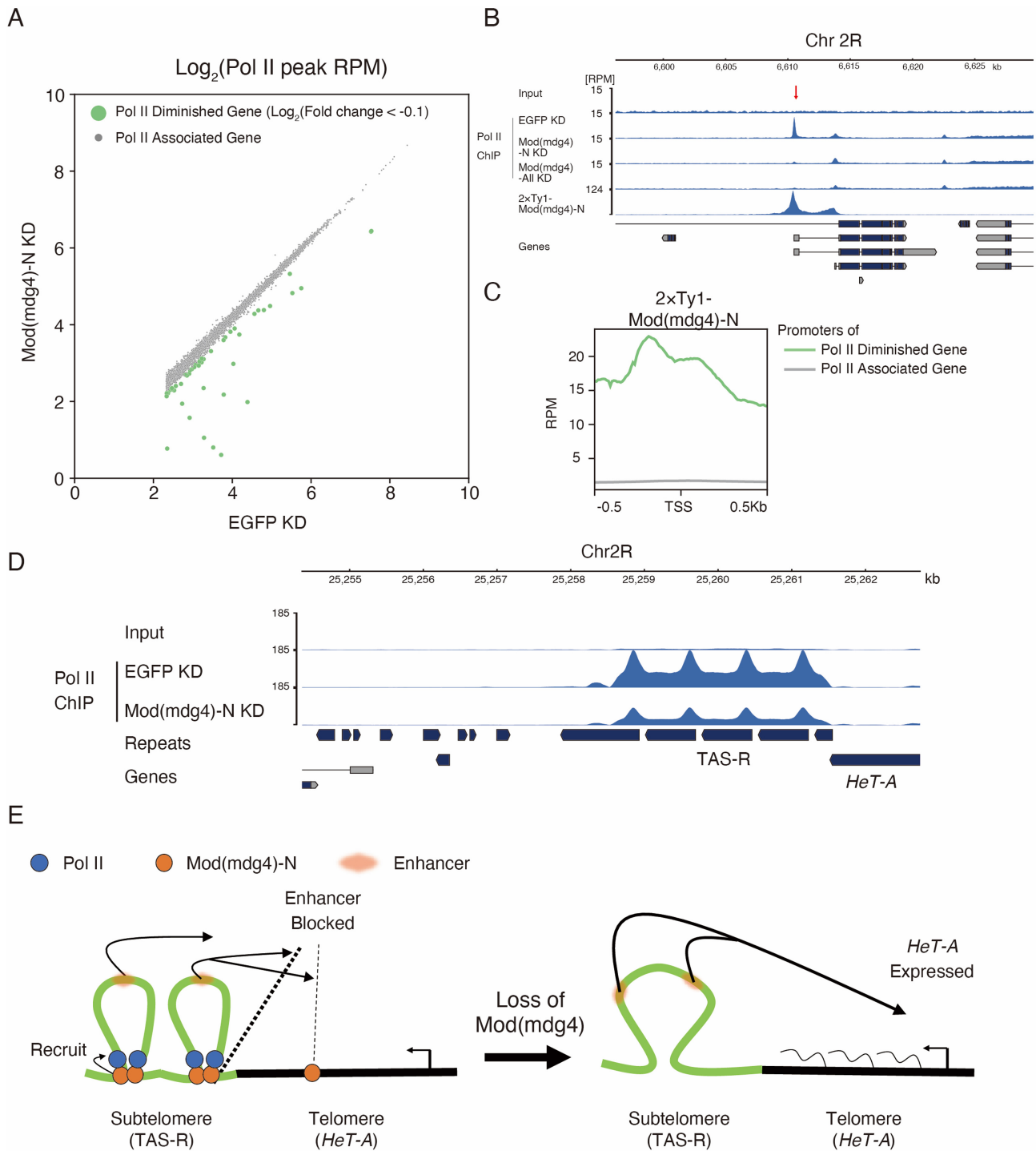
We further examined the molecular details of how Mod(mdg4)-N functions at its binding sites. Previous studies have demonstrated a link between RNA Pol II and insulation; some Pol II-enriched promoters function as

enhancer-blocking elements when these sequences are inserted between the enhancer and promoter (105,106), and promoters with strong Pol II accumulation physically interact with the insulator proteins (107). Consistent with these observations, approximately half of the Mod(mdg4) peaks were on promoters (Figure 3D). It is plausible that Mod(mdg4)-N acts as an insulator by regulating the accumulation of Pol II. To analyze the relationship between Mod(mdg4)-N and promoter-proximal pausing, we first re-analyzed public data for OSC Pol II ChIP (71) using the pausing index (Supplementary Figure S8A). The pausing index is the ratio of the Pol II ChIP-seq density in the region near the TSS ( $\pm 250$  bp) to that in the gene body (70). The higher this index, the greater the pausing observed at the promoter. The promoters with Mod(mdg4)-N had a significantly higher pausing index than all promoters (Supplementary Figure S8B), indicating that Mod(mdg4)-N tends to bind highly to Pol II-pausing promoters. Furthermore, TAS-R is similar to *Invader4* LTR; thus, it contains core promoter sequences, specifically Initiator (Inr) (108) and downstream promoter element (DPE) (109), according to the criteria described in previous research (110) (Supplementary Figure S9A). These results indicate the involvement of the Pol II machinery in the regulation of the subtelomeric enhancer-blocking function of Mod(mdg4)-N.

We next examined how the Pol II peak changes with the Mod(mdg4)-N KD using a Pol II ChIP-seq analysis. Mod(mdg4)-N bound only to selected promoters; therefore, global changes in the Pol II ChIP-seq signals were not observed with the Mod(mdg4)-N KD. However, some Pol II peaks were significantly diminished ( $\log_2(\text{fold change}) < -0.1$ ) upon Mod(mdg4)-N KD (Figure 6A, B). In contrast, almost no Pol II peaks became more pronounced upon Mod(mdg4)-N KD. We also observed that Mod(mdg4)-N ChIP-seq signals were highly enriched on these ‘Pol II Diminished Genes’ (Figure 6C), indicating that Mod(mdg4)-N had a Pol II-recruiting function at its binding site. Importantly, the TAS-R repeats at the subtelomeres were also associated with Pol II, and the KD of Mod(mdg4)-N resulted in a diminished accumulation of Pol II in this region (Figure 6D). In contrast, changes of Pol II accumulation were not observed at the *HeT-A* sites (Supplementary Figure S8C).

This observation implies that Pol II, regulated by Mod(mdg4)-N, may play a critical role in the enhancer-blocking activity. To test this, we investigated whether the loss of the TAS-R core promoter sequence attenuates its enhancer-blocking activity in the OSC. We designed a reporter in which the *Tj* enhancer acts on the DSCP, which transcribes EGFP:P2A:BlastR, and assessed the *Tj* enhancer-blocking activity of the TAS-R mutant sequences by the EGFP intensity (Supplementary Figure S9B). This reporter vector was integrated into chromatin using the PiggyBac system (Supplementary Figure S9C). To analyze the impact of the Pol II association on the enhancer-blocking activity, we compared two TAS-R sequence mutants for the enhancer reporter assay: the first mutant was a complete deletion of the core promoter sequences, Inr and DPE, from a TAS-R sequence; the second mutant was a replacement of the Inr sequence with a random sequence (Supplementary Figure S9A). Using this reporter assay, we observed the





**Figure 6.** RNA polymerase II is recruited to Mod(mdg4)-N binding sites including subtelomeric TAS-R repeats. **(A)** Scatter plot of maximum peak coverage between EGFP KD (x-axis) and Mod(mdg4)-N KD (y-axis). Both the x-axis and y-axis are shown with log<sub>2</sub> ratios. Peaks whose maximum values were under 5.0 at RPM in EGFP KD were removed from the analysis. Pol II-diminished genes (green) are defined as log<sub>2</sub>(fold changes) < -0.1, and other genes (Pol II-associated genes) whose Pol II level did not change with the Mod(mdg4)-N KD are shown with gray dots. **(B)** Coverage tracks showing examples of Mod(mdg4)-N localization and Pol II profile with EGFP KD, Mod(mdg4)-N KD, and Mod(mdg4)-all variants KD at the Pol II diminished gene locus (Chr2R:6.60Mb-6.63Mb). The lower panel indicates annotations of the transcripts from Flybase. The red arrow indicates the Pol II peak that diminished with KD of Mod(mdg4)-N or all variants. **(C)** The average profile showing enrichment of Mod(mdg4)-N ChIP-seq intensity around the Pol II peaks that diminished with the Mod(mdg4)-N KD (Pol II diminished genes; green) or the other genes (Pol II associated gene, gray). **(D)** Coverage tracks showing input and Pol II enrichment in the TAS-R regions with the EGFP KD and Mod(mdg4)-N KD. The lower panel indicates annotations of the transcripts from Flybase. **(E)** Summary of this research. Mod(mdg4)-N (orange) binds on both TAS-R and telomeric *HeT-A* retrotransposons. Mod(mdg4)-N represses enhancer activity at both TAS-R and *HeT-A*, although enhancer-blocking activity at TAS-R is stronger. Mod(mdg4)-N recruits Pol II (blue) on the subtelomeric TAS-R repeats, which assists in blocking the subtelomeric enhancer (red). Therefore, *HeT-A* is silenced in the normal state. Upon loss of Mod(mdg4)-N, enhancer-blocking activity is lost, leading to *HeT-A* expression.

enhancer-blocking activity of the two copies of TAS-R repeats (Supplementary Figure S9D), which is consistent with the live-imaging assay (Figure 5). Both the deletion and replacement of the TAS-R core promoter sequences resulted in an increased EGFP intensity compared with the wild-type TAS-R. This indicates that the enhancer-blocking activity weakened with the core promoter loss or replacement (Supplementary Figure S9D). Thus, Pol II recruitment to the TAS-R repeats can increase the enhancer-blocking activity, supporting the importance of Pol II recruitment to TAS-R by Mod(mdg4)-N. Altogether, we showed that a specific Mod(mdg4) variant represses the *HeT-A* retrotransposon by blocking enhancers in subtelomeric repeats, and the recruitment of Pol II can reinforce this function (Figure 6E).

## DISCUSSION

### Enhancer and insulator function of *Drosophila* subtelomeres

We have demonstrated Mod(mdg4)-N-mediated repression of *HeT-A* retrotransposon (Figures 1, 2), Mod(mdg4)-N binding at telomeric/subtelomeric regions and the existence of enhancer activity in TAS-R (Figures 3, 4), enhancer blocking activity of Mod(mdg4)-N-binding TAS-R and *HeT-A* sequences (Figure 5), and the importance of Pol II association in Mod(mdg4)-N-mediated enhancer-blocking (Figure 6). Based on these results, we propose a model in which the enhancer function of the TAS-R region is blocked by Mod(mdg4)-N, which in turn represses the expression of *HeT-A* at telomeres (Figure 6E).

We observed diminished accumulation of Pol II on TAS-R upon Mod(mdg4)-N depletion and deletion or mutation of core promoter sequences in TAS-R attenuated but did not completely abolish its insulator activity (Supplementary Figure S9). With this observation, we speculate that two mechanisms are involved in the enhancer-blocking function of Mod(mdg4)-N. The first is Pol II recruitment-dependent insulator activity (Figure 6, Supplementary Figure S9). In *Drosophila*, several insulator proteins exist, and some of these insulator proteins, including Mod(mdg4), are enriched near TSSs (89,111). Functionally, some of the insulator-binding promoters work as enhancer-blocking elements when these sequences are inserted between an enhancer and another promoter (105,106). Highly paused promoters are associated with insulator-binding sites (107), indicating that promoters can play an essential role in the insulator-mediated enhancer-blocking. Additionally, it has been proposed that regulation of chromatin conformation is important for enhancer-blocking activity (101,112). In fact, there is an example of the association of enhancer blocking activity with transcriptional regulation at promoters: M1BP is required for both Pol II recruitment and enhancer-blocking activity, and furthermore, transcriptional regulation by M1BP is coupled with chromatin conformation change (113,114). Thus, previous studies indicate a link between transcriptional machinery and enhancer-blocking activity. Taking these into account, Mod(mdg4)-N might cause chromatin conformation change mediated by transcriptional regulation at promoter sequences, which leads to enhancer-blocking activity at TAS-R. The second is

the biochemical nature of insulator proteins. Mod(mdg4) has a BTB/POZ domain at its N-terminal common region, which can form multimers with both itself and a BTB/POZ domain of Trl (Trithorax-like, a.k.a. GAGA Factor) (115,116). Therefore, the homophilic nature of the Mod(mdg4) BTB/POZ domain contributes to enhancer-blocking by dynamic loop formation between different Mod(mdg4)-N associated sites (101,104).

Overall, our results indicate that Mod(mdg4)-N is necessary for *HeT-A* repression by blocking subtelomeric enhancers. Interestingly, the enhancer-like function of subtelomeres is also reported in human, and transcription of TERRA (telomeric repeat-containing RNA) is promoted by insulator protein CTCF despite telomeric sequences in human and fly being completely different (14,117). This observation shows an evolutionarily conserved principle of enhancer-like functions at subtelomeres being modulated by insulator proteins. What drives this subtelomeric convergent evolution is still an open question.

### Specificity of Mod(mdg4)-N mediated enhancer-blocking activity at subtelomeric TAS-R and *HeT-A* repeats

We showed that Mod(mdg4)-N regulates gene expression almost exclusively at subtelomere/telomere regions (Figure 4B-C). This result suggests that Mod(mdg4)-N does not regulate its associated genes, but functions in a context-dependent manner, limiting its impact to telomeric regions. Also, since TAS-R repeats are the source of enhancer-blocking activity, the role of Mod(mdg4)-N in regulating telomeric transcription is limited to the telomeric sites with TAS-R repeats.

For Mod(mdg4)-N to block the enhancer, it is essential to have Mod(mdg4) located at certain genomic regions containing both enhancer and downstream promoter. We also found that Mod(mdg4)-N is responsible for recruiting Pol II at only limited genomic regions (Figure 6A). Additionally, we showed that Mod(mdg4)-N-mediated Pol II association within TAS-R plays an important role in enhancing the regulation of downstream genes (Supplementary Figure S9). Therefore, the requirement of Pol II association sites along with enhancer sites limits the enhancer-blocking activity of Mod(mdg4)-N to certain genomic locations. Another important requirement is the tandem-repeating nature of the Mod(mdg4)-N association sites. We demonstrated that multiple copies of TAS-R sites, capable of recruiting Mod(mdg4)-N, can induce an insulator function significantly stronger than that of the single copy of TAS-R (Figure 5B-F).

In summary, the key factors for inducing a strong enhancer-blocking effect, specifically at the TAS-R regions, are both the Pol II association sites within the TAS-R, capable of recruiting Pol II with the Mod(mdg4)-N association, and the highly repetitive nature of the TAS-R sequences. Overall, these results suggest that regulation on TAS-R is the major mechanism of Mod(mdg4)-mediated *HeT-A* expression, and Mod(mdg4)-N-binding *HeT-A* sequences are functional but have a minor contribution to *HeT-A* transcriptional regulation. This enhancer-blocking activity of Mod(mdg4)-N is a novel mechanism regulating *HeT-A* expression, ensuring genomic stability in *Drosophila*.

### Transcriptional regulation of telomeric retrotransposons

We showed that Mod(mdg4)-N represses *HeT-A* both *in vivo* and *in vitro* (Figures 1,2). Although Mod(mdg4) is ubiquitously expressed in the ovary (Figure 2A–E), Mod(mdg4)-N is especially important for the regulation of *HeT-A* in follicle cells where the PIWI–piRNA pathway does not target *HeT-A*. This result indicates that the Mod(mdg4)-mediated repression machinery is a compensatory mechanism for telomeric retrotransposon regulation. The PIWI–piRNA pathway directs heterochromatin formation on telomeres, and dysfunction of this pathway leads to a dramatic increase in *HeT-A* expression in the germline (38–41). Therefore, it is likely that even without Mod(mdg4)-N, the PIWI–piRNA pathway is strong enough to repress *HeT-A* expression in germline cells. The chromatin states of subtelomeres and telomeres are reported to differ among cell types, especially between germ cells and somatic cells (118). It is also possible that different chromatin states at the subtelomeres and telomeres prevent the binding of Mod(mdg4)-N in certain cell types.

We showed that subtelomeric enhancer activity is regulated by Mod(mdg4)-N associated with the subtelomeric region. This complex regulation may be due to the dynamic regulation of *HeT-A* expression in somatic cells, to enable expression of *HeT-A* under specific biological conditions. Also, this study highlights the transcriptional regulation of *HeT-A* associated with TAS-R, which harbors enhancer activity. *HeT-A* associated with TAS-L, without enhancer activity or Mod(mdg4) association, may be regulated by another mechanism. Multiple mechanisms regulating *HeT-A* are reported in a cell-type specific manner (33,36–46,119,120). We believe that future studies will elucidate interactions between these regulation mechanisms, and the biological significance of *HeT-A* being regulated by different mechanisms.

### DATA AVAILABILITY

GSE (ChIP-seq, RNA-seq): GSE176196.

GitHub (python Script): <https://doi.org/10.5281/zenodo.7232511>.

FACS rawdata: <http://flowrepository.org/id/FR-FCM-Z5Y3>.

### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

### ACKNOWLEDGEMENTS

We thank all Siomi lab members for experimental support and fruitful discussions, especially Akihiko Sakashita and Hirotsugu Ishizu for critical reading of the manuscript. Pol II antibody used in this study was a kind gift from Hiroshi Kimura (Tokyo Institute of Technology). siRNAs of each Mod(mdg4) variant were designed by Hidetoshi Tahara (Advanced Animal Model Support Platform). We thank Katrina Woolcock (Life Science Editors) for editing services.

*Author contributions:* C.T. designed and performed most of the experiments and analyzed the data. M.Y. and T.F.

performed live-imaging for the enhancer blocking assay in embryo. A.S. performed immunofluorescence experiment and fly fertility assay. Y.W.I. established antibody against Mod(mdg4) common region. S.K. and K.S. generated transgenic Mod(mdg4)-N mutant fly. Y.W.I., H.S. conceived this study. C.T., Y.W.I., H.S. wrote the paper with input from the other authors.

### FUNDING

Japan Science and Technology Agency [JPMJSP2123 to C.T., JPMJPR20E2 to Y.W.I., JPMJAX211J to M.Y.]; Japan Society of the Promotion of Science [22J12769 to C.T., 20K15710 to M.Y., 19H03154, 20H05357, 21H05742 to T.F., 16H06276, 19H05753 to H.S., 22H02547, 19H05268, 18H02421 to Y.W.I.]; Koyo Sakaguchi Memorial Keio University Medical Science Fund and The Keio University Doctorate Student Grant-in-Aid Program from Ushioda Memorial Fund (to C.T.); Sumitomo Foundation; Senri Life Science Foundation; Mitsubishi Foundation (to F.T.); Takeda Science Foundation (to F.T. and Y.W.I.).

*Conflict of interest statement.* None declared.

### REFERENCES

- De Lange, T. (2018) Shelterin-Mediated telomere protection. *Annu. Rev. Genet.*, **52**, 223–247.
- Greider, C.W. and Blackburn, E.H. (1985) Identification of a specific telomere terminal transferase activity in tetrahymena extracts. *Cell*, **43**, 405–413.
- Schoeftner, S. and Blasco, M.A. (2009) A ‘higher order’ of telomere regulation: telomere heterochromatin and telomeric RNAs. *EMBO J.*, **28**, 2323–2336.
- Baumann, P. and Cech, T.R. (2001) Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science*, **292**, 1171–1175.
- Ferreira, M.G. and Cooper, J.P. (2001) The fission yeast taz1 protein protects chromosomes from ku-dependent End-to-End fusions. *Mol. Cell*, **7**, 55–63.
- Nanavaty, V., Sandhu, R., Jehi, S.E., Pandya, U.M. and Li, B. (2017) Trypanosoma brucei RAP1 maintains telomere and subtelomere integrity by suppressing TERRA and telomeric RNA:DNA hybrids. *Nucleic Acids Res.*, **45**, 5785–5796.
- Pardo, B. and Marcand, S. (2005) Rap1 prevents telomere fusions by nonhomologous end joining. *EMBO J.*, **24**, 3117–3127.
- Song, X., Leehy, K., Warrington, R.T., Lamb, J.C., Surovtseva, Y.V. and Shippen, D.E. (2008) STN1 protects chromosome ends in *arabidopsisthaliana*. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 19815–19820.
- Van Steensel, B., Smogorzewska, A. and De Lange, T. (1998) TRF2 protects human telomeres from End-to-End fusions. *Cell*, **92**, 401–413.
- Azzalin, C.M., Reichenbach, P., Khoriauli, L., Giulotto, E. and Lingner, J. (2007) Telomeric repeat-containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science*, **318**, 798–801.
- Phillips-Cremins, J.E. and Corces, V.G. (2013) Chromatin insulators: linking genome organization to cellular function. *Mol. Cell*, **50**, 461–474.
- Beishline, K., Vladimirova, O., Tutton, S., Wang, Z., Deng, Z. and Lieberman, P.M. (2017) CTCF driven TERRA transcription facilitates completion of telomere DNA replication. *Nat. Commun.*, **8**, 2114.
- Stong, N., Deng, Z., Gupta, R., Hu, S., Paul, S., Weiner, A.K., Eichler, E.E., Graves, T., Fronick, C.C., Courtney, L. et al. (2014) Subtelomeric CTCF and cohesin binding site organization using improved subtelomere assemblies and a novel annotation pipeline. *Genome Res*, **24**, 1039–1050.
- Deng, Z., Wang, Z., Stong, N., Plasschaert, R., Moczan, A., Chen, H.-S., Hu, S., Wikramasinghe, P., Davuluri, R.V.,



- Bartolomei, M.S. *et al.* (2012) A role for CTCF and cohesin in subtelomere chromatin organization, TERRA transcription, and telomere end protection. *EMBO J.*, **31**, 4165–4178.
15. Osanai-Futahashi, M. and Fujiwara, H. (2011) Coevolution of telomeric repeats and telomeric repeat-specific Non-LTR retrotransposons in insects. *Mol. Biol. Evol.*, **28**, 2983–2986.
  16. Okazaki, S., Ishikawa, H. and Fujiwara, H. (1995) Structural analysis of TRAS1, a novel family of telomeric repeat-associated retrotransposons in the silkworm, *bombyx mori*. *Mol. Cell. Biol.*, **15**, 4545–4552.
  17. Biessmann, H., Mason, J.M., Ferry, K., d'Hulst, M., Valgeirsdottir, K., Traverse, K.L. and Pardue, M.L. (1990) Addition of telomere-associated HeT DNA sequences “heals” broken chromosome ends in *Drosophila*. *Cell*, **61**, 663–673.
  18. Levis, R.W., Ganesan, R., Houtchens, K., Tolar, L.A. and Sheen, F.M. (1993) Transposons in place of telomeric repeats at a *Drosophila* telomere. *Cell*, **75**, 1083–1093.
  19. Abad, J.P., De Pablos, B., Osogawa, K., De Jong, P.J., Martín-Gallardo, A. and Villasante, A. (2004) TAHRE, a novel telomeric retrotransposon from *Drosophila melanogaster*, reveals the origin of *Drosophila* telomeres. *Mol. Biol. Evol.*, **21**, 1620–1624.
  20. Pardue, M.L. and Debaryshe, P.G. (2002) In: Craig, N.L., Craigie, R., Gellert, M. and Lambowitz, A.M. (eds). *Mobile DNA II*. ASM Press, Washington, DC, pp. 870–890.
  21. McGurk, M.P., Dion-Côté, A.-M. and Barbash, D.A. (2021) Rapid evolution at the *Drosophila* telomere: transposable element dynamics at an intrinsically unstable locus. *Genet.*, **217**, iyaa027.
  22. George, J.A., Debaryshe, P.G., Traverse, K.L., Celniker, S.E. and Pardue, M.L. (2006) Genomic organization of the *Drosophila* telomere retrotransposable elements. *Genome Research*, **16**, 1231–1240.
  23. Gao, G., Walsler, J.-C., Beaucher, M.L., Morciano, P., Wesolowska, N., Chen, J. and Rong, Y.S. (2010) HipHop interacts with HOAP and HPI to protect *Drosophila* telomeres in a sequence-independent manner. *EMBO J.*, **29**, 819–829.
  24. Raffa, G.D., Siriaco, G., Cugusi, S., Ciapponi, L., Cenci, G., Wojcik, E. and Gatti, M. (2009) The *Drosophila* modigliani (moi) gene encodes a HOAP-interacting protein required for telomere protection. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 2271–2276.
  25. Komonyi, O., Schauer, T., Papai, G., Deak, P. and Boros, I.M. (2009) A product of the bicistronic *Drosophila melanogaster* gene CG31241, which also encodes a trimethylguanosine synthase, plays a role in telomere protection. *J. Cell Sci.*, **122**, 769–774.
  26. Saint-Leandre, B., Christopher, C. and Levine, M.T. (2020) Adaptive evolution of an essential telomere protein restricts telomeric retrotransposons. *Elife*, **9**, e60987.
  27. Lee, Y.C.G., Leek, C. and Levine, M.T. (2016) Recurrent innovation at genes required for telomere integrity in *Drosophila*. *Mol. Biol. Evol.*, **34**, 467–482.
  28. Cenci, G., Siriaco, G., Raffa, G.D., Kellum, R. and Gatti, M. (2003) The *Drosophila* HOAP protein is required for telomere capping. *Nat. Cell Biol.*, **5**, 82–84.
  29. Karpen, G.H. and Spradling, A.C. (1992) Analysis of subtelomeric heterochromatin in the *Drosophila* minichromosome Dp1187 by single p element insertional mutagenesis. *Genetics*, **132**, 737–753.
  30. Walter, M.F., Jang, C., Kasravi, B., Donath, J., Mechler, B.M., Mason, J.M. and Biessmann, H. (1995) DNA organization and polymorphism of a wild-type *Drosophila* telomere region. *Chromosoma*, **104**, 229–241.
  31. Bergman, C.M., Quesneville, H., Anxolabéhère, D. and Ashburner, M. (2006) Recurrent insertion and duplication generate networks of transposable element sequences in the *Drosophila melanogaster* genome. *Genome Biol.*, **7**, R112.
  32. Asif-Laidin, A., Delmarre, V., Laurentie, J., Miller, W.J., Ronsseray, S. and Teyssset, L. (2017) Short and long-term evolutionary dynamics of subtelomeric piRNA clusters in *Drosophila*. *DNA Res.*, **24**, 459–472.
  33. Morgunova, V., Akulenko, N., Radion, E., Olovnikov, I., Abramov, Y., Olenina, L.V., Shpiz, S., Kopytova, D.V., Georgieva, S.G. and Kalmykova, A. (2015) Telomeric repeat silencing in germ cells is essential for early development in *Drosophila*. *Nucleic Acids Res.*, **43**, 8762–8773.
  34. Morgunova, V., Kordyukova, M., Mikhaleva, E.A., Butenko, I., Pobeguts, O.V. and Kalmykova, A. (2021) Loss of telomere silencing is accompanied by dysfunction of polo kinase and centrosomes during *Drosophila* oogenesis and early development. *PLoS One*, **16**, e0258156.
  35. Saint-Leandre, B., Nguyen, S.C. and Levine, M.T. (2019) Diversification and collapse of a telomere elongation mechanism. *Genome Res.*, **29**, 920–931.
  36. Cacchione, S., Cenci, G. and Raffa, G.D. (2020) Silence at the end: how *Drosophila* regulates expression and transposition of telomeric retroelements. *J. Mol. Biol.*, **432**, 4305–4321.
  37. Iwasaki, Y.W., Siomi, M.C. and Siomi, H. (2015) PIWI-Interacting RNA: its biogenesis and functions. *Annu. Rev. Biochem.*, **84**, 405–433.
  38. Savitsky, M., Kwon, D., Georgiev, P., Kalmykova, A. and Gvozdev, V. (2006) Telomere elongation is under the control of the RNAi-based mechanism in the *Drosophila* germline. *Genes & Development*, **20**, 345–354.
  39. Senti, K.-A., Jurczak, D., Sachidanandam, R. and Brennecke, J. (2015) piRNA-guided slicing of transposon transcripts enforces their transcriptional silencing via specifying the nuclear piRNA repertoire. *Genes Dev.*, **29**, 1747–1762.
  40. Teo, R.Y.W., Anand, A., Sridhar, V., Okamura, K. and Kai, T. (2018) Heterochromatin protein 1a functions for piRNA biogenesis predominantly from pericentric and telomeric regions in *Drosophila*. *Nat. Commun.*, **9**, 1735.
  41. Shpiz, S., Olovnikov, I., Sergeeva, A., Lavrov, S., Abramov, Y., Savitsky, M. and Kalmykova, A. (2011) Mechanism of the piRNA-mediated silencing of *Drosophila* telomeric retrotransposons. *Nucleic Acids Res.*, **39**, 8703–8711.
  42. Savitsky, M., Kravchuk, O., Melnikova, L. and Georgiev, P. (2002) Heterochromatin protein 1 is involved in control of telomere elongation in *Drosophila melanogaster*. *Mol. Cell. Biol.*, **22**, 3204–3218.
  43. Raffa, G.D., Cenci, G., Siriaco, G., Goldberg, M.L. and Gatti, M. (2005) The putative *Drosophila* transcription factor woc is required to prevent telomeric fusions. *Mol. Cell*, **20**, 821–831.
  44. Satyaki, P.R.V., Cuykendall, T.N., Wei, K.H.C., Brideau, N.J., Kwak, H., Aruna, S., Ferree, P.M., Ji, S. and Barbash, D.A. (2014) The hmr and lhr hybrid incompatibility genes suppress a broad range of heterochromatic repeats. *PLoS Genet.*, **10**, e1004240.
  45. Font-Burgada, J., Rossell, D., Auer, H. and Azorin, F. (2008) *Drosophila* HP1c isoform interacts with the zinc-finger proteins WOC and Relative-of-WOC to regulate gene expression. *Genes Dev.*, **22**, 3007–3023.
  46. Zinshteyn, D. and Barbash, D.A. (2022) Stonewall prevents expression of ectopic genes in the ovary and accumulates at insulator elements in *D. melanogaster*. *PLoS Genet.*, **18**, e1010110.
  47. Kordyukova, M., Olovnikov, I. and Kalmykova, A. (2018) Transposon control mechanisms in telomere biology. *Curr. Opin. Genet. Dev.*, **49**, 56–62.
  48. George, J.A. and Pardue, M.-L. (2003) The promoter of the heterochromatic *Drosophila* telomeric retrotransposon, HeT-A, is active when moved into euchromatic locations. *Genetics*, **163**, 625–635.
  49. Zhang, L., Beaucher, M., Cheng, Y. and Rong, Y.S. (2014) Coordination of transposon expression with DNA replication in the targeting of telomeric retrotransposons in *Drosophila*. *EMBO J.*, **33**, 1148–1158.
  50. Zou, Z., Ohta, T., Miura, F. and Oki, S. (2022) ChIP-Atlas 2021 update: a data-mining suite for exploring epigenomic landscapes by fully integrating chip-seq, ATAC-seq and Bisulfite-seq data. *Nucleic Acids Res.*, **50**, W175–W182.
  51. Oki, S., Ohta, T., Shioi, G., Hatanaka, H., Ogasawara, O., Okuda, Y., Kawaji, H., Nakaki, R., Sese, J. and Meno, C. (2018) ChIP-Atlas: a data-mining suite powered by full integration of public chip-seq data. *EMBO Rep.*, **19**, e46255.
  52. Niki, Y., Yamaguchi, T. and Mahowald, A.P. (2006) Establishment of stable cell lines of *Drosophila* germ-line stem cells. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 16325–16330.
  53. Saito, K., Inagaki, S., Mituyama, T., Kawamura, Y., Ono, Y., Sakota, E., Kotani, H., Asai, K., Siomi, H. and Siomi, M.C. (2009) A regulatory circuit for piwi by the large maf gene traffic jam in *Drosophila*. *Nature*, **461**, 1296–1299.
  54. Kondo, S. and Ueda, R. (2013) Highly improved gene targeting by germline-specific cas9 expression in *Drosophila*. *Genetics*, **195**, 715–721.

55. Yokoshi, M., Segawa, K. and Fukaya, T. (2020) Visualizing the role of boundary elements in enhancer-promoter communication. *Mol. Cell*, **78**, 224–235.
56. Yokoshi, M., Kawasaki, K., Cambón, M. and Fukaya, T. (2021) Dynamic modulation of enhancer responsiveness by core promoter elements in living *Drosophila* embryos. *Nucleic Acids Res.* **50**, 92–107.
57. Greenfield, E.A. (2018) Polyethylene glycol fusion for hybridoma production. *Cold Spring Harb. Protoc.*, **2018**, <https://doi.org/10.1101/pdb.prot103176>.
58. Trcek, T., Lionnet, T., Shroff, H. and Lehmann, R. (2017) mRNA quantification using single-molecule FISH in *Drosophila* embryos. *Nature Protocols*, **12**, 1326–1348.
59. Takeuchi, C., Murano, K., Ishikawa, M., Okano, H. and Iwasaki, W.Y. (2022) Generation of stable *Drosophila* ovarian somatic cell lines using piggyBac system. *Methods Mol Biol.* **2509**, 143–153.
60. Venken, K.J.T., He, Y., Hoskins, R.A. and Bellen, H.J. (2006) P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in *D. melanogaster*. *Science*, **314**, 1747–1751.
61. Bischof, J., Maeda, R.K., Hediger, M., Karch, F. and Basler, K. (2007) An optimized transgenesis system for *Drosophila* using germ-line-specific  $\phi$ C31 integrases. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 3312–3317.
62. Ringrose, L. (2009) Transgenesis in *Drosophila melanogaster*. *Methods Mol Biol.*, **561**, 3–19.
63. Stasevich, T.J., Hayashi-Takanaka, Y., Sato, Y., Maehara, K., Ohkawa, Y., Sakata-Sogawa, K., Tokunaga, M., Nagase, T., Nozaki, N., McNally, J.G. *et al.* (2014) Regulation of RNA polymerase II activation by histone acetylation in single living cells. *Nature*, **516**, 272–275.
64. Haring, M., Offermann, S., Danker, T., Horst, I., Peterhansel, C. and Stam, M. (2007) Chromatin immunoprecipitation: optimization, quantitative analysis and data normalization. *Plant Methods*, **3**, 11.
65. Batki, J., Schnabl, J., Wang, J., Handler, D., Andreev, V.I., Stieger, C.E., Novatchkova, M., Lampersberger, L., Kauneckaitė, K., Xie, W. *et al.* (2019) The nascent RNA binding complex SFiNX licenses piRNA-guided heterochromatin formation. *Nat. Struct. Mol. Biol.*, **26**, 720–731.
66. Martin, Robert M., Rino, J., Carvalho, C., Kirchhausen, T. and Carmo-Fonseca, M. (2013) Live-Cell visualization of Pre-mRNA splicing with single-molecule sensitivity. *Cell Reports*, **4**, 1144–1155.
67. Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*, **17**, 10–12.
68. Machanick, P. and Bailey, T.L. (2011) MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics*, **27**, 1696–1697.
69. Lopez-Delisle, L., Rabbani, L., Wolff, J., Bhardwaj, V., Backofen, R., Grüning, B., Ramírez, F. and Manke, T. (2020) pyGenomeTracks: reproducible plots for multivariate genomic data sets. *Bioinformatics*, **37**, 422–423.
70. Gilchrist, D.A., Dos Santos, G., Fargo, D.C., Xie, B., Gao, Y., Li, L. and Adelman, K. (2010) Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. *Cell*, **143**, 540–551.
71. Sienski, G., Dönertas, D. and Brennecke, J. (2012) Transcriptional silencing of transposons by piwi and maelstrom and its impact on chromatin state and gene expression. *Cell*, **151**, 964–980.
72. Huang, D.W., Sherman, B.T. and Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, **4**, 44–57.
73. Jevitt, A., Chatterjee, D., Xie, G., Wang, X.-F., Otwell, T., Huang, Y.-C. and Deng, W.-M. (2020) A single-cell atlas of adult *Drosophila* ovary identifies transcriptional programs and somatic cell lineage regulating oogenesis. *PLoS Biol.*, **18**, e3000538.
74. Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J. *et al.* (2017) Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.*, **8**, 14049.
75. Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Khurana, J.S., Zhang, F., Schultz, N., Koppetsch, B.S., Nowosielska, A. *et al.* (2009) The *Drosophila* HPI homolog rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell*, **138**, 1137–1149.
76. Iwasaki, Y.W., Murano, K., Ishizu, H., Shibuya, A., Iyoda, Y., Siomi, M.C., Siomi, H. and Saito, K. (2016) Piwi modulates chromatin accessibility by regulating multiple factors including histone H1 to repress transposons. *Mol. Cell*, **63**, 408–419.
77. Lu, X., Wontakal, S.N., Emelyanov, A.V., Morcillo, P., Konev, A.Y., Fyodorov, D.V. and Skoultschi, A.I. (2009) Linker histone H1 is essential for *Drosophila* development, the establishment of pericentric heterochromatin, and a normal polytene chromosome structure. *Genes Dev.*, **23**, 452–465.
78. Vujatovic, O., Zaragoza, K., Vaquero, A., Reina, O., Bernués, J. and Azorín, F. (2012) *Drosophila melanogaster* linker histone dH1 is required for transposon silencing and to preserve genome integrity. *Nucleic Acids Res.*, **40**, 5402–5414.
79. Mongelard, F., Labrador, M., Baxter, E.M., Gerasimova, T.I. and Corces, V.G. (2002) Trans-splicing as a novel mechanism to explain interallelic complementation in *Drosophila*. *Genetics*, **160**, 1481–1487.
80. Thomas, S.E., Soltani-Bejnood, M., Roth, P., Dorn, R., Logsdon, J.M. Jr and McKee, B.D. (2005) Identification of two proteins required for conjunction and regular segregation of achiasmata homologs in *Drosophila* male meiosis. *Cell*, **123**, 555–568.
81. Gdula, D.A., Gerasimova, T.I. and Corces, V.G. (1996) Genetic and molecular analysis of the gypsy chromatin insulator of *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 9378–9383.
82. Gerasimova, T.I., Gdula, D.A., Gerasimov, D.V., Simonova, O. and Corces, V.G. (1995) A *Drosophila* protein that imparts directionality on a chromatin insulator is an enhancer of position-effect variegation. *Cell*, **82**, 587–597.
83. Cai, H. and Levine, M. (1995) Modulation of enhancer-promoter interactions by insulators in the *Drosophila* embryo. *Nature*, **376**, 533–536.
84. Sun, M.S., Weber, J., Blattner, A.C., Chaurasia, S. and Lehner, C.F. (2019) MNM and SNM maintain but do not establish achiasmata homolog conjunction during *Drosophila* male meiosis. *PLoS Genet.*, **15**, e1008162.
85. Büchner, K., Roth, P., Schotta, G., Krauss, V., Saumweber, H., Reuter, G. and Dorn, R. (2000) Genetic and molecular complexity of the position effect variegation modifier mod(mdg4) in *Drosophila*. *Genetics*, **155**, 141–157.
86. Kong, J., Han, H., Bergalet, J., Bouvrette, L.P.B., Hernández, G., Moon, N.-S., Vali, H., Lécuyer, É. and Lasko, P. (2019) A ribosomal protein S5 isoform is essential for oogenesis and interacts with distinct RNAs in *Drosophila melanogaster*. *Scientific Rep.*, **9**, 13779.
87. Jang, S., Lee, J., Mathews, J., Ruess, K., Williford, A.O., Rangan, P., Betrán, E. and Buszczak, M. (2021) The *Drosophila* ribosome protein S5 paralogs rps5b promotes germ cell and follicle cell differentiation during oogenesis. *Development*, **148**, dev199511.
88. Brinkman, A.B. and Stunnenberg, H.G. (2009) In: Ferguson-Smith, A.C., Greally, J.M. and Martienssen, R.A. (eds). *Epigenomics*. Springer, Dordrecht, pp. 438.
89. Nègre, N., Brown, C.D., Shah, P.K., Kheradpour, P., Morrison, C.A., Henikoff, J.G., Feng, X., Ahmad, K., Russell, S., White, R.A.H. *et al.* (2010) A comprehensive map of insulator elements for the *Drosophila* genome. *PLoS Genet.*, **6**, e1000814.
90. Jain, D., Baldi, S., Zabel, A., Straub, T. and Becker, P.B. (2015) Active promoters give rise to false positive ‘Phantom peaks’ in chip-seq experiments. *Nucleic Acids Res.*, **43**, 6959–6968.
91. Dorn, R. and Krauss, V. (2003) The modifier of mdg4 locus in *Drosophila*: functional complexity is resolved by trans splicing. *Genetics*, **117**, 165–177.
92. Beaster-Jones, L. and Okkema, P.G. (2004) DNA binding and in vivo function of *C.elegans* PEB-1 require a conserved FLYWCH motif. *J. Mol. Biol.*, **339**, 695–706.
93. Melnikova, L., Kostyuchenko, M., Molodina, V., Parshikov, A., Georgiev, P. and Golovnin, A. (2017) Multiple interactions are involved in a highly specific association of the Mod(mdg4)-67.2 isoform with the su(hw) sites in *Drosophila*. *Open Biol.*, **7**, 170150.
94. Shazman, S., Lee, H., Socol, Y., Mann, R.S. and Honig, B. (2013) OnTheFly: a database of *Drosophila melanogaster* transcription factors and their binding sites. *Nucleic Acids Res.*, **42**, D167–D171.
95. Cai, H.N. and Levine, M. (1997) The gypsy insulator can function as a promoter-specific silencer in the *Drosophila* embryo. *EMBO J.*, **16**, 1732–1741.
96. Arnold, C.D., Gerlach, D., Stelzer, C., Boryń, Ł.M., Rath, M. and Stark, A. (2013) Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science*, **339**, 1074–1077.

97. Thompson,P., Macfarlan,T. and Lorincz,M. (2016) Long terminal repeats: from parasitic elements to building blocks of the transcriptional regulatory repertoire. *Mol. Cell*, **62**, 766–776.
98. Bertrand,E., Chartrand,P., Schaefer,M., Shenoy,S.M., Singer,R.H. and Long,R.M. (1998) Localization of ASH1 mRNA particles in living yeast. *Mol. Cell*, **2**, 437–445.
99. Lucas,T., Ferraro,T., Roelens,B., Jose,Aleksandra, Coppey,M. and Dostatni,N. (2013) Live imaging of bicoid-dependent transcription in Drosophila embryos. *Curr. Biol.*, **23**, 2135–2139.
100. Garcia,H.G., Tikhonov,M., Lin,A. and Gregor,T. (2013) Quantitative imaging of transcription in living Drosophila embryos links polymerase activity to patterning. *Curr. Biol.*, **23**, 2140–2145.
101. Fukaya,T., Lim,B. and Levine,M. (2016) Enhancer control of transcriptional bursting. *Cell*, **166**, 358–368.
102. Bender,W., Akam,M., Karch,F., Beachy,P.A., Peifer,M., Spierer,P., Lewis,E.B. and Hogness,D.S. (1983) Molecular genetics of the bithorax complex in Drosophila melanogaster. *Science*, **221**, 23–29.
103. Ghosh,D., Gerasimova,T.I. and Corces,V.G. (2001) Interactions between the su(hw) and Mod(mdg4) proteins required for gypsy insulator function. *EMBO J.*, **20**, 2518–2527.
104. Kyrchanova,O., Maksimenko,O., Stakhov,V., Ivlieva,T., Parshikov,A., Studitsky,V.M. and Georgiev,P. (2013) Effective blocking of the white enhancer requires cooperation between two main mechanisms suggested for the insulator function. *PLoS Genet.*, **9**, e1003606.
105. Chopra,V.S., Cande,J., Hong,J.-W. and Levine,M. (2009) Stalled hox promoters as chromosomal boundaries. *Genes Dev.*, **23**, 1505–1509.
106. Kaushal,A., Mohana,G., Dorier,J., Özdemir,I., Omer,A., Cousin,P., Semenova,A., Taschner,M., Dergai,O., Marzetta,F. *et al.* (2021) CTCF loss has limited effects on global genome architecture in Drosophila despite critical regulatory functions. *Nat. Commun.*, **12**, 1011.
107. Liang,J., Lacroix,L., Gamot,A., Cuddapah,S., Queille,S., Lhoumaud,P., Lepetit,P., Martin,P.G.P., Vogelmann,J., Court,F. *et al.* (2014) Chromatin immunoprecipitation indirect peaks highlight long-range interactions of insulator proteins and pol II pausing. *Mol. Cell*, **53**, 672–681.
108. Purnell,B.A., Emanuel,P.A. and Gilmour,D.S. (1994) TFIID sequence recognition of the initiator and sequences farther downstream in Drosophila class II genes. *Genes Dev.*, **8**, 830–842.
109. Burke,T.W. and Kadonaga,J.T. (1996) Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes Dev.*, **10**, 711–724.
110. Shao,W., Alcantara,S.G.M. and Zeitlinger,J. (2019) Reporter-ChIP-nexus reveals strong contribution of the Drosophila initiator sequence to RNA polymerase pausing. *Elife*, **8**, e41461.
111. Schwartz,Y.B., Linder-Basso,D., Kharchenko,P.V., Tolstorukov,M.Y., Kim,M., Li,H.-B., Gorchakov,A.A., Minoda,A., Shanower,G., Alekseyenko,A.A. *et al.* (2012) Nature and function of insulator protein binding sites in the Drosophila genome. *Genome Res.*, **22**, 2188–2198.
112. West,A.G., Gaszner,M. and Felsenfeld,G. (2002) Insulators: many functions, many mechanisms. *Genes Dev.*, **16**, 271–288.
113. Li,J. and Gilmour,D.S. (2013) Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor. *EMBO J.*, **32**, 1829–1841.
114. Bag,I., Chen,S., Rosin,L.F., Chen,Y., Liu,C.-Y., Yu,G.-Y. and Lei,E.P. (2021) M1BP cooperates with CP190 to activate transcription at TAD borders and promote chromatin insulator activity. *Nat. Commun.*, **12**, 4170.
115. Bonchuk,A., Denisov,S., Georgiev,P. and Maksimenko,O. (2011) Drosophila BTB/POZ domains of “ttk group” can form multimers and selectively interact with each other. *J. Mol. Biol.*, **412**, 423–436.
116. Melnikova,L., Juge,F., Gruzdeva,N., Mazur,A., Cavalli,G. and Georgiev,P. (2004) Interaction between the GAGA factor and Mod(mdg4) proteins promotes insulator bypass in Drosophila. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 14806–14811.
117. Tutton,S., Azzam,G.A., Stong,N., Vladimirova,O., Wiedmer,A., Monteith,J.A., Beishline,K., Wang,Z., Deng,Z., Riethman,H. *et al.* (2016) Subtelomeric p53 binding prevents accumulation of DNA damage at human telomeres. *EMBO J.*, **35**, 193–207.
118. Radion,E., Morgunova,V., Ryazansky,S., Akulenko,N., Lavrov,S., Abramov,Y., Komarov,P.A., Glukhov,S.I., Olovnikov,I. and Kalmykova,A. (2018) Key role of piRNAs in telomeric chromatin maintenance and telomere nuclear positioning in Drosophila germline. *Epigenetics Chromatin*, **11**, 40.
119. Perrini,B., Piacentini,L., Fanti,L., Altieri,F., Chichiarelli,S., Berloco,M., Turano,C., Ferraro,A. and Pimpinelli,S. (2004) HP1 controls telomere capping, telomere elongation, and telomere silencing by two different mechanisms in Drosophila. *Mol. Cell*, **15**, 467–476.
120. Fanti,L., Giovinazzo,G., Berloco,M. and Pimpinelli,S. (1998) The heterochromatin protein 1 prevents telomere fusions in Drosophila. *Mol. Cell*, **2**, 527–538.