Active retrotransposons help maintain pericentromeric heterochromatin required for faithful cell division

Yajing Hao,[1,2,3,5] Dongpeng Wang,[1,2,5] Shuheng Wu,[1,2] Xiao Li,[3] Changwei Shao,[3] Peng Zhang,[1] Jia-Yu Chen,[3] Do-Hwan Lim,[3] Xiang-Dong Fu,[3] Runsheng Chen,[1,2,4] and Shunmin He[1,2]

[1]*Key Laboratory of RNA Biology, Center for Big Data Research in Health, Institute of Biophysics, Chinese Academy of Sciences, 100101, China;* [2]*University of Chinese Academy of Sciences, Beijing, 100049, China;* [3]*Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, California 92093-0651, USA;* [4]*Guangdong Geneway Decoding Bio-Tech Corporation Limited, Foshan 528316, China*

Retrotransposons are populated in vertebrate genomes, and when active, are thought to cause genome instability with potential benefit to genome evolution. Retrotransposon-derived RNAs are also known to give rise to small endo-siRNAs to help maintain heterochromatin at their sites of transcription; however, as not all heterochromatic regions are equally active in transcription, it remains unclear how heterochromatin is maintained across the genome. Here, we address these problems by defining the origins of repeat-derived RNAs and their specific chromatin locations in *Drosophila* S2 cells. We demonstrate that repeat RNAs are predominantly derived from active *gypsy* elements and processed by Dcr-2 into small RNAs to help maintain pericentromeric heterochromatin. We also show in cultured S2 cells that synthetic repeat-derived endo-siRNA mimics are sufficient to rescue Dcr-2-deficiency-induced defects in heterochromatin formation in interphase and chromosome segregation during mitosis, demonstrating that active retrotransposons are required for stable genetic inheritance.

[Supplemental material is available for this article.]

Eukaryotic genomes contain both gene-rich and gene-poor regions, corresponding to euchromatin and heterochromatin, respectively. Heterochromatin can be further divided into two classes: facultative, which is dynamic and marked by H3K27me3, and constitutive, which is largely stable and marked by H3K9me2/3 (Grewal and Jia 2007). Constitutive heterochromatin is predominantly associated with centromeric and pericentromeric regions, telomeres, and internal genomic regions (Saksouk et al. 2015). Constitutive heterochromatin plays vital roles in genome organization (Csink and Henikoff 1996), suppression of recombination to protect genome integrity (Grewal and Klar 1997), and stable genetic inheritance during development and differentiation (Allshire et al. 1995; Peters et al. 2001; Becker et al. 2016), underscoring the biological importance of constitutive heterochromatin.

Regarding the formation/maintenance of constitutive heterochromatin (hereafter referred to generally as heterochromatin), our current knowledge is largely derived from studies in fission yeast and flies (Grewal and Jia 2007; Holoch and Moazed 2015b). Transcription is required for initiating heterochromatin formation, even though the eventual fate is to shut down transcription. In fission yeast, initial repeat-derived transcripts are amplified by an RNA-dependent RNA polymerase (RdRP). Resultant double-stranded RNAs (dsRNAs) are next processed by Dicer to produce small interfering RNAs (siRNAs), which are loaded onto Ago1 to form the RNA-induced transcription silencing (RITS) complex to target nascent repeat RNAs. RITS recruits a key histone methyltransferase Clr4 to generate H3K9me2/3, which then attracts its reader Swi6, together inducing a series of RNA–protein and protein–protein interactions to mediate both initial deposition and spreading of H3K9me2/3 to neighboring sequences (Volpe et al. 2002; Verdel et al. 2004). *Drosophila* appears to follow a similar scheme except using the piRNA system to process and amplify repeat-derived RNAs to actively repress retrotransposition in germline (Vagin et al. 2006; Halic and Moazed 2009; Muerdter et al. 2013).

Given this general conceptual framework, there are multiple questions that remain to be answered. First, heterochromatin is still dynamic, rather than completely inert, raising the question of how transcription is restarted and whether heterochromatin maintenance depends on local transcription in all regions in need of being patched up. Second, in principle, some repeat-derived RNAs may also be capable of acting in *trans*, as suggested by apparent crosstalk between a reporter gene with one copy localized near a pericentromeric region of one chromosome and the other in its native euchromatic context of another chromosome in fission yeast (Yu et al. 2018). However, it remains unclear whether this principle applies to heterochromatin maintenance on all pericentromeric regions. Third, both fission yeast and *Drosophila* germ cells are equipped with an RNA amplification system, but such system is lacking in somatic cells of flies and mammals (Stein et al. 2003), begging the question of how somatic cells

[5]These authors contributed equally to this work.
Corresponding authors: xdfu@ucsd.edu, chenrs@sun5.ibp.ac.cn, heshunmin@ibp.ac.cn
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.256131.119.

30:1570–1582 Published by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/20; www.genome.org

meet this supply/demand dilemma. Last, but not least, Dicer deficiency has been reported to cause chromosome missegregation during mitosis in both flies and mammals (Pek and Kai 2011; Huang et al. 2015), but it has remained unclear whether such phenotype results from impaired production of some sort of repeat-derived siRNAs or other function(s) of Dicer in the nucleus. In the current study, we aim to address these important questions.

Results

Strategy for genome-wide assignment of multimapped RNA and DNA reads

We recently developed a technology called global RNA–DNA interaction sequencing (GRID-seq), which employs a bivalent linker to ligate RNA and DNA on fixed nuclei, followed by selection and cleavage of linker-ligated products with MmeI to generate "mated" RNA and DNA for deep sequencing (Li et al. 2017). To control for the specificity in RNA–DNA mating, we generated a GRID-seq library on mixed human and *Drosophila* S2 cells, thus enabling the construction of a background model by using cross-species reads (i.e., fly RNA ligated to human DNA and vice versa). We generated two concordant GRID-seq libraries on S2 cells (Supplemental Table S1; Supplemental Fig. S1A), with both showing a high percentage (65%–75%) of RNA reads mapped to repeat-derived transcripts, but a much smaller fraction (10%–20%) of DNA reads assigned to repeat-rich genomic loci (Supplemental Fig. S1B). To fully utilize these multimapped reads, we took a previously established ShortStack strategy to make assignment to specific transcripts and genomic locations based on the local density of uniquely mapped reads (Axtell 2013). As illustrated (Fig. 1A), we first assigned each RNA read to unique transcripts according to FlyBase (Drysdale and FlyBase Consortium 2008) or repeat-derived transcripts based on RepeatMasker (Smit et al. 2013–2015), and each DNA read to genomic fragments, generated by AluI, an enzyme used to fragment the fly genome during GRID-seq library construction. For multimapped RNA or DNA reads, we distributed them according to the relative density of uniquely mapped reads (Supplemental Fig. S1C).

We next filtered out singular RNA–DNA mates and low-density mates below a set threshold and subtracted the background based on our mix library (see Supplemental Methods), which is quite significant in many accessible chromatin regions (Supplemental Fig. S1D). After these data processing steps, retained RNA–DNA mates show high consistency between the two independent GRID-seq libraries, as indicated by predominant common mates associated with both unique (Fig. 1B) and repeat-derived transcripts (Fig. 1C). This consistency is also reflected at the quantitative levels of individual common RNA–DNA mates (Fig. 1D,E), thus enabling us to rely on these common mates to generate the final RNA–DNA interactome for downstream analysis. After assigning multimapped RNA and DNA reads, most gaps around repeat-rich DNA regions were "filled" to similar levels, as compared to adjacent unique regions across the fly genome (Supplemental Fig. S2A).

Validation of mapping results with an independent data set

It is critical to validate our mapping strategy, even though ShortStack has been a widely used strategy to dynamically assign multimapped reads. To this end, we utilized the data generated by ChAR-seq (Bell et al. 2018), a strategy similar to GRID-seq except longer RNA and DNA reads were generated by sonication after linker ligation. Random fragmentation is evident from the relatively even size distribution on the DNA side, and in contrast, small RNA populations are significantly enriched on the RNA side (Supplemental Fig. S2B), suggesting the extensive association of small RNAs with chromatin (see below). Compared to GRID-seq that generates mated RNA–DNA pairs, ChAR-seq tends to trade off relatively longer reads with a large fraction of unmated RNA or DNA reads from sequenced libraries. Moreover, it was not optimal to use the ChAR-seq data in the first place for several reasons: (1) GRID-seq libraries were generated on S2 cells where many other types of genomic data are available for comparison (Supplemental Table S2), whereas ChAR-seq libraries were produced on a less commonly used *Drosophila* cell line (CME-W1-cl8+); (2) the vast majority of ChAR-seq reads was from a single library (Supplemental Table S1), thus prohibitive to assessing internal data reproducibility; and most importantly, (3) one of our GRID-seq libraries was constructed on mixed fly and human cells, thus permitting the use of cross-species RNA–DNA mates to build a background for nonspecific RNA–DNA interactions, which is missing from the existing ChAR-seq libraries. Nevertheless, the available ChAR-seq data with longer RNA and DNA reads provided an independent data set to evaluate our strategy for assigning multimapped RNA and DNA reads.

We first observed an overall high Spearman's correlation between the ChAR-seq and GRID-seq data sets ($R = 0.75$ for RNA reads; $R = 0.62$ for DNA reads) across the fly genome (Supplemental Fig. S2C). However, when focused on repeat-enriched AluI DNA bins, the correlation was quite modest at both RNA ($R = 0.37$) (left panel in Fig. 1F) and DNA ($R = 0.23$) (left panel in Fig. 1G) levels. A population of DNA reads (those in the lower right of Fig. 1G) was scored by ChAR-seq, but less by GRID-seq, likely due to the higher mapping power of the former. The correlation was improved when comparing uniquely mapped RNA or DNA reads from ChAR-seq with uniquely plus multimapped RNA or DNA reads from GRID-seq ($R = 0.70$ for RNA and $R = 0.48$ for DNA) (middle panels in Fig. 1F and Fig. 1G, respectively). As expected, after including multimapped reads from ChAR-seq using ShortStack, the correlation was further improved ($R = 0.84$ for RNA and $R = 0.53$ for DNA) (right panels in Fig. 1F and Fig. 1G, respectively), which becomes comparable to the overall correlation (Supplemental Fig. S2C). The progressive improvement is also evident on representative genomic regions, showing gained RNA (Fig. 1H) or DNA (Fig. 1I) signals, each in a repeat-rich region (dashed boxes), both of which were detectable with uniquely mapped ChAR-seq reads, but missing from uniquely mapped GRID-seq reads, and then became visible after assigning multimapped GRID-seq reads. Together, these data validate our computational strategy to assign multimapped RNA and DNA sequencing reads to the fly genome.

Preferential interaction of distinct RNA classes with eu- versus heterochromatin

As reported earlier (Li et al. 2017), among 4856 nonrepeat chromatin-associated RNAs, most showed interactions with DNA near their sites of transcription (Supplemental Fig. S3A). We also noted that, among different RNA classes, most annotated snoRNAs and snRNAs were not only expressed but also engaged in interactions with chromatin in S2 cells, whereas virtually no pre-miRNA was detected with chromatin (Supplemental Fig. S3B). Additionally, we identified 230 repeat-derived RNAs with significant interactions with chromatin, including all rRNAs, satellite DNA-transcribed RNAs, and >70% of long interspersed nuclear element (LINE)- and long terminal repeat (LTR)-derived RNAs

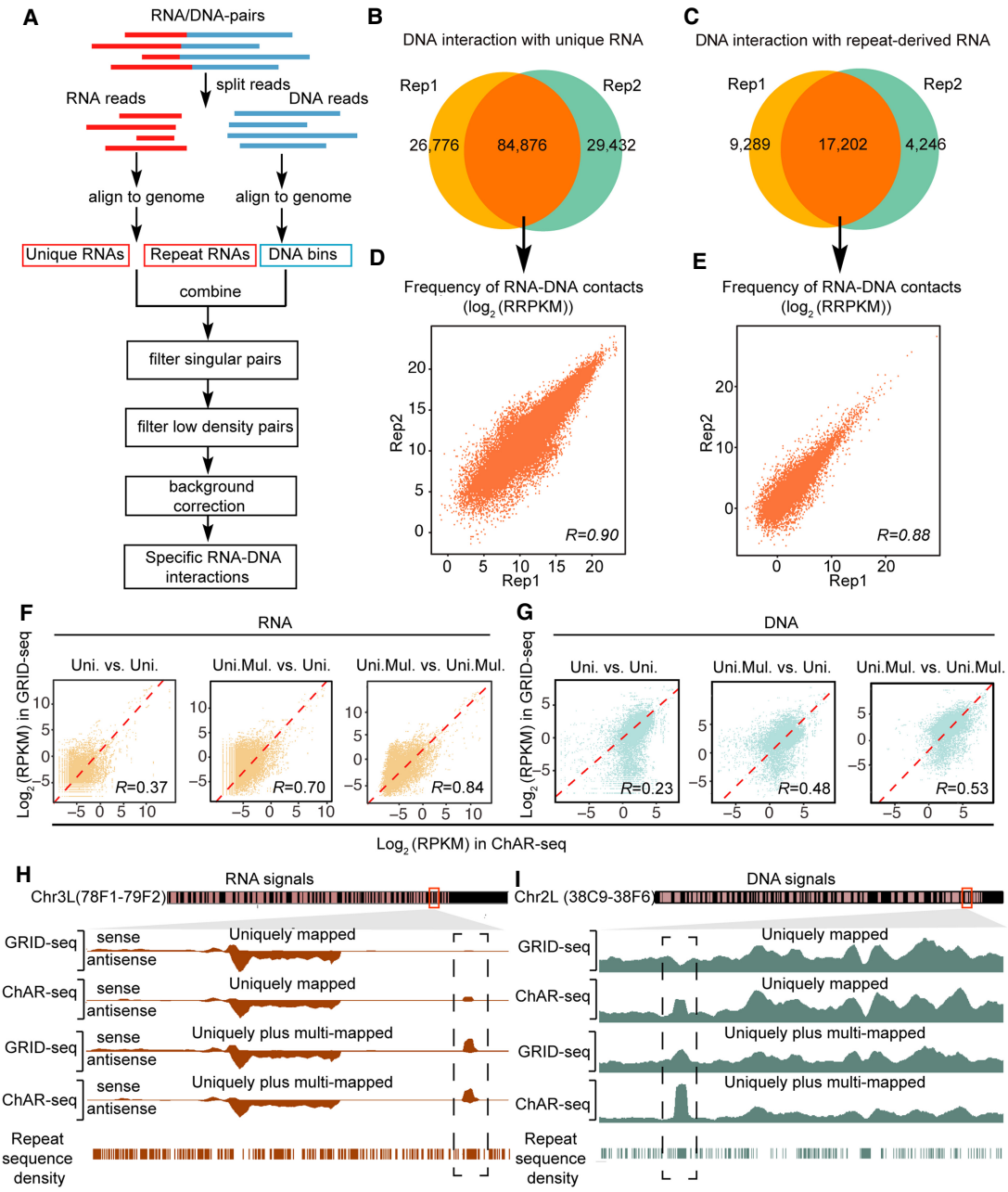

Figure 1. Strategy for assigning multimapped RNA and DNA reads. (*A*) Schematic presentation of the GRID-seq data processing pipeline. (*B*,*C*) Overlapped RNA–DNA mates containing annotated unique (*B*) or repeat-derived (*C*) RNA between the two independent GRID-seq libraries. (*D*,*E*) Quantitative analysis of commonly identified RNA–DNA mates containing annotated unique RNA (*D*) or repeat-derived RNA (*E*) between the two independent GRID-seq libraries. (RRPKM) Read counts per kilobase of RNA per kilobase of DNA per million. (*F*,*G*) Comparison between GRID-seq and ChAR-seq with uniquely or uniquely plus multimapped RNA (*F*) or DNA (*G*) in repeat-enriched AluI-generated DNA bins. (Uni.) Uniquely mapped reads, (Uni.Mul.) uniquely plus multimapped reads. (*H*,*I*) Transcribed RNA signals (*H*) or RNA-contacted DNA signals (*I*) obtained with different mapping strategies on representative genomic regions.

(Supplemental Fig. S3C). Percentage-wise, the majority of RNA species was from LTR (45%), simple repeat (28%), and LINE (14%) classes (Fig. 2A). We then displayed these 230 repeat-derived RNAs across the fly genome and observed some of them concentrating on pericentromeric regions on Chromosomes 2 and 3 (large dashed boxes in Fig. 2B). A subset of these repeat-derived RNAs also bound Chromosome 4 (small dashed boxes in Fig. 2B), which is known to be predominantly heterochromatic (Sun et al. 2000).

We next linked RNA–DNA interactions to critical chromatin features based on published epigenetic profiles (Supplemental Table S2). As expected, nonrepeat RNAs showed the highest association with H3K27ac (Fig. 2C), exhibiting coincidental peak summits (Fig. 2D), whereas repeat-derived RNAs displayed the greatest preference for H3K9me3 (Fig. 2E,F). Conversely, we sorted AluI-generated DNA fragments according to the levels of associated repeat-derived RNAs, observing that, among the top 1000 repeat

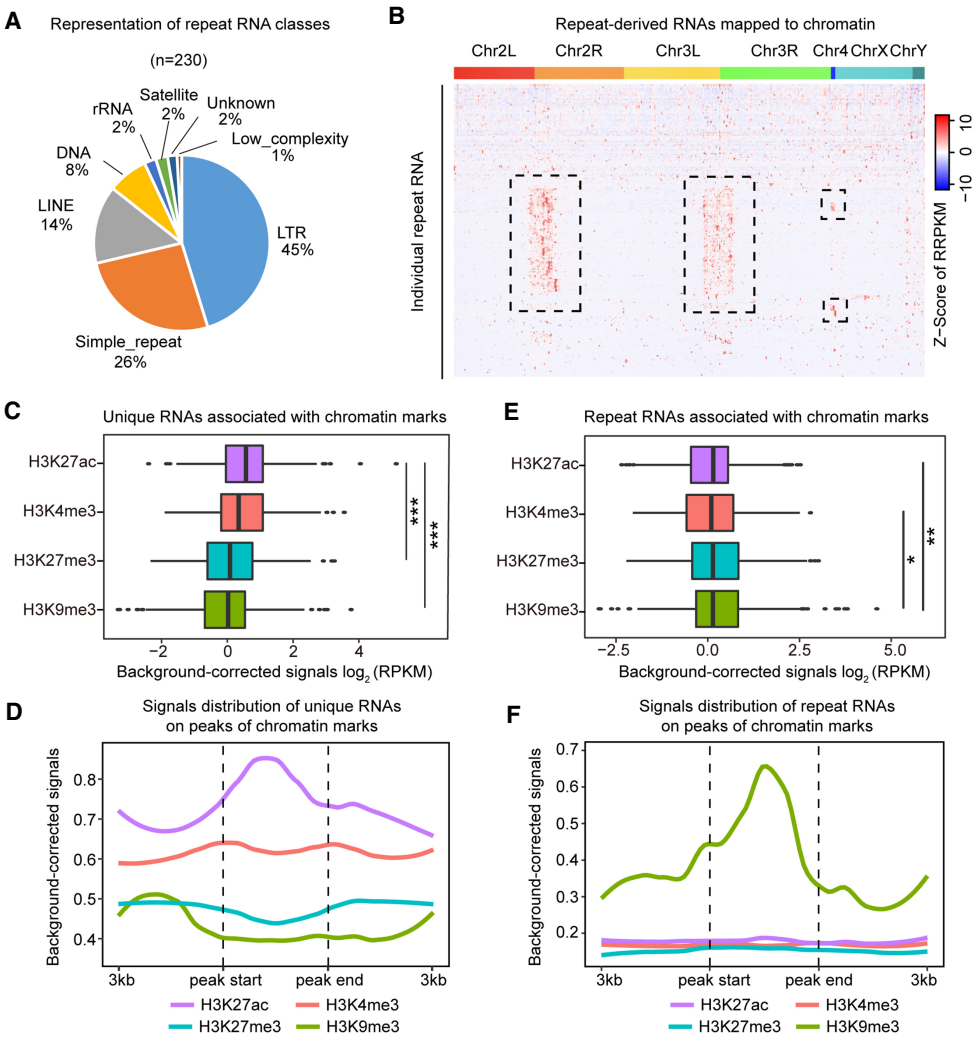

Figure 2. Interaction of distinct RNA classes with eu- versus heterochromatin. (*A*) Different repeat RNA classes represented by 230 DNA-bound repeat-derived RNAs. (*B*) Heat map showing the distribution of individual 230 repeat-derived RNAs across the *Drosophila* genome in S2 cells. (Row) Individual chromatin-associated repeat-derived RNAs, (column) AluI DNA bins, (boxed regions) repeat-derived RNAs that showed preferential binding to constitutive heterochromatin in pericentromeric regions. (*C*) Association of chromatin marks with background-corrected signals of unique RNAs. (***) $P<0.001$ (unpaired Student's *t*-test). (*D*) Distribution of unique RNA interaction signals around chromatin mark peaks. (*E*) Association of chromatin marks with background-corrected signals of repeat-derived RNAs. (*) $P<0.05$, (**) $P<0.01$ (unpaired Student's *t*-test). (*F*) Distribution of repeat RNA interaction signals around chromatin mark peaks.

RNA-associated DNA bins, H3K9me3 was the dominant signal on these DNA regions (Supplemental Fig. S3D), thus suggesting a general trend where nonrepeat RNAs tend to interact with euchromatin and repeat-derived RNAs with heterochromatin. While somewhat anticipated from the vast literature, we now obtained critical information on which specific repeat-derived RNAs are more prevalent than others in interacting with specific heterochromatic regions in the fly genome, thus laying a critical foundation to investigate their relative contributions to the initiation and/or maintenance of heterochromatin.

Prevalent association of *gypsy*-derived RNAs with constitutive heterochromatin

We next intersected the density of individual repeat-derived RNAs with markers for constitutive heterochromatin characterized by the coordinated ChIP-seq signals for H3K9me3 and its reader Su (var)205 (Fig. 3A). By determining the colocalization coefficient for each of the 230 repeat-derived RNAs between their DNA interaction frequencies and relative densities of H3K9me3 and Su(var) 205 signals in 1Mb-DNA bins of the fly genome, we identified 79 repeat-derived RNAs, including two rRNAs with Pearson's correlation coefficient >0.3 (red dots in Fig. 3A). The association of rRNAs with heterochromatin agrees with the observations that transcriptionally inert centromere-proximal regions tend to be organized around the nucleolus (Quinodoz et al. 2018) and that some rRNA-derived fragments may also contribute to gene silencing (Jain et al. 2016). Excluding rRNAs, we named the rest of heterochromatin-enriched RNAs as CHARRs (Constitutive Heterochromatin-Associated Repeat-derived RNAs). These CHARRs appear to show exclusive association with constitutive heterochromatin, with little colocalization with the facultative chromatin marker H3K27me3 (Supplemental Fig. S4A), as illustrated with three specific CHARRs (Supplemental Fig. S4B). These CHARRs were

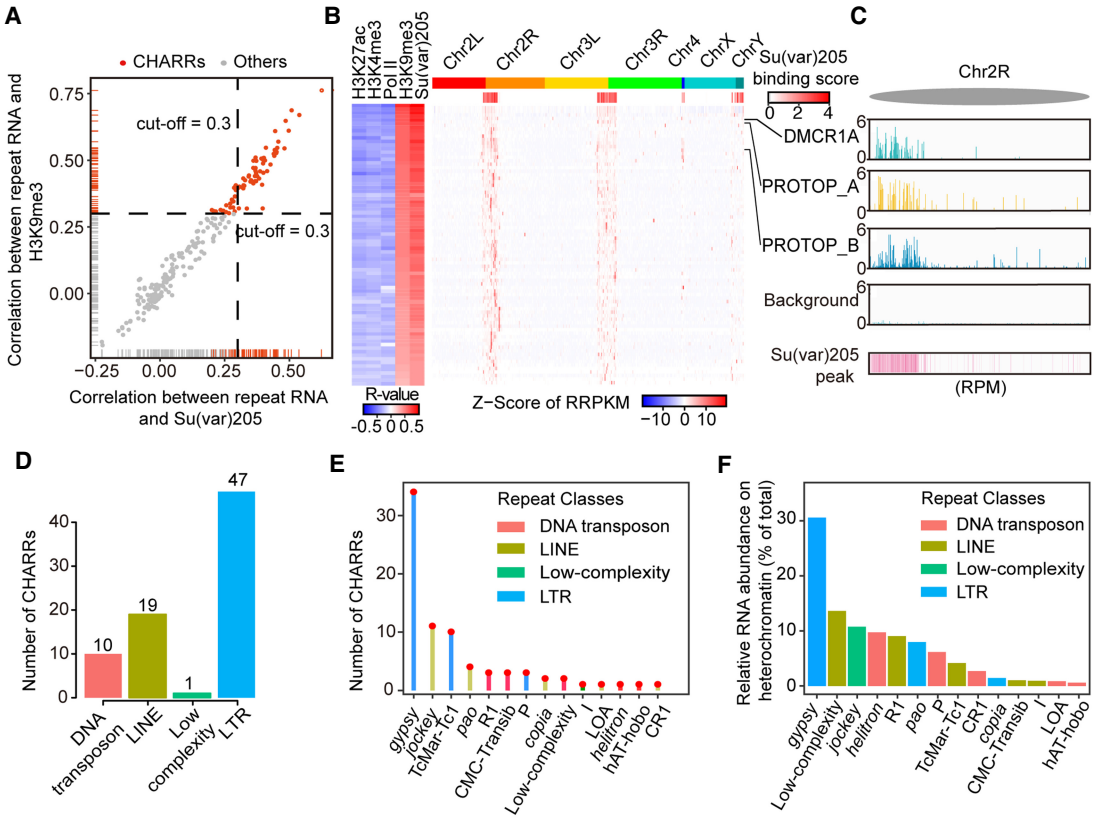

Figure 3. Prevalent association of *gypsy*-derived RNAs with constitutive heterochromatin. (*A*) Scatterplot of colocalization coefficients between repeat-derived RNA signals on DNA and the levels of H3K9me3 (*y*-axis) or HP1 (*x*-axis) in S2 cells. A threshold of 0.3 was chosen for both chromatin marks (red lines). Dashed lines were used to differentiate CHARRs (red dots) from other repeat-derived RNAs (gray dots). (*B*) *Left*: Heat map of colocalization of each CHARR with individual chromatin marks or Pol II ChIP-seq signals. *Right*: Heat map of chromatin interaction of each CHARR with AluI-generated DNA bins, which were merged from 100 continuous AluI DNA bins. HP1 binding signals in these DNA bins are displayed *below* the schematic presentation of *Drosophila* chromosomes. Three representative CHARRs are labeled on the *right*. (*C*) Chromatin interaction signals of three representative CHARRs, DMCR1A, PROTOP_A, and PROTOP_B on a genomic region in Chromosome 2R in comparison with nonspecific background based on human RNA mapped to fly DNA and HP1 binding density. All signals were scaled to reads per million. (*D*,*E*) The number of repeat classes (*D*) or subfamilies (*E*) associated with CHARRs. (*F*) The relative RNA abundance (% of total) of CHARRs on constitutive heterochromatin and specific subfamilies to which they belong. Colors show the RNA classes to which individual CHARR subfamilies belong.

insignificantly associated with msl-3 (Supplemental Fig. S4C,D), a key component of the silencing complex involved in X inactivation (Xi) in *Drosophila*, consistent with predominant facultative heterochromatin on Xi (Franke and Baker 1999).

The 77 CHARRs were predominantly associated with the pericentromere of Chromosomes 2 and 3 where their interactions with DNA positively correlated to heterochromatin markers (e.g., H3K9me3 and Su(var)205) and negatively to euchromatin markers (e.g., H3K27ac and H3K4me3), as well as RNAP II ChIP-seq signals (Fig. 3B), as highlighted with DMCR1A, PRTOP_A, and PRTOP_B on the right arm of Chromosome 2 (Fig. 3C). We next determined RNA class, genomic origin, and relative abundance for each CHARR. The largest RNA class corresponds to LTR ($n=47$) and the second to LINE repeats ($n=19$) (Fig. 3D). The majority of RNA species from these two classes of retrotransposons belong to the *gypsy* and *jockey* subfamilies (Fig. 3E). By ranking individual CHARRs according to their relative abundance on heterochromatin and summing the collective abundance according to specific RNA subclasses, we found that active *gypsy* family members were top contributors to CHARRs on heterochromatin (Fig. 3F). This suggests a major role of *gypsy*-derived RNAs in heterochromatin formation/maintenance in *Drosophila* S2 cells.

Cis- and *trans*-acting repeat-derived RNAs on chromatin

The RNA–DNA interactome permitted, for the first time, a determination of both the source of repeat-derived RNAs and their locations on specific chromatin regions. Given the predominant mode of nonrepeat RNAs to act in *cis* (*cis* defined by mated RNA and DNA reads mapped to the same chromosomes, as opposed to *trans* defined by mated RNA and DNA reads mapped to different chromosomes), we asked whether this might also apply to repeat-derived RNAs. We noted that rRNA-derived RNAs all interacted with heterochromatin regions near the loci of their transcription where CHARRs also predominantly bound (Supplemental Fig. S5A,B), and in contrast, a typical CHARR *gypsy*4_I-int appeared to engage in both in *cis* and *trans* interactions (Supplemental Fig. S5C).

This prompted us to examine global CHARR–chromatin interactions, finding their prevalent intra-chromosomal interactions but a significant degree of inter-chromosomal interactions as well (Fig. 4A). On individual CHARRs, about half (46.8%) of the CHARRs were engaged in intra-chromosomal interactions, whereas the other half (53.2%) were actively involved in both intra- and inter-chromosomal interactions (Fig. 4B). This is illustrated with

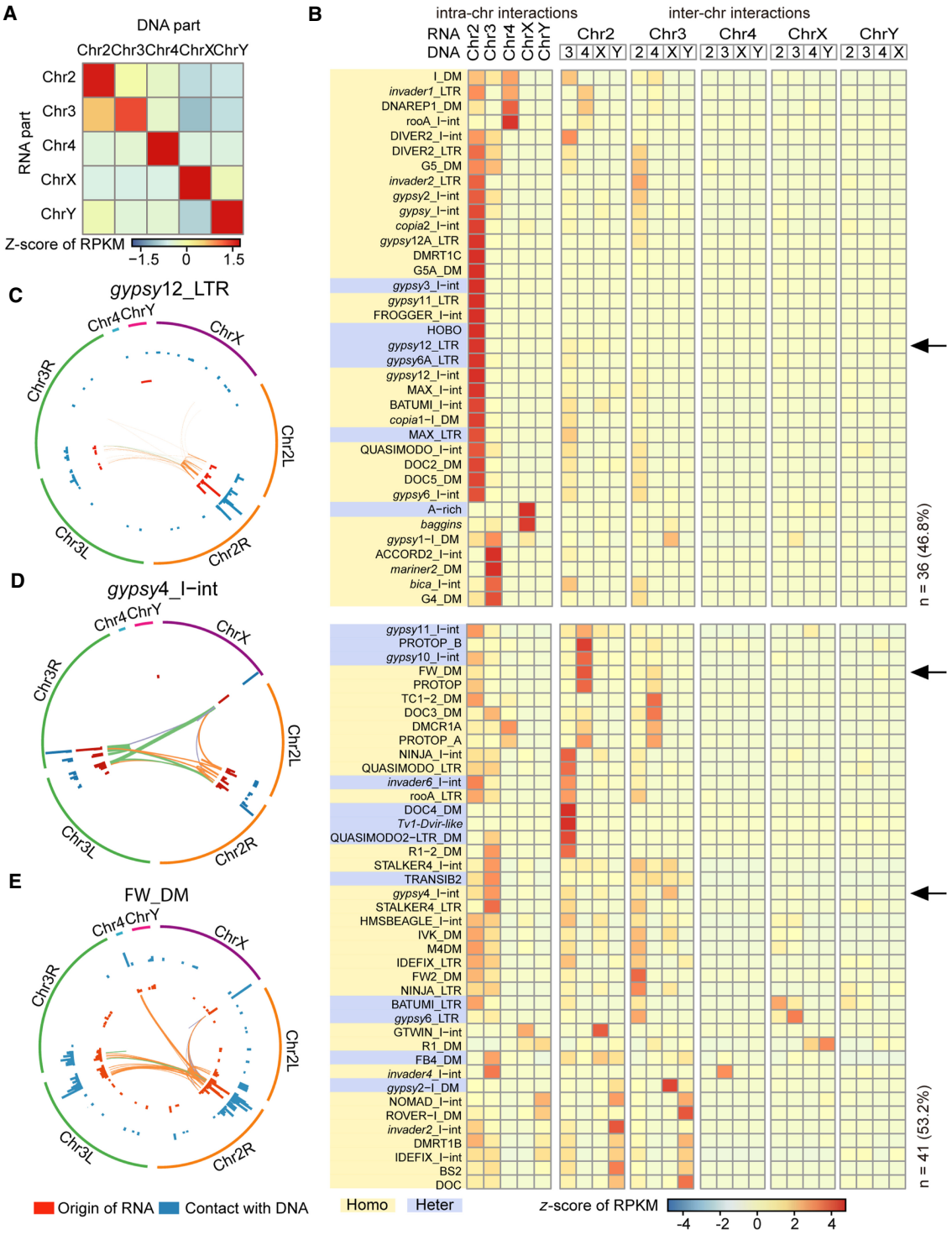

Figure 4. *Cis*- and *trans*-acting repeat-derived RNAs on chromatin. (*A*) Heat map of RNA–DNA interaction scores for 77 CHARRs on the same or different chromosomes. *Trans*-chromosome interactions are referred to as RNA interacting with DNA in chromosomes other than those from which the RNA is transcribed. (*B*) Assignment of each of the 77 CHARRs to either intra- or inter-chromosomal interactions in S2 cells and sorting of the data by unsupervised clustering. *Top*: RNA derived from different chromosomes (first row) and its DNA contacts in the same or different chromosomes (second row). The RNA–DNA interaction intensities were indicated by the *z*-scores according to the color key at the *bottom*. Arrows point to the three specific CHARRs *gypsy*12_LTR, *gypsy*4_I-int, and FW_DM, as individually illustrated in panels *C*–*E*. All CHARRs are either highlighted with yellow or blue to indicate their preferences for homotropic or heterotropic interactions, as described later. (*C*–*E*) The origins of three representative CHARRs and their DNA contacts, as shown by the Circos plots for *gypsy*12_LTR (C), *gypsy*4_I-int (*D*), and FW_DM (*E*). In each of these plots, the *inner* (red) track indicates the origins of individual RNAs and the *outer* (blue) track shows where the RNAs interact with DNA in 100-kb DNA windows. The heights of the signals correspond to reads per 100-kb window per millions. Lines specify intra- or inter-chromosomal interactions, with width suggesting the relative interaction levels in each case. The line color indicates the origin of chromatin-interacting RNAs.

three representative CHARRs using Circos plots (Krzywinski et al. 2009) where *gypsy12*_LTR preferentially interacted with DNA on the same chromosome (Fig. 4C), whereas both *gypsy4*_I-int and FW_DM were engaged in both intra- and inter-chromosomal interactions (Fig. 4D,E). We also noted that individual CHARRs selectively bound Hi-C defined "B" domains in pericentromeric regions (Supplemental Fig. S5D,E).

Tendency for *trans*-acting RNAs to supplement *cis*-acting RNAs on chromatin

During our analysis of *cis*- and *trans*-acting CHARRs, we further noted that DNA loci expressing high levels of CHARRs tended to associate with less *trans*-acting RNAs (which are related repeat-derived RNA species transcribed from other chromosomes), and the converse was also true, as exemplified on a specific region in Chr 2R, where the *Tv1-Dvir-like* RNA (a *gypsy* family member from the LTR class) was mainly transcribed from region 1 (Fig. 5A, second track), most of which contacted DNA locally around the transcribing locus (Fig. 5A, third track). *Trans*-acting RNAs (transcribed from *Tv1-Dvir-like*-related repeat species from other chromosomes) were mostly mapped to region 2, where little *Tv1-Dvir-like* RNA was produced, and region 3, where a much lower level of the *Tv1-Dvir-like* RNA was transcribed compared to region 1 (Fig. 5A, fourth track). We quantified these results by dividing all active *Tv1-Dvir-like* loci into three groups according to their levels of transcription (bottom 25%, middle 50%, and top 25%) (Fig. 5B) and then determined the ratio of *trans*-acting over total RNAs for each group. We observed a reverse correlation between local RNA production and the percentage of association with *trans*-acting RNAs (bars, Fig. 5B). This also applied to another repeat RNA, *baggins* (a LOA family member from the LINE class) (Fig. 5C,D). To determine whether this represents a general trend, we extended the analysis to all CHARRs and found that most followed this rule (Fig. 5E), suggesting that highly transcribed CHARRs supply RNAs in *trans* to interact with DNA regions with less transcriptional activities.

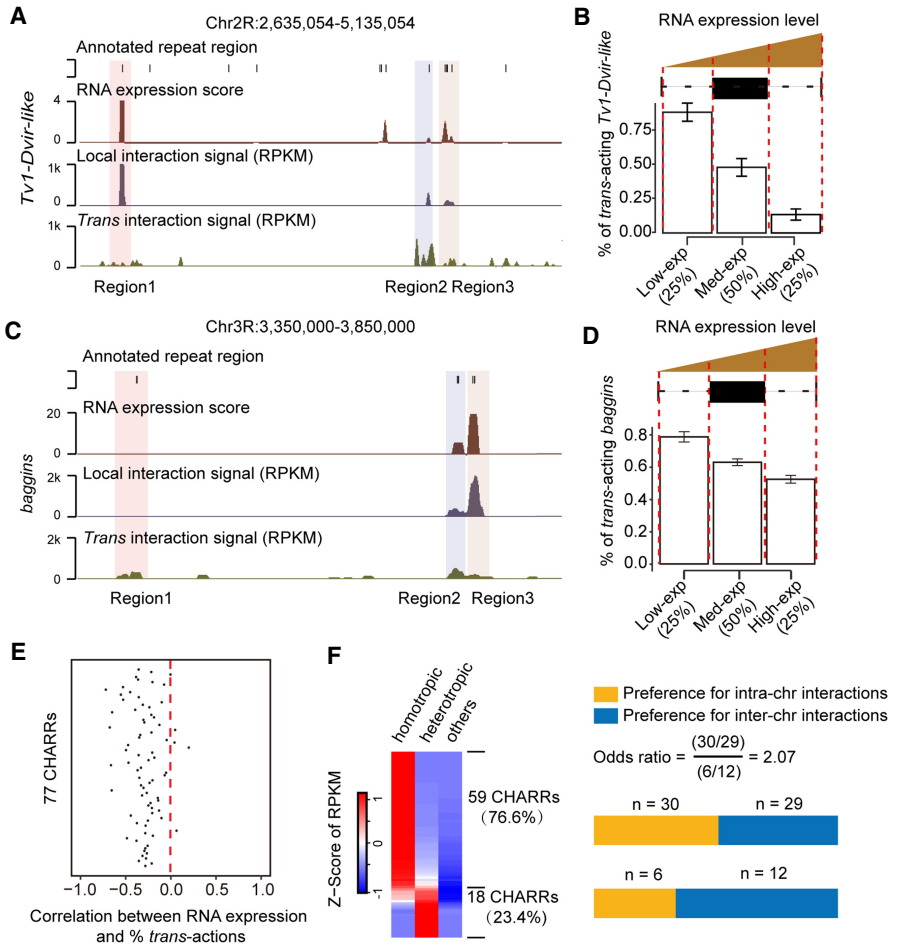

Figure 5. Reverse correlation between local transcription and association with *trans*-acting RNAs on chromatin. (*A*) A representative genomic region showing *Tv1-Dvir-like* transcription loci and interactions with *Tv1-Dvir-like* subfamily RNAs produced either locally or from other chromosomes (*trans*-acting RNAs). Three annotated *Tv1-Dvir-like* transcription regions are indicated at the *bottom*. (*B*) *Tv1-Dvir-like* subfamily-derived RNAs were segregated into three groups according to their levels of transcription (bottom 25%, middle 50%, and top 25%). Bars indicate the percentage of *trans*-acting *Tv1-Dvir-like* RNAs on the transcription loci in each group. (*C*,*D*) Similar analysis and illustration for another typical CHARR *baggins* as in *A* and *B*. (*E*) Pearson's correlation for individual CHARRs between their expression and the percentage of associated *trans*-acting RNAs from the same subfamilies. (*F*) Tendencies of CHARRs engaging homotropic versus heterotropic interactions based on normalized reads per kilobase per million (*left*). The two classes of CHARRs were further separated into those with preference for inter- (yellow) or intra- (blue) chromosomal interactions (*right*), as indicated in Figure 4.

Given different degrees of individual CHARRs in interacting with DNA, we next characterized the underlying DNA sequences that might specify such RNA–DNA interactions. If a given CHARR specifically interacted with a DNA region that harbors the same repeat sequence within a 1-kb window, we called it a "homotropic" interaction; if the interacting DNA region contains distinct repeat sequence, we then referred it to as a "heterotropic" interaction, as illustrated with *baggins* and *Tv1-Dvir-like* RNAs (Supplemental Fig. S5F,G). All DNA interaction regions that contain no repeat sequence were classified as "others." According to these definitions, we found that 59 CHARRs (yellow-labeled in Fig. 4B) were engaged in homotropic interactions, whereas the remaining 18 (blue-labeled in Fig. 4B) were more involved in heterotropic interactions, but none showed significant interactions with DNA that contains no repeat sequence (Fig. 5F, left). Relative to CHARRs with preference for homotropic interactions, more CHARRs with a higher tendency to engage in heterotropic interactions were more involved in inter-chromosomal interactions, as indicated by a significant odds ratio (Fig. 5F, right). These observations imply that CHARRs with preference for homotropic interactions might facilitate initiating heterochromatin formation, whereas those for heterotropic interactions might be important for heterochromatin spreading as well as maintenance.

Evidence for CHARR-derived endo-siRNAs on chromatin

CHARRs may function on chromatin either as nascent RNAs or processed RNAs or both. Previous results show most heterochromatin-bound CHARR-derived RNAs are involved in *trans* interactions. This indicates some of CHARRs may perform functions as processed transcripts rather than nascent transcripts. Because repeats can generate endo-siRNAs, we examined whether CHARR-derived RNAs have critical features of endo-siRNAs. As CHARRs may be prior processed into endo-siRNA or fragmented into small RNA of various sizes during library construction, we first characterized the size distribution of ChAR-seq RNA reads in 1-nt bins to compare CHARR-derived RNAs (red) with those from euchromatin (blue) or other repeats (green) and found that CHARR transcripts have a high tendency to generate small-length RNAs in the size ranges of 18–23 nt and 27–38 nt (Fig. 6A). Using euchromatin-derived RNAs to establish the level of RNAs that resulted from random RNA fragmentation during ChAR-seq library construction, we further calculated the fold-enrichment of RNAs from CHARRs and other repeats, observing that CHARRs have a higher tendency to generate 21-nt RNAs compared to other repeat transcripts, although such 21-nt RNAs are clearly embedded with other RNA populations (Fig. 6B). These analyses provide initially suggestive evidence for the association of a small fraction of CHARR-derived endo-siRNAs on chromatin.

Next, we took advantage of the existing GRO-seq data on S2 cells (Supplemental Table S2) to observe that, contrary to rRNAs that were predominantly transcribed in the sense direction (Supplemental Fig. S6A), CHARRs were among the most abundant bidirectionally transcribed RNAs (Supplemental Fig. S6B). In contrast, the majority of nonrepeat transcripts were transcribed from the sense strand (Supplemental Fig. S6C), with the exception of some lncRNAs (Supplemental Fig. S6D). Bidirectionally transcribed repeat RNAs may thus provide dsRNA substrates for further processing into endo-siRNAs to function in heterochromatin formation/maintenance (Fagegaltier et al. 2009; Volpe and Martienssen 2011).

We then determined how these CHARR-derived small RNAs were generated. In contrast to piRNA-mediated heterochromatin formation in germline, Dcr-2 has been reported to be specifically devoted to endo-siRNA processing in fly somatic cells (Czech et al. 2008; Ghildiyal et al. 2008; Kawamura et al. 2008; Okamura et al. 2008). To determine whether CHARRs depended on Dcr-2 for their efficient processing and thus expression, we took advantage of the existing small RNA-seq data to compare their expression levels between wild-type, *Dcr-2* knockout (ko), and *Dcr-2* rescued female fly heads (Kandasamy and Fukunaga 2016). We found that 70 out of 77 CHARRs were down-regulated in *Dcr-2* ko cells, which were rescued upon *Dcr-2* re-expression (Supplemental Fig. S7A). These CHARR-derived, Dcr-2-dependent small RNAs were predominantly 21 nt in length (orange) compared to slightly longer RNAs among all reads in the library (gray), most of which likely represent abundant miRNAs (Supplemental Fig. S7B).

These CHARR-derived endo-siRNAs are assembled into AGO2 complexes, as our analysis on S2 cells revealed that 15%–18% small RNA-seq reads from the total AGO2 IPed samples corresponded to Dcr-2 regulated CHARRs, and in comparison, only 0.52% was associated with AGO1, which is known to predominantly function in the miRNA pathway in the fly (Supplemental Fig. S7C). Furthermore, ~81% of AGO2-associated reads from Dcr-2-regulated CHARRs are 21 nt in size (Supplemental Fig. S7D). Collectively, these data provide evidence for processing of CHARRs into 21-nt endo-siRNAs that are incorporated into AGO2-containing complexes.

Rescuing global heterochromatin defects in Dcr-2-deficient S2 cells

We further confirmed Dcr-2-dependent processing of several specific CHARRs into endo-siRNAs in S2 cells by northern blotting. Efficient *Dcr-2* knockdown (kd) (Supplemental Fig. S8A) drastically reduced the levels of small RNAs derived from *gypsy*2-I_DM and *gypsy*4_I-int (Fig. 6C) as well as DMCR1A and DOC (Supplemental Fig. S8B). As CHARRs are predominantly enriched in heterochromatic regions, a fraction of which shows endo-siRNAs properties, we next wished to test whether these CHARRs may regulate heterochromatin, in part, through endo-siRNAs. Because many CHARRs appear to supply RNAs in *trans*, we asked whether or not the Dcr-2-deficiency-induced heterochromatin defects could be "rescued" with small RNAs derived from CHARRs. For this purpose, we chemically synthesized a pool of endo-siRNA mimics based on a representative subset of CHARRs (DMCR1A, FB4_DM, FW_DM, *gypsy*4_I-int, DOC, *gypsy2*-I_DM, and I_DM) (Supplemental Table S3) and transfected this pool into S2 cells depleted of Dcr-2. Relative to mock-treated cells and *GFP* knockdown cells, *Dcr-2* kd reduced H3K9me3 as expected (Supplemental Fig. S8C,D), and in comparison with scrambled siRNA, the CHARR-derived siRNA pool effectively restored this heterochromatin marker, while the levels of H3K27me3 and H3K4me3 remained constant (Fig. 6D; Supplemental Fig. S8C). This was also evident at the immunocytochemical level by staining for H3K9me3 (Fig. 6E; Supplemental Fig. S8D). These data demonstrated the sufficiency of *trans*-acting endo-siRNA mimics derived from CHARRs to rescue heterochromatin defects in Dcr-2-deficient S2 cells.

To demonstrate the contribution to repeat-derived endo-siRNAs to heterochromatin maintenance genome-wide, we next performed ChIP-seq for H3K9me3 in comparison with H3K4me3 and H3K27me3 in response to *Dcr-2* kd with or without treatment with CHARR-derived endo-siRNA mimics in S2 cells. Our ChIP-seq data sets were comparable to the profiles of these histone marks published earlier on the same cell type (Supplemental Fig. S9A, B). Upon *Dcr-2* kd, we detected a global decrease of ChIP-seq signals for H3K9me3, but not for H3K4me3 and H3K27me3 (Fig. 6F). The effects were also evident on individual H3K9me3-marked genomic loci (cf. the first two tracks in Fig. 6G), suggesting that Dcr-2 is functionally required for maintaining heterochromatin genome-wide.

The next question was whether these endo-siRNA mimics could rescue heterochromatin defects, and if so, whether the rescue required their targeting specificity. Because those endo-siRNA mimics were designed to target 10 representative CHARRs, we analyzed the H3K9me3 ChIP-seq signals on 1-kb-binned genomic regions that show homology with at least one of the CHARR-derived siRNA mimics (Fig. 6H) in comparison with genomic regions that showed H3K9me3 ChIP-seq signals but with <50% homology to any of those CHARR-derived siRNA mimics (Fig. 6I). Indeed, the specific endo-siRNA mimics, but not scrambled siRNA, effectively rescued H3K9me3 ChIP-seq signals on CHARRs target genomic regions (cf. lane 2 vs. 4 in Fig. 6G) but was modest at best (likely due to a remaining degree of heterotropic interactions) on regions not targeted by CHARRs (cf. lane 2 vs. 4 in Fig. 6I) ($P = 0.054$). As expected, little H3K4me3 and H3K27me3 ChIP-seq signals were detected in H3K9me3-marked genomic loci. These general trends were also illustrated on two representative genome loci for

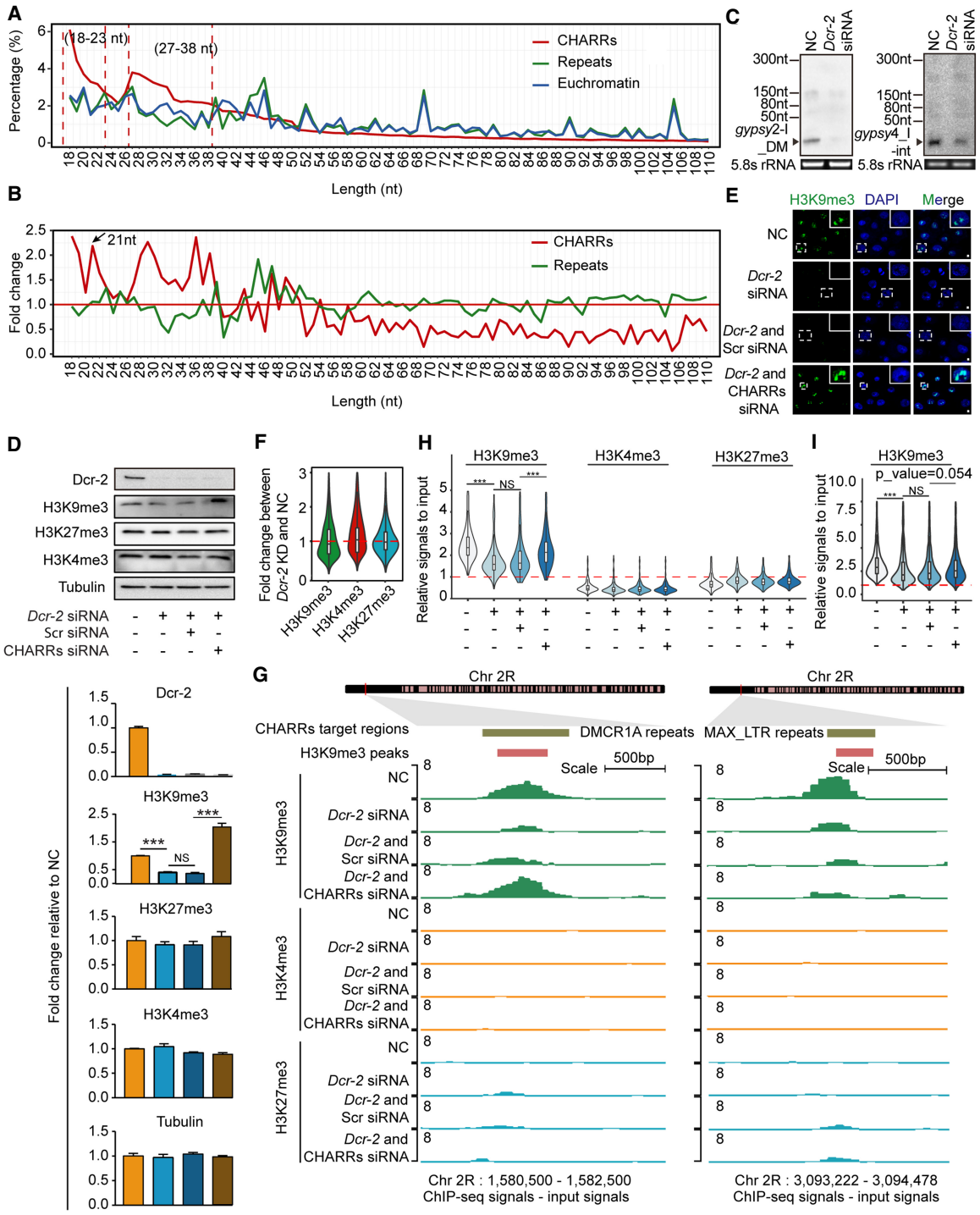

Figure 6. *Trans*-acting repeat-derived RNAs for heterochromatin maintenance. The length distribution in 1-nt bins based on the combined ChAR-seq data set (*A*), and at each bin, reads from CHARRs (red) or other repeats (green) were normalized against those from euchromatin RNAs (*B*). (*C*) Confirmation of *Dcr-2*-dependent expression of *gypsy*2-I_DM (*left*) and *gypsy*4_I-int (*right*) expression by northern blotting analysis. (*D*) Western blotting analysis of Dcr-2, H3K9me3, H3K27me3, H3K4me3, and Tubulin in S2 cells in response to siRNA-mediated knockdown of *Dcr-2* and rescue with a transfected pool of CHARR-derived synthetic siRNAs. Quantified data are shown *below* as fold-change (FC) relative to mock-treated cells (NC) in lane *1*. Data are presented as mean ± SEM ($n = 3$ biological replicates). (***) $P < 0.001$, (NS) not significant (unpaired Student's *t*-test). (*E*) H3K9me3 detected by immunocytochemistry in S2 cells treated with different combinations of siRNAs, as indicated. (Green) H3K9me3 signals, (blue) DAPI. Scale bar, 2 μm. (*F*) Violin plot for fold-change of H3K9me3, H3K4me3, and H3K27me3 signals on targeted peaks in response to *Dcr-2* knockdown. (KD) Knockdown. (*G*) H3K9me3 ChIP-seq signals on representative genomic loci, one corresponding to a CHARR-target locus (*left*) and a non-CHARR-target locus (*right*) in response to *Dcr-2* knockdown, complemented with either scrambled or derived siRNAs. ChIP-seq signals for euchromatin (H3K4me3) or facultative heterochromatin (H3K27me3) were shown for comparison. (*H*,*I*) H3K9me3 signals on targeted (*H*) or nontargeted (*I*, with <50% sequence homology to any of derived siRNAs) regions in response to *Dcr-2* knockdown, complemented with either scrambled or derived siRNAs. (***) $P < 0.001$ (Student's *t*-test). H3K27me3 and H3K4me3 binding signals were displayed for comparison.

CHARRs targets (Fig. 6G, left; Supplemental Fig. S10A) and regions not targeted by CHARRs (Fig. 6G, right; Supplemental Fig. S10B). We further performed ChIP-qPCR analysis on two internal repeat-rich regions and two pericentromeric regions upon transfection of two specific CHARR-derived siRNAs and observed sequence complementarity-dependent rescues in the respective regions (Supplemental Fig. S10C,D). Collectively, these data demonstrated that *trans*-acting endo-siRNA mimics derived from CHARRs are able to bypass the functional requirement of Dcr-2 for maintaining heterochromatin homeostasis on their target regions in S2 cells.

CHARR-derived siRNAs required for faithful chromosome segregation in S2 cells

Knockdown of *Dcr-2* and *AGO2* has been shown to cause significant defects in H3K9me3 signals, which we also confirmed (see Supplemental Fig. S8C,D), and heterochromatin formation (Deshpande et al. 2005; Peng and Karpen 2007) as well as chromosome segregation during cell division in S2 cells (Pek and Kai 2011). Because Dcr-2 and AGO2 are key components in the siRNA pathway, it is reasonable to extrapolate that endo-siRNAs processed by Dcr-2 and loaded on AGO2 are responsible for the phenotype, but direct evidence for this critical conclusion has been lacking. Given the coverage of CHARRs on most pericentromeric regions in the fly genome (see Fig. 3B) and the incorporation of CHARR-derived endo-siRNAs in AGO2 complexes (see Supplemental Fig. S7C,D), we asked whether CHARR-derived endo-siRNAs were also able to rescue the chromosome segregation defects. To this end, we first confirmed that either *Dcr-2* or *AGO2* kd-induced cell cycle defects, and, as expected, we detected G1-S arrest in both cases (Fig. 7A,B; Supplemental Fig. S11A). We found CHARR-derived endo-siRNAs, but not scrambled siRNA, were able to rescue the cell cycle defects in *Dcr-2* kd cells (Fig. 7A), but not in *AGO2* kd cells (Fig. 7B). We also tested smaller pools consisting of one or five endo-siRNAs and found that smaller pools only partially rescued the mitotic defects (Supplemental Fig. S11B), suggesting the requirement for the collective action of multiple CHARR-derived endo-siRNAs. These data are fully in line with the requirement of Dcr-2 for processing CHARRs into small endo-siRNAs, which could be bypassed by the transfected endo-siRNA mimics, although these mimics still need AGO2 to execute their functions in S2 cells.

We closely examined the mitotic defects in *Dcr-2* kd cells, noting both misaligned chromosomes and lagged as well as missegregated chromosomes (Fig. 7C; Supplemental Fig. S12A) in about equal frequencies (Fig. 7D). Transfection of CHARR-derived endo-siRNA mimics, but not scrambled siRNA, into these Dcr-2-deficient S2 cells was sufficient to rescue all of these defects (Fig. 7D; Supplemental Fig. S12A). In contrast, we observed similar mitotic defects in AGO2-deficient S2 cells without or with transfection of the CHARR-derived endo-siRNAs (Fig. 7E,F; Supplemental Fig. S12B). Taken together, these results provide evidence for active retrotransposon RNAs to preserve pericentromeric heterochromatin homeostasis, which is required for cell cycle progression via the Dcr-2/AGO2-mediated siRNA pathway in S2 cells.

Discussion

Genetic and biochemical experiments in fission yeast and *Drosophila* have laid a general conceptual framework in understanding the formation/maintenance of heterochromatin (Ekwall et al. 1995; Kellum and Alberts 1995). While heterochromatin is, in general, prohibitive to transcription, mounting evidence suggests that transcription is actually required to initiate heterochromatin formation. This so-called nascent RNA model (Holoch and Moazed 2015a) also represents a puzzle in envisioning how heterochromatin could be actively maintained. As a matter of fact, repeat-derived RNAs are now known to be prevalent on chromatin in vertebrates (Hall et al. 2014), implying their regulatory functions at the chromatin levels.

The consensus in the field is that heterochromatin-associated RNAs are mostly derived from active retrotransposons and simply repeats. However, the nature of such repetitive sequences has made it difficult to determine their origins and destinations in the genome. We have now addressed this fundamental problem by using the newly elucidated RNA–DNA interactome in *Drosophila*. We show that various active retrotransposons, especially those from the *gypsy* family of the LTR class, produce a large amount of repeat RNAs with the ability to act in both *cis* and *trans* on chromatin. A fraction of these repeats-derived RNAs is characteristic of endo-siRNAs. The demonstrated sufficiency of transfected endo-siRNA mimics in rescuing heterochromatin defects in Dcr-2-deficient S2 cells strongly suggests the contribution of *trans*-acting endo-siRNAs to heterochromatin formation/maintenance, as established in fission yeast. Further work is clearly needed to test specific complexes involved and other mechanistic aspects to extend the model to metazoans. As DNA loci with more active local transcription seem to inversely correlate to their ability to attract *trans*-acting RNAs from related repeat species, RNAs released from those more active loci may supply extra RNAs to act in *trans* on less transcribed loci. This "community" act of repeat-derived RNAs may thus help "patch up" certain damaged heterochromatic regions in the cell.

Dicer (DICER1 in mammals and Dcr-2 in fly) and AGO2 have been shown to regulate mitosis in both fly and mammalian cells (Pek and Kai 2011; Huang et al. 2015), but it has remained unclear how these siRNA machineries function in the cell cycle control. Our data now suggest that the mitotic defects likely result from compromised heterochromatin in pericentromeric regions and show for the first time that a transfected pool of CHARR-derived endo-siRNA mimics is sufficient to not only restore heterochromatin but also rescue the mitotic defects in Dcr-2-deficient S2 cells. This experiment was not attempted earlier, likely because of the assumption that Dcr-2 deficiency can also compromise siRNA loading into AGO2, as siRNA biogenesis and loading appears to be coupled by Dcr-2/R2D2 in the fly (Lee et al. 2004; Liu et al. 2006). Our data demonstrate that transfected endo-siRNA mimics are clearly functional in Dcr-2-deficient cells, implying sufficient loading of synthetic siRNAs into AGO2 under Dcr-2 shortage conditions at least in cultured S2 cells. Here, it is also important to note that Dcr-2/AGO2 is not essential in development (Lee et al. 2004; Okamura et al. 2004), indicating that its critical role in cell cycle progression might be partially compensated in live animals. Moreover, the more visible phenotype might also be due to continuous and rapid cell division of S2 cells in culture, which is not the case with many cell types in live animals.

The function of repeat-derived small RNAs in maintaining heterochromatin for stable genetic inheritance in somatic cells emphasizes a key and immediate benefit of active retrotransposons to the genome, although retrotransposons have been traditionally viewed as mutagens to cause genome instability. In *Drosophila* germ cells, an RNA amplification mechanism has been evolved to maximally suppress this mutagen function of retrotransposons.

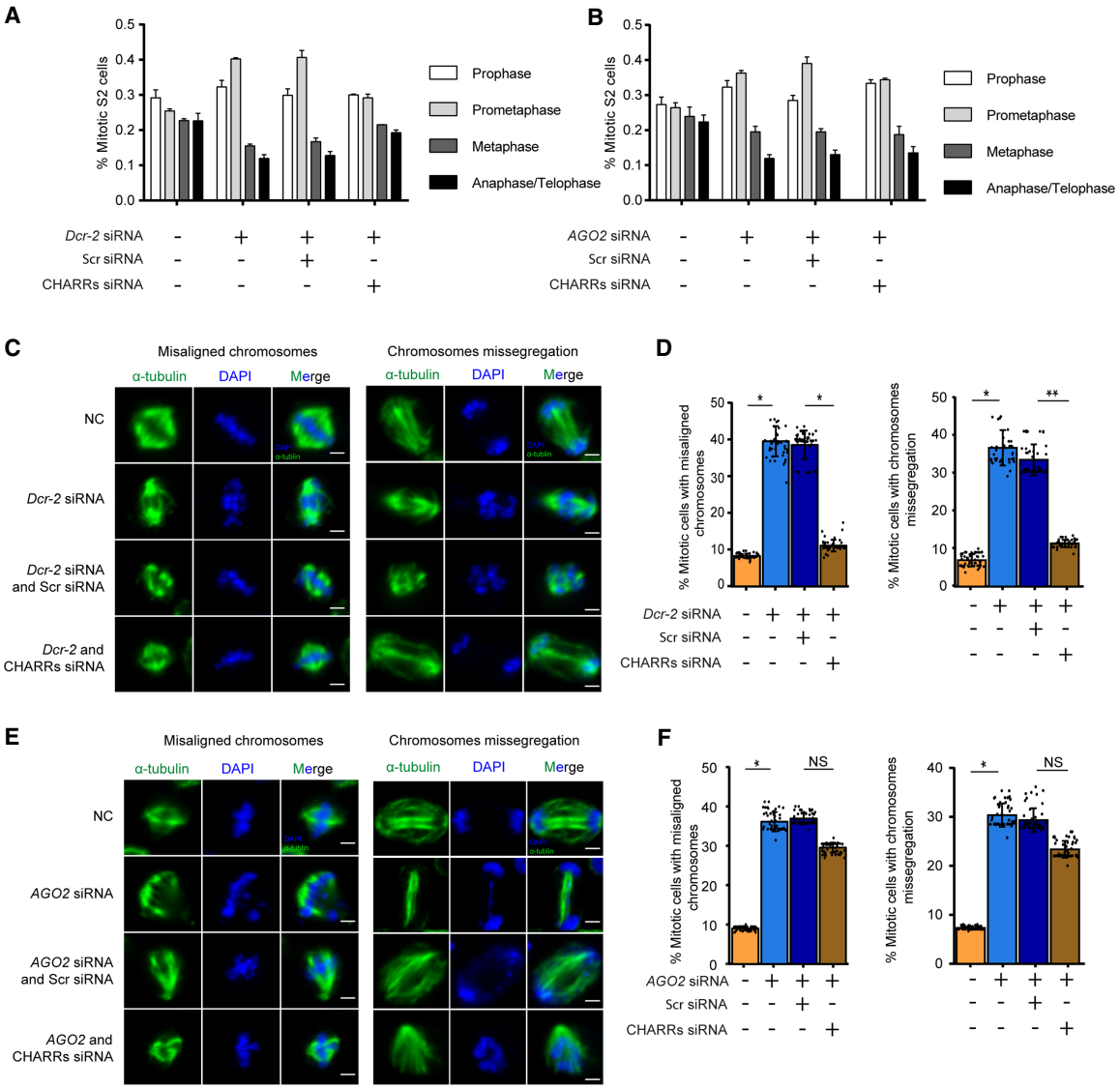

Figure 7. CHARRs can rescue *Dcr-2* knockdown- but not *AGO2* knockdown-induced cell division defects. (*A*,*B*) Percentages of S2 cells at each stage of mitosis in response to knockdown of *Dcr-2* (*A*) or *AGO2* (*B*) and treatment with either scrambled or derived siRNAs. $n = 50$ for each experiment. (*C*) *Dcr-2* knockdown-induced chromosomes misalignment (*left*) and missegregation (*right*) and rescue by derived siRNAs, but not scrambled siRNA. (Green) Stained α-tubulin, (blue) DAPI. Scale bar, 2 μm. (*D*) Percentages of S2 cells at metaphase exhibiting misaligned chromosomes (*left*) and anaphase exhibiting lagging chromosomes (*right*) at different experimental conditions. $n = 50$ for each condition. (*) $P < 0.05$, (**) $P < 0.01$ (multiple group Student's *t*-test). (*E*,*F*) Similar to *C* and *D* except on *AGO2* knockdown cells. $n = 50$ for each condition. (*) $P < 0.05$, (NS) not significant (multiple group Student's *t*-test).

Despite the lack of such RNA amplification mechanism in somatic cells, we now show that the siRNA machinery is quite active in diminishing full-length retrotransposon transcripts and to ensure the supply of repeat-derived small RNAs for stable genetic inheritance during cell division. As higher eukaryotic genomes are populated with enormous amounts of repeat sequences, we suggest that some of those repeat DNA sequences may have important functions, while most others are fossils of genome evolution.

Methods

Cell lines and cell culture conditions

S2 cells, negative for mycoplasma contamination, were cultured under sterile conditions at 26°C in Schneider medium (Thermo Fisher Scientific) containing 10% heat-inactivated fetal bovine serum and 100 μg/mL penicillin-streptomycin. Specific assay kits have been provided at Supplemental Table S4.

Alignment of GRID-seq reads to the *Drosophila* genome

GRID-seq raw reads were split into RNA and DNA reads according to the designed bivalent linker. Trimmomatic (Bolger et al. 2014) was used to remove adapter sequences and filter low-quality reads by using the parameters *MINLEN: 18* and *SLIDINGWINDOW: 2:20*. Filtered reads were aligned to the *Drosophila* genome (genome version :dm6) with ShortStack (Axtell 2013) using the parameter *–m 400* for DNA reads and *–m 200* for RNA reads, respectively. Multiple mapped reads were weighted based on the frequencies of neighbor uniquely mapped reads (Supplemental Fig. S1C).

Unmapped reads were cleaned with SAMtools (Li et al. 2009) using the parameter *–F 4*, and the weighted score of each read was recorded in the fifth column of a BED file. Annotation and processing of RNA and DNA reads from the GRID-seq libraries and comparison with ChAR-seq libraries as well as with specific chromatin marks are detailed in Supplemental Methods.

siRNA-mediated knockdown of *Dcr–2* and *AGO2*

Transfection was performed as described (Rogers and Rogers 2008), using the "bathing" method for siRNA delivery. Cells were counted, pelleted, and resuspended at $1–5 \times 10^6$ cells/mL in serum-free media. We added ~10–30 µg dsRNA to each well in the six-well tissue culture plate to obtain the final concentration of 25–50 nM. About 1 mL of cells were seeded in each well in the six-well plate and incubated at room temperature for 30 min, followed by the addition of 3 mL complete media with 10% FBS to each well. This process was repeated every other day three times before harvesting the cells for downstream assays. Mock-treated cells (all reagents except dsRNA) and *GFP* kd cells served as controls for *Dcr-2* and *AGO2* kd with validated dsRNAs (Pek and Kai 2011; Taliaferro et al. 2013). The *GFP* dsRNA template was PCR-amplified from pEGFP-N1 using *GFP* siRNA forward and reverse primers. Primers of individual dsRNAs are listed in Supplemental Table S3. Specific plasmid and assay kits are provided in Supplemental Table S4.

ChIP-seq library construction and data analysis

DNA libraries were constructed using the NEBNext Ultra II DNA Library Prep kit (New England Biolabs) following the manufacturer's recommendations, as detailed in Supplemental Methods.

Analysis of mitotic defects

Media and fetal bovine sera were batch-tested for support of normal cell growth and RNAi efficiency. For kd, specific synthetic siRNAs were added to cell culture in 24-well plates. After siRNA treatment for 4 d, cells were resuspended and transferred to glass-bottom, 24-well plates (Cellvis) and allowed to adhere for 2.5 h before fixation (Goshima et al. 2007). Cells were fixed in 4% paraformaldehyde for 10 min, permeabilized with 0.1% Triton X-100 in PBS for 5 min, and incubated overnight at 4°C with anti-α-tubulin (1:1000; ab7291 Abcam) in PBS containing 0.1% Triton X-100 and 0.5 mg/mL BSA, followed by staining with secondary antibodies and DAPI (1 µg/mL). For rescue, siRNA treatment was performed as above followed by transfection of endo-siRNAs 2 d later. Immunostained specimens were imaged under a Zeiss LSM-700 confocal laser scanning microscope, using a 63× 1.4 NA oil immersion objective to achieve high resolution. We imaged two channels (DAPI, AF488) at typically 10–20 sites/well to obtain 50 metaphase cells/well on average. Specific assay kits are specified in Supplemental Table S4.

Quantification and statistical analysis

Statistical parameters are reported either in individual figures or corresponding figure legends. Quantified data are, in general, presented as bar/line plots, with the error bar representing mean ± SEM, or box plots showing the median (middle line), first and third quartiles (box boundaries), and furthest observation or 1.5 times of the interquartile (end of whisker). All statistical analyses were done in R (R Core Team 2018).

Data access

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE134307. All custom scripts in this study are provided in the Supplemental Code.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

We thank Dr. Yang Yu for providing S2 cells. This work was supported by grants from the National Natural Science Foundation of China (91940306, 31871294, 31520103905) and the National Key R&D Program of China (2016YFC0901702, 2016YFC09 010002) to R.S.C. and S.M.H. The first author, Y.J.H., initiated this work during graduate study and completed the project as a postdoc at UC, San Diego, supported by National Institutes of Health grants (HG004659 and GM052872) to X.-D.F.

Author contributions: Conceptualization, X.-D.F., R.S.C., and S.M.H.; methodology development and data analysis, Y.J.H.; generation of the GRID-seq data in S2 cells, X.L.; experimental design and execution, Y.J.H., D.P.W., and S.H.W.; data interpretation and discussion, Y.J.H., X.-D.F., R.S.C., S.M.H., D.P.W., S.H.W., P.Z., J.-Y.C., C.W.S., and D.-H.L.; paper writing, Y.J.H., S.M.H., and X.-D.F.

References

Allshire RC, Nimmo ER, Ekwall K, Javerzat JP, Cranston G. 1995. Mutations derepressing silent centromeric domains in fission yeast disrupt chromosome segregation. *Genes Dev* **9:** 218–233. doi:10.1101/gad.9.2.218

Axtell MJ. 2013. ShortStack: comprehensive annotation and quantification of small RNA genes. *RNA* **19:** 740–751. doi:10.1261/rna.035279.112

Becker JS, Nicetto D, Zaret KS. 2016. H3K9me3-dependent heterochromatin: Barrier to cell fate changes. *Trends Genet* **32:** 29–41. doi:10.1016/j.tig.2015.11.001

Bell JC, Jukam D, Teran NA, Risca VI, Smith OK, Johnson WL, Skotheim JM, Greenleaf WJ, Straight AF. 2018. Chromatin-associated RNA sequencing (ChAR-seq) maps genome-wide RNA-to-DNA contacts. *eLife* **7:** e27024. doi:10.7554/eLife.27024

Bolger AM, Lohse M, Usadesl B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30:** 2114–2120. doi:10.1093/bioinformatics/btu170

Csink AK, Henikoff S. 1996. Genetic modification of heterochromatic association and nuclear organization in *Drosophila*. *Nature* **381:** 529–531. doi:10.1038/381529a0

Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel JA, Sachidanandam R, et al. 2008. An endogenous small interfering RNA pathway in *Drosophila*. *Nature* **453:** 798–802. doi:10.1038/nature07007

Deshpande G, Calhoun G, Schedl P. 2005. *Drosophila* argonaute-2 is required early in embryogenesis for the assembly of centric/centromeric heterochromatin, nuclear division, nuclear migration, and germ-cell formation. *Genes Dev* **19:** 1680–1685. doi:10.1101/gad.1316805

Drysdale R, FlyBase Consortium. 2008. Flybase: a database for the *Drosophila* research community. *Methods Mol Biol* **420:** 45–59. doi:10.1007/978-1-59745-583-1_3

Ekwall K, Javerzat JP, Lorentz A, Schmidt H, Cranston G, Allshire R. 1995. The chromodomain protein Swi6 - a key component at fission yeast centromeres. *Science* **269:** 1429–1431. doi:10.1126/science.7660126

Fagegaltier D, Bougé AL, Berry B, Poisot E, Sismeiro O, Coppée JY, Théodore L, Voinnet O, Antoniewski C. 2009. The endogenous siRNA pathway is involved in heterochromatin formation in *Drosophila*. *Proc Natl Acad Sci* **106:** 21258–21263. doi:10.1073/pnas.0809208105

Franke A, Baker BS. 1999. The *rox1* and *rox2* RNAs are essential components of the compensasome, which mediates dosage compensation in *Drosophila*. *Mol Cell* **4:** 117–122. doi:10.1016/S1097-2765(00)80193-8

Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, Xu J, Kittler EL, Zapp ML, Weng Z, et al. 2008. Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* **320:** 1077–1081. doi:10.1126/science.1157396

Goshima G, Wollman R, Goodwin SS, Zhang N, Scholey JM, Vale RD, Stuurman N. 2007. Genes required for mitotic spindle assembly in *Drosophila* S2 cells. *Science* **316:** 417–421. doi:10.1126/science.1141314

Grewal SI, Jia S. 2007. Heterochromatin revisited. *Nat Rev Genet* **8:** 35–46. doi:10.1038/nrg2008

Grewal SI, Klar AJ. 1997. A recombinationally repressed region between *mat2* and *mat3* loci shares homology to centromeric repeats and regulates directionality of mating-type switching in fission yeast. *Genetics* **146:** 1221–1238.

Halic M, Moazed D. 2009. Transposon silencing by piRNAs. *Cell* **138:** 1058–1060. doi:10.1016/j.cell.2009.08.030

Hall LL, Carone DM, Gomez AV, Kolpa HJ, Byron M, Mehta N, Fackelmayer FO, Lawrence JB. 2014. Stable C_0T-1 repeat RNA is abundant and is associated with euchromatic interphase chromosomes. *Cell* **156:** 907–919. doi:10.1016/j.cell.2014.01.042

Holoch D, Moazed D. 2015a. RNA-mediated epigenetic regulation of gene expression. *Nat Rev Genet* **16:** 71–84. doi:10.1038/nrg3863

Holoch D, Moazed D. 2015b. Small-RNA loading licenses Argonaute for assembly into a transcriptional silencing complex. *Nat Struct Mol Biol* **22:** 328–335. doi:10.1038/nsmb.2979

Huang C, Wang X, Liu X, Cao S, Shan G. 2015. RNAi pathway participates in chromosome segregation in mammalian cells. *Cell Discov* **1:** 15029. doi:10.1038/celldisc.2015.29

Jain R, Iglesias N, Moazed D. 2016. Distinct functions of Argonaute Slicer in siRNA maturation and heterochromatin formation. *Mol Cell* **63:** 191–205. doi:10.1016/j.molcel.2016.05.039

Kandasamy SK, Fukunaga R. 2016. Phosphate-binding pocket in Dicer-2 PAZ domain for high-fidelity siRNA production. *Proc Natl Acad Sci* **113:** 14031–14036. doi:10.1073/pnas.1612393113

Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, Okada TN, Siomi MC, Siomi H. 2008. *Drosophila* endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* **453:** 793–797. doi:10.1038/nature06938

Kellum R, Alberts BM. 1995. Heterochromatin protein 1 is required for correct chromosome segregation in *Drosophila* embryos. *J Cell Sci* **108:** 1419–1431.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res* **19:** 1639–1645. doi:10.1101/gr.092759.109

Lee YS, Nakahara K, Pham JW, Kim K, He Z, Sontheimer EJ, Carthew RW. 2004. Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117:** 69–81. doi:10.1016/S0092-8674(04)00261-2

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25:** 2078–2079. doi:10.1093/bioinformatics/btp352

Li X, Zhou B, Chen L, Gou LT, Li H, Fu XD. 2017. GRID-seq reveals the global RNA-chromatin interactome. *Nat Biotechnol* **35:** 940–950. doi:10.1038/nbt.3968

Liu X, Jiang F, Kalidas S, Smith D, Liu Q. 2006. Dicer-2 and R2D2 coordinately bind siRNA to promote assembly of the siRISC complexes. *RNA* **12:** 1514–1520. doi:10.1261/rna.101606

Muerdter F, Guzzardo PM, Gillis J, Luo Y, Yu Y, Chen C, Fekete R, Hannon GJ. 2013. A genome-wide RNAi screen draws a genetic framework for transposon control and primary piRNA biogenesis in *Drosophila*. *Mol Cell* **50:** 736–748. doi:10.1016/j.molcel.2013.04.006

Okamura K, Ishizuka A, Siomi H, Siomi MC. 2004. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev* **18:** 1655–1666. doi:10.1101/gad.1210204

Okamura K, Chung WJ, Ruby JG, Guo H, Bartel DP, Lai EC. 2008. The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs. *Nature* **453:** 803–806. doi:10.1038/nature07015

Pek JW, Kai T. 2011. DEAD-box RNA helicase Belle/DDX3 and the RNA interference pathway promote mitotic chromosome segregation. *Proc Natl Acad Sci* **108:** 12007–12012. doi:10.1073/pnas.1106245108

Peng JC, Karpen GH. 2007. H3k9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability. *Nat Cell Biol* **9:** 25–35. doi:10.1038/ncb1514

Peters AH, O'Carroll D, Scherthan H, Mechtler K, Sauer S, Schöfer C, Weipoltshammer K, Pagani M, Lachner M, Kohlmaier A, et al. 2001. Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107:** 323–337. doi:10.1016/S0092-8674(01)00542-6

Quinodoz SA, Ollikainen N, Tabak B, Palla A, Schmidt JM, Detmar E, Lai MM, Shishkin AA, Bhat P, Takei Y, et al. 2018. Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell* **174:** 744–757.e24. doi:10.1016/j.cell.2018.05.024

R Core Team. 2018. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/.

Rogers SL, Rogers GC. 2008. Culture of *Drosophila* S2 cells and their use for RNAi-mediated loss-of-function studies and immunofluorescence microscopy. *Nat Protoc* **3:** 606–611. doi:10.1038/nprot.2008.18

Saksouk N, Simboeck E, Déjardin J. 2015. Constitutive heterochromatin formation and transcription in mammals. *Epigenetics Chromatin* **8:** 3. doi:10.1186/1756-8935-8-3

Smit AFA, Hubley R, Green P. 2013–2015. *RepeatMasker Open-4.0*. http://www.repeatmasker.org.

Stein P, Svoboda P, Anger M, Schultz RM. 2003. RNAi: mammalian oocytes do it without RNA-dependent RNA polymerase. *RNA* **9:** 187–192. doi:10.1261/rna.2860603

Sun FL, Cuaycong MH, Craig CA, Wallrath LL, Locke J, Elgin SC. 2000. The fourth chromosome of *Drosophila melanogaster*: interspersed euchromatic and heterochromatic domains. *Proc Natl Acad Sci* **97:** 5340–5345. doi:10.1073/pnas.090530797

Taliaferro JM, Aspden JL, Bradley T, Marwha D, Blanchette M, Rio DC. 2013. Two new and distinct roles for *Drosophila* Argonaute-2 in the nucleus: alternative pre-mRNA splicing and transcriptional repression. *Genes Dev* **27:** 378–389. doi:10.1101/gad.210708.112

Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD. 2006. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313:** 320–324. doi:10.1126/science.1129333

Verdel A, Jia S, Gerber S, Sugiyama T, Gygi S, Grewal SI, Moazed D. 2004. RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303:** 672–676. doi:10.1126/science.1093686

Volpe T, Martienssen RA. 2011. RNA interference and heterochromatin assembly. *Cold Spring Harb Perspect Biol* **3:** a003731. doi:10.1101/cshperspect.a003731

Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, Martienssen RA. 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297:** 1833–1837. doi:10.1126/science.1074973

Yu R, Wang X, Moazed D. 2018. Epigenetic inheritance mediated by coupling of RNAi and histone H3K9 methylation. *Nature* **558:** 615–619. doi:10.1038/s41586-018-0239-3

Received August 19, 2019; accepted in revised form September 22, 2020.