# Trading Genome Vulnerability for Stable Genetic Inheritance: Active Retrotransposons Help Maintain Pericentromeric Heterochromatin Required for Faithful Cell Division

Yajing Hao[1,2,3,4], Dongpeng Wang[1,2,4], 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, *], Shunmin He[1, *], and Runsheng Chen[1, *]

[1]Key Laboratory of RNA Biology, Institute of Biophysics, Chinese Academy of Sciences, 100101, China

[2]University of Chinese Academy of Sciences, Beijing, 100101, China

[3]Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA 92093-0651, USA

[4] These authors contributed equally

*Corresponding authors:

Xiang-Dong Fu (xdfu@ucsd.edu); Shunmin He (heshunmin@ibp.ac.cn); Runsheng Chen (chenrs@sun5.ibp.ac.cn)

# ABSTRACT

Retrotransposons are extensively populated in vertebrate genomes, which, 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 attack these problems by defining the origins of repeat-derived RNAs and their specific chromatin registers in *Drosophila* S2 cells. We demonstrate that repeat RNAs are predominantly derived from active Gypsy elements, and upon their processing by Dicer-2, these endo-siRNAs act in *cis* and *trans* to help maintain pericentromeric heterochromatin. Remarkably, we show that synthetic repeat-derived siRNAs are sufficient to rescue Dicer-2 deficiency-induced defects in heterochromatin formation in interphase and chromosome segregation during mitosis, thus demonstrating that active retrotransposons are actually required for stable genetic inheritance.

# INTRODUCTION

Eukaryotic genomes contain both gene-rich and gene-poor regions, respectively corresponding to euchromatin and heterochromatin. 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 predominately associated with centromeric and pericentromeric regions, telomeres, and retrotransposons(Lippman et al. 2004). Constitutive heterochromatin plays important roles in genome organization in the nucleus(Avner and Heard 2001; Zhang et al. 2019), 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). These critical biological functions underscore the importance of repeat-rich sequences underneath constitutive heterochromatin, which used to be referred to as "junk" DNA in the genome. In fact, besides their potential contribution to genome evolution, it has been unclear whether active retrotransposons have any immediate benefit to an organism.

Regarding the formation and maintenance of constitutive heterochromatin (thereafter generally referred to as heterochromatin), our current knowledge is largely derived from elegant genetic and biochemical studies in fission yeast and *Drosophila melanogaster*(Tschiersch et al. 1994; Grewal and Jia 2007; Holoch and Moazed 2015b). The most striking aspect of the emerging theme is that 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 are next processed by Dicer to produce small interfering RNAs (siRNAs), which are then loaded onto Ago1 to form the RNA-induced transcription silencing (RITS) complex to target nascent repeat RNA. RITS recruits a key histone methyltransferase Clr4 (Su(var)3-9 in flies and SUV39H1 in humans) to generate H3K9me2/3, which then attracts its reader Swi6 (HP1 in flies and humans), 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 melanogaster* appears to follow a similar scheme except using the piRNA system to process and amplify repeat-derived RNAs to establish heterochromatin to actively repress retrotransposition in the germline(Vagin et al. 2006; Halic and Moazed 2009; Muerdter et al. 2013; Iwasaki et al. 2015).

While the general conceptual framework for heterochromatin formation has been well established, there are multiple puzzles that remain to be solved. 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 to be patched up. Second, in principle, some repeat-derived RNAs may also be capable of acting in *trans*, as suggested by a recent analysis of 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 generally applies to heterochromatin maintenance on all pericentromeric regions, and if

78 so, what is the relative contribution of *cis*- versus *trans*-acting RNAs to such maintenance. Third,
79 both fission and *Drosophila* germ cells are equipped with an RNA amplification system (RNA-
80 dependent RNA polymerase in fission yeast and the ping-pong cycle in the fly), but such system
81 is lacking in somatic cells of flies and mammals(Stein et al. 2003). The question is how somatic
82 cells meet this supply/demand dilemma. Forth, fission yeast uses siRNA to drive heterochromatin
83 formation and maintenance, but fly enlists the piRNA pathway for such purpose in the germline.
84 This raises the question of whether different organisms may use distinct machineries for processing
85 repeat-derived RNAs. Last, but not least, Dicer deficiency has been reported to cause chromosome
86 mis-segregation during mitosis in both flies and mammals(Pek and Kai 2011; Huang et al. 2015),
87 but it has remained unclear whether such dramatic phenotype results from impaired production of
88 some sort of endo-siRNAs or other function(s) of Dicer in the nucleus.

89 Addressing the above questions would require critical information on where repeat-rich
90 transcripts are generated and where these repeat RNAs target specific loci in the genome, which
91 has been a challenging problem. Recent technological innovations have made it possible to
92 comprehensively elucidate the RNA-chromatin interactome, which has been applied to multiple
93 cell types, including *Drosophila* S2 cells(Li et al. 2017; Sridhar et al. 2017; Bell et al. 2018), but
94 rich information on repeat-derived RNAs has not yet been explored. We have now utilized this
95 information to define the interaction of repeat-derived RNAs with chromatin across the fly genome.
96 Our analysis reveals that endo-siRNAs are mostly derived from the Gypsy family of
97 retrotransposons, which are able to act in *cis* and *trans* to balance local transcription to help

maintain pericentromeric heterochromatin. Strikingly, we show that a pool of synthetic endo-siRNA mimics is sufficient to rescue chromosome segregation defects in Dicer-deficient cells. These findings reveal that active retrotransposons are functionally required for maintaining heterochromatin to ensure stable genetic inheritance during cell cycle.

# RESULTS

## Strategy for genome-wide assignment of multi-mapped RNA and DNA reads

We recently developed a technology called global RNA-DNA interaction sequencing (GRID-seq) to detect chromatin-associated RNAs and their respective binding sites genome-wide. GRID-seq employs a bivalent linker to ligate to RNA in one end and fragmented DNA in the other end on fixed nuclei followed by selection and cleavage of linker ligated products with a type IIS restriction enzyme (Mme I) to generate "mated" RNA and DNA (both ~20nt in length) for deep sequencing(Li et al. 2017). To control for the specificity in RNA-DNA mating, we also generated a GRID-seq library on mixed human and fly cells, thus enabling the construction of a critical background model by using cross-species reads (i.e. fly RNA ligated to human DNA and vice versa). Using these datasets, we previously explored uniquely mapped RNA-DNA mates to reveal nascent RNA-covered "transcription hubs" where specific promoters and enhancers are interconnected in the nucleus(Li et al. 2017). Given such high-quality data, we herein explored the biological significance of multi-mapped RNA and DNA reads in *Drosophila* S2 cells, taking advantage of high density reads on the much smaller genome of fly relative to humans.

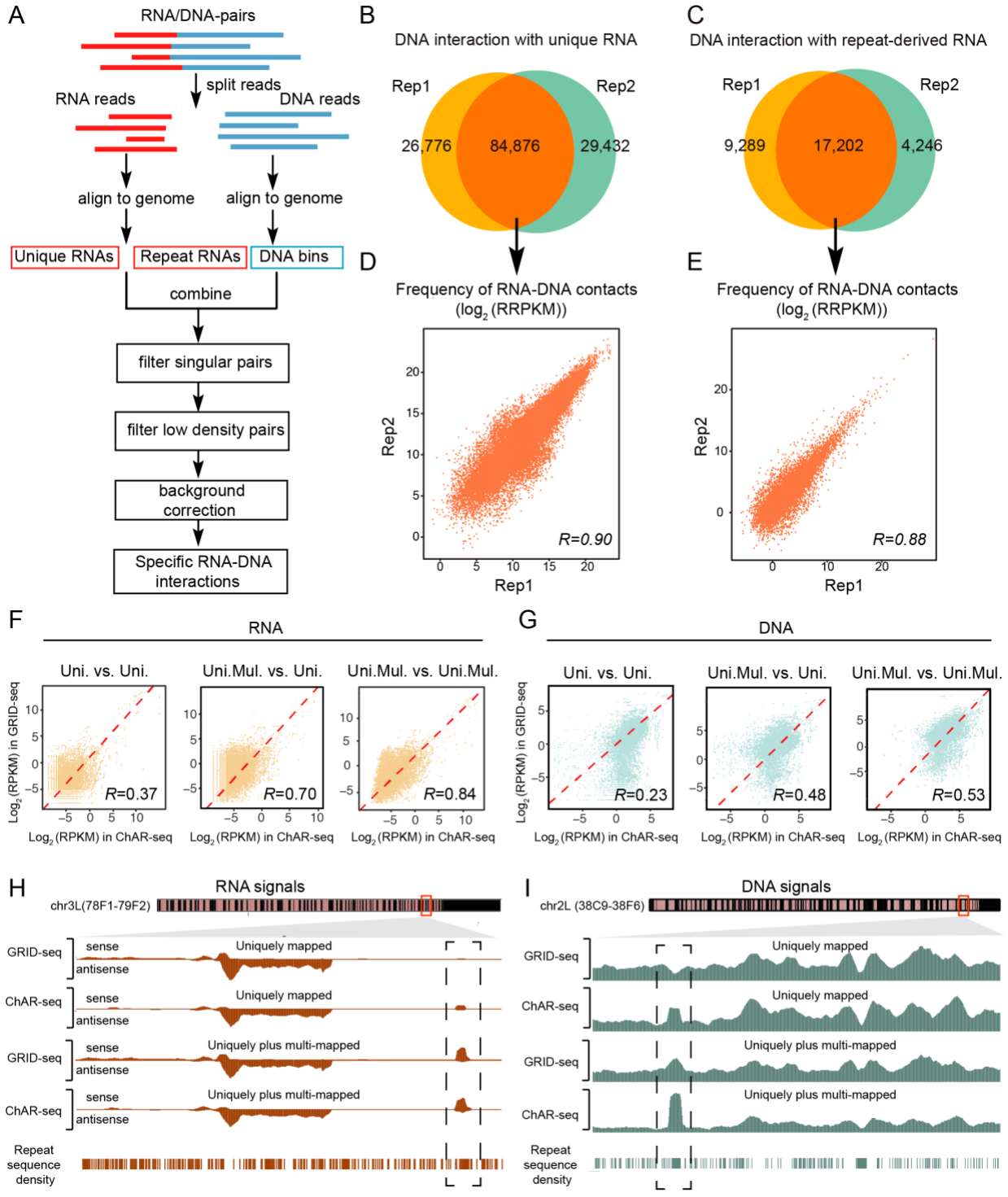



**Figure 1. Strategy for assigning multi-mapped RNA and DNA reads**

(A) Schematic presentation of the GRID-seq data processing pipeline. (B and C) Overlapped RNA-DNA mates containing annotated unique (B) or repeat-derived (C) RNA between the two independent GRID-seq libraries.(D and 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 and G) Comparison between GRID-seq and ChAR-seq with uniquely or uniquely plus multi-mapped RNA (F) or DNA (G) in repeat-enriched Alu I-generated DNA bins. Uni.: uniquely mapped reads. Uni.Mul.: uniquely plus multi-mapped reads. (H and I) Transcribed RNA signals (H) or RNA-contacted DNA signals (I) obtained with different mapping strategies on representative genomic regions.

The two independently generated GRID-seq libraries on S2 cells (Supplementary Table 1) were of high global concordance between RNA or DNA reads (Supplementary Fig. 1A). Interestingly, both replicates showed a high percentage (65% to 75%) of RNA reads that were mapped to repeat-derived transcripts, but a much smaller fraction (10% to 20%) of DNA reads were assigned to repeat-rich genomic loci (Supplementary Fig. 1B). To fully utilize these multi-mapped RNA and DNA 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 aligned each RNA read to genome and assigned it to annotated unique RNA transcripts according to the FlyBase database(Drysdale 2008) or repeat-derived transcripts based on the RepeatMasker database (http://www.repeatmasker.org), and each DNA read to genomic fragments(Roberts et al. 2015), generated by Alu I, a restriction enzyme used to fragment the fly genome during library construction. For multi-mapped reads, we distributed them to individual RNA transcripts or DNA fragments according to the relative density of uniquely mapped reads (Supplementary Fig. 1C).

We next filtered out singular RNA-DNA mates and low-density mates below a set threshold according to the Poisson distribution of all mates across the genome (see Online

146 METHODS). Additionally, we also subtracted the background based on human RNA signals
147 mapped to fly DNA loci from our fly/human GRID-seq library. The density of such background
148 reads is quite significant in many accessible chromatin regions (Supplementary Fig. 1D), as
149 detailed earlier(Li et al. 2017). After these data processing steps, retained RNA-DNA mates show
150 high consistency between the two independent GRID-seq libraries, as indicated by predominant
151 common mates associated with both annotated unique RNAs (Fig. 1B) and repeat-derived
152 transcripts (Fig. 1C). This consistency is also reflected at the quantitative levels of individual
153 common RNA-DNA mates (Fig. 1D,1E), thus enabling us to rely on these common mates to
154 generate the final RNA-DNA interactome for downstream analysis. Notably, after assigning multi-
155 mapped RNA and DNA reads, most gaps around repeat-rich DNA regions were "filled" to the
156 similar levels, as compared to adjacent unique regions across the fly genome (Supplementary Fig.
157 2A).
158
159 **Validation of mapping results with an independent dataset**
160 It is critical to validate our mapping strategy, even though ShortStack has been generally
161 accepted as a strategy to dynamically assign multi-mapped reads in a given genome. For this
162 purpose, we utilized the data generated by ChAR-seq, a strategy similar to GRID-seq except longer
163 RNA and DNA reads were generated by sonication after linker ligation (Supplementary Fig.
164 2B)(Bell et al. 2018). Compared to GRID-seq that generates predominantly mated RNA-DNA
165 pairs, ChAR-seq tends to trade off relative longer reads with a large fraction of unmated RNA or

166 DNA reads from sequenced libraries. Moreover, it was not optimal to use the ChAR-seq data in
167 the first place for several reasons: (i) GRID-seq libraries were generated on *Drosophila* S2 cells
168 where many other types of genomic data are available for comparison (see Supplementary Table 2),
169 whereas ChAR-seq libraries were produced on a less commonly used *Drosophila* cell line (CME-
170 W1-cl8+), (ii) the vast majority of ChAR-seq reads was from a single library (see Supplementary
171 Table 1), thus prohibitive to assessing internal data reproducibility, and most importantly, (iii) one
172 of our GRID-seq libraries was constructed on mixed fly and human cells, thus permitting the use
173 of cross-species RNA-DNA mates to build a background for non-specific RNA-DNA interactions,
174 which is missing from the existing ChAR-seq libraries. Nevertheless, the available ChAR-seq data
175 with longer RNA and DNA reads provided an independent dataset to evaluate our strategy for
176 assigning multi-mapped RNA and DNA reads, despite the fact that GRID-seq and ChAR-seq
177 libraries were derived from different fly cell types.

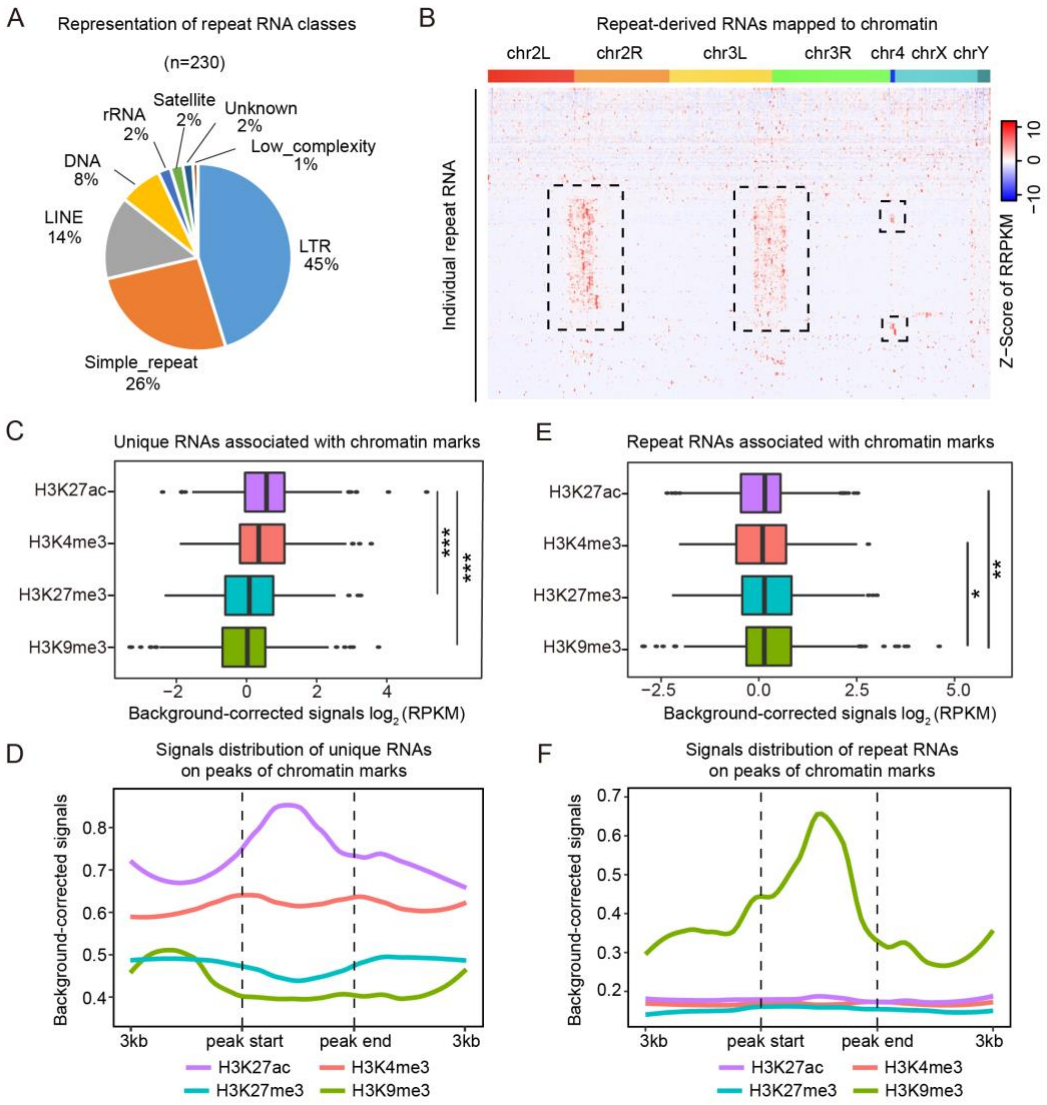
178 We first compared between ChAR-seq and GRID-seq data, observing an overall high
179 Spearman correlation (R=0.75 for RNA reads; R=0.62 for DNA reads) across the reference fly
180 genome (Supplementary Fig. 2C). However, when focused on repeat-enriched Alu I-generated
181 DNA bins, the correlation was quite modest at the levels of both RNA (left panel in Fig. 1F, R=0.37)
182 and DNA (left panel in Fig. 1G, R=0.23). Notably, a population of DNA reads (distributed in lower
183 right in Fig. 1G) was scored by ChAR-seq, but less by GRID-seq, likely due to the higher mapping
184 power of the former. Interestingly, the correlation was dramatically improved when comparing
185 uniquely mapped RNA or DNA reads from ChAR-seq with uniquely plus multi-mapped RNA or

DNA reads from GRID-seq (middle panels in Fig. 1F, R=0.70 for RNA and Fig. 1G, R=0.48 for DNA), suggesting that many assigned multi-mapped GRID-seq reads matched uniquely mapped ChAR-seq reads. As expected, after also assigning multi-mapped reads from ChAR-seq using ShortStack, the correlation was further improved when compared between uniquely plus multi-mapped RNA or DNA reads in both cases (right panels in Fig. 1F, R=0.84 for RNA and Fig. 1G, R=0.53 for DNA), which becomes comparable to the overall correlation across the genome (see Supplementary Fig. 2C, the remaining differences likely, at least in part, result from different cell types used to construct GRID-seq and ChAR-seq libraries). 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, missing from uniquely mapped GRID-seq reads, and then became visible after assigning multi-mapped GRID-seq reads. Together, these data validate our computational strategy to assign multi-mapped RNA and DNA reads to the genome.

**Preferential interaction of distinct RNA classes with eu- versus hetero-chromatin**

Having maximally utilized both uniquely and multi-mapped RNA and DNA reads from our GRID-seq libraries, we next wished to investigate how different classes of RNA might differentially interact with DNA in the fly genome. We were first focused on annotated non-repeat RNAs with respect to their interactions with the genome. As reported earlier(Li et al. 2017), among 4,856 non-repeat RNAs associated with chromatin, most showed interactions with DNA near their

206 sites of transcription (Supplementary Fig. 3A). Upon further subdividing different classes of such
207 RNAs, we noted that most annotated snoRNAs and snRNAs were not only expressed but also
208 extensively engaged in interactions with chromatin in *Drosophila* S2 cells, whereas virtually no
209 pre-miRNA was associated with chromatin (Supplementary Fig. 3B). The intermediate levels of
210 lncRNAs, mRNAs, and tRNAs on chromatin probably reflected wide expression ranges of these
211 RNA species, and thus, only those with sufficient expression were detectable on chromatin.




212

**Figure 2. Interaction of distinct RNA classes with eu- versus hetero-chromatin**

(A) Different repeat RNA classes represented by 230 DNA bound repeat-derived RNAs. (B) Heatmap 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<0.05, **p<0.01, *** 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.

In contrast to the chromatin binding patterns of annotated non-repeat RNAs, we also identified 230 repeat-derived RNA species that showed significant interactions with chromatin, which include all rRNAs, satellite DNA-transcribed RNAs (including those from simple repeats), and >70% of LINE and LTR-derived RNA species (Supplementary Fig. 3C). 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 two general patterns, one scattering across the genome and the other concentrating on pericentromeric regions in chromosome 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 various published epigenetic profiles (Supplementary Table 2). As expected, annotated non-repeat RNAs showed the highest association with H3K27ac (Fig. 2C), exhibiting coincidental peak summits (Fig. 2D). In contrast, repeat-derived RNAs displayed the greatest preference for H3K9me3 (Fig.

238 2E,F). Conversely, we sorted the restriction enzyme Alu I-generated DNA fragments according to
239 the levels of associated repeat-derived RNAs, observing that, among top 1,000 repeat RNA-
240 associated DNA bins, H3K9me3 was the dominant signal on these DNA regions (Supplementary
241 Fig. 3D). Together, these data suggest a general trend in which annotated non-repeat RNAs tend
242 to interact with euchromatic regions marked by H3K27ac, whereas repeat-derived RNAs prefer to
243 associate with heterochromatin marked by H3K9me3. While this is somewhat anticipated from the
244 vast literature, we now obtained critical information on which specific repeat-derived RNAs are
245 more prevalent than others in interacting with specific heterochromatic regions in the fly genome.
246 This established the critical foundation to investigate their relative contributions to the initiation
247 and/or maintenance of heterochromatin in *Drosophila* cells.

248

249 **Prevalent association of Gypsy-derived RNAs with constitutive heterochromatin**

250 We next asked which repeat-derived RNAs were more prevalent than others on
251 heterochromatin. We intersected the density of individual repeat-derived RNAs with markers for
252 constitutive heterochromatin characterized by the coordinated ChIP-seq signals for H3K9me3 and
253 its reader HP1 (Fig. 3A). By determining the co-localization coefficient for each of the 230 repeat-
254 derived RNAs between their DNA interaction frequencies and relative densities of H3K9me3 and
255 HP1 signals in 1Mb DNA bins of the fly genome, we identified 79 repeat-derived RNAs, including
256 two rRNAs, that showed the *Pearson* correlation coefficient of >0.3 (red dots in Fig. 3A). The
257 association of rRNAs with heterochromatin agrees with a recent observation that transcriptionally

inert centromere-proximal regions tend to be organized around the nucleolus in 3D genome(Quinodoz et al. 2017). However, since rRNAs are assembled into ribosomes, rather than processed into small RNAs, it is unlikely that they contribute to heterochromatin functions. Excluding rRNAs, we named the rest of heterochromatin-enriched RNAs as CHARRs (Constitutive Heterochromatin-Associated Repeats-derived RNAs). These CHARRs appear to show exclusive association with constitutive heterochromatin, as none of them exhibited significant co-localization with the facultative chromatin marker H3K27me3 (Supplementary Fig. 4A), as illustrated with three specific CHARRs (Supplementary Fig. 4B). Furthermore, these CHARRs showed little association with MSL (Supplementary Fig. 4C,D), a key component of the silencing complex involved in X inactivation in *Drosophila*, consistent with the nature of predominant facultative heterochromatin formed on X-inactivation (Xi)(Baker et al. 1994; Franke and Baker 1999).

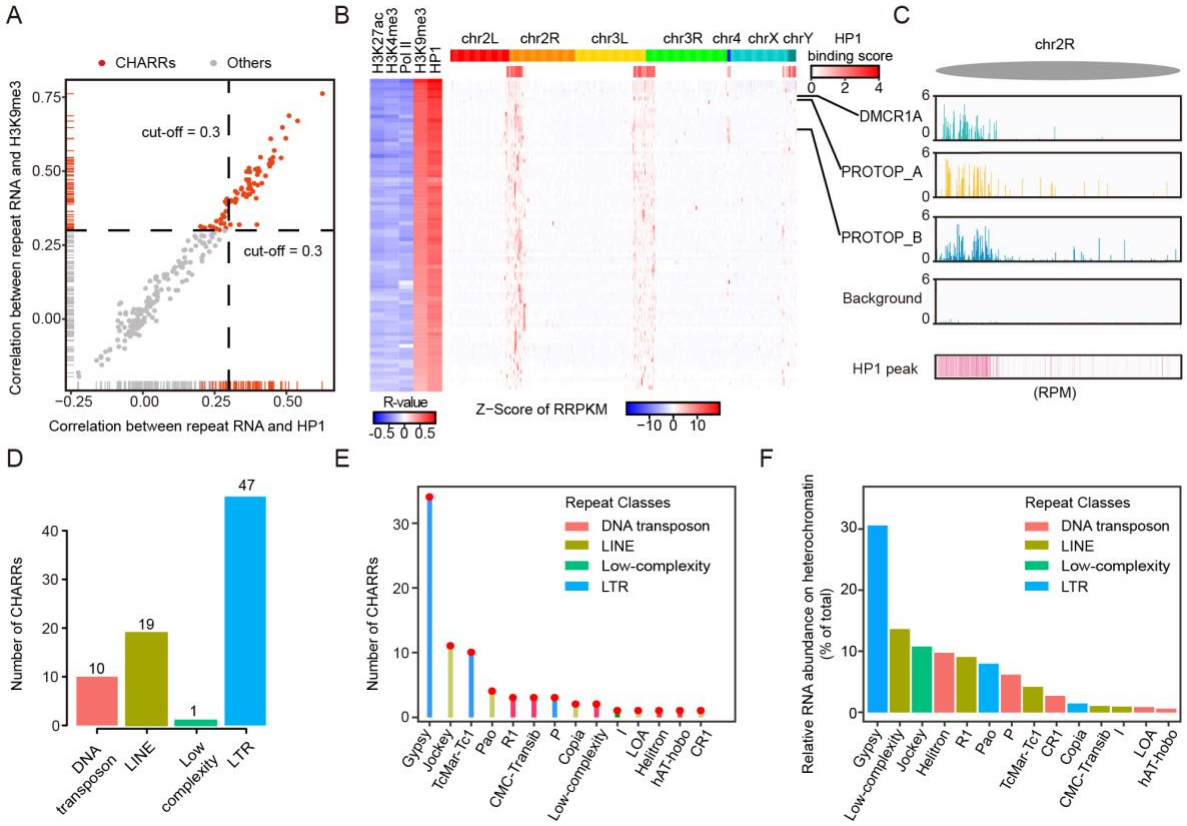



**Figure 3. Prevalent association of Gypsy-derived RNAs with constitutive heterochromatin**

(A) Scatterplot of co-localization 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 (grey dots). (B) Left: Heatmap of co-localization of each CHARR with individual chromatin marks or Pol II ChIP-seq signals. Right: Heatmap of chromatin interaction of each CHARR with Alu I-generated DNA bins, which were merged from 100 continuous Alu I 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 non-specific background based on human RNA mapped to fly DNA and HP1 binding density. All signals were scaled to reads per million. (D and E) The number of repeat classes (D) or sub-families (E) associated with CHARRs. (F) The relative RNA abundance (% of total) of CHARRs on constitutive heterochromatin and specific subfamilies they belong to. Colors show the RNA classes individual CHARR subfamilies belong to.

The 77 CHARRs we identified correspond to those predominantly associated with pericentromere of chromosome 2 and 3 where their interactions with DNA positively correlated to heterochromatin markers (e.g. H3K9me3 and HP1) and negatively to euchromatin markers (e.g. H3K27ac and H3K4me3) as well as RNAP II ChIP-seq signals (Fig. 3B). This is further highlighted with 3 specific repeat-derived RNAs (e.g. DMCR1A，PRTOP_A and PRTOP_B) on the right arm of chromosome 2 (chr2R, Fig. 3C). We next determined RNA class, genomic origins, and relative abundance for each CHARR. Interestingly, the largest RNA class of identified CHARRs corresponds to LTR (n=47), while the second largest class to LINE repeats (n=19) (Fig. 3D). The majority of RNA species from these two classes of retrotransposons belongs to the Gypsy and Jockey subfamilies, respectively (Fig. 3E). Most importantly, 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 the overall RNA signals 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 to determine both the source of repeat-derived RNAs and their registers on specific chromatin regions. Given the predominant mode of annotated non-repeat RNAs that act in *cis* (*cis* defined by mated RNA and DNA reads mapped to the same chromosomes, as supposed to trans defined by mated RNA and DNA reads

mapped to different chromosomes), we asked whether this also applied 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 (Supplementary Fig. 5A,B). This suggests multiple active rRNA transcription sites are in close spatial proximity with adjacent pericentromeric regions, as recently observed based on proximity ligation(Quinodoz et al. 2017). Importantly, we also identified CHARRs on multiple non-pericentromeric regions, suggesting their potential interactions with chromatin in both *cis*- and *trans*-modes.

We therefore determined the origins of CHARRs and their collective interactions with chromatin at the chromosomal levels. This analysis revealed their extensive chromatin interactions not only within the same chromosomes, but also across different chromosomes (Fig. 4A), suggesting a significant degree of *trans*-interactions. This prompted us to examine individual CHARRs to segregate their *cis*- and *trans*-interactions by first assigning their origins of transcription, and then determining their linkage to DNA on the same (intra) or different (inter) chromosomes. Interestingly, we found that about half (46.8%) of the CHARRs were predominantly 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 further illustrated with 3 representative CHARRs. Specifically, Gypsy12_LTR preferentially interacted with DNA on the same chromosome (Fig. 4C); Gypsy4_I-int was engaged in both intra- and inter-chromosomal interactions (Fig. 4D); and FW_DM seemed to mainly act in *trans* on other

chromosomes (Fig. 4E). It is also interesting to note that individual CHARRs all selectively bound Hi-C defined "B" domains in pericentromeric regions (Supplementary Fig. 5C,D).

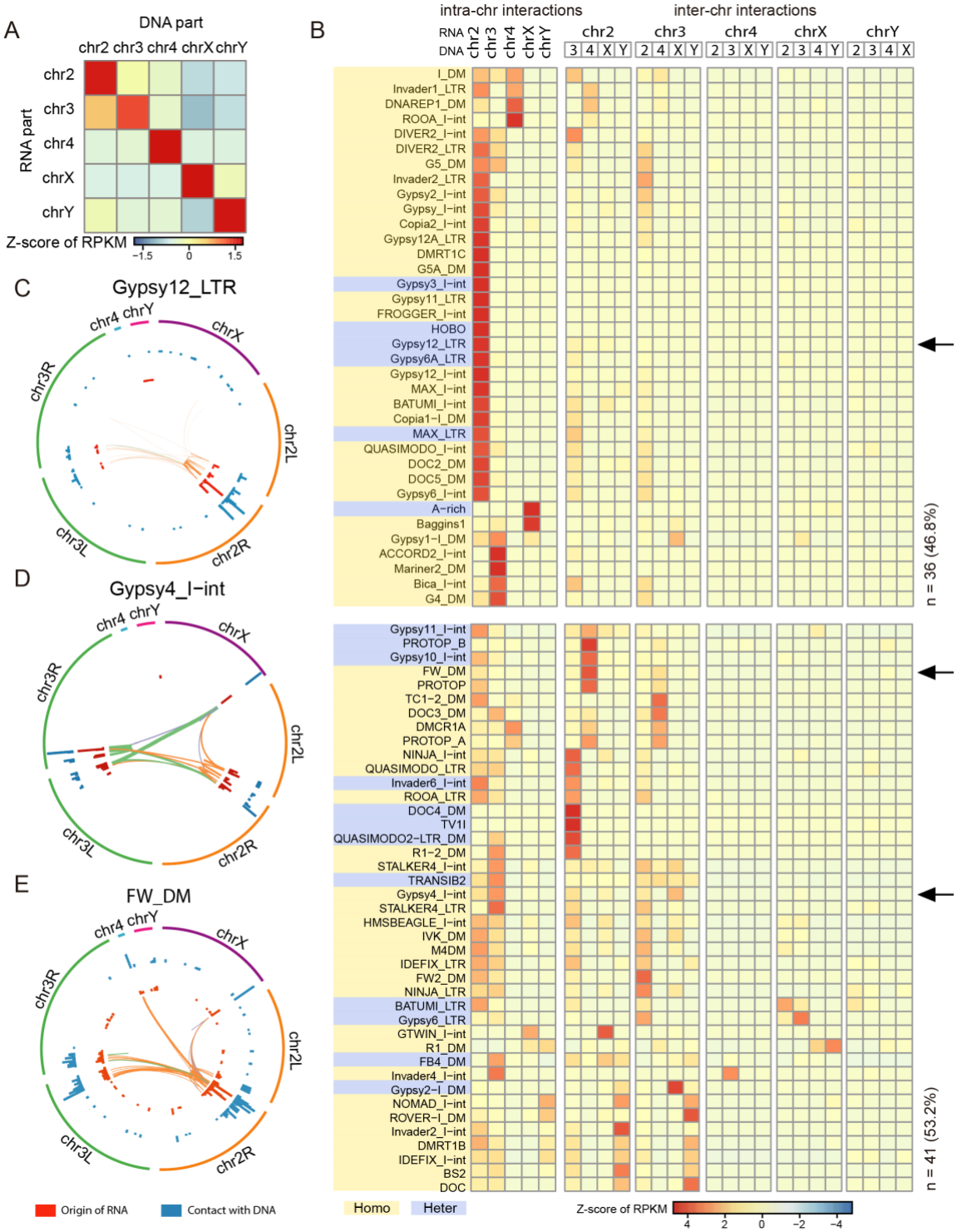


**Figure 4. *Cis*- and *trans*-acting repeat-derived RNAs on chromatin**

(A) Heatmap of RNA-DNA interaction scores for 77 CHARRs on the same or different chromosomes. *Trans*-chromosome interactions are referred to as RNA interacts with DNA in chromosomes other the RNA is transcribed from. (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 Gypsy12_LTR, Gypsy4_I-int and FW_DM, as individually illustrated in panel C, D, and E. All CHARRs are either highlighted with yellow or blue to indicate their preferences for homotropic or heterotropic interactions as described later. (C, D, E) The origins of three representative CHARRs and their DNA contacts, as shown by the Circos plots for Gypsy12_LTR (C), Gypsy4_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 100kb DNA windows. The heights of the signals correspond to Reads per 100kb window per millions. Lines specify intra- or inter-chromosomal interactions with width indicating the relative interaction levels in each case. The line color indicates the origin of chromatin-interacting RNAs.

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

During our analysis of *cis*- and *trans*-acting CHARRs, we further noted an interesting phenomenon where DNA loci expressing high levels of CHARRs tended to associate with less *trans*-acting RNAs (related repeat-derived RNA species transcribed from other chromosomes) and the converse was also true, as exemplified on a specific region in chromosome 2R where TV1I 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). Interestingly, *trans*-acting RNAs (transcribed from TV1I-related repeat species from other chromosomes) were mostly mapped to region 2 where little TV1I RNA was produced and 3

where a much lower level of the TV1I RNA was transcribed compared to region 1 (Fig. 5A, third track). We quantified these results by dividing all active TV1I loci into 3 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 *cis*- plus *trans*-acting RNAs for each group and observed a reverse correlation between local RNA production and the percentage of association with *trans*-acting RNAs (bars, Fig. 5B). This phenomenon also applied to another repeat RNA, Baggins1 (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, with a few exceptions, most followed this rule (Fig. 5E). Together, these observations suggest that highly transcribed CHARRs supply RNAs in *trans* to interact with DNA regions with less transcription activities.

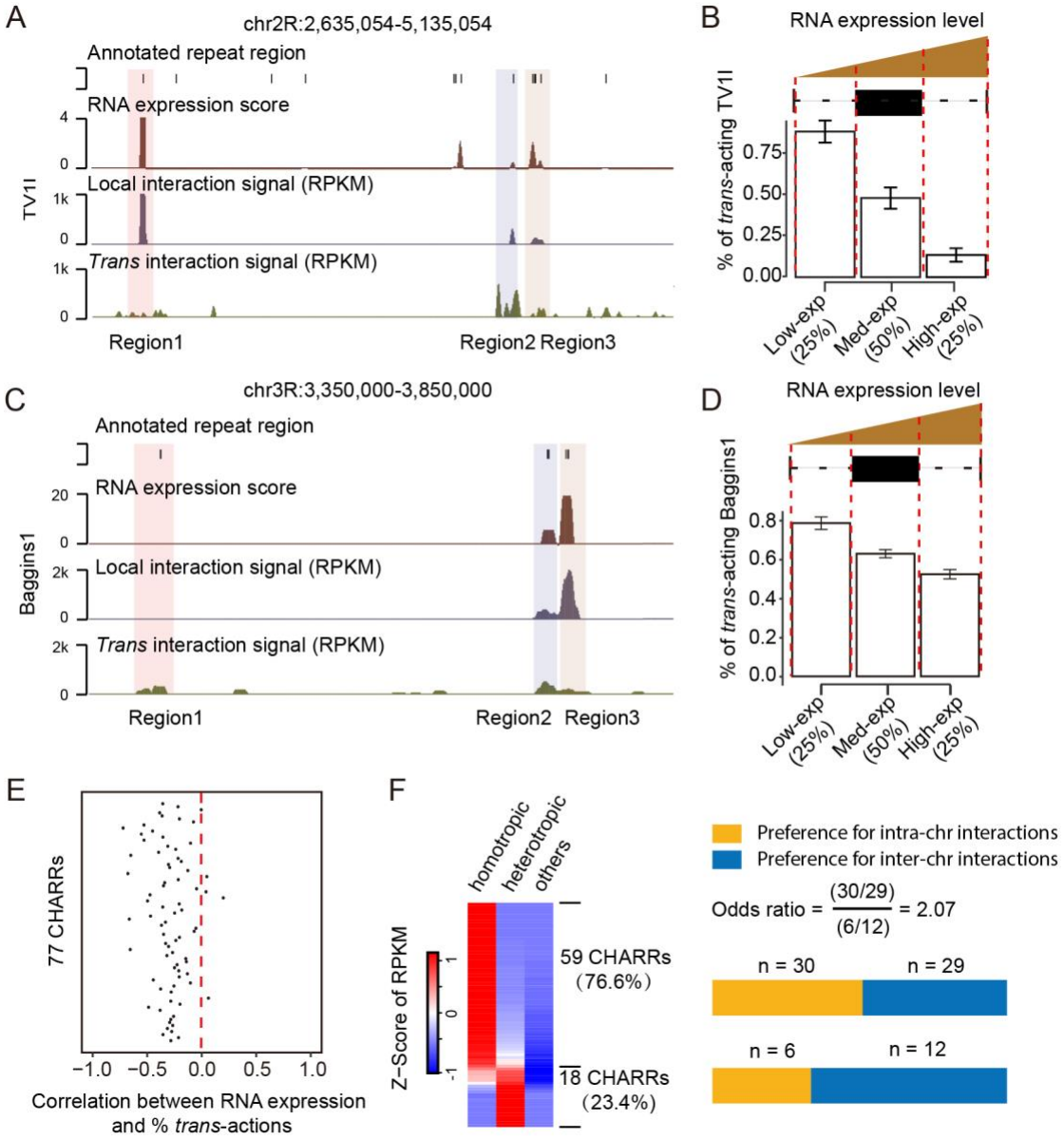



**Figure 5. Reverse correlation between local transcription and association with *trans*-acting RNAs on chromatin**

(A) A representative genomic region showing TV1I transcription loci and interactions with TV1I subfamily RNAs produced either locally or from other chromosomes (*trans*-acting RNAs). Three annotated TV1I transcription regions are indicated at bottom. (B) TV1I 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 TV1I RNAs on the transcription loci in each group. (C and D) Similar analysis and illustration for another typical CHARR Baggins 1 as in A and B.(E) Pearson correlation scores 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 Fig. 4.

Given the observation that some CHARRs preferred to interact with DNA near their sites of transcription while others showed both *cis*- and *trans*-chromosomal interactions, 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 1kb window, we called it "homotropic" interactions. If the interacting DNA region contains distinct repeat sequence (e.g. those encoding for RNAs of different classes or different subfamily members), we then referred it to as "heterotropic" interactions. 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 predominantly engaged in homotropic interactions, whereas the remaining 18 (blue-labeled in Fig. 4B) were more involved in heterotropic interactions, and none showed significant interactions with DNA that contain no repeat sequence (Fig. 5F, left). We further noted that relative to CHARRs with preference for homotropic interactions, more CHARRs with 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 may facilitate initiating heterochromatin formation, whereas those for heterotropic interactions may play more important roles in heterochromatin spreading as well as maintenance.

400

401 **Dicer-2 processed repeat-derived RNAs for heterochromatin maintenance**

402 Most retrotransposons are known to be bi-directionally transcribed. To verify this, we
403 took advantage of the existing GRO-seq data on *Drosophila* S2 cells (Supplementary Table 2),
404 showing that both sense and anti-sense repeat-derived transcripts were indeed represented in
405 relatively equal abundance, with rRNAs served as control for predominantly sense transcription
406 (Supplementary Fig. 6A). Interestingly, but not necessarily surprisingly, the identified CHARRs
407 were among the most abundant repeat-derived RNAs. This was true regardless of different RNA
408 classes (Supplementary Fig. 6B), which is in contrast to annotated non-repeat transcripts, the
409 majority of which was transcribed from the sense strand (Supplementary Fig. 6C), with exception
410 of some long intergenic non-coding RNAs (Supplementary Fig. 6D).

411 Bi-directionally transcribed repeat RNAs may thus provide dsRNA substrates for further
412 processing into endo-siRNAs to function in heterochromatin formation/maintenance, as
413 demonstrated earlier(Fagegaltier et al. 2009; Volpe and Martienssen 2011). Furthermore, in
414 contrast to piRNA-mediated heterochromatin formation in germline, Dicer-2 has been reported to
415 be specifically devoted to endo-siRNA processing in fly somatic cells(Lee et al. 2004; Pham et al.
416 2004; Czech et al. 2008; Ghildiyal et al. 2008). To determine whether the CHARRs we identified
417 all depended on Dicer-2 for their efficient processing and thus expression, we took advantage of
418 the existing small RNA-seq data (Supplementary Table 2) to compare their expression levels
419 between wild-type, Dicer-2 knockout, and Dicer-2 rescued S2 cells(Kandasamy and Fukunaga

2016). We found that almost all CHARRs were down regulated in response to Dicer-2 knockout and rescued in Dicer-2 re-expressed cells (Supplementary Fig. 7A). Note that much higher expression of those endo-siRNAs in Dicer-2 re-expressed cells likely resulted from Dicer-2 overexpression. Importantly, the RNA-seq reads of total CHARRs from wild-type S2 cells were distributed between 18 to 25nt in length, consistent with their processing into endo-siRNAs (Supplementary Fig. 7B). These endo-siRNAs appeared to have assembled into Ago2-containing complexes, as 54 (75%) CHARRs could be identified in the published Ago2 RIP data (Supplementary Fig. 7C). Therefore, re-analysis of these published data on CHARRs strongly suggests that Dicer-2 is responsible for processing CHARRs into endo-siRNAs to help maintain heterochromatin in S2 cells.

To directly test this hypothesis, we performed Dicer-2 knockdown in S2 cells (Supplementary Fig. 7D, E) and confirmed drastic reduction of small RNAs derived from DMCR1A or DOC by Northern blotting (Fig. 6A). Because many CHARRs were able to supply RNAs in *trans*, as we documented in the present study, we asked whether or not the associated phenotype previously detected on heterochromatin 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, FB_DM, Gypsy4_I-inU, DOC, Gypsy2-I and I_DM, Supplementary Table 3) and transfected this pool into S2 cells depleted of Dicer-2. We found that relative to wild-type cells, Dicer-2 knockdown reduced H3K9me3 as expected, and importantly, the CHARR-derived siRNA pool, but not scrambled siRNA, effectively

440 restored this heterochromatin marker, while the levels of H3K27me3 and H3K4me3 remained
441 constant under these experimental conditions (Fig. 6B, quantified below based on triplicated
442 experiments). This was also evident at the immunocytochemical level by staining for H3K9me3
443 (Fig. 6C). These data suggest that the CHARR-derived siRNA mimics were able to rescue
444 heterochromatin defects upon transfection (thus only acting in *trans*) into Dicer-2 deficient S2
445 cells.

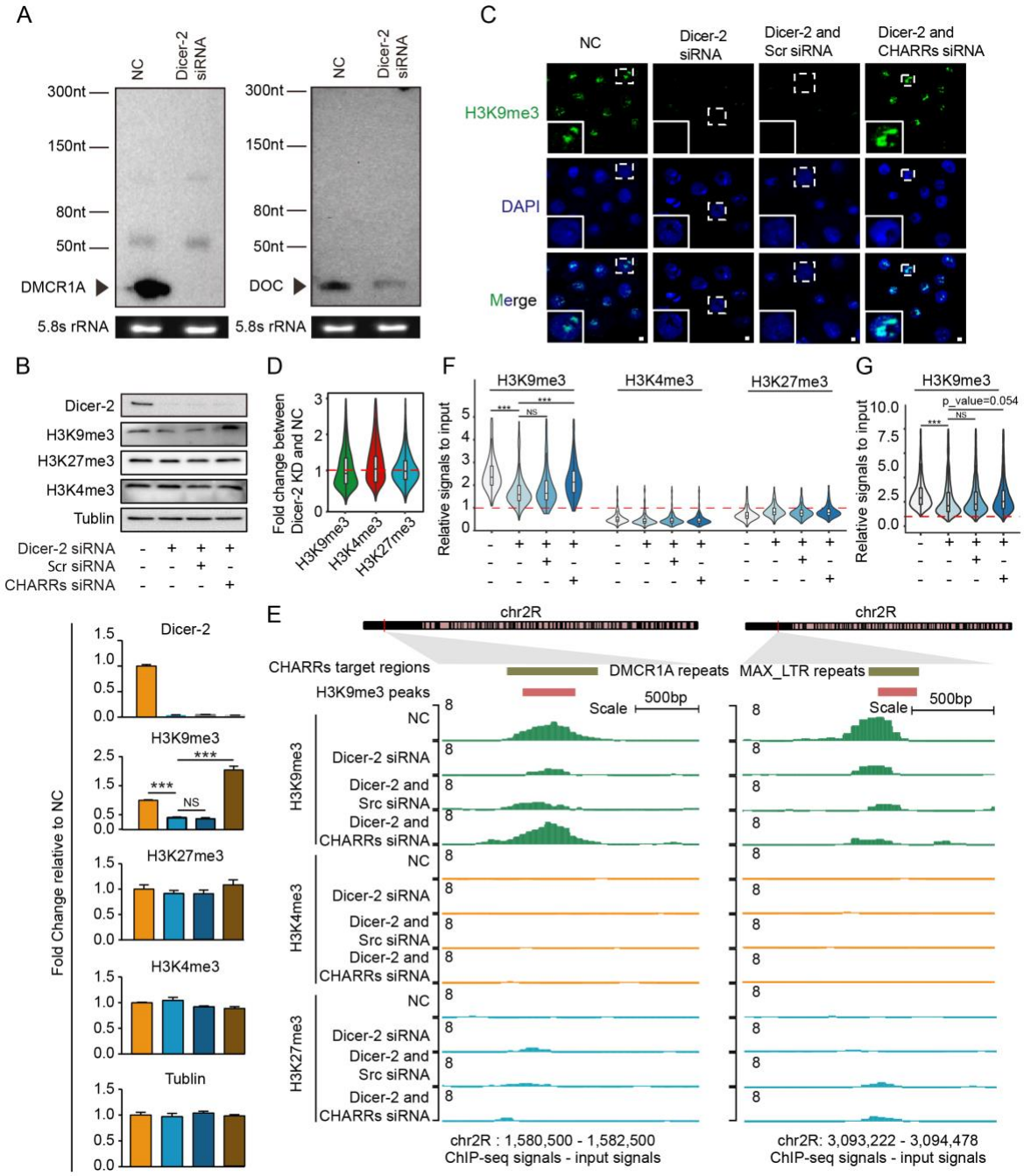



**Figure 6. Trans-acting repeat-derived RNAs for heterochromatin maintenance**

(A) Confirmation of Dicer-2 dependent expression of DMCR1A (left) and DOC (right) expression by Northern blotting analysis.(B) Western blotting analysis of Dicer-2, H3K9me3, H3K27me3, H3K4me3 and Tubulin in S2 cells in response to siRNA-mediated knockdown of Dicer-2 and rescue with a transfected pool of CHARR-derived synthetic siRNAs. Quantified data are shown below as fold-change (FC) relative to input (NC) in lane 1. Data are presented as mean ± SEM (n

453 = 3 biological replicates). *p<0.05, **p< 0.01, ***p<0.001. (unpaired Student's t test).(C)
454 H3K9me3 detected by immunocytochemistry in S2 cells treated with different combinations of
455 siRNAs, as indicated. Green, H3K9me3 signals; blue, DAPI. Scale bar, 2μm.(D) Violin plot for
456 fold-change of H3K9me3, H3K4me3 and H3K27me3 signals on CHARRs-targeted peaks in
457 response to Dicer-2 knockdown. KD: knockdown. (E) H3K9me3 ChIP-seq signals on
458 representative genomic loci, one corresponding to a CHARR-target locus (left) and a non-
459 CHARR-target locus (right) in response to Dicer-2 knockdown, complemented with either
460 scrambled or CHARRs-derived siRNAs. ChIP-seq signals for euchromatin (H3K4me3) or
461 facultative heterochromatin (H3K27me3) were shown for comparison.

462 **Rescuing global heterochromatin defects in Dicer-2 deficient cells**

463 To demonstrate the contribution to repeat-derived endo-siRNAs to heterochromatin
464 maintenance genome-wide, we next performed ChIP-seq for H3K9me3 in comparison with
465 H3K4me3 and K3K27me3 in response to Dicer-2 knockdown with or without treatment with
466 CHARR-derived siRNA mimics in S2 cells. Our ChIP-seq datasets were comparable to the profiles
467 of these histone marks published earlier on the same cell type, demonstrating the quality of our
468 data (Supplementary Fig. 8A, B). Upon Dicer-2 knockdown, we detected global decrease of the
469 H3K9me3 ChIP-seq signals, while the ChIP-seq signals for H3K4me3 and H3K27me3 were not
470 affected (Fig. 6D). The effects were also evident on individual H3K9me3-marked genomic loci
471 (see first two tracks in Fig. 6E), suggesting that Dicer-2 is functionally required for maintaining
472 heterochromatin genome-wide.

473 The next question was whether CHARR-derived siRNA mimics were able to rescue
474 heterochromatin defects, and if so, whether the rescue required their targeting specificity. Because
475 those siRNA mimics were designed to target 10 representative CHARRs (Supplementary Table
476 3), we thus analyzed the H3K9me3 ChIP-seq signals on 1kb-binned genomic regions that show

477 homology with at least one of the CHARR-derived siRNA mimics (Fig. 6F) in comparison with
478 genomic regions that showed H3K9me3 ChIP-seq signals but with <50% homology with any of
479 those CHARR-derived siRNA mimics (Fig. 6G). Indeed, we found that CHARR-derived siRNA
480 mimics, but not scrambled siRNA, effectively rescued H3K9me3 ChIP-seq signals on CHARRs
481 target genomic regions (compare lane 2 vs. 4 in Fig. 6F), but modest at best (likely due to a
482 remaining degree of heterotropic interactions) on none CHARRs target regions (compare lane 2
483 vs. 4, p=0.054, in Fig. 6G). As expected, little H3K4me3 and H3K27me3 ChIP-seq signals were
484 detected in H3K9me3-marked genomic loci. These general trends were also illustrated on two
485 representative genome loci for CHARRs targets (Fig. 6E, left and Supplementary Fig. 9A) and
486 none CHARRs targets (Fig. 6E, right and Supplementary Fig. 9B). Collectively, these data
487 demonstrated that a pool of *trans*-acting CHARR-derived siRNA mimics was able to bypass the
488 functional requirement of Dicer-2 for maintaining heterochromatin homeostasis on their target
489 regions in *Drosophila* S2 cells.

490

491 **CHARR-derived endo-siRNAs required for faithful chromosome segregation**

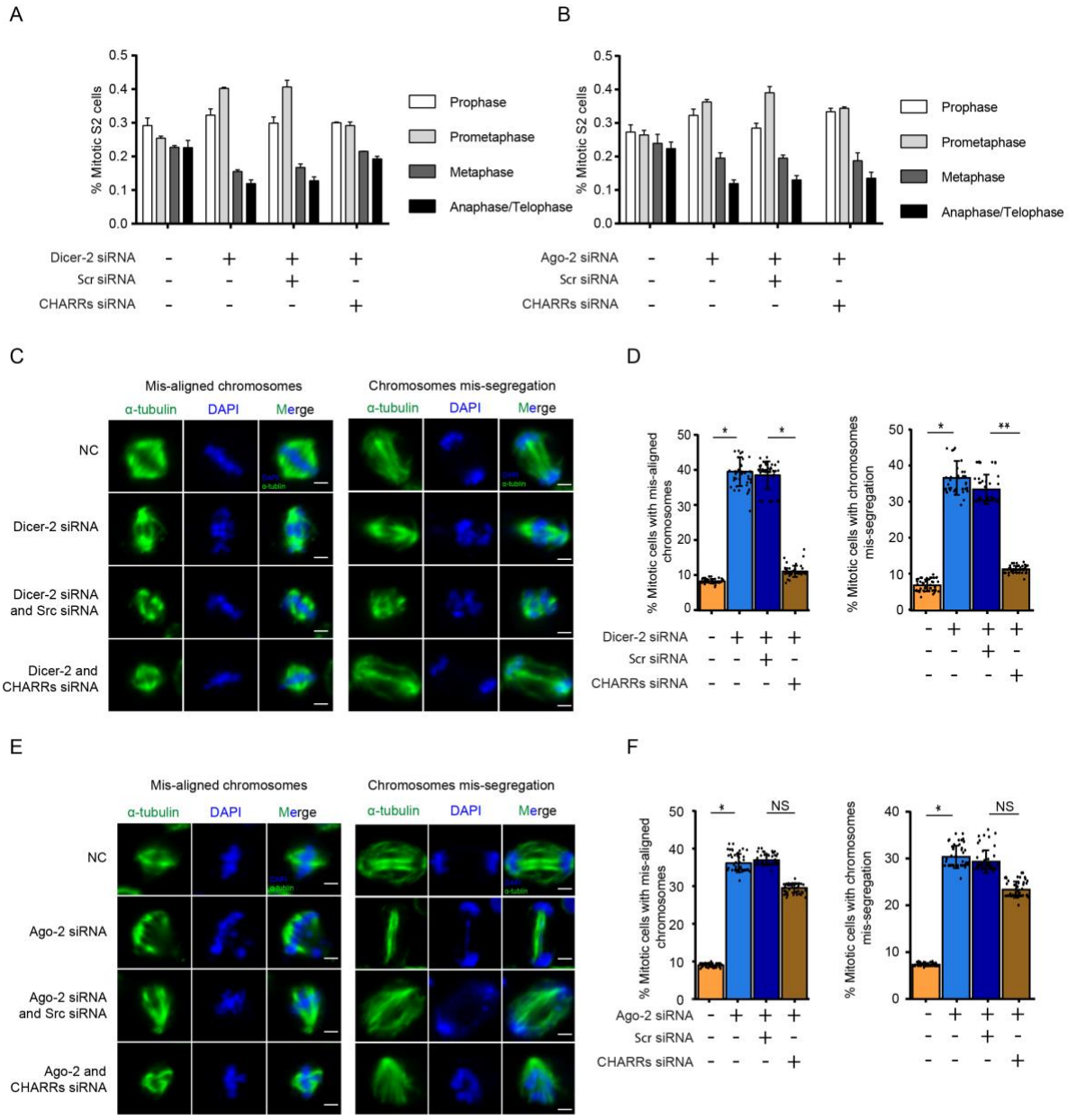
492 It has been previously documented that knockdown of Dicer-2 and Ago-2 caused
493 significant defects in chromosome segregation during cell division in S2 cells(Pek and Kai 2011).
494 Because Dicer-2 and Ago-2 are key siRNA pathway components, it is reasonable to extrapolate
495 that endo-siRNAs processed by Dicer-2 and loaded on Ago-2 are responsible for the phenotype,
496 but direct evidence for this critical conclusion has been lacking. Given that the CHARRs we

497 identified cover most of pericentromeric regions to maintain the heterochromatin status in the fly
498 genome (see Fig. 3B) and most of these CHARRs-derived were bound by Ago-2 (see
499 Supplementary Fig. 7C), we took advantage of the ability of CHARRs-derived siRNAs to rescue
500 most heterochromatin defects to ask whether these siRNA mimics were also able to rescue the
501 chromosome segregation defects. To this end, we first confirmed that knockdown of either Dicer-
502 2 or Ago-2 caused cell cycle defects, and as expected, we detected G1-S arrest in both cases (Fig.
503 7A, B and Supplementary Fig. 10). Importantly, we found that CHARRs-derived siRNAs, but not
504 scrambled siRNA, were able to rescue the cell cycle defects in Dicer-2 knockdown cells (Fig. 7A),
505 but not Ago-2 knockdown cells (Fig. 7B). These data are fully in line with the requirement of
506 Dicer-2 for processing CHARRs into small endo-siRNAs, which could be bypassed by the
507 transfected siRNA mimics, although these siRNA mimics would still need Ago-2 to execute their
508 functions in the cell.

509 We closely examined the mitotic defects in Dicer-2 knockdown cells, noting both mis-
510 aligned chromosomes and lagged as well as mis-segregated chromosomes (Fig. 7C and
511 Supplementary Fig. 11A) in about equal frequencies (Fig. 7D). Importantly, transfection of
512 CHARRs-derived siRNA mimics into these Dicer-2 deficient cells, but not scrambled siRNA,
513 were sufficient to rescue these defects (Fig. 7D and Supplementary Fig. 11A). In contrast, we
514 observed similar mitotic defects in Ago-2 knockdown cells, but the CHARRs-derived siRNAs
515 failed to correct the phenotype (Fig. 7E and Supplementary Fig. 11B). Taken together, these results
516 provide unequivocal evidence for active retrotransposon RNAs to preserve pericentromeric

517 heterochromatin homeostasis and thus cell cycle progression through the Dicer-2/Ago-2 mediated

518 endo-siRNA pathway.




519

520 **Figure 7. CHARRs could rescue Dicer-2 knockdown, but not Ago-2 knockdown, induced**

521 **cell division defects**

522 (A and B) Percentages of S2 cells at each stage of mitosis in response to knockdown of Dicer-2

523 (A) or Ago-2 (B) and treatment with either scrambled or CHARRs-derived siRNAs. n=50 for each

experiment. (C) Dicer-2 knockdown-induced chromosomes mis-alignment (left) and mis-segregation (right) and rescue by CHARRs-derived siRNAs, but not scrambled siRNA. Green, stained α-tubulin; blue, DAPI. Scale bar, 2μm.(D) Percentages of S2 cells at metaphase exhibiting mis-aligned chromosomes (left) and anaphase exhibiting lagging chromosomes (right) at different experimental conditions. n=50 for each condition. *p<0.05, **p<0.01, NS: not significant (multiple group Student's t test). (E and F) Similar to C and D except on Ago-2 knockdown cells. n=50 for each condition. *p<0.05, **p<0.01, NS: not significant (multiple group Student's t test).

## DISCUSSION

Genetic and biochemical experiments in fission and *Drosophila* have laid a general conceptual framework in understanding the formation and maintenance of constitutive heterochromatin in centromeric and pericentromeric regions(Ekwall et al. 1995; Kellum and Alberts 1995; Grewal and Jia 2007). Interestingly, however, 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(Buhler et al. 2006; Holoch and Moazed 2015a) also creates a puzzle in envisioning how heterochromatin could be actively maintained. As a matter of fact, it has been reported that repeat-derived RNAs are prevalent on chromatin in vertebrate cells from *Drosophila* to humans(Hall et al. 2014), implying that there is no shortage of repeat-derived RNAs to help maintain heterochromatin. The central question is what is the nature of these RNAs and where they come from?

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

547 attacked this fundamental problem by using the newly elucidated RNA-DNA interactome in
548 *Drosophila* somatic cells(Li et al. 2017). Taking advantage of such high density interactome, we
549 now show that various active retrotransposons, especially those from the Gypsy family of the LTR
550 class, produce a large amount of repeat RNAs that are selectively associated with pericentromeric
551 regions in the fly genome. Interestingly, while a subset of those repeat-derived RNAs tends to act
552 in *cis*, another subset appears to interact with DNA in both *cis* and *trans* modes. Particularly
553 interesting is a general trend we have uncovered, where DNA loci with more active local
554 transcription seem to inversely correlate to their ability to attract *trans*-acting RNAs from related
555 repeat species. Therefore, RNAs released from those more active loci may supply extra RNAs to
556 act in *trans* on less transcribed loci. This "community" act of repeat-derived RNAs may thus help
557 "patch up" certain heterochromatic regions "damaged" by chromatin remodeling activities in the
558 cell, such as that catalyzed by the H3K9me3 demethylase, thereby ensuring heterochromatin
559 homeostasis for stable epigenetic inheritance.

560 In fission yeast, repeat-derived RNAs are amplified by an RNA-dependent RNA
561 polymerase, which may supply a population of *trans*-acting RNAs to help initiate and/or maintain
562 heterochromatin in other genomic loci(Yu et al. 2018). A different mechanism is employed in
563 *Drosophila* germline where repeat-rich transcripts are processed and amplified by the piRNA
564 machinery to ensure efficient silencing of retrotransposons to protect the genome integrity as well
565 as to help maintain heterochromatin in centromeric and pericentromeric regions to ensure accurate
566 chromosome segregation during cell division in gonad(Muerdter et al. 2013; Iwasaki et al. 2015).

567 The problem is that such RNA amplification mechanism does not seem to exist in somatic cells of
568 flies and mammals. The endo-siRNA pathway has clearly been implicated in *Drosophila* somatic
569 cells(Czech et al. 2008; Ghildiyal et al. 2008), although it remains unclear whether such endo-
570 siRNA pathway also operates in somatic cells of mammals, which might result from less attention
571 paid on processing of repeat-derived RNAs by the conserved siRNA machinery known to be
572 highly active in mammals. In any case, at least in *Drosophila* S2 cells, we now document that such
573 endo-siRNA pathway is responsible for generating small RNAs that can act in *cis* and *trans* to help
574 maintain heterochromatin homeostasis in pericentromeric regions. We envision a similar
575 mechanism that may also operate in somatic cells in mammals, which requires future investigation
576 once the RNA-DNA interactome of much higher density becomes available.

577 Knockdown of Dicer (Dicer-1 in mammals and Dicer-2 in fly) and Ago-2 has been shown
578 to cause mitotic defects in both fly and mammalian cells(Pek and Kai 2011; Huang et al. 2015),
579 but it has been unclear whether these small RNA machineries act through their traditional functions
580 or through some new mechanism(s) in the nucleus. We now show that a pool of synthetic repeat-
581 derived siRNA mimics is able to rescue all measurable cell cycle defects in *Drosophila* S2 cells,
582 which is fully compatible to the central role of the endo-siRNA pathway in ensuring cell cycle
583 progression through maintaining pericentromeric heterochromatin. We speculate the conversed
584 function of this pathway in mammals, although different classes of retrotransposons are likely
585 involved in different organisms.

The production of repeat-derived small RNAs to help maintain heterochromatin for stable genetic inheritance in somatic cells suggests a key and immediate benefit of active retrotransposons for the genome. This realization is interesting because retrotransposons have been traditionally viewed as mutagens in the genome, although their relatively random actions may facilitate genome evolution in the long run. In *Drosophila* germ cells, an RNA amplification mechanism has been evolved to maximally suppress this mutagen function of retrotransposons to ensure genome integrity. We now show that despite the lack of such RNA amplification mechanism in somatic cells, the endo-siRNA machinery is still quite active to ensure the supply of repeat-derived small RNAs. Therefore, without worrying about the responsibility to transmit genetic materials to offspring, somatic cells may tolerate potential genome vulnerability in exchange for stable genetic inheritance during development and differentiation. As higher eukaryotic genomes are populated with enormous amounts of repeat sequences, we suggest that some of those "junk" DNA sequences actually have important functions while most others are fossil of genome evolution.

# METHODS

**Cell Lines and Cell Culture Conditions**

S2 cells were cultured under sterile conditions at 26°C in Schneider medium (Invitrogen) containing 10% heat-inactivated fetal bovine serum and 100 µg/ml penicillin-streptomycin. S2 cell lines are negative for mycoplasma contamination.

**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 (Supplementary Fig. 1C). 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 bed file.

**Annotation of RNA and DNA reads**

Annotated unique RNAs of *Drosophila* was downloaded from FlyBase(Drysdale 2008) and the repeat sequences from RepeatMasker track in the UCSC genome browser(Jurka et al. 2005). RNA reads were annotated to unique and repeat RNAs using IntersectBed tool with the parameter *-f 1.0*, *-split* and *–s*. The *Drosophila* genome was scanned with EMBOSS (Rice et al. 2000) based on Alu I restriction sites from REBASE (Roberts et al. 2015) to generate AluI DNA bins. IntersectBed tool was used to connect Alu I DNA bins using the parameter *-f 1.0*.

**Assigning RNA-DNA interactions**

Multiple read pairs that have the same mapped RNA and DNA loci associated with the same PCR primer sequences were considered PCR duplicated, and thus were counted only once. The following equation was used to compute RNA-DNA interactions in each AluI DNA bin:

$$Num_{(interaction)} = \sum_{i=1}^{n} (R_i * D_i)$$

Where *n* is the number of assigned read pairs. For each read pair *i*, $R_i$ is the contribution score of this read to the RNA part and $D_i$ is the contribution score of this read to the interacting AluI DNA bin. $R_i = 1$, if the RNA part of the read pair is uniquely mapped, or equals to a fraction based on the weighted score.

**Construction of non-specific background using mixed GRID-seq libraries**

A mixed GRID-seq library from human MDA-MB-231 and *Drosophila* S2 cells was used to construct the non-specific RNA-DNA interaction profile. RNA reads, which were only mapped to the human genome (hg38) using Bowtie(Langmead et al. 2009) with the parameter *–n 0*, were kept. Their mated DNA reads were processed using ShortStack, as described above. The human RNA signals within each 1kb DNA bin were normalized to one million, which was further smoothed by a moving window that includes 5 upstream and 5 downstream bins. The final coverage of the 1kb DNA bin *i* is:

$$Cov_i = \frac{1}{11}\sum_{i-5}^{i+5}\left(\frac{10^6 * \sum_m Read_{im}}{N}\right)$$

where *m* is the number of reads mapped to the 1k DNA bin *i* and *N* is the total read number mapped to the *Drosophila* genome.

To make DNA binding scores comparable between Alu I binned versus 1kb binned genome, RNA binding signals in each 1k DNA bin were converted to RNA binding signals in each AluI DNA bin by first dividing each 1kb DNA bin into 1000 1bp bins to calculate the signal in each small bin based on $\frac{Cov_{bin}}{1000}$. IntersectBed tool with the parameter *–f 1.0* was used to compute signals in each Alu I DNA bin by summing the signals from all 1bp bins in the fragment. Finally, signals in AluI DNA bins were all normalized to signals per 1kb:

$$Score_{AluI\ DNA\ bin\ j\ in\ mixed\ library} = \left(10^3 * \sum_{m=1}^{Len_j} Cov_m\right) \Big/ Len_j$$

where $Len_j$ is the length of the AluI DNA bin *j*.

**Filtering singular and background to identify specific RNA-DNA interactions**

To support specific RNA-DNA interactions, we required at least two RNA-DNA mates for each RNA transcript in a given Alu I DNA bin. We also developed two background models to simulate Poisson distribution of RNA binding signals on DNA. The first was based on uniform distribution of individual RNAs on DNA, based on which we estimated the background score:

$$Score_{interaction\ i-j} = (Len_j * N_i) \Big/ Len_g$$

where the $Len_g$ is the length of the genome， $Len_j$ is the length of AluI DNA bin *j* and $N_i$ is total signals for RNA *i*. This score was further normalized according to length in each AluI DNA bin, and the resulting RPK (reads per kilobase) was used as the $\lambda_{BG\_i}$ value to obtain the Poisson distribution of this RNA-engaged genomic interactions and to calculate the p-value for such interactions. The ratio of $Num_{interaction\ i-j}$ over $Score_{interaction\ i-j}$ was reported as the fold_change (FC) above the background. We also developed a second background model with that data deduced with RNA signals from the mixed library. We first calculated the non-specific interaction score ($Score_{AluI\ DNA\ bin\ j\ in\ mix\ library}$) based on human-derived RNA binding signals, and then used this score as the $\lambda_{BG\_mix}$ value to obtain the Poisson distribution of human RNA-engaged genomic interactions and to calculate the p-value for such interactions. The ratio of length and sequencing depth normalized $Num_{interaction\ i-j}$ (in RPKM) over $\lambda_{BG\_mix}$ was reported as the fold_change (FC). RNA-DNA interactions that met the requirement of p-value <0.05 and the fold_change (FC) >2 based on both background models were considered specific and thus retained for further analysis.

**Data normalization for comparison between GRID-seq libraries and different RNAs within the same libraries**

The interaction RNA-DNA score for each RNA is affected by the sequencing depth in different GRID-seq libraries, the length of each Alu I DNA bin, and the length of each RNA. To enable comparison among different libraries and different RNAs within the same libraries, we normalized these variables according to:

$$RRPKM_{ij} = {(Num_{ij} * 10^6 * 10^3 * 10^3)}\Big/{(TotalReadCounts * Len\_D_j * Len\_R_i)}$$

where $Num_{ij}$ is the interaction score of RNA *i* with Alu I DNA bin *j*, *TotalReadCounts* is the sequencing depth of individual libraries, $Len\_D_j$ is the length of Alu I DNA bin *j* and $Len\_R_i$ is the length of RNA *i*.

**Comparison between GRID-seq and ChAR-seq datasets**

ChAR-seq raw data were downloaded from the GEO database (Supplementary Table 2). PCR duplicates were removed using Clumpify with default parameters. All five independent ChAR-seq libraries were combined followed by adapter trimming and filtering low-quality reads with Trimmomatic using the parameters *MINLEN: 36, LEADING:3 TRAILING:3* and *SLIDINGWINDOW: 4:15*. Filtered ChAR-Seq reads were split into paired RNA and DNA reads, and then processed as described above for the GRID-seq data. FeatureCounts was used to calculate reads in 1kb DNA bins and then converted to RPKM using edgeR package in R language. For validation for assigning multi-mapped reads, we only considered 1kb DNA bins that contain newly assigned multi-mapped reads from the GRID-seq dataset.

**Comparison of RNA-DNA interactions in relationship with chromatin marks**

Public ChIP-seq data from S2 cells were downloaded from the SRA database (Supplementary Table 2). Fastq-dump were used to covert raw SRA data to Fastq. Quality control and data processing were similar to the procedures for processing the GRID-seq data with parameters adjusted to *MINLEN: 36* and *SLIDINGWINDOW: 4:20* according to the read length. Filtered reads were mapped to the reference *Drosophila* genome (dm6) using STAR (Dobin et al. 2013) with the parameters: *--outFilterScoreMinOverLread 0.1*, *–outFilterMatchNminOverLread 0.1*, *--alignIntronMax 1* and *--alignEndsType EndToEnd*. The wig files of each chromatin mark ChIP-seq dataset was obtained through STAR using the parameters: *–outWigTypewiggle read1_5p*, *--outWigStrandUnstranded* and *--outWigNormRPM*. Enriched peaks were detected by MACS2 with input data as control. Top 500 peaks were used for further analysis.

**Identification of repeat RNA on constitutive heterochromatin**

The ChIP-seq data for constitutive heterochromatin markers (H3K9me3 and HP1) were obtained from the GEO database (Supplementary Table 2) and the RPKM values were calculated on 10kb

DNA bins. The Pearson correlation score was calculated on each DNA bin containing binding signals from repeat RNAs. Repeat sequences, except rRNAs, were defined as CHARRs if the correlation score with H3K9me3 or HP1 is >0.3.

**Analysis of CHARR-DNA interactions relative to Hi-C defined compartments**

Public Hi-C data from S2 cells were downloaded (Supplementary Table 2) and processed with Trimmomatic using default settings plus trim tool in Homer using the parameter *-3 AAGCTT*. Trimmed reads with length of ≥38nt were mapped to the *Drosophila* genome using end-to-end alignment model provided by Bowtie2 (Ref (Langmead and Salzberg 2012)). We discarded potential PCR duplicates as well as reads with no useful information, including (1) read pairs separated <1.5× of the sequenced insert fragment length, (2) reads from 10kb regions containing >5× of the average coverage, (3) read pairs lacking restriction sites at the 3' end of either read within the estimated fragment length, (4) reads with their ends resulting from self-ligation with adjacent restriction fragments. We then used the filtered dataset to perform PCA analysis with HOMER, using runHiCpca.pl with the parameters *–res 10kb* and *–superRes 20kb* and compartments analysis using runHiCpca.pl and findHiCCompartments.pl. Finally, CHARR-DNA interactions signals and Hi-C compartments were intersected with IntersectBed from BedTools. The length of Hi-C compartment A and B was normalized to 1Mb.

**Processing GRO-seq, small RNA-seq and Ago2 RIP data from S2 cells**

All data were downloaded from public databases (Supplementary Table 2). GRO-seq data were processed as with the ChIP-seq data for chromatin marks. Annotated unique and repeat RNA species were used to calculate individual transcription scores by using FeatureCounts and EdgeR.

Small RNA-seq data were similarly processed as above using adjusted Trimmomatic parameters: *SLIDINGWINDOW:4:24* and *MINLEN:18*. SortMeRNA(Kopylova et al. 2012) were used to filter ribosomal RNAs with default setting. We also used ShortStack to map both uniquely and multi-mapped reads to obtain the RPM value for each CHARR.

Reads for the downloaded Ago2 RIP data were blasted to identify repeat RNA-derived sequences using the parameters *-outfmt 6* and *-word_size 7*. Reads will be kept for further analysis if the mismatch number is under 3 and the mapping region of the reads equals to the whole length of the reads. We then counted the reads number for each repeat sequence.

**siRNAs-mediated knockdown of Dicer-2 and Ago-2 and immunofluorescence**

Transfection was performed, as described(Rogers and Rogers 2008), using the "bathing" method for siRNA delivery. Cells were counted, pelleted and resuspend at 1-5x$10^6$ cells/ml in serum free media. We added ~10-30 µg dsRNA to each well in the 6-well tissue culture plate to obtain the final concentration of 25-50 nM. About 1 ml of cells were seeded each well in the 6-well plate and

incubate 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 for three times before harvesting the cells for downstream assays. Sequences of individual siRNAs were listed in Supplementary Table 3.

For immunostaining, S2 cells were washed in 1× PBS and fixed in 4% paraformaldehyde (pH-7.2) for 10 min at room temperature. After washing four times, cells were permeabilized with 0.1% Triton X100 in 1× PBS for 5 min at room temperature. Permeabilized cells were then incubated in 1% normal goat serum in 1× PBST for 30 min at room temperature, and then with the primary antibody (α-rabbit H3K9me3 1:500) and secondary antibodies (1:400) both in blocking buffer (3% BSA, 1% goat serum in PBST), each for 1h at room temperature. Cells were washed three times at room temperature, each for 5 min with 1× PBST. After mounting on coverslip with DAPI, cells were examined under Zeiss LSM-700 confocal laser scanning microscope.

**ChIP-seq library construction and data analysis**

DNA libraries were constructed using the NEBNext® Ultra™ II DNA Library Prep kit (NEB, USA) following manufacturer's recommendations. After end repair, 5´ phosphorylation. and dA-tailing of purified DNA fragments, NEBNext adaptors with hairpin loop structure were ligated and library fragments were purified with SPRIselect sample purification beads (NEB, USA). After ensuring the quality of libraries on the Agilent Bioanalyzer 2100 system, individual libraries were sequenced on the Illumina HiSeq X Ten platform to generate 150 bp paired-end.

Adapters and low-quality reads were filtered to obtain clean reads and clean reads mapped to the reference *Drosophila* genome (dm6) using Bowtie2. The wig files of each chromatin mark ChIP-seq dataset was obtained by using bamCoverage tools from Deeptools(Ramirez et al. 2016) with the parameters*: --binSize 1000, --normalizeTo1x 142573017 and --ignoreForNormalization chrM.* Enriched peaks were detected by MACS2 with input data as control using the parameters *-f BAMPE --nomodel --keep-dup all --broad.*

**Analysis of mitotic defects**

Medium and fetal bovine serum were batch tested for support of normal cell growth and RNAi efficiency. For the knockdown group, specific synthetic siRNA was added to cell culture in 24-well plates. After siRNA treatment for 4 days, cells were resuspended and transferred to glass-bottom, 24-well plates (Cellvis) and allowed to adhere for 2.5 hrs before fixation, as described(Goshima et al. 2007). Cells were fixed in 4% paraformaldehyde for 10 min, permeabilized with 0.1% Triton X100 in PBS for 5 min and incubated overnight at 4˚ C with anti-α-tubulin (ab7291 from abcam; 1:1000) in PBS containing 0.1% Triton and 0.5 mg/ml BSA, followed by staining with secondary antibodies and DAPI (1 µg/ml). For the rescue group, siRNAi treatment was performed as above followed by transfection of CHARRs siRNA 2 days later.

Immunostained specimens were imaged under a Zeiss LSM-700 confocal laser scanning microscope, using a 63x 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.

**Western blotting, RT-qPCR and Northern blotting**

Total cell lysate containing 15-25 μg protein from S2 cell cultures were fractionated by SDS-PAGE, immunoblotted and probed with specific antibodies. After incubation with peroxidase-conjugated secondary antibodies (1:5000; Abcam), blots were developed with Supersignal West Pico Chemiluminescent Substrate (Pierce) and exposed to film (SAGECREATION, MiNiChemi). Signal intensity for Dicer-2, H3K9me3, H3K27me3, H3K4me3 and Tublin was quantified using ImageJ.

For RNA quantification, total RNA was extracted, and genomic DNA was removed with DNase I (Roche, 04716728001). First-strand cDNA was generated with the SuperScript III Fist-Stand using random hexamers. The expression levels of RNAs were quantified on Rotor-Gene Q (QIAGEN) and normalized against GAPDH mRNA. PCR primers sequences were listed in Supplementary Table 3.

About 30 μg of total RNA isolated with Trizol was loaded into each lane of agarose gel and blotted onto membrane with Chemiluminescent Nucleic Acid Detection Module (Thermo Fisher) according to manufacturer's instruction. RNA probes were labeled by *in vitro* transcription of plasmids with T7 RNA polymerase (Promega) in the presence of Biotin RNA labeling mix (Roche). The primers used were listed in Supplementary Table 3.

## Quantification and statistical analysis

Statistical parameters were reported either in individual figures or corresponding figure legends. Quantified data were in general presented as bar/line plots, with the error bar representing mean ± SEM, or boxplot 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. Wherever asterisks are used to indicate the statistical significances, *stands for $p < 0.05$; ** for $p < 0.01$, *** for $p < 0.001$ and NS for not significant.

## Data Access

All sequencing and processed files were deposited to Gene Expression Omnibus under accession number GSE134307. To review GEO accession GSE134307: Go to https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE134307 , then enter token qzexqwwmbbujhol into the box. All software and the datasets used in this study are available from the corresponding author, Xiang-Dong Fu (xdfu@ucsd.edu), upon reasonable request.

## ACKNOWLEDGEMENTS

We thank Dr. Yang Yu for providing S2 cells. This work was supported by grants from the National Key R&D Program of China (2016YFC0901702, 2016YFC09010002) and the National Natural Science Foundation of China (31871294, 31520103905) to R.S.C. and S.M.H., and NIH grants (HG004659, GM049369 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. and X.D.F.

## DISCLOSURE DECLARATION

The authors declare no competing interests.

## 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.

Avner P, Heard E. 2001. X-chromosome inactivation: counting, choice and initiation. *Nat Rev Genet* **2**: 59-67.

Axtell MJ. 2013. ShortStack: comprehensive annotation and quantification of small RNA genes. *Rna* **19**: 740-751.

Baker BS, Gorman M, Marin I. 1994. Dosage compensation in Drosophila. *Annu Rev Genet* **28**: 491-521.

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**.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114-2120.

862 Buhler M, Verdel A, Moazed D. 2006. Tethering RITS to a nascent transcript initiates RNAi-
863 and heterochromatin-dependent gene silencing. *Cell* **125**: 873-886.

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

867 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras
868 TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15-21.

869 Drysdale R. 2008. FlyBase : a database for the Drosophila research community. *Methods in*
870 *molecular biology (Clifton, NJ)* **420**: 45-59.

871 Ekwall K, Javerzat JP, Lorentz A, Schmidt H, Cranston G, Allshire R. 1995. The Chromodomain
872 Protein Swi6 - a Key Component at Fission Yeast Centromeres. *Science* **269**: 1429-1431.

873 Fagegaltier D, Bouge AL, Berry B, Poisot E, Sismeiro O, Coppee JY, Theodore L, Voinnet O,
874 Antoniewski C. 2009. The endogenous siRNA pathway is involved in heterochromatin
875 formation in Drosophila. *Proceedings of the National Academy of Sciences of the United*
876 *States of America* **106**: 21258-21263.

877 Franke A, Baker BS. 1999. The rox1 and rox2 RNAs are essential components of the
878 compensasome, which mediates dosage compensation in Drosophila. *Molecular Cell* **4**:
879 117-122.

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

883 Goshima G, Wollman R, Goodwin SS, Zhang N, Scholey JM, Vale RD, Stuurman N. 2007.
884 Genes required for mitotic spindle assembly in Drosophila S2 cells. *Science* **316**: 417-
885 421.

886 Grewal SI, Jia S. 2007. Heterochromatin revisited. *Nat Rev Genet* **8**: 35-46.

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

890 Halic M, Moazed D. 2009. Transposon silencing by piRNAs. *Cell* **138**: 1058-1060.

Hall LL, Carone DM, Gomez AV, Kolpa HJ, Byron M, Mehta N, Fackelmayer FO, Lawrence JB. 2014. Stable C0T-1 repeat RNA is abundant and is associated with euchromatic interphase chromosomes. *Cell* **156**: 907-919.

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

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

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

Iwasaki YW, Siomi MC, Siomi H. 2015. PIWI-Interacting RNA: Its Biogenesis and Functions. *Annu Rev Biochem* **84**: 405-433.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and genome research* **110**: 462-467.

Kandasamy SK, Fukunaga R. 2016. Phosphate-binding pocket in Dicer-2 PAZ domain for high-fidelity siRNA production. *Proc Natl Acad Sci U S A* **113**: 14031-14036.

Kellum R, Alberts BM. 1995. Heterochromatin Protein-1 Is Required for Correct Chromosome Segregation in Drosophila Embryos. *J Cell Sci* **108**: 1419-1431.

Kopylova E, Noe L, Touzet H. 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**: 3211-3217.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**: 357-359.

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

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.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

921 Li X, Zhou B, Chen L, Gou LT, Li H, Fu XD. 2017. GRID-seq reveals the global RNA-
922 chromatin interactome. *Nat Biotechnol* **35**: 940-950.

923 Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K, Mittal
924 V, May B, Kasschau KD et al. 2004. Role of transposable elements in heterochromatin
925 and epigenetic control. *Nature* **430**: 471-476.

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

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

931 Peters AH, O'Carroll D, Scherthan H, Mechtler K, Sauer S, Schofer C, Weipoltshammer K,
932 Pagani M, Lachner M, Kohlmaier A et al. 2001. Loss of the Suv39h histone
933 methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**:
934 323-337.

935 Pham JW, Pellino JL, Lee YS, Carthew RW, Sontheimer EJ. 2004. A Dicer-2-dependent 80s
936 complex cleaves targeted mRNAs during RNAi in Drosophila. *Cell* **117**: 83-94.

937 Quinodoz SA, Ollikainen N, Tabak B, Palla A, Schmidt JM, Detmar E, Lai M, Shishkin A, Bhat
938 P, Trinh V et al. 2017. doi:10.1101/219683.

939 Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dundar F,
940 Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data
941 analysis. *Nucleic Acids Res* **44**: W160-165.

942 Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software
943 Suite. *Trends in genetics : TIG* **16**: 276-277.

944 Roberts RJ, Vincze T, Posfai J, Macelis D. 2015. REBASE--a database for DNA restriction and
945 modification: enzymes, genes and genomes. *Nucleic Acids Res* **43**: D298-299.

946 Rogers SL, Rogers GC. 2008. Culture of Drosophila S2 cells and their use for RNAi-mediated
947 loss-of-function studies and immunofluorescence microscopy. *Nature protocols* **3**: 606-
948 611.

949 Sridhar B, Rivas-Astroza M, Nguyen TC, Chen W, Yan Z, Cao X, Hebert L, Zhong S. 2017.
950 Systematic Mapping of RNA-Chromatin Interactions In Vivo. *Curr Biol* **27**: 610-612.

951 Stein P, Svoboda P, Anger M, Schultz RM. 2003. RNAi: Mammalian oocytes do it without
952 RNA-dependent RNA polymerase. *Rna* **9**: 187-192.

953 Sun FL, Cuaycong MH, Craig CA, Wallrath LL, Locke J, Elgin SC. 2000. The fourth
954 chromosome of Drosophila melanogaster: interspersed euchromatic and heterochromatic
955 domains. *Proc Natl Acad Sci U S A* **97**: 5340-5345.

956 Tschiersch B, Hofmann A, Krauss V, Dorn R, Korge G, Reuter G. 1994. The protein encoded by
957 the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains
958 of antagonistic regulators of homeotic gene complexes. *EMBO J* **13**: 3822-3831.

959 Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD. 2006. A distinct small RNA
960 pathway silences selfish genetic elements in the germline. *Science* **313**: 320-324.

961 Verdel A, Jia S, Gerber S, Sugiyama T, Gygi S, Grewal SI, Moazed D. 2004. RNAi-mediated
962 targeting of heterochromatin by the RITS complex. *Science* **303**: 672-676.

963 Volpe T, Martienssen RA. 2011. RNA interference and heterochromatin assembly. *Cold Spring*
964 *Harbor perspectives in biology* **3**: a003731.

965 Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, Martienssen RA. 2002. Regulation of
966 heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**:
967 1833-1837.

968 Yu R, Wang X, Moazed D. 2018. Epigenetic inheritance mediated by coupling of RNAi and
969 histone H3K9 methylation. *Nature* **558**: 615-619.

970 Zhang YJ, Guo L, Gonzales PK, Gendron TF, Wu Y, Jansen-West K, O'Raw AD, Pickles SR,
971 Prudencio M, Carlomagno Y et al. 2019. Heterochromatin anomalies and double-
972 stranded RNA accumulation underlie C9orf72 poly(PR) toxicity. *Science* **363**.

973