

RESEARCH ARTICLE

Sex-specific variation in R-loop formation in *Drosophila melanogaster*Timothy J. Stanek^{1,2}, Weihuan Cao¹, Rohan M Mehra¹, Christopher E. Ellison^{1*}

1 Department of Genetics, Human Genetics Institute of New Jersey, Rutgers, The State University of New Jersey, Piscataway, New Jersey, United States of America, **2** Department of Pathology, Robert Wood Johnson Medical School, Piscataway, New Jersey, United States of America

* chris.ellison@rutgers.edu

Abstract

R-loops are three-stranded nucleotide structures consisting of a DNA:RNA hybrid and a displaced ssDNA non-template strand. Previous work suggests that R-loop formation is primarily determined by the thermodynamics of DNA:RNA binding, which are governed by base composition (e.g., GC skew) and transcription-induced DNA superhelicity. However, R-loops have been described at genomic locations that lack these properties, suggesting that they may serve other context-specific roles. To better understand the genetic determinants of R-loop formation, we have characterized the *Drosophila melanogaster* R-loop landscape across strains and between sexes using DNA:RNA immunoprecipitation followed by high-throughput sequencing (DRIP-seq). We find that R-loops are associated with sequence motifs that are G-rich or exhibit G/C skew, as well as highly expressed genes, tRNAs, and small nuclear RNAs, consistent with a role for DNA sequence and torsion in R-loop specification. However, we also find motifs associated with R-loops that are A/T-rich and lack G/C skew as well as a subset of R-loops that are enriched in polycomb-repressed chromatin. Differential enrichment analysis reveals a small number of sex-biased R-loops: while non-differentially enriched and male-enriched R-loops form at similar genetic features and chromatin states and contain similar sequence motifs, female-enriched R-loops form at unique genetic features, chromatin states, and sequence motifs and are associated with genes that show ovary-biased expression. Male-enriched R-loops are most abundant on the dosage-compensated X chromosome, where R-loops appear stronger compared to autosomal R-loops. R-loop-containing genes on the X chromosome are dosage-compensated yet show lower MOF binding and reduced H4K16ac compared to R-loop-absent genes, suggesting that H4K16ac or MOF may attenuate R-loop formation. Collectively, these results suggest that R-loop formation in vivo is not fully explained by DNA sequence and topology and raise the possibility that a distinct subset of these hybrid structures plays an important role in the establishment and maintenance of epigenetic differences between sexes.

OPEN ACCESS

Citation: Stanek TJ, Cao W, Mehra RM, Ellison CE (2022) Sex-specific variation in R-loop formation in *Drosophila melanogaster*. PLoS Genet 18(6): e1010268. <https://doi.org/10.1371/journal.pgen.1010268>

Editor: Brian Oliver, UNITED STATES

Received: February 3, 2022

Accepted: May 22, 2022

Published: June 10, 2022

Copyright: © 2022 Stanek et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Sequencing data are available via the NCBI Sequence Read Archive (SRA), BioProject (PRJNA800016). All custom scripts generated for this project can be found on GitHub at https://github.com/Ellison-Lab/Adults_DRIPseq. Source data are provided with this paper and via zenodo [<https://doi.org/10.5281/zenodo.5899005>].

Funding: This research was supported by a National Institute of General Medical Sciences R01 award (GM140160) to C.E.E. and K12 award (GM093854) to T.J.S. The funders had no role in

study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

R-loops are DNA:RNA hybrid structures that act as important regulators of gene expression and genomic stability and whose dysregulation can contribute to diseases such as neurological disorders and cancer. Here, we utilize DNA:RNA immunoprecipitation followed by high-throughput sequencing (DRIP-seq) to assess the sex-specific variability in R-loop formation in *Drosophila melanogaster* adults. Most R-loops are found at simple repeats in regions of high transcriptional activity such as active chromatin states, 5'UTRs, tRNAs, and topologically associating domain boundaries. In both sexes, we find that R-loops are more common on the X chromosome compared to autosomes, likely due to an increased density of X-linked simple repeats that favor R-loop formation. While R-loops are largely conserved between sexes, we uncover a small but significant subset of sex-biased R-loops. Female-enriched R-loops are associated with genes showing ovary-biased expression and form at unique genome features, compared to other R-loops. Male-enriched R-loops occur preferentially at dosage-compensated genes on the X chromosome, yet surprisingly, show reduced levels of the marker of dosage compensation, H4K16ac, raising the possibility that the H4K16ac histone modification may attenuate R-loop formation. Our identification of sex-biased R-loops suggests a specialized role for these structures in establishing and maintaining sex-specific epigenetic programs.

Introduction

Within the nucleus, the mechanical processes driving transcription must strike a balance between providing the cell sufficient transcripts for survival and the inherent danger to genome stability via induction of torsional stress. One mechanism by which cells regulate transcription and relieve said stress is the formation of R-loops. R-loops form when RNA invades double-stranded DNA and binds the template strand, creating a DNA:RNA hybrid and displacing the non-template strand. R-loops have been associated with a variety of biological processes and are implicated in essential aspects of gene regulation as well as genome stability [1,2]. Persistent dysregulation of R-loop maintenance can result in replication stress, DNA double-strand breaks, and chromosomal rearrangements that contribute to diseases such as neurological disorders [3] and cancer [4–7].

On a mechanistic level, there is emerging evidence that R-loop formation may be primarily driven by a combination of DNA sequence and DNA topology. At the sequence level, previous work has shown that G-rich sequences and sequences exhibiting GC skew are prone to R-loop formation [8]. DNA:RNA basepairing is more energetically favorable than DNA:DNA basepairing for G-rich and G/A-rich sequences [9], whereas transcription of linear DNA molecules exhibiting GC skew or CpG islands has been shown to lead to R-loop formation in vitro [10].

In terms of DNA topology, R-loops are known to form in response to transcription and replication-induced torsional stress in double-stranded DNA [11]. R-loop formation significantly absorbs negative superhelicity upstream of advancing polymerase complexes, functioning as a complement to DNA topoisomerase I and DNA gyrase in managing torsional stress [12–15]. Subsequent resolution of R-loops releases this stored negative superhelicity, inducing local changes such as strand separation or histone binding and potentially priming gene promoters for successive rounds of Pol II binding and firing. Such releases of superhelicity have also been shown to facilitate long-range changes in chromatin architecture such as altered promoter-enhancer contacts and loop extrusion [16,17]. Together, these prior studies suggest that

R-loops are most likely to form at thermodynamically favorable regions of the genome, which are largely denoted by the base composition or the torsional state of a particular locus.

A relationship between R-loops and DNA torsion is further supported by multiple studies that document a correlation between R-loop formation and high rates of transcription, which leads to negative supercoiling upstream of the translocating polymerase. For example, in yeast strains lacking RNase H activity, R-loops have been detected at Pol III-transcribed genes, such as tRNAs and small nuclear RNAs, likely due to their high expression levels [18]. Furthermore, R-loops were found to be enriched at genes proximal to topologically associating domain (TAD) boundaries, which are known to be highly transcribed [19]. However, rather than being passive byproducts of transcription, there is evidence that R-loops are involved in specific mechanisms of gene regulation. R-loop formation may aid in Pol II pausing at transcriptional start sites [20] and promote transcriptional termination by stalling the Pol II complex and mediating access of exonucleases for 3' cleavage of polyA sites [21–23]. In murine embryonic stem cells and *Drosophila* embryos, R-loops have been shown to play a role in Polycomb repressive complex 1 (PRC1) and Polycomb repressive complex 2-mediated repression of Polycomb group (PcG) target genes [24,25]. R-loops have additionally been shown to form *in trans* at circular RNAs (circRNAs) to regulate splicing factor recruitment [26] and DNA repair [27], and their formation at DNA double-strand breaks and short telomere repeats regulates Rad51-mediated homology-dependent repair [28,29]. These prior findings suggest that R-loops regulate a variety of nuclear processes, and their formation is both versatile and context-specific.

In this study, we investigate the determinants of R-loop formation in *Drosophila* using DNA:RNA Immunoprecipitation followed by high-throughput sequencing (DRIP-seq). We characterize genome features and sequence motifs that are associated with R-loops, and we compare the location and strength of R-loops between males and females to address whether transcriptional levels are major determinants of R-loop formation. We also assess whether hypertranscription of dosage-compensated genes is associated with increased R-loop formation in males. Overall, we find a consistent positive association between gene expression level and R-loop formation within both sexes. Furthermore, for female-enriched R-loops, we find a significant association between sex-biased gene expression and sex-biased R-loop formation. By contrast, we find that male-biased R-loops are not associated with male-biased gene expression, but instead are enriched on the X chromosome and are associated with dosage-compensated genes. While male-biased and non-biased R-loops are associated with similar sequence motifs, genome features, and chromatin states, female-biased R-loops inhabit unique regions of the genome. These results suggest that, while high transcription levels may play a role in R-loop formation, they are not sufficient to determine the R-loop landscape in *Drosophila*. Other genetic and epigenetic factors must also be involved, lending support to the multifaceted and context-specific nature of these genomic features.

Results

DRIP-sequencing

To determine the characteristics that contribute to natural variation in R-loop formation, we performed DRIP-seq separately for adult males and females in two strains of *D. melanogaster* from the *Drosophila* Genetic Reference Panel (DGRP): DGRP379 and DGRP732. Standard DRIP-seq protocols employ a restriction enzyme cocktail to fragment chromatin. However, previous studies have shown that this approach biases R-loop enrichment for specific euchromatic regions of the genome [30]. To correct for this bias, we adopted a modified protocol using a specialized sonication-based fragmentation procedure that has been shown to both

preserve and isolate R-loops within genomic sequences insensitive to enzymatic digestion [31] (see [Methods](#)).

Paired-end DRIP-seq reads were aligned to primary autosomes and the X chromosome (for sequencing library information, quality control, and alignment statistics, see [S1 Table](#)). The dot chromosome was excluded from our analysis due to its high repeat density, heterochromatin content, and poor mappability [32]. After DRIP-seq peak calling, we compared peak density among chromosome arms and found a striking depletion of X-linked R-loops in males and enrichment of X-linked R-loops in females ([Fig 1A](#)). Previous work has shown that the number of identifiable ChIP-seq peaks scales with sequencing depth, without a clear saturation in many cases [33]. It is therefore possible that the depletion of DRIP-seq peaks on the male X chromosome is due to reduced sequencing depth compared to the male autosomes (and all female chromosomes), despite the fact that DRIP signal was quantified relative to an input control. In order to test this prediction, we downsampled both the male and female datasets so

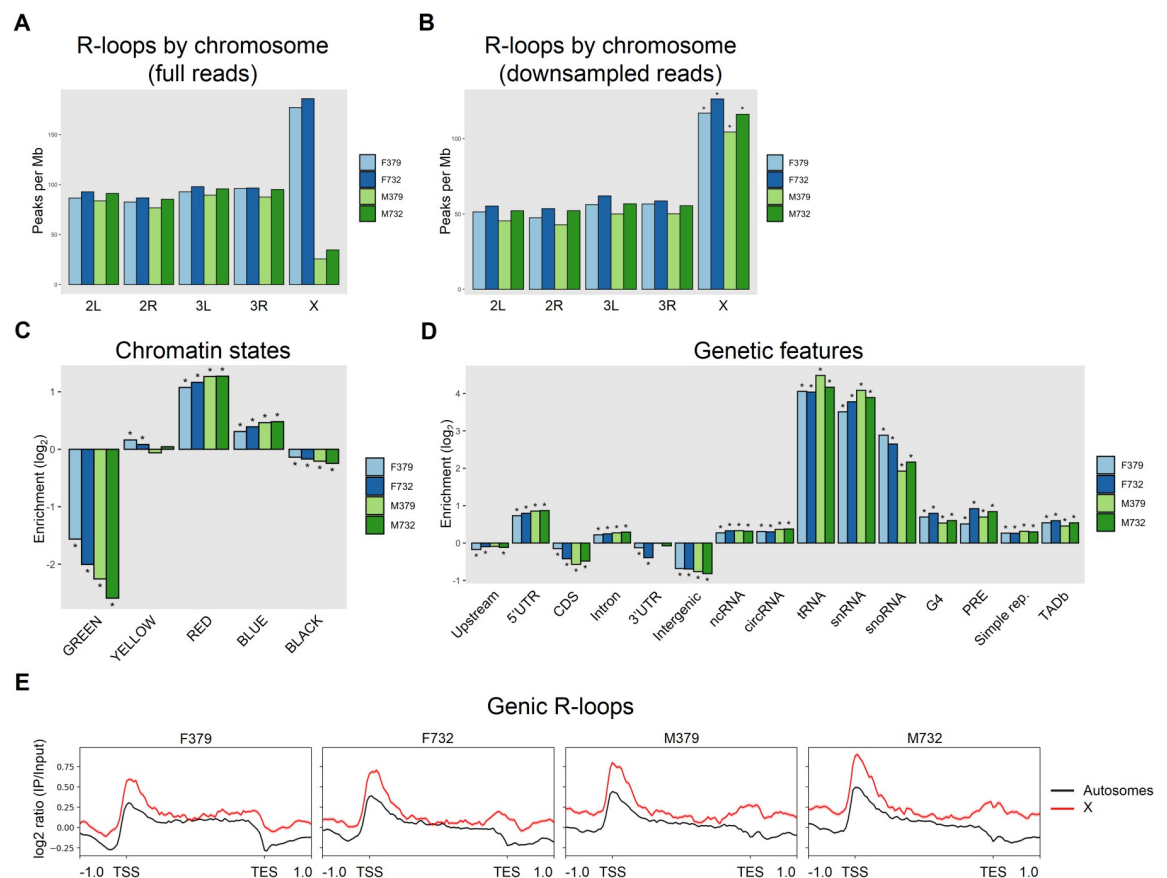


Fig 1. R-loop identification & feature enrichment in *D. melanogaster* adults. Whole adult flies from strains DGRP379 and DGRP732 were separated by sex and subjected to DRIP-sequencing to detect R-loops. (A) DRIP-seq peak density across chromosomes for the full male and female datasets (i.e. without downsampling) shows an apparent depletion of X-linked R-loops in males. (B) Downsampling all female DRIP-seq reads and autosomal male DRIP-seq reads so that all chromosome arms have similar sequencing depths shows that X-linked R-loops are enriched in both males and females (Binomial test, * = $p < 2.2e-16$). (C) R-loop formation at chromatin states as described in [34]. (D) R-loop formation at various genetic features. (E) Metaprofiles of R-loop signal across protein-coding genes, autosomes versus X chromosome. For Panels (C) and (D): R-loop enrichment is shown as the observed number of DRIP-seq peaks overlapping each feature (or chromatin state) divided by the expected number of peaks (see [Methods](#)). P-values were calculated via a Permutation Test with Benjamini-Hochberg correction for multiple comparisons, * = corrected $p < 0.05$. For Panel (E): the solid lines represent the mean DRIP-seq signal within each metagene bin, and the shading represents the standard error of the mean.

<https://doi.org/10.1371/journal.pgen.1010268.g001>

that all chromosomes had similar sequencing coverage (see [Methods](#), [S1 Table](#)). The downsampled data resulted in a smaller number of peaks identified overall (as expected due to reduced sequencing depth from downsampling). However, more than twice as many peaks were identified on the male X chromosome, suggesting that the depletion of male X-linked R-loops observed in the full dataset is an artifact of lower sequencing coverage ([Fig 1B](#)). We therefore used the downsampled data for all further analyses, which resulted in 7645 to 9269 high-confidence, reproducible DRIP-seq peaks for each sample (see [Methods](#), [Fig 1B](#), and [S2 Table](#)).

R-loop locations and feature enrichment

We next analyzed our high-confidence R-loop peaks for enrichment at various chromatin states [[34](#)] ([Fig 1C](#)) and genomic features ([Fig 1D](#)). R-loops are enriched primarily within the transcriptionally active RED chromatin state and are depleted in GREEN HP1-associated heterochromatin and the gene-poor BLACK repressive state. R-loops are also enriched in BLUE Polycomb-associated heterochromatin ([Fig 1C](#)), in agreement with previous studies reporting the role of R-loops in PRC-mediated repression [[24,25](#)]. Also consistent with previous studies, R-loops are enriched at 5'UTRs and introns and show no enrichment in upstream and intergenic regions ([Fig 1D](#)), highlighting their known role in Pol II-mediated transcription [[8,20,35](#)]. Surprisingly, R-loops are not enriched at 3'UTRs, in contrast with the reported association of R-loops with transcriptional termination [[21,23](#)]. R-loops are also enriched at various classes of noncoding RNAs, including circRNAs, small nuclear and small nucleolar RNAs, and tRNAs, likely due to high levels of transcription at these loci [[18](#)]. Similarly, we find that R-loops are enriched at G-quadruplexes (G4), which have been associated with open chromatin and high transcription [[36](#)], as well as TAD boundaries, where increased gene expression has been observed relative to genes within TADs [[19](#)]. Finally, we found that R-loops are enriched at simple repeats (sr), consistent with previous findings [[37](#)]. These patterns of R-loop enrichment and depletion are consistent between autosomes and the X chromosome ([S1A Fig](#)), supporting the broad role of R-loop formation in active transcription in males and females.

Metagene analysis of gene-associated R-loops reveals that at autosomal genes, R-loops appear most abundant just downstream of the transcriptional start site (TSS) and are absent from the 3'UTR, while X chromosome-associated R-loops show increased signal both across the entire gene region and at the 3'UTR ([Figs 1E and S1B](#)). These distinct profiles are observed in both sexes, suggesting that increased R-loop signal on the X chromosome occurs independently of dosage compensation-associated hypertranscription. Examination of R-loop peaks by chromosome shows that, while autosomal R-loops are present at similar levels across all samples, there is a significant enrichment of R-loops on the X chromosomes in both sexes (Binomial test $P < 2.2e-16$) ([Fig 1B](#)). These data indicate that R-loops tend to be associated with specific genome features and chromatin states, and their location is largely conserved between individual strains and across sexes, with the X chromosome consistently showing higher R-loop DRIP signal and peak density relative to autosomes.

Differential enrichment and motif analysis

Despite fewer overall R-loop peaks identified in males, gene-associated R-loop signal is higher in males compared to females on both autosomes and the X chromosome ([Figs 1E and S1B](#)). To assess sex-specific variation in R-loop formation, we subjected R-loop peaks to differential enrichment analysis using Diffbind [[38](#)]. Principal component analysis (PCA) of Diffbind peaks shows that 60% of the variation in R-loop location is captured by PC1 and PC2, which are associated with strain and sex, respectively ([Fig 2A](#)). Comparing R-loop profiles by sex via

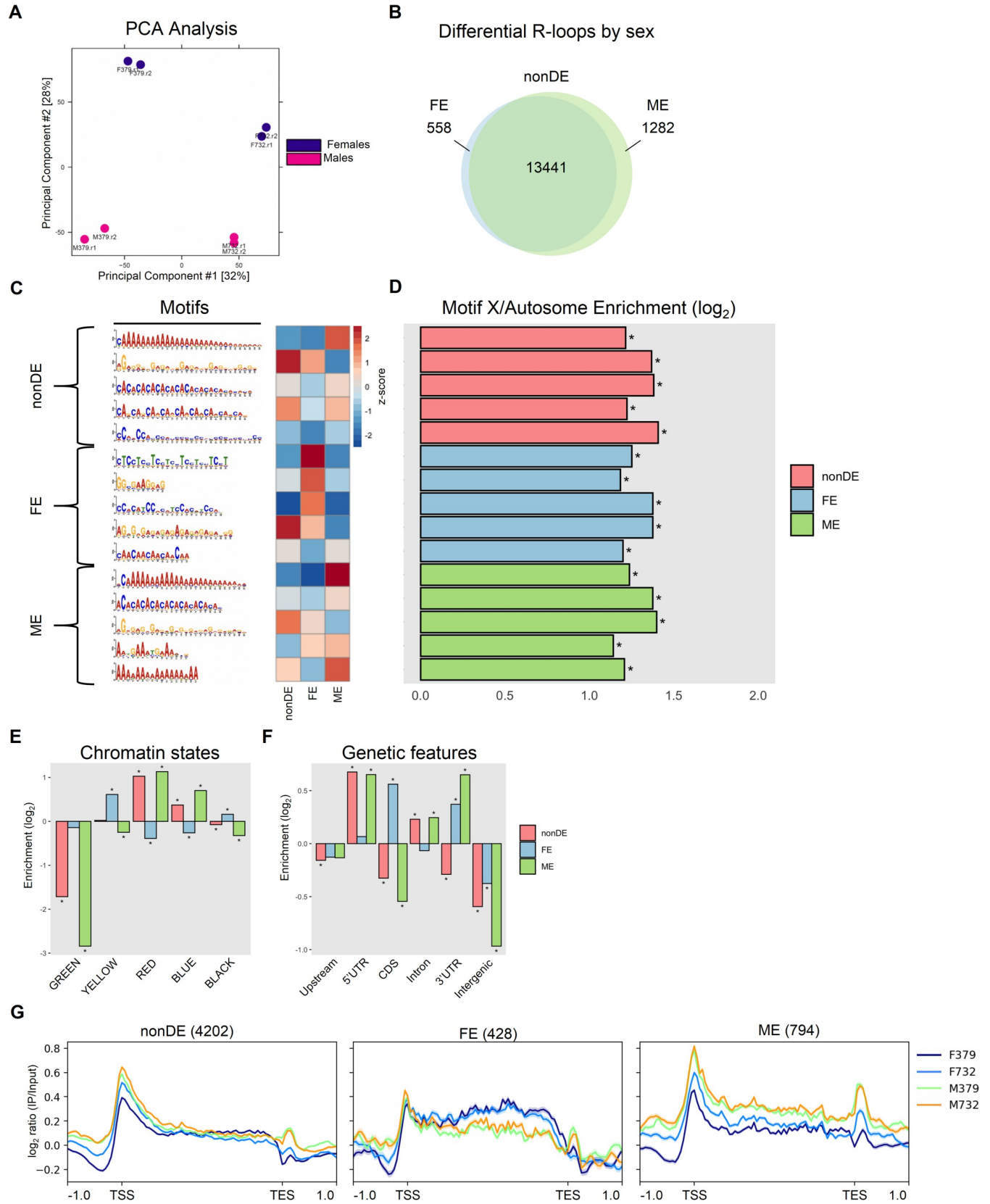


Fig 2. Differential enrichment and motif analysis of R-loops. (A) PCA analysis of DRIP conditions. (B) Venn diagram of non-differentially enriched (nonDE) and sex-biased (Female Enriched [FE] and Male Enriched [ME]) R-loops as identified by DiffBind. (C) STREME motif analysis by DE group; the top 5 motifs from each DE group are represented graphically (left), with z-score enrichment for each motif across DE groups plotted in the heatmap (right). (D) Motif enrichment on the X chromosome versus autosomes, plotted as log₂ motifs per Mb. Binomial test, * = $p < 0.001$. (E) R-loop formation at chromatin states as described in [34]. (F) R-loop formation at various genetic features. (G) Metaprofiles of R-loop signal at genes within each DE group. For Panels (E) and (F): R-loop enrichment is shown as the observed number of DRIP-seq peaks overlapping each feature (or chromatin state) divided by the expected number of peaks (see Methods). P-values were calculated via a Permutation Test with Benjamini-Hochberg correction for multiple comparisons, * = corrected $p < 0.05$. For Panel (G): the solid lines represent the mean DRIP-seq signal within each metagene bin, and the shading represents the standard error of the mean.

<https://doi.org/10.1371/journal.pgen.1010268.g002>

DiffBind, we identified 13,441 shared R-loops (non-differentially enriched, nonDE), 558 R-loops enriched in females (female-enriched, FE) and 1,282 R-loops enriched in males (male-enriched, ME) (Fig 2B). Overall, ~12% of R-loops are sex-biased, with more than twice as many male-biased R-loops compared to female-biased R-loops.

Motif analysis across these differentially enriched R-loop-containing loci reveals several interesting aspects of sex-biased R-loops. First, many of the most enriched motifs across all three DE groups comprise simple repeats (Fig 2C). Although some motifs contain classical GC skew [8,21] ((CCMM)_n in the nonDE group and ‘GGCGAAGGAG’ and (CTC)_n in the FE group), several enriched motifs lack such skew and instead display no skew ((CACA)_n in the nonDE and ME groups and (AGAG)_n in the FE group) or AT-skew (poly-A tracts in the nonDE and ME groups) (Fig 2C). R-loop formation at poly-A tracts specifically has been linked to high gene expression [31]. These variations in R-loop sequence favorability have been observed between model organisms, where R-loops preferentially form at GC-rich sequences in mammals, at AT-rich sequences in *Arabidopsis*, and at both sequence classes in yeast [8,21,39,40]. Interestingly, the nonDE and ME groups share three of their five top enriched motifs, whereas only one motif is shared between the nonDE and FE groups (Fig 2C). All motifs identified are significantly enriched on the X chromosome compared to autosomes (Fig 2D), consistent with our observed enrichment of R-loops on the X chromosome (Fig 1B).

The unique motifs found at FE R-loops suggest that they may occupy genomic features distinct from the nonDE and ME R-loops. Assigning these sex-biased R-loops to known chromatin states [34] reveals similar enrichments between nonDE and ME R-loops (Fig 2E). By contrast, FE R-loops show the opposite pattern of enrichment or depletion in nearly every chromatin state, with enrichment primarily in the YELLOW active state and the BLACK repressive state (Fig 2E). Genetic feature analysis similarly shows that FE R-loops preferentially form at loci distinct from the nonDE and ME groups, most notably within the CDS of genes (Fig 2F). Metagene analysis of these differentially enriched R-loops at genic loci further support the uniqueness of the FE group: the DRIP signal at FE genes is depleted from the TSS and concentrated across the CDS, opposite from that seen for the nonDE and ME groups (Figs 2G and S2A). Furthermore, gene ontology analysis reveals that ME R-loops form at genes associated with developmental and regulatory processes, whereas FE R-loops form at genes associated with translation and biosynthesis, including a large number of ribosomal proteins (S3 Table).

Given the relationship between high rates of transcription and R-loop formation, we next sought to determine whether the ME and FE R-loops occur at genes that show sex-biased expression patterns. We performed RNA-seq from whole flies for the same four samples used for DRIP-seq and found that, surprisingly, ME and FE R-loop-containing genes show similar expression patterns in both sexes (S2B Fig). To address the possibility that the ME and FE genes are differentially expressed in specific tissues, we used the FlyAtlas database to investigate their expression patterns in ovaries, testes, and brain, tissues where sex-biased gene expression has previously been characterized (S2C Fig) [41–43]. In contrast to our whole fly

expression data, genes containing female-enriched R-loops are expressed at significantly higher levels in ovaries compared to genes with male-enriched R-loops and conserved R-loops (S2C Fig, left panel). Conversely, genes containing male-enriched R-loops are expressed at significantly lower levels in the testes compared to genes with female-enriched R-loops and conserved R-loops (S2C Fig, center panel). In the brain, genes from all three categories (FE, ME, and nonDE) are expressed at similar levels (S2C Fig, right panel). Taken together, these observations suggest that sex-specific utilization of R-loops in *Drosophila* is only partly explained by sex-biased gene expression.

Because the gonads make up a large percentage of the adult body, we reasoned that the DRIP signal from whole adult flies is likely similar to that from gonads alone. To test this prediction, we performed DRIP-seq in ovaries dissected from adult females and found strong, highly significant correlations between the female whole fly samples and ovary samples for both nonDE and FE peaksets from the adult data (Spearman's $\rho > 0.72$ and $P < 2.2e-16$ in all comparisons, S2D Fig).

X chromosome-specific R-loop enrichment

Given the presence of sex-specific R-loop enrichment in *Drosophila* adults, we explored the possibility that these differences were associated with the sex chromosome and dosage compensation. As noted above, R-loops form with increased frequency on the X chromosome compared to autosomes in both males and females (Fig 1B). However, when focusing only on differentially enriched R-loops, those found on the X chromosome are enriched in males significantly more than the general X enrichment seen across all R-loops (Fig 3A and 3B). By contrast, the DE R-loops in females show no X-enrichment (Fig 3B). Given the established relationship between R-loops and transcription, this observation suggests an association of R-loops with the dosage compensation mechanism of the male X chromosome [44,45].

R-loop-containing genes are expressed at higher levels than genes with no detectable R-loops on both autosomes and the X chromosome (Fig 3C), consistent with transcription-induced R-loop formation [18,31]. On the X chromosome, this association is further supported by the significantly closer proximity of nonDE and ME R-loops to chromosomal entry sites (CES) (Fig 3D), which help initiate and propagate histone acetylation associated with the dosage compensation complex (DCC). However, analysis of a publicly available dataset of DCC machinery and its cognate modifications [46,47] shows that, in male salivary glands, both the histone acetyltransferase males-absent on the first protein (MOF) and the histone mark H4K16ac are depleted at ME genes relative to both nonDE genes and no R-loop genes (Fig 3E and 3F), despite showing evidence of dosage compensation based on gene expression (S3A Fig). These results are consistent with a positive association between transcription and R-loop formation that is possibly attenuated by high levels of the H4K16ac histone modification.

To further explore the relationship between R-loop formation and MOF/H4K16ac levels, we extended our comparison above to females. In females, the MSL complex is absent due to translational repression of *msh-2* by Sex-lethal (Sxl) [48]. Instead, the MOF-containing NSL complex acts to deposit H4K16ac at actively transcribed genes [49]. We therefore used the female R-loop data to determine whether a negative relationship exists between the presence of R-loops and MOF and H4K16ac levels, similar to what we observed on the male X chromosome. All R-loop classes (i.e., non-DE, FE, and ME) show significantly lower enrichment of MOF and H4K16ac compared to expressed genes with no R-loops, on both the autosomes and X chromosome (Fig 3G and 3H), providing additional support that R-loops are less likely to form at genomic regions with high levels of MOF/H4K16ac.

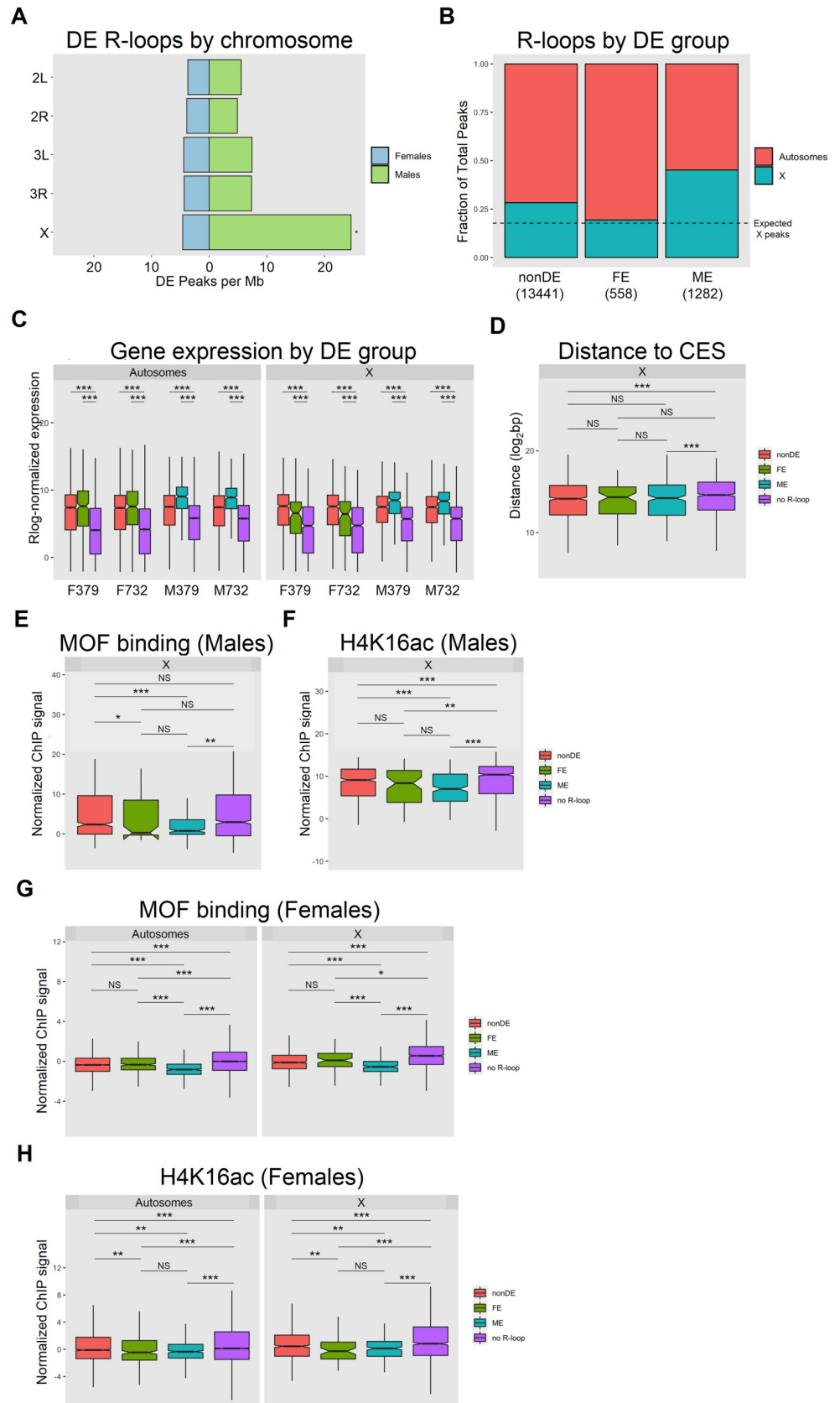


Fig 3. X chromosome-specific R-loop enrichment. (A) Differentially enriched R-loops by chromosome, plotted as R-loop peaks per Mb. Binomial test, * = $p < 0.001$. (B) Differentially enriched R-loop frequency on autosomes versus X chromosome, plotted as a fraction of total R-loops per DE group. (C) Gene expression analysis of R-loop-containing genes by DE group versus R-loop-absent (no R-loop) genes, on autosomes and the X chromosome, plotted as rlog-normalized expression. Wilcoxon test, *, **, *** = $p < 0.05, 0.01, 0.001$. (D) Distance to chromosomal entry site (CES) on the X chromosome across DE groups, plotted in \log_2 base pairs (bp). Wilcoxon test, *** = $p < 0.001$. (E) MOF binding and (F) H4K16ac enrichment on the X chromosome in third-instar larva male salivary glands across DE groups. Wilcoxon test, **, *** = $p < 0.01, 0.001$. (G) MOF binding and (H) H4K16ac enrichment on autosomes and the X chromosome in third-instar larva female salivary glands across DE groups. Wilcoxon test, *, **, *** = $p < 0.05, 0.01, 0.001$.

<https://doi.org/10.1371/journal.pgen.1010268.g003>

Discussion

Our assessment of natural R-loop variation between sexes has revealed multiple insights. First, R-loops are largely associated with transcriptionally active loci: they are found in chromatin states with either broad or specific transcriptional programs. More specifically, R-loops form proximal to the TSS, at multiple classes of ncRNAs and G-quadruplexes, and at TAD boundaries where transcriptionally active loci have been shown to reside [19]. Mechanistically, this enrichment is supported by the role R-loops play in relieving transcription-mediated torsional stress [11,14,16]. To our surprise, we failed to observe R-loop enrichment at 3'UTRs in *Drosophila*, despite previous studies implicating R-loop formation in transcriptional termination [22,23]. At some of these features, such as circRNAs, R-loop formation can occur *in trans* [26,27], yet whether the circRNA-associated R-loops detected in this study also form *in trans* remains to be determined. In addition to an association between R-loops and overall gene transcription, our observation of R-loop enrichment within the BLUE Polycomb-regulated chromatin state and at PREs in *Drosophila* supports previously established roles that R-loops play in Polycomb-mediated gene repression [24,25,50].

Our differential enrichment analysis confirms the strong conservation of the R-loop landscape between individuals and sexes, in line with previous comparisons across human and murine cell lines [30]. Additionally, the majority of R-loop-associated DNA motifs that we identify are simple repeats. The *Drosophila* X chromosome has previously been shown to be enriched for simple repeats, in general, compared to autosomes [51], suggesting that the sequence content of this chromosome may explain, at least in part, our observation of increased R-loop formation on the X. However, we note that R-loop density is roughly 2-fold higher on the X compared to the autosomes whereas the R-loop motifs all show less than 1.5-fold enrichment on the X, suggesting that other factors may also be involved.

Despite the overall conservation of the R-loop landscape between sexes, we also observe a subset of sex-biased R-loops, suggesting some level of specialization for their formation and function. The enrichment of nearly half of all male-enriched R-loops on the male X chromosome likely reflects its hypertranscribed, dosage-compensated state. Furthermore, the relationship between transcription and R-loop formation is maintained even within the X chromosome: R-loop-containing genes on the male X exhibit higher gene expression levels than X-linked genes lacking R-loops. It seems paradoxical, then, that canonical markers of active transcription deposited by the MSL and NSL complexes are reduced at R-loop-containing genes compared to genes with no R-loops. One possible explanation is that H4K16 acetylation may subtly disfavor R-loop formation. Indeed, previous studies have established a role for the DCC and H4K16ac in reducing negative superhelicity and disordering dosage-compensated chromatin to encourage DNA binding protein activity and Pol II loading [52,53]. This reduction in negative superhelicity could make R-loop formation less energetically favorable at highly acetylated genes or genes strongly bound by the DCC. However, such a relationship does not mean these two modifications should be mutually exclusive. Instead, the propensity

for either R-loops or histone acetylation to relieve transcription-associated superhelicity is likely affected by multiple aspects of the local chromatin environment.

For the genes associated with female-enriched R-loops, their ovary-biased expression raises the possibility that the FE R-loops form specifically in ovaries, yet the function of these R-loops remains elusive. Their unique motifs and association with ribosomal protein and translation-related genes distinguish them from the nonDE and ME groups. The sequence-specific transcription factor binding protein (M1BP) regulates transcription of ribosomal protein genes [54] and is maternally deposited and highly expressed in early embryos [55], but whether this contributes directly to the enrichment of FE R-loops at ribosomal and translation-associated genes remains to be seen. More locally, at genic loci, the distribution of FE R-loops across the gene body diverges from the typical TSS enrichment observed in the other DE groups. Additional scrutiny of these intragenic R-loops is required to determine their function in comparison with the more typical promoter and terminator-associated R-loops. Previous studies have demonstrated a role for R-loops regulating histone modifications and chromatin remodeling complex binding [56–58], raising the possibility that these female-enriched R-loops, rather than forming in response to DNA superhelicity, instead serve a distinct and context-specific regulatory function.

In summary, this work provides insight into the genome features, sequence motifs, and natural variation of the R-loop landscape in *Drosophila*. Our results are consistent with transcription rate, DNA torsion, and base composition being important determinants of R-loop formation. However, none of these properties fully explains the sex-biased R-loops that we identify, suggesting that other genetic or epigenetic mechanisms are involved in their formation. Further study of these male and female-biased R-loops will provide insight into their role in the establishment and maintenance of epigenetic differences between sexes.

Materials and methods

S1-DRIP-seq

As R-loops are known to be sensitive to sonication-induced degradation [30,31], we digested purified chromatin with S1 nuclease to remove the non-template strand prior to sonication, which has been shown to protect R-loop integrity through the sonication process [31]. Purification and sequencing of R-loops was performed as described in [31], with modifications. Briefly, whole adult DGRP379 and DGRP732 flies were separated by sex and homogenized, and genomic DNA (gDNA) was extracted using the DNEasy Blood & Tissue Kit (Qiagen). Extracted gDNA was digested with S1 nuclease to remove the non-template strand of DNA:RNA hybrids and sonicated with a Covaris S2 (Covaris) to an average fragment size of 100–300 bp. Consistent fragment size distribution across samples was confirmed via capillary electrophoresis (Agilent Fragment Analyzer, S4 Table). R-loops were immunoprecipitated with the S9.6 antibody (EMD Millipore) conjugated to Dynabeads Protein A (ThermoFisher), eluted with 1% SDS, and purified with the ChIP DNA Clean & Concentrator kit (Zymo Research). Illumina libraries from IP and Input samples were prepared with the DNA SMARTer ThruPLEX DNA-Seq kit (Takara Bio) and SMARTer DNA Unique Dual Index kit (Takara Bio).

For DRIP-seq of ovaries, ovaries were dissected from 2 to 10-days-old w1118 females (90 females per replicate) and homogenized in cold PBS. Genomic DNA was extracted and processed for S1-DRIP-seq as described above.

R-loops alignment and peak-calling

Reads were trimmed with trimmomatic [59] with the following options: “PE -phred33 ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:8:TRUE LEADING:15 TRAILING:15

SLIDINGWINDOW:3:10 MINLEN:36”, and aligned to the *D. melanogaster* reference genome assembly FlyBase version 6 [60,61] using bowtie2 [62] with the following options: “--no-mixed --no-discordant --dovetail --phred33 -X 1000”.

The ENCODE group has found that there is a consistent relationship between the number of ChIP-seq peaks identified and sequencing depth, without a clear saturation in most cases, due to an increased ability to identify low-affinity sites with increased sequencing depth [33]. Reads aligned to all female chromosomes and all male autosomes were therefore downsampled by 50% using samtools [63] with the following options: “-b -s 0.5”. High-confidence R-loop peaks were called using MACS2 [64] with the following options: “callpeak -f BAMPE -g dm -B -p 1e-3 -t IP.bam -c Input.bam”. To ensure reproducibility of R-loop loci between individual replicates of each condition, we employed the Irreproducibility Discovery Rate (IDR) framework [65] to identify a set of high-confidence DRIP peaks from all peaks called by MACS2. Pseudoreplicate and self-pseudoreplicate ratios confirmed that the shared peaks identified in each condition were reproducible (S2 Table). We assessed the chromosomal enrichment of R-loops by counting the number of peaks present on each of the 5 major chromosome arms of *D. melanogaster*, for each of our DRIP-seq samples. To control for the differences in length among the chromosome arms, we normalized the peak counts by chromosome length (in millions of basepairs, Mb). To determine whether the increased number of peaks observed on the X chromosome was statistically significant, we used a binomial test to test the null hypothesis that there is no difference in peak density between the X chromosome and autosomes.

Features overlap

DRIP peaks from each sample were intersected with genomic features and chromatin states using bedtools intersect [66] and plotted as log₂ enrichment of observed counts/expected counts, where observed counts comprised the total number of overlaps between a specific peakset and each feature or chromatin state, and expected counts comprised the average number of overlaps between each feature or chromatin state and 10000 iterations of shuffled R-loop peaks using bedtools shuffle with the “-chrom” option to preserve peak width and chromosomal location; reads aligning to tRNAs were counted only once per unique tRNA gene sequence. For each feature type, two-sided permutation test P-values were calculated as the proportion of permutations showing the same or more overlaps as the observed counts (i.e., enrichment) or the same or fewer overlaps as the observed counts (i.e., depletion). P-values were then adjusted for multiple hypothesis testing using the Benjamini-Hochberg correction. Genomic coordinates for gene, tRNA, snRNA, and snoRNA features were derived from FlyBase genome annotations [67]. Genomic coordinates for custom features were derived from the following studies: chromatin states [34], circRNAs [68], Polycomb-responsive elements [25], simple repeats previously identified from Repeatmasker [69], G-quadruplexes identified using pqsfinder (min_score = 52) [70], and strain-specific TAD boundaries identified from previously published Hi-C data [71] using HiC-Explorer [72].

Differential enrichment analysis

Differential R-loops were identified using DiffBind [38]. Briefly, IDR-called peaks were used to create a consensus peakset (bUseSummarizeOverlaps = TRUE, summits = FALSE) composed of DRIP-seq peaks found in at least two of the four samples (i.e., F379, F732, M379, M732). Sex-biased R-loops were subsequently identified using this consensus peakset (bContrasts = TRUE, adjusted p-value < 0.1).

Metagene analysis

Metaprofiles of expressed R-loop-containing genes were generated using deeptools [73] computeMatrix with the following options: “scale-regions --transcriptID mrna --skipZeros -p 20 -b 1000 -a 1000 --regionBodyLength 3000 --binSize 50”, followed by plotProfile with the following options: “--perGroup --plotType se”.

Motif identification and enrichment

Motif analysis was performed using STREME (MEME suite) [74] with the following p-value thresholds across DE groups: nonDE -pvt 1e-10, FE -pvt 1e-2, ME -pvt 1e-2. For comparison of motif enrichment across enrichment groups, the top five motifs from each group were analyzed using *gimme maelstrom* from GimmeMotifs [75] using the “--no-filter” option. To assess autosome-vs-X chromosome enrichment of these identified motifs, FIMO (MEME suite) [76] was used to identify all occurrences of each DNA motif on the 5 major chromosome arms. The ratio of motif occurrences per Mb on the X chromosomes versus autosomes was used to determine motif enrichment. Statistical significance was determined via a binomial test of the null hypothesis that there is no difference in motif density between the X chromosome and autosomes.

Adult females versus ovaries read coverage comparison

Female adult DRIP-seq and ovaries DRIP-seq read coverage at loci from nonDE and FE peaksets was compared using deeptools [73] multiBamSummary with the following options: “--genomeChunkSize 129941135 --outRawCounts”. Output of raw read counts per loci was normalized to total read count per sample and plotted as reads per megabase (Mb). Statistical significance was determined using Spearman correlation coefficient.

RNA-seq and gene expression analysis

We used ~10 whole adult DGRP379 and DGRP732 males and females. Flies were homogenized with an electric pestle in DNA/RNA Shield solution (Zymo Research). Homogenized tissue was digested with Proteinase K and RNA was purified with the Zymo Quick-RNA Plus Kit (Zymo Research). Ribosomal RNAs were removed using siTools rRNA depletion Kit (Galen Laboratory Supplies) and MyOne Streptavidin C1 Dynabeads (ThermoFisher) (#65001). Ribosomal RNA-depleted RNA was purified using the RNA Clean and Concentrator-5 kit (Zymo Research). Illumina libraries were generated using NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB).

RNA sequencing reads were first aligned to the FlyBase r6.27 rRNA sequences using HISAT2 [77]. Non-ribosomal sequences were subsequently aligned to FlyBase r6.27 transcript sequences using htseq-ct [78]. Counts were filtered to include only expressed transcripts using DESeq2 [79] ($\text{rowSums}(\text{DESeqDataSetFromHTSeqCount}) \geq 1$), which were subsequently normalized using rlog transformation ($\text{blind} = \text{TRUE}$) in DESeq2 [79].

FlyAtlas microarray expression analysis

Microarray expression data were downloaded from <http://flyatlas.org/>. Genes were separated by differential enrichment group: “nonDE,” “FE,” “ME,” and “no R-loop.” The minimum gene expression threshold for each tissue was determined by detectable expression in at least two of four microarrays (columns “OvaryPresent,” “TestisCall,” or “BrainPresent” ≥ 2).

Supporting information

S1 Fig. Related to Fig 1, R-loop identification & feature enrichment in *D. melanogaster* adults. (A) R-loop formation at various genetic features on autosomes (upper panel) and the X chromosome (lower panel). R-loop enrichment is shown as the observed number of DRIP-seq peaks overlapping each feature (or chromatin state) divided by the expected number of peaks (see [Methods](#)). P-values were calculated via a Permutation Test with Benjamini-Hochberg correction for multiple comparisons, * = corrected $p < 0.05$. (B) Metaprofiles of R-loop signal across protein-coding genes (from [Fig 1E](#)), overlapped by condition, grouped by chromosome. The solid lines represent the mean DRIP-seq signal within each metagene bin and the shading represents the standard error of the mean.
(TIF)

S2 Fig. Related to Fig 2, Differential enrichment and motif analysis of R-loops. (A) Metaprofiles of R-loop signal across protein-coding genes (from [Fig 2G](#)), separated by condition and by DE group. The solid lines represent the mean DRIP-seq signal within each metagene bin, and the shading represents the standard error of the mean. (B) Gene expression analysis of R-loop-containing genes by sex, separated by DE group versus R-loop-absent (no R-loop) genes across all chromosomes, autosomes, and the X chromosome, plotted as rlog-normalized expression. (C) Microarray gene expression analysis from the FlyAtlas [41] of R-loop-containing genes across tissues, separated by DE group. Wilcoxon test, **,*** = $p < 0.01, 0.001$. (D) DRIP-seq normalized read coverage of whole female flies and ovaries. The black line in each plot represents a slope of 1 with intersect at 0. Spearman's $\rho = 0.72, 0.76$ and $p < 2.2e-16, 2.2e-16$ [nonDE peaks-F379, F732], Spearman's $\rho = 0.74, 0.79$ and $p < 2.2e-16, 2.2e-16$ [FE peaks-F379, F732].
(TIF)

S3 Fig. Related to Fig 3, X chromosome-specific R-loop enrichment. (A) Male-to-female gene expression ratios by DE group on autosomal and X chromosome genes. Wilcoxon test, **,*** = $p < 0.01, 0.001$.
(TIF)

S1 Table. DRIP-seq sequencing statistics and down-sampling by chromosome.
(XLSX)

S2 Table. IDR analysis determines the rate of reproducibility between replicates.
(XLSX)

S3 Table. GO Enrichment analysis of sex-biased R-loops.
(XLSX)

S4 Table. DRIP-seq libraries Agilent Fragment Analyzer statistics.
(XLSX)

Acknowledgments

We gratefully acknowledge the Office of Advanced Research Computing (OARC) at Rutgers, The State University of New Jersey for providing access to the Amarel cluster and associated research computing resources that have contributed to the results reported here.

Author Contributions

Conceptualization: Timothy J. Stanek, Christopher E. Ellison.

Data curation: Timothy J. Stanek.

Formal analysis: Timothy J. Stanek, Rohan M Mehra, Christopher E. Ellison.

Funding acquisition: Christopher E. Ellison.

Investigation: Timothy J. Stanek, Rohan M Mehra, Christopher E. Ellison.

Methodology: Timothy J. Stanek, Weihuan Cao, Christopher E. Ellison.

Project administration: Christopher E. Ellison.

Resources: Timothy J. Stanek, Weihuan Cao, Christopher E. Ellison.

Software: Timothy J. Stanek, Rohan M Mehra.

Supervision: Christopher E. Ellison.

Validation: Timothy J. Stanek.

Visualization: Timothy J. Stanek.

Writing – original draft: Timothy J. Stanek, Christopher E. Ellison.

Writing – review & editing: Timothy J. Stanek, Christopher E. Ellison.

References

1. Niehrs C, Luke B. Regulatory R-loops as facilitators of gene expression and genome stability. *Nat Rev Mol Cell Biol.* 2020; 21(3):167–78. Epub 20200131. <https://doi.org/10.1038/s41580-019-0206-3> PMID: 32005969; PubMed Central PMCID: PMC7116639.
2. Crossley MP, Bocek M, Cimprich KA. R-Loops as Cellular Regulators and Genomic Threats. *Mol Cell.* 2019; 73(3):398–411. <https://doi.org/10.1016/j.molcel.2019.01.024> PMID: 30735654; PubMed Central PMCID: PMC6402819.
3. Sagie S, Toubiana S, Hartono SR, Katzir H, Tzur-Gilat A, Havazelet S, et al. Telomeres in ICF syndrome cells are vulnerable to DNA damage due to elevated DNA:RNA hybrids. *Nat Commun.* 2017; 8:14015. Epub 20170124. <https://doi.org/10.1038/ncomms14015> PMID: 28117327; PubMed Central PMCID: PMC5286223.
4. Stork CT, Bocek M, Crossley MP, Sollier J, Sanz LA, Chedin F, et al. Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage. *Elife.* 2016; 5. Epub 20160823. <https://doi.org/10.7554/eLife.17548> PMID: 27552054; PubMed Central PMCID: PMC5030092.
5. Bersani F, Lee E, Kharchenko PV, Xu AW, Liu M, Xega K, et al. Pericentromeric satellite repeat expansions through RNA-derived DNA intermediates in cancer. *Proc Natl Acad Sci U S A.* 2015; 112(49):15148–53. Epub 20151102. <https://doi.org/10.1073/pnas.1518008112> PMID: 26575630; PubMed Central PMCID: PMC4679016.
6. Arora R, Lee Y, Wischnewski H, Brun CM, Schwarz T, Azzalin CM. RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. *Nat Commun.* 2014; 5:5220. Epub 20141021. <https://doi.org/10.1038/ncomms6220> PMID: 25330849; PubMed Central PMCID: PMC4218956.
7. Sciamanna I, De Luca C, Spadafora C. The Reverse Transcriptase Encoded by LINE-1 Retrotransposons in the Genesis, Progression, and Therapy of Cancer. *Front Chem.* 2016; 4:6. Epub 20160211. <https://doi.org/10.3389/fchem.2016.00006> PMID: 26904537; PubMed Central PMCID: PMC4749692.
8. Ginno PA, Lott PL, Christensen HC, Korf I, Chedin F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell.* 2012; 45(6):814–25. Epub 20120301. <https://doi.org/10.1016/j.molcel.2012.01.017> PMID: 22387027; PubMed Central PMCID: PMC3319272.
9. Huppert JL. Thermodynamic prediction of RNA-DNA duplex-forming regions in the human genome. *Mol Biosyst.* 2008; 4(6):686–91. Epub 20080430. <https://doi.org/10.1039/b800354h> PMID: 18493667.
10. Stolz R, Sulthana S, Hartono SR, Malig M, Benham CJ, Chedin F. Interplay between DNA sequence and negative superhelicity drives R-loop structures. *Proc Natl Acad Sci U S A.* 2019; 116(13):6260–9. Epub 20190308. <https://doi.org/10.1073/pnas.1819476116> PMID: 30850542; PubMed Central PMCID: PMC6442632.

11. Chedin F, Benham CJ. Emerging roles for R-loop structures in the management of topological stress. *J Biol Chem*. 2020; 295(14):4684–95. Epub 20200227. <https://doi.org/10.1074/jbc.REV119.006364> PMID: 32107311; PubMed Central PMCID: PMC7135976.
12. Drolet M, Bi X, Liu LF. Hypernegative supercoiling of the DNA template during transcription elongation in vitro. *J Biol Chem*. 1994; 269(3):2068–74. PMID: 8294458.
13. Phoenix P, Raymond MA, Masse E, Drolet M. Roles of DNA topoisomerases in the regulation of R-loop formation in vitro. *J Biol Chem*. 1997; 272(3):1473–9. <https://doi.org/10.1074/jbc.272.3.1473> PMID: 8999816.
14. Masse E, Drolet M. Escherichia coli DNA topoisomerase I inhibits R-loop formation by relaxing transcription-induced negative supercoiling. *J Biol Chem*. 1999; 274(23):16659–64. <https://doi.org/10.1074/jbc.274.23.16659> PMID: 10347234.
15. Masse E, Phoenix P, Drolet M. DNA topoisomerases regulate R-loop formation during transcription of the *rrnB* operon in Escherichia coli. *J Biol Chem*. 1997; 272(19):12816–23. <https://doi.org/10.1074/jbc.272.19.12816> PMID: 9139742.
16. Drolet M. Growth inhibition mediated by excess negative supercoiling: the interplay between transcription elongation, R-loop formation and DNA topology. *Mol Microbiol*. 2006; 59(3):723–30. <https://doi.org/10.1111/j.1365-2958.2005.05006.x> PMID: 16420346.
17. Racko D, Benedetti F, Dorier J, Stasiak A. Transcription-induced supercoiling as the driving force of chromatin loop extrusion during formation of TADs in interphase chromosomes. *Nucleic Acids Res*. 2018; 46(4):1648–60. <https://doi.org/10.1093/nar/gkx1123> PMID: 29140466; PubMed Central PMCID: PMC5829651.
18. El Hage A, Webb S, Kerr A, Tollervey D. Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria. *PLoS Genet*. 2014; 10(10): e1004716. Epub 20141030. <https://doi.org/10.1371/journal.pgen.1004716> PMID: 25357144; PubMed Central PMCID: PMC4214602.
19. An L, Yang T, Yang J, Nuebler J, Xiang G, Hardison RC, et al. OnTAD: hierarchical domain structure reveals the divergence of activity among TADs and boundaries. *Genome Biol*. 2019; 20(1):282. Epub 20191218. <https://doi.org/10.1186/s13059-019-1893-y> PMID: 31847870; PubMed Central PMCID: PMC6918570.
20. Chen L, Chen JY, Zhang X, Gu Y, Xiao R, Shao C, et al. R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. *Mol Cell*. 2017; 68(4):745–57 e5. Epub 20171102. <https://doi.org/10.1016/j.molcel.2017.10.008> PMID: 29104020; PubMed Central PMCID: PMC5957070.
21. Ginno PA, Lim YW, Lott PL, Korf I, Chedin F. GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome Res*. 2013; 23(10):1590–600. Epub 20130718. <https://doi.org/10.1101/gr.158436.113> PMID: 23868195; PubMed Central PMCID: PMC3787257.
22. Skourti-Stathaki K, Kamieniarz-Gdula K, Proudfoot NJ. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature*. 2014; 516(7531):436–9. Epub 20141005. <https://doi.org/10.1038/nature13787> PMID: 25296254; PubMed Central PMCID: PMC4272244.
23. Skourti-Stathaki K, Proudfoot NJ, Gromak N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol Cell*. 2011; 42(6):794–805. <https://doi.org/10.1016/j.molcel.2011.04.026> PMID: 21700224; PubMed Central PMCID: PMC3145960.
24. Skourti-Stathaki K, Torlai Triglia E, Warburton M, Voigt P, Bird A, Pombo A. R-Loops Enhance Polycomb Repression at a Subset of Developmental Regulator Genes. *Mol Cell*. 2019; 73(5):930–45 e4. Epub 20190129. <https://doi.org/10.1016/j.molcel.2018.12.016> PMID: 30709709; PubMed Central PMCID: PMC6414425.
25. Alecki C, Chiwara V, Sanz LA, Grau D, Arias Perez O, Boulier EL, et al. RNA-DNA strand exchange by the *Drosophila* Polycomb complex PRC2. *Nat Commun*. 2020; 11(1):1781. Epub 20200414. <https://doi.org/10.1038/s41467-020-15609-x> PMID: 32286294; PubMed Central PMCID: PMC7156742.
26. Conn VM, Hugouvieux V, Nayak A, Conos SA, Capovilla G, Cildir G, et al. A circRNA from SEPAL-LATA3 regulates splicing of its cognate mRNA through R-loop formation. *Nat Plants*. 2017; 3:17053. Epub 20170418. <https://doi.org/10.1038/nplants.2017.53> PMID: 28418376.
27. Xu X, Zhang J, Tian Y, Gao Y, Dong X, Chen W, et al. CircRNA inhibits DNA damage repair by interacting with host gene. *Mol Cancer*. 2020; 19(1):128. Epub 20200824. <https://doi.org/10.1186/s12943-020-01246-x> PMID: 32838810; PubMed Central PMCID: PMC7446195.
28. Cohen S, Puget N, Lin YL, Clouaire T, Aguirrebengoa M, Rocher V, et al. Senataxin resolves RNA:DNA hybrids forming at DNA double-strand breaks to prevent translocations. *Nat Commun*. 2018; 9(1):533.

- Epub 20180207. <https://doi.org/10.1038/s41467-018-02894-w> PMID: 29416069; PubMed Central PMCID: PMC5803260.
29. D'Alessandro G, Whelan DR, Howard SM, Vitelli V, Renaudin X, Adamowicz M, et al. BRCA2 controls DNA:RNA hybrid level at DSBs by mediating RNase H2 recruitment. *Nat Commun.* 2018; 9(1):5376. Epub 20181218. <https://doi.org/10.1038/s41467-018-07799-2> PMID: 30560944; PubMed Central PMCID: PMC6299093.
 30. Halasz L, Karanyi Z, Boros-Olah B, Kuik-Rozsa T, Sipos E, Nagy E, et al. RNA-DNA hybrid (R-loop) immunoprecipitation mapping: an analytical workflow to evaluate inherent biases. *Genome Res.* 2017; 27(6):1063–73. Epub 20170324. <https://doi.org/10.1101/gr.219394.116> PMID: 28341774; PubMed Central PMCID: PMC5453320.
 31. Wahba L, Costantino L, Tan FJ, Zimmer A, Koshland D. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes Dev.* 2016; 30(11):1327–38. <https://doi.org/10.1101/gad.280834.116> PMID: 27298336; PubMed Central PMCID: PMC4911931.
 32. Riddle NC, Elgin SCR. The *Drosophila* Dot Chromosome: Where Genes Flourish Amidst Repeats. *Genetics.* 2018; 210(3):757–72. <https://doi.org/10.1534/genetics.118.301146> PMID: 30401762; PubMed Central PMCID: PMC6218221.
 33. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglu S, et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* 2012; 22(9):1813–31. <https://doi.org/10.1101/gr.136184.111> PMID: 22955991; PubMed Central PMCID: PMC3431496.
 34. Filion GJ, van Bommel JG, Braunschweig U, Talhout W, Kind J, Ward LD, et al. Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell.* 2010; 143(2):212–24. Epub 20100930. <https://doi.org/10.1016/j.cell.2010.09.009> PMID: 20888037; PubMed Central PMCID: PMC3119929.
 35. Aguilera A, Garcia-Muse T. R loops: from transcription byproducts to threats to genome stability. *Mol Cell.* 2012; 46(2):115–24. <https://doi.org/10.1016/j.molcel.2012.04.009> PMID: 22541554.
 36. Lago S, Nadai M, Cernilogar FM, Kazerani M, Dominiguez Moreno H, Schotta G, et al. Promoter G-quadruplexes and transcription factors cooperate to shape the cell type-specific transcriptome. *Nat Commun.* 2021; 12(1):3885. Epub 20210623. <https://doi.org/10.1038/s41467-021-24198-2> PMID: 34162892; PubMed Central PMCID: PMC8222265.
 37. Zeng C, Onoguchi M, Hamada M. Association analysis of repetitive elements and R-loop formation across species. *Mob DNA.* 2021; 12(1):3. Epub 20210120. <https://doi.org/10.1186/s13100-021-00231-5> PMID: 33472695; PubMed Central PMCID: PMC7818932.
 38. Stark RB G. DiffBind: differential binding analysis of ChIP-Seq peak data 2011.
 39. Xu W, Xu H, Li K, Fan Y, Liu Y, Yang X, et al. The R-loop is a common chromatin feature of the Arabidopsis genome. *Nat Plants.* 2017; 3(9):704–14. Epub 20170828. <https://doi.org/10.1038/s41477-017-0004-x> PMID: 28848233.
 40. Chan YA, Aristizabal MJ, Lu PY, Luo Z, Hamza A, Kobor MS, et al. Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-chip. *PLoS Genet.* 2014; 10(4):e1004288. Epub 20140417. <https://doi.org/10.1371/journal.pgen.1004288> PMID: 24743342; PubMed Central PMCID: PMC3990523.
 41. Chintapalli VR, Wang J, Dow JA. Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nat Genet.* 2007; 39(6):715–20. <https://doi.org/10.1038/ng2049> PMID: 17534367.
 42. Grath S, Parsch J. Sex-Biased Gene Expression. *Annu Rev Genet.* 2016; 50:29–44. Epub 20160826. <https://doi.org/10.1146/annurev-genet-120215-035429> PMID: 27574843.
 43. Khodursky S, Svetec N, Durkin SM, Zhao L. The evolution of sex-biased gene expression in the *Drosophila* brain. *Genome Res.* 2020; 30(6):874–84. Epub 20200618. <https://doi.org/10.1101/gr.259069.119> PMID: 32554780; PubMed Central PMCID: PMC7370887.
 44. Lucchesi JC, Kuroda MI. Dosage compensation in *Drosophila*. *Cold Spring Harb Perspect Biol.* 2015; 7(5). Epub 20150501. <https://doi.org/10.1101/cshperspect.a019398> PMID: 25934013; PubMed Central PMCID: PMC4448616.
 45. Samata M, Akhtar A. Dosage Compensation of the X Chromosome: A Complex Epigenetic Assignment Involving Chromatin Regulators and Long Noncoding RNAs. *Annu Rev Biochem.* 2018; 87:323–50. Epub 20180418. <https://doi.org/10.1146/annurev-biochem-062917-011816> PMID: 29668306.
 46. Conrad T, Cavalli FM, Holz H, Hallaceli E, Kind J, Ilik I, et al. The MOF chromobarrel domain controls genome-wide H4K16 acetylation and spreading of the MSL complex. *Dev Cell.* 2012; 22(3):610–24. <https://doi.org/10.1016/j.devcel.2011.12.016> PMID: 22421046.
 47. Lee H, Oliver B. Non-canonical *Drosophila* X chromosome dosage compensation and repressive topologically associated domains. *Epigenetics Chromatin.* 2018; 11(1):62. Epub 20181024. <https://doi.org/10.1186/s13072-018-0232-y> PMID: 30355339; PubMed Central PMCID: PMC6199721.

48. Beckmann K, Grskovic M, Gebauer F, Hentze MW. A dual inhibitory mechanism restricts msl-2 mRNA translation for dosage compensation in *Drosophila*. *Cell*. 2005; 122(4):529–40. <https://doi.org/10.1016/j.cell.2005.06.011> PMID: 16122421.
49. Lam KC, Muhlpfordt F, Vaquerizas JM, Raja SJ, Holz H, Luscombe NM, et al. The NSL complex regulates housekeeping genes in *Drosophila*. *PLoS Genet*. 2012; 8(6):e1002736. Epub 20120614. <https://doi.org/10.1371/journal.pgen.1002736> PMID: 22723752; PubMed Central PMCID: PMC3375229.
50. Chen PB, Chen HV, Acharya D, Rando OJ, Fazio TG. R loops regulate promoter-proximal chromatin architecture and cellular differentiation. *Nat Struct Mol Biol*. 2015; 22(12):999–1007. Epub 20151109. <https://doi.org/10.1038/nsmb.3122> PMID: 26551076; PubMed Central PMCID: PMC4677832.
51. Gallach M, Arnau V, Marin I. Global patterns of sequence evolution in *Drosophila*. *BMC Genomics*. 2007; 8:408. Epub 20071109. <https://doi.org/10.1186/1471-2164-8-408> PMID: 17996078; PubMed Central PMCID: PMC2180185.
52. Dunlap D, Yokoyama R, Ling H, Sun HY, McGill K, Cugusi S, et al. Distinct contributions of MSL complex subunits to the transcriptional enhancement responsible for dosage compensation in *Drosophila*. *Nucleic Acids Res*. 2012; 40(22):11281–91. Epub 20121009. <https://doi.org/10.1093/nar/gks890> PMID: 23047951; PubMed Central PMCID: PMC3526317.
53. Cugusi S, Ramos E, Ling H, Yokoyama R, Luk KM, Lucchesi JC. Topoisomerase II plays a role in dosage compensation in *Drosophila*. *Transcription*. 2013; 4(5):238–50. <https://doi.org/10.4161/trns.26185> PMID: 23989663.
54. Baumann DG, Gilmour DS. A sequence-specific core promoter-binding transcription factor recruits TRF2 to coordinately transcribe ribosomal protein genes. *Nucleic Acids Res*. 2017; 45(18):10481–91. <https://doi.org/10.1093/nar/gkx676> PMID: 28977400; PubMed Central PMCID: PMC5737516.
55. Omura CS, Lott SE. The conserved regulatory basis of mRNA contributions to the early *Drosophila* embryo differs between the maternal and zygotic genomes. *PLoS Genet*. 2020; 16(3):e1008645. Epub 20200330. <https://doi.org/10.1371/journal.pgen.1008645> PMID: 32226006; PubMed Central PMCID: PMC7145188.
56. Beckedorff FC, Ayupe AC, Crocci-Souza R, Amaral MS, Nakaya HI, Soltys DT, et al. The intronic long noncoding RNA ANRASSF1 recruits PRC2 to the RASSF1A promoter, reducing the expression of RASSF1A and increasing cell proliferation. *PLoS Genet*. 2013; 9(8):e1003705. Epub 20130822. <https://doi.org/10.1371/journal.pgen.1003705> PMID: 23990798; PubMed Central PMCID: PMC3749938.
57. Boque-Sastre R, Soler M, Oliveira-Mateos C, Portela A, Moutinho C, Sayols S, et al. Head-to-head anti-sense transcription and R-loop formation promotes transcriptional activation. *Proc Natl Acad Sci U S A*. 2015; 112(18):5785–90. Epub 20150422. <https://doi.org/10.1073/pnas.1421197112> PMID: 25902512; PubMed Central PMCID: PMC4426458.
58. Gibbons HR, Shaginuova G, Kim LC, Chapman N, Spurlock CF 3rd, Aune TM. Divergent lncRNA GATA3-AS1 Regulates GATA3 Transcription in T-Helper 2 Cells. *Front Immunol*. 2018; 9:2512. Epub 20181029. <https://doi.org/10.3389/fimmu.2018.02512> PMID: 30420860; PubMed Central PMCID: PMC6215836.
59. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014; 30(15):2114–20. Epub 20140401. <https://doi.org/10.1093/bioinformatics/btu170> PMID: 24695404; PubMed Central PMCID: PMC4103590.
60. Hoskins RA, Carlson JW, Wan KH, Park S, Mendez I, Galle SE, et al. The Release 6 reference sequence of the *Drosophila melanogaster* genome. *Genome Res*. 2015; 25(3):445–58. Epub 20150114. <https://doi.org/10.1101/gr.185579.114> PMID: 25589440; PubMed Central PMCID: PMC4352887.
61. Thurmond J, Goodman JL, Strelets VB, Attrill H, Gramates LS, Marygold SJ, et al. FlyBase 2.0: the next generation. *Nucleic Acids Res*. 2019; 47(D1):D759–D65. <https://doi.org/10.1093/nar/gky1003> PMID: 30364959; PubMed Central PMCID: PMC6323960.
62. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012; 9(4):357–9. Epub 20120304. <https://doi.org/10.1038/nmeth.1923> PMID: 22388286; PubMed Central PMCID: PMC3322381.
63. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021; 10(2). <https://doi.org/10.1093/gigascience/giab008> PMID: 33590861; PubMed Central PMCID: PMC7931819.
64. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008; 9(9):R137. Epub 20080917. <https://doi.org/10.1186/gb-2008-9-9-r137> PMID: 18798982; PubMed Central PMCID: PMC2592715.
65. Li QH, Brown JB, Huang HY, Bickel PJ. Measuring Reproducibility of High-Throughput Experiments. *Ann Appl Stat*. 2011; 5(3):1752–79. <https://doi.org/10.1214/11-Aoas466> WOS:000300382500003.

66. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26(6):841–2. Epub 20100128. <https://doi.org/10.1093/bioinformatics/btq033> PMID: 20110278; PubMed Central PMCID: PMC2832824.
67. Larkin A, Marygold SJ, Antonazzo G, Attrill H, Dos Santos G, Garapati PV, et al. FlyBase: updates to the *Drosophila melanogaster* knowledge base. *Nucleic Acids Res*. 2021; 49(D1):D899–D907. <https://doi.org/10.1093/nar/gkaa1026> PMID: 33219682; PubMed Central PMCID: PMC7779046.
68. Westholm JO, Miura P, Olson S, Shenker S, Joseph B, Sanfilippo P, et al. Genome-wide analysis of *drosophila* circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation. *Cell Rep*. 2014; 9(5):1966–80. Epub 20141126. <https://doi.org/10.1016/j.celrep.2014.10.062> PMID: 25544350; PubMed Central PMCID: PMC4279448.
69. Shah K, Cao W, Ellison CE. Adenine Methylation in *Drosophila* Is Associated with the Tissue-Specific Expression of Developmental and Regulatory Genes. *G3 (Bethesda)*. 2019; 9(6):1893–900. Epub 20190605. <https://doi.org/10.1534/g3.119.400023> PMID: 30988038; PubMed Central PMCID: PMC6553526.
70. Hon J, Martinek T, Zendulka J, Lexa M. pqsfinder: an exhaustive and imperfection-tolerant search tool for potential quadruplex-forming sequences in R. *Bioinformatics*. 2017; 33(21):3373–9. <https://doi.org/10.1093/bioinformatics/btx413> PMID: 29077807.
71. Ellison CE, Cao W. Nanopore sequencing and Hi-C scaffolding provide insight into the evolutionary dynamics of transposable elements and piRNA production in wild strains of *Drosophila melanogaster*. *Nucleic Acids Res*. 2020; 48(1):290–303. <https://doi.org/10.1093/nar/gkz1080> PMID: 31754714; PubMed Central PMCID: PMC6943127.
72. Ramirez F, Bhardwaj V, Arrigoni L, Lam KC, Gruning BA, Villaveces J, et al. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun*. 2018; 9(1):189. Epub 20180115. <https://doi.org/10.1038/s41467-017-02525-w> PMID: 29335486; PubMed Central PMCID: PMC5768762.
73. Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 2016; 44(W1):W160–5. Epub 20160413. <https://doi.org/10.1093/nar/gkw257> PMID: 27079975; PubMed Central PMCID: PMC4987876.
74. Bailey TL. STREME: Accurate and versatile sequence motif discovery. *Bioinformatics*. 2021. Epub <https://doi.org/10.1093/bioinformatics/btab203> PMID: 33760053. PubMed Central PMCID: PMC8479671.
75. van Heeringen SJ, Veenstra GJ. GimmeMotifs: a de novo motif prediction pipeline for ChIP-sequencing experiments. *Bioinformatics*. 2011; 27(2):270–1. Epub 20101115. <https://doi.org/10.1093/bioinformatics/btq636> PMID: 21081511; PubMed Central PMCID: PMC3018809.
76. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinformatics*. 2011; 27(7):1017–8. Epub 20110216. <https://doi.org/10.1093/bioinformatics/btr064> PMID: 21330290; PubMed Central PMCID: PMC3065696.
77. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015; 12(4):357–60. Epub 20150309. <https://doi.org/10.1038/nmeth.3317> PMID: 25751142; PubMed Central PMCID: PMC4655817.
78. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015; 31(2):166–9. Epub 20140925. <https://doi.org/10.1093/bioinformatics/btu638> PMID: 25260700; PubMed Central PMCID: PMC4287950.
79. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014; 15(12):550. <https://doi.org/10.1186/s13059-014-0550-8> PMID: 25516281; PubMed Central PMCID: PMC4302049.