R-loop Mapping and Characterization During Drosophila Embryogenesis Reveals Developmental Plasticity in R-loop Signatures

Alexander Munden[1], **Mary Lauren Benton**[2], **John A. Capra**[3], **Jared T. Nordman**[1,*]

[1] Department of Biological Sciences, Vanderbilt University, Nashville, TN 37212, USA

[2] Department of Computer Science, Baylor University, Waco, TX 76706, USA

[3] Bakar Computational Health Sciences Institute and Department of Epidemiology and Biostatistics, University of California, San Francisco, CA 94103, USA

Abstract

R-loops are involved in transcriptional regulation, DNA and histone post-translational modifications, genome replication and genome stability. To what extent R-loop abundance and genome-wide localization is actively regulated during metazoan embryogenesis is unknown. Drosophila embryogenesis provides a powerful system to address these questions due to its well-characterized developmental program, the sudden onset of zygotic transcription and available genome-wide data sets. Here, we measure the overall abundance and genome localization of R-loops in early and late-stage embryos relative to Drosophila cultured cells. We demonstrate that absolute R-loop levels change during embryogenesis and that RNaseH1 catalytic activity is critical for embryonic development. R-loop mapping by strand-specific DRIP-seq reveals that R-loop localization is plastic across development, both in the genes which form R-loops and where they localize relative to gene bodies. Importantly, these changes are not driven by changes in the transcriptional program. Negative GC skew and absolute changes in AT skew are associated with R-loop formation in Drosophila. Furthermore, we demonstrate that while some chromatin binding proteins and histone modifications such as H3K27me3 are associated with R-loops throughout development, other chromatin factors associated with R-loops in a developmental specific manner. Our findings highlight the importance and developmental plasticity of R-loops during Drosophila embryogenesis.

Correspondence to Jared T. Nordman: jared.nordman@vanderbilt.edu (J.T. Nordman).

Keywords

Chromatin; Epigenetics; RNA

Introduction

R-loops are a three-stranded nucleic acid structure canonically formed when nascent RNA from transcription reanneals to the template DNA strand, resulting in a displaced single strand of DNA.[1] R-loops were initially identified at the highly transcribed *18S* and *28S* sequences within the rDNA locus of *Drosophila melanogaster*.[2,3] More recent studies have demonstrated that R-loops are critical for a diverse set of biological processes.[4,5] In fact, genome-wide R-loop mapping studies have revealed that R-loops are abundant in eukaryotes and can occupy 10% or more of the genome.[6-17] While R-loops were identified over 40 years ago, their physiological relevance remained elusive for many years.

R-loops are found in all domains of life and their formation is often conserved across cell types and even species.[18] Deciphering the function of R-loops, however, has been challenging due to their diverse and sometimes contradictory roles in genome function. R-loops are essential for initiation of replication in plasmids and promote mitochondrial genome stability.[19,20] In contrast, R-loops can block replication fork progression and promote genome instability in an orientation-specific manner.[21,22] While potentially causing double-strand breaks at head-on replication-transcription conflicts, R-loops can promote recombination and double strand break repair.[23,24] R-loops also have diverse roles in transcription and chromatin function. In mammalian cells, R-loops have been shown to regulate both histone and DNA methylation at promoter regions.[12,14] While R-loops are often associated with histone modifications correlated with active transcription, recent work has shown that R-loops can help recruit the Polycomb complex to target loci to promote transcriptional silencing.[25,26] Genome-wide R-loop mapping studies in yeast, plants and mammalian cultured cells have identified factors such as DNA sequence, DNA topology and histone modifications associated with R-loop formation.[14,27,28] R-loop mapping studies in plants and mammalian cells have further revealed that R-loop formation can be dynamic as a function of development.[8,10,29] The extent of R-loop plasticity in other metazoans has yet to be defined. Studying R-loops in the context of development could provide insight into the functional roles R-loops play in establishing developmental-specific changes in chromatin structure, function and transcriptional programs.

Drosophila provide a well-established developmental system to interrogate R-loop plasticity during development. At the earliest stages of Drosophila embryogenesis, rapid cell proliferation is driven by maternally stockpiled proteins and RNA.[30] Approximately two hours after fertilization, zygotic genome activation is triggered and the transcription of over 3000 genes necessary for growth and differentiation are induced in a process known as the maternal-to-zygotic transition (MZT).[31,32] Prior to the MZT, cells are largely undifferentiated and have abbreviated cell cycles.[33] After the MZT, however, the cell cycle slows and cells become differentiated as morphogenesis proceeds.[34] The changes in cell cycle programs, the onset of zygotic gene activation and cell differentiation

during embryogenesis provide a unique system to interrogate whether R-loop formation or resolution impacts embryogenesis and the extent to which, if any, R-loop position and properties change as a function of development.

In this study, we measured R-loop abundance and position in Drosophila embryos and cultured cells. We show that absolute R-loop levels change during embryogenesis and resolution of R-loops is essential for embryogenesis. We mapped R-loops at near base-pair resolution in 2–3 hour embryos (immediately after the MZT), late-stage embryos (14–16 hours after fertilization) and cultured S2 cells, which are derived from late-stage embryos. We show that, while some sites of R-loop formation are constant during development, there is extensive R-loop plasticity during Drosophila development. Furthermore, we were able to demonstrate changes in the localization of R-loops across gene bodies and the role AT and GC skew play in Drosophila R-loop formation. By leveraging data available through modENCODE and other publicly available datasets, we were able to identify specific histone modifications and chromatin binding proteins associated with R-loop formation in Drosophila and the role active transcription has on R-loop formation. Importantly, developmental-specific R-loops are not driven by transcriptional changes, emphasizing the role that chromatin and R-loop binding proteins play in regulating R-loop formation. Our work establishes Drosophila as a powerful developmental model system to study R-loop biology.

Results

R-loop abundance is developmentally regulated and R-loop homeostasis is necessary for development

To determine if R-loop abundance and genomic location are regulated throughout development, we turned to the powerful Drosophila embryogenesis system. For our analysis, we chose embryos at two distinct time points: 2–3 hours after egg laying (AEL) and 14–16 hours AEL (Figure 1(A)). The 2–3 hour time point corresponds with the onset of the maternal-to-zygotic transition (MZT) occurring during nuclear cleavage cycle 14.[35] This time point represents the onset of zygotic transcription and allows us to draw upon the wealth of scientific literature that has previously been published, including time-matched modENCODE datasets. The wide-scale activation of zygotic transcription at this time point should provide the first opportunity for R-loop formation during development. To complement this developmental stage, we chose 14–16 hour AEL embryos to understand how R-loop formation might differ in differentiated cells with a more mature chromatin environment and a transcription program characterized by cell-type-specific maintenance.[36-38] S2 cells, an established Drosophila cell culture line derived from late-stage embryos, were used to determine how R-loops might differ between embryos and cultured cells, where the majority of R-loop research has been conducted.[39]

To begin, we asked whether the absolute levels of R-loops are influenced by development. To this end, genomic DNA was extracted from each sample and spotted onto a nitrocellulose membrane and probed with the S9.6 antibody, which recognizes RNA:DNA hybrids.[40] S2 cells and 2–3 h embryos showed similar amounts of S9.6 signal, while DNA from 14-16 h embryos showed a significant decrease in S9.6 signal (Figure 1(B)). To ensure the

S9.6 signal stems from R-loops, we pretreated control samples with RNase H1, which degrades the RNA moiety of a RNA:DNA hybrid. The S9.6 antibody has some specificity to double-stranded RNA and Drosophila embryos are known to contain dsRNA.[41] In fact, in the RNase H1 treated control samples we initially detected some signal with the S9.6 antibody, which was completely eliminated by pretreatment with RNase III. Therefore, for all R-loop assays we pretreat our samples with RNase III to ensure S9.6 signal isn't due to dsRNA.

Next, we asked whether perturbing R-loop homeostasis affects embryogenesis. *rnh1* mutants survive into larval development, which suggests that *rnh1* and R-loop processing may be dispensable during embryogenesis.[42] More likely, however, *rnh1* mutant embryos survive from maternal stockpiles of RNase H1. To circumvent this, we generated flies that overexpress a GFP-tagged, nuclear localized version of Drosophila RNase H1 or a catalytically dead version of the same protein (RNase H1[CD]). To ensure that the RNase H1 proteins were maternally deposited and present at the earliest stages of embryogenesis, we used the pUASz expression system coupled with the maternal triple driver.[43,44] After confirming that the GFP was observable by Western blot (Supplemental Figure 1(A)), we performed a hatch rate assay to determine if perturbing RNaseH1 catalytic activity affects embryogenesis. We observed a consistent but statistically insignificant hatching defect in the RNase H1 overexpression embryos (Figure 1(C)). The RNase H1[CD] expressing embryos, however, had a ~25% failure to hatch rate, which was significantly different from the wild-type and the RNase H1 overexpression controls. To determine the effect overexpression of RNase H1 or RNase H1[CD] constructs have on absolute R-loop levels, we measured bulk R loop levels from 2-6 h embryos expressing these constructs. While there wasn't a significant reduction in R-loop levels upon RNaseH1 overexpression, R-loop levels increased upon overexpression of the RNaseH1[CD] mutant, suggesting the catalytic dead mutant blocks the processing of R-loops even in the presence of endogenous RNaseH1 (Supplemental Figure 1 (B)). We cannot rule out the possibility that some RNase H1 catalytic activity remains in the RNase H1[CD] overexpression strain from the endogenous RNase H1. We think it is likely, however, that the excess catalytically inactive protein outcompetes endogenous RNase H1 at sites of R-loop formation. Overall, we conclude that the absolute abundance of R-loops changes during development and that RNase H1 catalytic activity is likely important for R-loop resolution and embryonic development.

R-loop position and properties are influenced during development

While the absolute abundance of R-loops changes during development, we wanted to determine how R-loop position throughout the genome changes during Drosophila development. Genome-wide R-loop mapping during Drosophila development would allow us to ask if R-loop formation is hardwired into the genome and driven only by cell-type-specific transcription, or, more interestingly, is R-loop formation plastic during development changing independent of sequence composition and transcription status. To address this question, we performed D̲NA:R̲NA immunoprecipitation on sonicated nucleic acids followed by strand-specific sequencing of the DNA strand (ssDRIP-seq) in S2 cells, 2–3 h and 14–16 h embryos (Figure 2(A)).[9] We initially tried DNA-RNA immunoprecipitation followed by cDNA conversion coupled to high-throughput sequencing (DRIPc-seq).[18] When

conducted in Drosophila, however, we found high levels of RNA contamination in the final sequencing results (data not shown). Even with the ssDRIP-seq method, it was necessary to pre-treat genomic DNA preps with RNase A and RNase III as Drosophila embryos are stockpiled with RNA.

ssDRIP-seq of embryos and S2 cells revealed strand-specific signal that was sensitive to RNase H1- pretreatment, and showed cell-typespecific R-loop formation (Figure 2(B and C)). Providing validity to our data sets, biological replicates were highly correlated (Supplemental Figure 2(A)) and our ssDRIP data sets correlated with recently published ssDRIP-seq data sets in Drosophila S2 cells and embryos as expected based on the similar but different time points (Supplemental Figure 2(B)) (2–3 h and 14–16 h vs. 2–6 h and 10–14 h embryos) and known differences in R-loop mapping between different labs.[26,45] Furthermore, our S2 data sets were highly correlated with two R-loop mapping studies performed in cultured cells, but not correlated with an R-loop data set generated for a spike-in control and not used in mapping studies (Supplemental Figure 2(C-D)).[13,46] We validated several sites using DRIP-qPCR to confirm our sequencing results (Supplemental Figure 2(F-G)). Taken together, these data indicate that our ssDRIP signal reflects high quality and robust RNA:DNA hybrid mapping throughout the genome and that ssDRIP is a robust method to map sites of R-loop formation in Drosophila.

To map the precise location of R-loops throughout the genome and allow us to compare both quantitative and qualitative properties of R-loops, we used MACS to define R-loop peaks. Peaks were called separately against the input samples and RNase H1 treated controls, and only overlapping peaks were kept for analysis. Using this criterion, we identified 27,646, 22,581 and 29,801 peaks in S2 cells, 2–3 h and 14–16 h, respectively, which occupied between 8.3 and 12.5% of the genome. The overlap of sense and antisense R-loops had similar ratios (Supplemental Figure 2(E)). R-loop peak size was similar between sample types with a median of approximately 500 bp, but R-loops could occupy zones up to 10 kb in size (Figure 2(D)). Out of the 51,916 total unique R-loop peaks identified between all samples, 12.9% were common to all sample types, 28.3% were present in at least two samples and 58.8% were specific to an individual sample (Figure 2(E)).

Since ssDRIP allows for strand-specific annotation, we characterized R-loops relative to strand-specific genomic features. Relative to transcription units, 55–60% of R-loops occur in sense to transcription in S2 cells and 2–3 h embryos, whereas ~ 15% of R-loops are antisense (Figure 2(F)). In all samples, 25–30% of the R-loops form in unannotated regions of the genome. Next, we used Pavis to annotate R-loop signal relative to genomic features.[47] In all samples, we found that ~50% of R-loops mapped to introns or exons (Figure 2(G)). This is expected given that a significant fraction of R-loops should be produced from coding regions. GO term analysis of R-loop forming genes revealed that R-loops preferentially form in genes associated with RNA Pol II-dependent transcription and sample-specific R-loops form in genes associated with sample-specific development (Supplemental Table 2). Taken together, these results demonstrate that R-loop signal across Drosophila development is dynamic.

R-loop enrichment at transcription units changes during development

In mammals, R-loops are known to preferentially form at transcription start sites (TSS), gene bodies and transcription termination sites (TTS).[18,48] To ask if this pattern of R-loop formation is similar in Drosophila, and whether it changes during development, we measured R-loop abundance across gene bodies in our developmental samples. We then generated metaplots using strand-specific data for all time points. S2 cells and 2–3 h embryos display a very similar pattern of R-loop formation, with a strong peak at the TSS and continued signal over the gene body (Figure 3(A)), which is similar to R-loop positions in other metazoans.[18] In 2–3 h embryos and S2 cells, there was a greater signal for sense R-loop over the gene body, as would be expected given that the majority of R-loops are generated during transcription. Antisense R-loop signal was prevalent at the TSS and close to the TTS in 2–3 h embryos (Figure 3(A)). Interestingly, there is a depletion of R-loops immediately after the TTS in 2–3 h embryos and S2 cells (Figure 3(A)). The 14–16 h embryos, however, have a significantly different pattern altogether. In 14–16 h embryos, we observed the most abundant signal within and around the TSS and TTS regions with a relative reduced signal within the gene body (Figure 3(A)). The enrichment of R-loops at the TTS in 14–16 h embryos was not driven by differences in R-loop forming genes between the samples as R-loop forming genes are similar between 2–3 h and 14–16 h embryos (Supplemental Figure 3(B)). In the 14–16 h embryos, however, both sense and antisense R-loops have similar levels throughout the transcription unit (Figure 3(A)). Taken together, we conclude that R-loop enrichment at transcription units is not hardwired into the genome, but can be dynamic as a function of development.

Given that the absolute levels and relative position of R-loops can change between developmental states in Drosophila, we wanted to assess the contribution DNA sequence composition has on R-loop formation in Drosophila. Unlike in mouse and human cells, Drosophila lack high GC content at the TSS. In fact, GC content decreases relative to the gene body in Drosophila (Figure 3(B)). We asked if R-loop forming genes differ in their GC content relative to genes that lack R-loops. We found that genes with and without R-loops have a near-identical GC content along the gene body (Figure 3(B)). While overall GC content is not different in R-loop positive or negative genes, GC and AT skew has been shown to be a contributing factor to R-loop formation.[14] To test if GC or AT skew is associated with R-loop formation in Drosophila, we measured the AT/GC skew directly over all identified R-loops. This analysis revealed a striking transition from positive to negative AT skew at the center of our combined R-loop signal. This is mirrored by a less dramatic transition from negative to positive GC skew centered at the combined R-loop signal (Figure 3(C)). Interestingly, developmental-specific R-loops had AT/GC skew profiles that were distinct from all R-loops combined (Figure 3(C), Supplemental Figure 3(A)).

We also calculated GC and AT skew for R-loop forming and deficient genes in all samples. Stronger negative GC skew at the TSS was observed in R-loop forming genes relative to genes that fail to form R-loops (Figure 3(D)). Specifically, AT skew at the TSS transitioned from positive skew in R-loop deficient genes to negatively skew in R-loop forming genes. At the TTS, there is a strong positive AT skew around the TTS in both R-loop positive and negative genes (Figure 3(D)). Negative GC skew is stronger at the TSS in R-loop forming

genes. This analysis reveals a correlation between altered AT skew and negative GC skew in R-loop forming genes, suggesting that AT/GC skew could contribute to R-loop formation in Drosophila. Due to the strong presence of R-loops at promoters and TSS in this and other R-loop mapping studies, we examined the AT and GC skew specifically at the promoter or TSS regions to determine if they were driving the overall skew. AT and GC skew was calculated for R-loops at promoter and TSS regions versus every other R-loop peak for each cell type (Supplemental Figure 3(B)). AT and GC skew at promoter and TSS regions was similar to overall skew, though this varied in S2 cells. Together, we conclude that while AT and GC skew could facilitate R-loop formation, developmental-specific R-loop formation is not likely driven by changes in AT or GC skew. This suggests that transcription, chromatin environment or other factors could contribute to cell type specific R-loop formation.

To test whether any specific DNA sequence motifs are associated with R-loop formation, we searched for motifs enriched in the set of all Drosophila R-loops. Two motifs stood out as an order of magnitude more significantly enriched that any others: a polyadenine tract and a polypurine tract (Figure 3(D), Supplemental Figure 3(B) for the entire table). This indicates that polypurine tracts are conducive to R-loop formation, which is consistent with the known thermodynamic stability of RNA:DNA hybrid formation in purine-rich template sequences.[49]

-Common and cell-typespecific chromatin features associated with R-loops

R-loops are associated with activating chromatin marks such as H3K4me2/3 and H3K9ac and, to a lesser extent, with repressive chromatin marks such as H3K27me3.[18] Chromatin marks associated with R-loops, however, vary depending on species. One possibility is that there are marks that are universally associated with R-loop formation whereas some chromatin marks could associate with R-loops in a developmental-specific manner. To answer this question, we leveraged time-matched ChIP-seq modENCODE datasets for S2 cells, 2–4 h embryos (ChIP-chip and ChIP-seq) and 14–16 h embryos. To quantitatively determine if chromatin marks were positively or negatively associated with R-loops, we evaluated the probability of R-loops overlapping a variety of histone modifications and chromatin-associated proteins by chance using a peak shuffling bootstrap procedure (see Materials and Methods). The available chromatin proteins vary for each sample, but there are 10 chromatin or histone markers common in all three developmental samples (Figure 4(A)). Several factors that are associated with transcriptional activation and have been previously shown to be associated with R-loops, are enriched at R-loops in S2 cells and 2–3 hour embryos (Figure 4(A), Supplemental Figure 4). Additionally, repressive chromatin marks such as Polycomb complex subunits and H3K27me3 are enriched in all samples, which is consistent with recent work linking R-loops to transcriptional repression (Figure 4(A), Supplemental Figure 4).[25,26]

We asked which marks are consistently associated with R-loops (positively or negatively) across development and which factors are developmental specific. We found that the repressive mark H3K27me3 was positively associated with R-loops in all developmental samples, highlighting the link between R-loops and transcriptional repression (Figure 4(B)). Interestingly, we identified factors (H3K4me2 and ZW5) that were enriched in

one developmental sample but not in others (Figure 4(B)). These results suggest while some factors are associated with R-loops regardless of development state, other factors are associated with R-loops in a developmentally-specific manner.

R-loop formation as a function of transcription

In this study, we have noted distinctive changes in R-loop formation across development. One possibility is that these changes are driven by developmental-specific changes in the transcription program. As embryos are stockpiled with maternally deposited RNA and RNA-seq is an indirect readout of active transcription, we turned to previously published and time-matched GRO-seq datasets in S2 cells and 2–2.5 h embryos, respectively.[50,51] Unfortunately, time-matched GRO or PRO-seq datasets do not exist for 14–16 h embryos. We converted GRO-seq signal to FPKM for each annotated transcript in the Drosophila transcriptome. Then, we compared the GRO-seq value of all R-loop-containing genes to genes devoid of R-loops. In S2 cells, R-loop positive and negative genes had a similar median FPKM value by GRO-seq (Figure 5(A)). R-loop-containing genes in 2–3 h embryos, however, revealed a different paradigm. R-loop positive genes had a significantly higher expression level than R-loop negative genes (Figure 5(C)).

To ask if R-loop-containing genes were over or underrepresented with genes that have high or low expression levels, we binned GRO-seq FPKM values into quartiles and asked what fraction of R-loop containing genes fell within each expression quartile (Figure 5(B, D)). In S2 cells, R-loop containing genes were slightly overrepresented in the highest expression quartile and, to a lesser extent, in the lowest expression quartile (Figure 5(B)). In 2–3 h embryos, however, R-loops were significantly overrepresented in the highest expression quartile and underrepresented from the lowest expression quartile (Figure 5(D)). While analyzing this data, we also found the number of R-loops forming sites per gene was correlated with transcriptional activity (Figure 5(E)). We observe a consistent increase in the average number of R-loops per gene as transcriptional activity increases (Figure 5(E)). The increase in the average number of R-loops per gene could represent multiple R-loops within a given gene or larger R-loop zones allowing R-loops to form over a larger target region.

One explanation for developmental-specific R-loop formation is that specificity is driven by developmental-specific transcription status. To test this, we compared expression level of genes that exhibit R-loops only in S2 cell or only in 2–3 h embryos (Figure 5(F)). If active transcription drives the changes in R-loop formation, we would expect R-loop positive genes that are unique to 2–3 h embryos would have significantly higher expression level in 2–3 h embryos relative to S2 cells, and vice-versa. The median difference of GRO-seq values in developmental-specific R-loop-containing genes, however, is approximately zero with a normal distribution (Figure 5(F)). Therefore, we conclude that active transcription is not a driver of developmental-specific R-loop formation and that factors such as chromatin state or R-loop-specific proteins drive these differences.

We asked if the chromatin signature of R-loops in highly expressed genes differs from the signature of R-loops in transcriptionally repressed genes. To this end, we selected R-loops in the highest expression quartile and lowest expression quartile from S2 cells (Figure 5(B)). Next, we used the random shuffling method to identify chromatin-associated

factors enriched at R-loops derived from highly and lowly expressed genes. This analysis revealed that the chromatin signature of R-loops in highly and lowly expressed genes are distinct (Figure 5(G)). For example, R-loops in highly expressed genes are enriched for active chromatin marks (e.g. H3K27ac and H3K4me2; Figure 5(G)). In contrast, repressive chromatin marks such as H3K27me3 are enriched at R-loops derived from lowly expressed genes. We repeated the same analysis with the 2–3 h embryo time point and noticed a striking difference; both active chromatin marks and repressive chromatin marks were associated with highly expressed genes (Figure 5(H)). Given the differentiation state of cells in the early embryo, this would suggest that R-loops can be associated with poised or bivalent genes.[52]

R-loops have the potential to trigger ATR activation at the MZT

The onset of zygotic transcription at the MZT is associated with RPA accumulation at the 5′ end of genes and activation of the ATR-mediated DNA damage checkpoint response.[53] Delaying the onset of zygotic transcription delays the activation of ATR (Mei41 in Drosophila), indicating that replication-transcription conflicts drive the activation of the DNA damage response at the MZT.[53,54] It is unknown, however, what aspect of the replication-transcription conflict triggers ATR activation at the MZT. If genome instability at the MZT was at least partially due to R-loops, we would predict to see an enrichment of RPA at R-loop forming sequences in 2–3 h embryos. Qualitatively, we see overlap between RPA and R-loops in 2–3 h embryos (Figure 6(A)). We tested the significance of this overlap by using the random shuffling method previously described. Quantitatively, we observe a significant enrichment of RPA at R-loop forming sequences in the 2–3 h embryo. Importantly, there was an even more substantial enrichment of RPA at R-loop peaks that are unique to 2–3 h embryos (Figure 6(B)). Further supporting the hypothesis that R-loops could be partially responsible for the transcription-induced genome instability at the MZT, only R-loops from the 2–3 h sample were enriched at RPA binding sites (Figure 6(B)). This data suggests that R-loops could contribute to the transcription-induced DNA damage that occurs in the absence of ATR at the MZT. We do note, however, that the RPA ChIP-seq data comes from a time point ~20 minutes earlier in development than the time point we chose for R-loop mapping.[53] Given this caveat, we think it is even more notable that significant overlap of RPA and R-loops is observed in this analysis.

Discussion

By mapping R-loops in a developing organism, we have been able to provide new insight into the role that DNA sequence, active transcription and chromatin associated factors has on R-loop formation. While previous R-loop mapping and genome-wide analysis of R-loop metabolism across development has been performed in plants and mammalian cultured cells,[10,29,55] we present a functional characterization of R-loops during Drosophila embryogenesis. The benefit of a developmental approach to studying R-loop formation is that it allows the distinction between factors that are stably linked to R-loop formation from those that are developmental specific. This has the potential to identify key molecules and processes that could drive R-loop formation and resolution during development and disease.

One surprising finding is that the absolute level of R-loops changes during embryogenesis. This is unlikely due to changes in transcription during development as the stages of embryogenesis used in this study are similarly active. This suggests that there is an active mechanism which prevents R-loop formation or resolves active R-loops during later stages of Drosophila embryogenesis. The importance of R-loop processing during development is further highlighted by the observation that RNase H1 catalytic activity is necessary to prevent hatching defects in Drosophila embryos. Interestingly, overexpression of catalytically active and inactive RNaseH1 do not have the same effect. One possible explanation for this is that maternally deposited RNaseH1 is highly active in the embryo. Therefore, additional RNaseH1 has no further effect on R-loop levels. Overexpression of catalytic inactive RNaseH1, however, could bind to RNA:DNA hybrids and block RNaseH1-mediated processing of R-loops. This would have the potential to drive replication-transcription conflicts and genome instability in the developing embryo.

Consistent with R-loops as a driving force of genome instability during embryogenesis, we have found an enrichment of R-loops at potential sites of replication fork stalling in the early embryo. Given that we see an enrichment of R-loops and RPA specifically in the 2–3 h embryo sample, our data suggests that R-loops could contribute to ATR activation at the MZT. It is interesting to note, however, that we do not observe RPA accumulation at all sites of R-loop formation. Therefore, there must be something unique about the R-loops associated with RPA accumulation at this time point. Perhaps these R-loops represent sites of head-on conflicts. Alternatively, hyper stable R-loops could drive chromatin or transcriptional changes that negatively impact embryogenesis.[12] Further work will be required to distinguish between these and other possibilities.

Specific DNA sequence biases are associated with R-loop formation.[14,27] While we found that overall GC content is the same for R-loop positive and negative genes, AT and GC skew were associated with R-loop forming sequences. Interestingly, this skew varied as a function of the transcription unit.[14,56] G4 quadraplex forming regions with high GC skew on the non-template strand are associated with R-loop formation.[14,56] Additionally, R-loops can modulate DNA methylation at CpG islands in promoter regions.[14] Unlike in plants and mammals, however, Drosophila lack wide-scale DNA methylation.[57] Therefore, Drosophila allows the uncoupling between R-loop formation and DNA methylation, which could explain why R-loops are associated with a higher AT skew than GC skew in Drosophila. Similar to mammalian cells, we see a transition to positive GC skew at the center of R-loops peaks. What's unique to Drosophila, however, is the drastic transition from positive to negative AT skew at the center of R-loop peaks. These biases in AT and GC skew could create a thermodynamically stable environment for R-loop formation and resolution. Similar to other organisms, we have found several polypurine motifs associated with R-loops. Again, this likely reflects the thermodynamic stability associated with RNA: DNA hybrids at purine-rich sequences.[7][49] One interesting observation in Drosophila is that the R-loop signal relative to the transcription unit can vary as a function of development. The most significant difference is in 14–16 h embryos where R-loops are broadly enriched at the TSS and the TTS but not the gene body in comparison to 2–3 h embryos or S2 cells. This difference does not appear to be driven by AT or GC skew. We propose that a combination of

factors such as transcription status, chromatin marks and R-loop binding proteins drive these changes in R-loop formation during development.

We have found that R-loops are positively and negatively associated with specific histone modifications and chromatin associated factors. Many of the factors we analyzed in Drosophila have been shown to be enriched or depleted in other systems, including mammalian cells.[18,58,59] More importantly, however, factors associated with R-loops can change as a function of development. For example, R-loops in 14–16 h embryos lose their association with common activating histone marks such as H3K4me3 and H3K36me2/3. In contrast, H3K27me3 is enriched at R-loops in all developmental states. Therefore, it is critical to assay multiple cell types or developmental states before concluding that a chromatin factor is correlated with R-loop formation.

The link between R-loops, transcription state, histone marks and chromatin associated factors has been seen in other organisms.[18] In Drosophila, we see a consistent relationship between active and repressive chromatin marks, signified by enrichment in both H3K27ac and H3K27me3, and R-loop formation. This is supported by the association of R-loops with both highly active and silent genes in both embryos and cultured cells. Our work, and that of others, identify R-loops associated with transcriptionally active and inactive genes.[25] This suggests that, at least in Drosophila, there may exist at least two classes of R-loops. R-loops that form as a byproduct of active transcription and R-loops that function in a repressive capacity to prevent transcription within repressive chromatin domains. This would be consistent with recent work demonstrating that R-loops facilitate silencing by the Polycomb complex.[25,26] Understanding how different categories of R-loops maintain their identity will be an exciting challenge. For example, how do cells know which R-loops should function in a repressive manner versus those that function as activators? The question of whether R-loops help establish a chromatin state or are a function of it remains an outstanding question in R-loop biology.

Mapping of R-loops has been performed in a variety of organisms ranging from yeast, worms, plants, and mammalian cultured cells. While there are factors and processes that are consistently associated with R-loops across organisms, there are also key differences. For example, in plants there are low levels of R-loops at gene terminators compared to other organisms and high accumulation of antisense R-loops that regulate specific loci.[29,60] In contrast, mammalian cells exhibit R-loops at promoters and TTS and the number of antisense R-loops are much more limited.[18] The fact that Drosophila exhibit changes in antisense R-loop signal across the gene body depending on developmental state highlights the importance of examining R-loops in a developmental context. Drosophila provides a powerful model to understand key properties of R-loop biology in the context of unperturbed metazoan development. Here, we demonstrate that R-loop formation within the same genomic sequence can vary as a function of development. Our work suggests that a combination of transcription, chromatin-associated factors and sequence elements drive differential R-loop formation during development. Therefore, Drosophila provides a powerful model to understand, mechanistically, the factors responsible for R-loop formation and resolution to execute specific developmental programs.

Material and Methods

S9.6 antibody

A hybridoma cell line producing the S9.6 antibody was purchased through ATCC (product #HB-8730). The cell line was grown under recommended conditions. The S9.6 antibody was purified on a protein G column using the GE aKTA system and run over a desalting column for buffer exchange into PBS to obtain a final concentration of 1 mg/mL. The antibody was aliquoted and stored at −80 °C. A fresh aliquot was used for every ssDRIP-seq experiment.

RNase H1 overexpression

Drosophila RNase H1 was cloned from RNA derived from Oregon R embryos. RNA was converted into cDNA, PCR amplified, and cloned into the pUASz vector with a C-terminal GFP tag.[43] The A isoform was chosen as the isoform B isn't detected in Drosophila tissues.[61] The mitochondrial localization start site was converted to AAA to ensure RNase H1-GFP would only be present in the nucleus. The catalytically dead version of RNase H1 (D201N) was made by site-directed mutagenesis (Agilent QuickChange Lightning). Plasmids were injected into an *attP2* containing stock (BestGene) for site-specific integration.

Hatch rate assay

For the overexpression experiments, homozygous RNase H1 males were crossed with unmated female homozygous for the maternal triple driver (MTD, Bloomington Stock 31777) to drive expression early in embryogenesis. Male Oregon R flies were crossed with MTD females as a control. Progeny were transferred to bottles with a grape juice agar plate with wet yeast for embryo collection. 100 unhatched embryos were carefully moved to a fresh grape juice plate and incubated overnight at 25 °C. After 36 hours, unhatched embryos were counted. This was repeated three times each from two separate crosses.

Cell culture

S2 cells were obtained directly from the Drosophila Genomic Resource Center (DGRC). Cells were confirmed negative for mycoplasma contamination via PCR. Cells were grown at 25 °C in Schneider's Drosophila Medium with 10% heat-inactivated FBS (Gemini Bio Products) and 100 U/mL of Penicillin/Streptomycin (Fisher Scientific).

Embryo collection and staging

Oregon R flies were expanded into population cages containing grape juice plates supplemented with wet yeast. Population cages were kept at 25 °C in a humidified room and plates were changed daily. Before embryo collections, flies were precleared for at least one hour to minimize the number of late-stage embryos. Embryos were collected and aged at 25 °C to obtain embryos that were 2–3 or 14–16 hours old. After aging and collection, embryos were dechorionated in 50% bleach for 2 minutes and thoroughly rinsed in water. Embryos were flash frozen in liquid nitrogen and kept at −80 °C until ready to use. An aliquot of embryos was taken from each batch before freezing to verify staging. For this, embryos were fixed in heptane and 2% paraformaldehyde for 20 minutes with shaking,

devitellinized in methanol, washed with methanol and rehydrated in PBS + 0.1% Triton X-100 overnight. Embryos were stained with DAPI and mounted in Vectashield medium (Vector Labs). Images were acquired on a Nikon Ti-E inverted microscope with a Zyla sCMOS digital camera.

Genomic DNA purification and RNase treatment

Genomic DNA purification is based on Alecki et al., 2020.[26] For genomic DNA isolation from S2 cells, cells were collected at 70–80% confluency, washed once in PBS, resuspended in TE with 0.5% SDS and 100 μg/mL proteinase K and incubated at 37 °C overnight. Embryos were devitellinized in heptane and methanol, rinsed thoroughly in PBS and incubated in 50 mM Tris-HCl pH 8.0, 100 mM EDTA, 100 mM NaCl, 0.5% SDS, and 5 mg/ml proteinase K for 3 hours at 50 °C. At this point, cells and embryos were processed the same. Extracts were purified with phenol:chloroform, and DNA was precipitated with sodium acetate and ethanol. DNA was spooled using a glass pipette and transferred to 70% ethanol. After several washes in ethanol, the DNA was air dried and resuspended in TE. To degrade free RNA, samples were incubated with 100 μg of RNase A with 500 mM NaCl for 1 hour at 37 °C. RNase A was degraded by spiking in 100 μg/mL proteinase K and incubated for an additional 45 minutes. Samples were cleaned with phenol:chloroform, precipitated with sodium acetate and ethanol, and resuspended in TE. Samples were diluted to 100 ng/μL and sonicated in a Bioruptor Plus for 8 cycles (30″ on/90″ off) on low power. 10 μg of nucleic acid was digested with 5 μL RNase H1 (NEB) at 37 °C for 16 hours and 10 μg was mock digested without RNase H1. Both samples had 1 μL of RNase III added (Thermo Fisher). After phenol: chloroform purification and precipitation, samples were immediately used for DRIP or slot blot experiments.

Slot blot

Hybond Nylon membrane (Amersham) was pre-soaked in TE and a slot blot apparatus was assembled according to manufacturer's instructions (Bio-Rad). Samples with matching RNase H1-digested controls were added to the blot in decreasing amounts, and nucleic acids were crosslinked to the membrane with a Strategene UV Stratalinker 1800 using the auto crosslink setting. Blots were blocked in milk, incubated with S9.6 (1:2,000) followed by mouse-HRP and imaged in a Bio-Rad Chemidoc MP. After imaging the R-loops, blots were stripped and re-probed using a dsDNA-specific antibody (Abcam ab27156) at 1:20,000. Intensities were measured with ImageJ,[62] and normalized intensity was obtained by dividing the S9.6 signal by the dsDNA signal.[63] A standard plot was made for each sample and antibody, and samples were chosen for analysis when their intensity was linear.

DRIP-qPCR and ssDRIP-seq

DRIP was carried out as described in Ginno et al. 2012.[14] Briefly, 4.4 μg of DNA was resuspended in 500 μL of TE. 10% was taken for the input sample. DRIP binding buffer was added to each sample (10 mM sodium phosphate, 140 mM NaCl, 0.05% Triton X-100 final concentration) and 20 μL of 1 mg/mL S9.6 was added to each DRIP reaction. After overnight incubation at 4 °C, 50 μL of pre-washed protein G Dynabeads (Life Technologies) were added to the extract. After 2 hours at 4 °C, beads with captured nucleic acid were washed in 1x DRIP binding buffer 5 times and eluted in 50 mM Tris, 10 mM EDTA, 0.5%

SDS with proteinase K at 50 °C for 45 minutes. Nucleic acid in the eluate was purified with phenol:chloroform, precipitated and resuspended in 10 mM Tris. For DRIP-qPCR, 1 μL of nucleic acid was diluted 1:10 in water and mixed with 10 μL SSoAdvanced Universal Sybr (Bio-Rad). Primers were added to a final concentration of 250 nM each. A list of primers used in this study can be found in Supplemental Table 1. qPCR was carried out on a Bio-Rad CFX96 Touch instrument using the following protocol: 98 °C heat denaturation for 60″ followed by 40 cycles of 98 °C for 15″ and 60 °C for 30″. A heat denaturation was included to monitor the purity of the reaction products. For ssDRIP, nucleic acid was sonicated in a Bioruptor Plus for 8 cycles at high power (30″ on/30″ off) to 250 bp. Libraries were constructed with the Accel-NGS 1S Plus DNA Library Kit according to the manufacturer's instruction (Swift Biosciences 10024). Barcoded libraries were sequenced using an Illumina Novaseq for 150 bp PE reads.

Bioinformatics

Alignment and peak calling.—Fastq files were initially trimmed of adapters using Trimmomatic v0.3.8.[64] Each paired read was trimmed 10 base-pairs at the 3′ end to eliminate the additional low complexity from the library preparation kit. Reads for sequencing were mapped to the Drosophila genome (dm6) using bowtie2 version 2.3.4.1 using the –very-sensitive-local setting.[65] Duplicates were marked using picard MarkDuplicates v2.17.10, and stranded bam files were created using samtools as described in Xu and Sun et al. 2017.[9,66] Stranded bam files were used to generate ssDRIP peaks with callpeaks from MAC2 v2.1.2.[67] The RNase H1 pretreated DRIP file was used as control, peak calling was done on the 2 replicates in paired-end mode, with –keep-dup = auto and effective genome size for Drosophila dm6. A small number of peaks mapped to both strands as determined with bed-tools.[68] Peaks that mapped to both strands and had reciprocal overlap of 90% as determined by bedtools intersect were removed from down-stream analysis. Bam files were combined and 50 million reads were randomly selected for visualization. Stranded reads were visualized using deeptools bamCoverage using --binSize 50 bp, --ignoreForNormalization chrY chrM, and --normalizeUsing RPKM.[69] Pearson correlation plots were created using deeptools multiBamCoverage and plotCorrelation with default settings, 1 kb windows and the mitochondrial genome excluded.

ssDRIP-seq analysis.—Peak annotation was performed using Pavis to the dm6 genome with up- and downstream regions set to 5 kb.[47] Overall sense and antisense R-loops were determined via bedtools intersect with strandedness against the Refseq Drosophila transcriptome, downloaded from UCSC genome browser. Metagene plots were made with the Deeptools software package, using computeMatrix and plotProfile. For computeMatrix, scale-regions or reference-point as appropriate, with a 1 kb region size and 500 bp up- and down-stream of the start and end site, respectively. For options -binSize was 50 and the mean was plotted. For plotProfile, 'add standard error' was added to Plot type. –yMin and –yMax were chosen to be the same for both sense and antisense to aid in visualization.

Gene Ontology enrichment analysis of R-loop containing genes was performed with PANTHER, with Fisher's exact test and using the Bonferroni correction for multiple testing.[70-72]

GRO-seq FPKM counts were determined with HOMER analyzeRepeats.pl using S2 datasets from Core and Lis et al. 2012 and GRO-seq data on 2–2.5 embryos from Saunders and Ashe et al. 2013.[50,51,73] R-loops peaks were split into 2 files containing their + and − peaks and annotation of R-loop peaks was done with HOMER software package using annotatePeaks.pl against dm6 and requiring the appropriate strandedness.[73] R-loops mapping to transcripts were extracted from the HOMER annotation, and GROseq values for these transcripts was determined using custom R scripts. Plots summarizing these data were created in Prism 9.

Functional genomic data from modENCODE.—We downloaded histone modification peaks and transcription factor binding sites identified by ChIP-chip or ChIP-seq in Drosophila from ModENCODE (Table 1).[74] We considered samples assayed in S2 cells and at two developmental time-points (2-4hr, 14-16hr). These were chosen to match the ssDRIP timepoints.

Chromatin associated factor enrichment in R-loops.—For each ChIP-chip or ChIP-seq marker with a matching DRIP timepoint, we calculated the number of overlapping base-pairs (bp) between the marker and the R-loop peaks. We used permutation-based approach to determine whether the observed amount of overlap was more or less than expected by chance. Briefly, we calculated an empirical *p* value for the observed amount of overlap by comparing the number of overlapping bp to a null distribution. We obtained the null distribution by randomly shuffling length-matched regions throughout the genome and calculating the amount of overlap in each permutation. The *p*-values are adjusted for multiple testing using the Bonferroni method.

When permuting, we matched the length distribution of the shuffled peaks to the original set of peaks, and excluded all gap and blacklisted regions from consideration (dm3; version 1).[75] Peaks called from DRIP were lifted over to dm3 for this analysis. For peaks obtained from ChIP-chip data, we required that the shuffled peaks maintained both the overall length distribution and the probe density of the original peak. We reshuffled any peaks that fell more than 2 standard deviations (approx. 0.03) away from the original probe density until at least 99% of the original peaks were appropriately matched. We performed 1000 permutations for each marker and R-loop pair.

For the general analyses, we maintained the location of the R-loop peaks and shuffled the locations of the histone modification or transcription factor binding peaks. For a secondary analysis, we examined a subset of R-loops quantified specifically in the TTS and 3′ UTR. For this set of R-loops, we maintained the R-loop location within the TTS/3′ UTR and shuffled the chromatin markers.

Calculation of AT- and GC-skew in R-loops.—We calculated GC and AT skew over the entire Drosophila genome (dm6). GC skew was calculated for 50 bp windows tiled across the annotation regions as $S_i = \frac{(G_i - C_i)}{G_i + C_i}$.[76]

In the equation, G_i repressnts the frequency of guanine nucleotides and C_i represents the frequency of cytosine nucleotides in the window i. The range of GC skew for a window (S_i) spans from −1 to 1. AT Skew was calculated in the same way. The resulting GC and AT skew was converted to a bigwig file, and the value across each set of genomic regions was calculated using the computeMatrix function from deeptools and visualized using plotProfile (For computeMatrix, scale-regions or reference-point as appropriate, with 500 bp up- and down-stream of the start and end site, respectively. For options -binSize was 50 and the mean was plotted. For plotProfile, 'add standard error' was added to Plot type. –yMin and –yMax were chosen to be the same for AT and GC skew to aid in visualization).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

DATA AVAILABILITY

All data has been deposited into GEO under the record GSE185403.

References

1. Aguilera A, García-Muse T, (2012). R Loops: From Transcription Byproducts to Threats to Genome Stability. Mol. Cell 46, 115–124. [PubMed: 22541554]

2. White RL, Hogness DS, (1977). R loop mapping of the 18S and 28S sequences in the long and short repeating units of *Drosophila melanogaster* rDNA. Cell 10, 177–192. [PubMed: 402221]

3. Glover DM, Hogness DS, (1977). A novel arrangement of the 18S and 28S sequences in a repeating unit of drosophila melanogaster rDNA. Cell 10, 167–176. [PubMed: 402220]

4. Chédin F, (2016). Nascent Connections: R-Loops and Chromatin Patterning. Trends Genet.. 32, 828–838. [PubMed: 27793359]

5. Skourti-Stathaki K, Proudfoot NJ, (2014). A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes Dev. 28, 1384–1396. [PubMed: 24990962]

6. Dumelie JG, Jaffrey SR, (2018). Defining the location of promoter-associated R-loops at near-nucleotide resolution using bisDRIP-seq. eLife 6, e28306

7. Wahba L, Costantino L, Tan FJ, Zimmer A, Koshland D, (2016). S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. Genes Dev. 30, 1327–1338. [PubMed: 27298336]

8. Fang Y, Chen L, Lin K, Feng Y, Zhang P, Pan X, Sanders J, Wu Y, Wang X, Su Z, et al. , (2019). Characterization of functional relationships of R-loops with gene transcription and epigenetic modifications in rice. Genome Res. 29, 1287–1297. [PubMed: 31262943]

9. Xu W, Xu H, Li K, Fan Y, Liu Y, Yang X, Sun Q, (2017). The R-loop is a common chromatin feature of the Arabidopsis genome. Nature Plants 3, 704–714. [PubMed: 28848233]

10. Yan P, Liu Z, Song M, Wu Z, Xu W, Li K, Ji Q, Wang S, Liu X, Yan K, et al. , (2020). Genome-wide R-loop Landscapes during Cell Differentiation and Reprogramming. Cell Rep. 32, 107870 [PubMed: 32640235]

11. Chen L, Chen J-Y, Zhang X, Gu Y, Xiao R, Shao C, Tang P, Qian H, Luo D, Li H, et al. , (2017). R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. Mol. Cell 68, 745–757.e5. [PubMed: 29104020]

12. Chen PB, Chen HV, Acharya D, Rando OJ, Fazzio TG, (2015). R loops regulate promoter-proximal chromatin architecture and cellular differentiation. Nature Struct. Mol. Biol 22, 999–1007. [PubMed: 26551076]

13. Crossley MP, Bocek MJ, Hamperl S, Swigut T, Cimprich KA, (2020). qDRIP: a method to quantitatively assess RNA–DNA hybrid formation genome-wide. Nucleic Acids Res. e84 [PubMed: 32544226]

14. Ginno PA, Lott PL, Christensen HC, Korf I, Chédin F, (2012). R-Loop Formation Is a Distinctive Characteristic of Unmethylated Human CpG Island Promoters. Mol. Cell 45, 814–825. [PubMed: 22387027]

15. Tan-Wong SM, Dhir S, Proudfoot NJ, (2019). R-Loops Promote Antisense Transcription across the Mammalian Genome. Mol. Cell 76, 600–616.e6. [PubMed: 31679819]

16. Chan YA, Aristizabal MJ, Lu PYT, Luo Z, Hamza A, Kobor MS, Stirling PC, Hieter P, (2014). Genome-Wide Profiling of Yeast DNA:RNA Hybrid Prone Sites with DRIP-Chip. PLoS Genet. 10, e1004288 [PubMed: 24743342]

17. Liu Y, Liu Q, Su H, Liu K, Xiao X, Li W, Sun Q, Birchler JA, Han F, (2021). Genome-wide mapping reveals R-loops associated with centromeric repeats in maize. Genome Res. 31, 1409–1418. [PubMed: 34244230]

18. Sanz LA, Hartono SR, Lim YW, Steyaert S, Rajpurkar A, Ginno PA, Xu X, Chédin F, (2016). Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. Mol. Cell 63, 167–178. [PubMed: 27373332]

19. Dasgupta S, Masukata H, Tomizawa J, (1987). Multiple mechanisms for initiation of ColE1 DNA replication: DNA synthesis in the presence and absence of ribonuclease H. Cell 51, 1113–1122. [PubMed: 2446774]

20. Silva S, Camino LP, Aguilera A, (2018). Human mitochondrial degradosome prevents harmful mitochondrial R loops and mitochondrial genome instability. Proc. Natl. Acad. Sci 115, 11024–11029. [PubMed: 30301808]

21. Hamperl S, Bocek MJ, Saldivar JC, Swigut T, Cimprich KA, (2017). Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. Cell 170, 774–786. [PubMed: 28802045]

22. Lang KS, Hall AN, Merrikh CN, Ragheb M, Tabakh H, Pollock AJ, Woodward JJ, Dreifus JE, Merrikh H, (2017). Replication-Transcription Conflicts Generate R-Loops that Orchestrate Bacterial Stress Survival and Pathogenesis. Cell 170, 787–799. [PubMed: 28802046]

23. Stork CT, Bocek M, Crossley MP, Sollier J, Sanz LA, Chédin F, Swigut T, Cimprich KA, (2016). Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage. eLife, 5.

24. Ouyang J, Yadav T, Zhang J-M, Yang H, Rheinbay E, Guo H, Haber DA, Lan L, Zou L, (2021). RNA transcripts stimulate homologous recombination by forming DR-loops. Nature 594, 283–288. [PubMed: 33981036]

25. Skourti-Stathaki K, Triglia ET, Warburton M, Voigt P, Bird A, Pombo A, (2019). R-Loops Enhance Polycomb Repression at a Subset of Developmental Regulator Genes. Mol. Cell 73, 930–945. [PubMed: 30709709]

26. Alecki C, Chiwara V, Sanz LA, Grau D, Pérez OA, Boulier EL, Armache K-J, Chédin F, Francis NJ, (2020). RNA-DNA strand exchange by the Drosophila Polycomb complex PRC2. Nature Commun. 11, 1781. [PubMed: 32286294]

27. Stolz R, Sulthana S, Hartono SR, Malig M, Benham CJ, Chedin F, (2019). Interplay between DNA sequence and negative superhelicity drives R-loop structures. Proc. Natl. Acad. Sci 116, 6260–6269. [PubMed: 30850542]

28. Hage AE, French SL, Beyer AL, Tollervey D, (2010). Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. Genes Dev. 24, 1546–1558. [PubMed: 20634320]

29. Xu W, Li K, Li S, Hou Q, Zhang Y, Liu K, Sun Q, (2020). The R-loop Atlas of Arabidopsis Development and Responses to Environmental Stimuli. Plant Cell 32, 888–903. [PubMed: 32075864]

30. Tadros W, Lipshitz HD, (2009). The maternal-to-zygotic transition: a play in two acts. Development 136, 3033–3042. [PubMed: 19700615]

31. Hamm DC, Harrison MM, (2018). Regulatory principles governing the maternal-to-zygotic transition: insights from Drosophila melanogaster. R. Soc. Open Biol 8, 180–183.

32. Harrison MM, Li X-Y, Kaplan T, Botchan MR, Eisen MB, (2011). Zelda Binding in the Early Drosophila melanogaster Embryo Marks Regions Subsequently Activated at the Maternal-to-Zygotic Transition. PLoS Genet. 7, e1002266 [PubMed: 22028662]

33. Foe VE, Alberts BM, (1983). Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in Drosophila embryogenesis. J. Cell Sci 61, 31–70. [PubMed: 6411748]

34. Farrell JA, O'Farrell PH, (2014). From Egg to Gastrula: How the Cell Cycle Is Remodeled During the Drosophila Mid-Blastula Transition. Annu. Rev. Genet 48, 1–26. [PubMed: 25036377]

35. Blythe SA, Wieschaus EF, (2015). Chapter Four Coordinating Cell Cycle Remodeling with Transcriptional Activation at the Drosophila MBT. Elsevier Inc..

36. Bonnet J, Lindeboom RGH, Pokrovsky D, Stricker G, Çelik MH, Rupp RAW, Gagneur J, Vermeulen M, Imhof A, Müller J, (2019). Quantification of Proteins and Histone Marks in Drosophila Embryos Reveals Stoichiometric Relationships Impacting Chromatin Regulation. Dev. Cell 51 , 632–644. [PubMed: 31630981]

37. Bowman SK, Deaton AM, Domingues H, Wang PI, Sadreyev RI, Kingston RE, Bender W, (2014). H3K27 modifications define segmental regulatory domains in the Drosophila bithorax complex. eLife 3, e02833 [PubMed: 25082344]

38. Smith AV, Orr-Weaver TL, (1991). The regulation of the cell cycle during Drosophila embryogenesis: the transition to polyteny. Development 112, 997–1008. [PubMed: 1935703]

39. Schneider I, (1972). Cell lines derived from late embryonic stages of Drosophila melanogaster. J. Embryol. Exp. Morph 27, 353–365. [PubMed: 4625067]

40. Boguslawski SJ, Smith DE, Michalak MA, Mickelson KE, Yehle CO, Patterson WL, Carrico RJ, (1986). Characterization of monoclonal antibody to DNA · RNA and its application to immunodetection of hybrids. J. Immunol. Methods 89, 123–130. [PubMed: 2422282]

41. Hartono SR, Malapert A, Legros P, Bernard P, Chédin F, Vanoosthuyse V, (2018). The Affinity of the S9.6 Antibody for Double-Stranded RNAs Impacts the Accurate Mapping of R-Loops in Fission Yeast. J. Mol. Biol 430, 272–284. [PubMed: 29289567]

42. Filippov V, Filippova M, Gill S, (2001). Drosophila RNase H1 is essential for development but not for proliferation. Mol. Genet. Genomics 265, 771–777. [PubMed: 11523794]

43. DeLuca SZ, Spradling AC, (2018). Efficient Expression of Genes in the Drosophila Germline Using a UAS-Promoter Free of Interference by Hsp70 piRNAs. Genetics 209, 381–387. [PubMed: 29669732]

44. Rørth P, (1998). Gal4 in the Drosophila female germline. Mech. Dev 78, 113–118. [PubMed: 9858703]

45. Chédin F, Hartono SR, Sanz LA, Vanoosthuyse V, (2021). Best practices for the visualization, mapping, and manipulation of R-loops. EMBO J., e106394. [PubMed: 33411340]

46. Bayona-Feliu A, Casas-Lamesa A, Reina O, Bernués J, Azorín F, (2017). Linker histone H1 prevents R-loop accumulation and genome instability in heterochromatin. Nature Commun. 8, 283. [PubMed: 28819201]

47. Huang W, Loganantharaj R, Schroeder B, Fargo D, Li L, (2013). PAVIS: a tool for Peak Annotation and Visualization. Bioinformatics 29, 3097–3099. [PubMed: 24008416]

48. Skourti-Stathaki K, Kamieniarz-Gdula K, Proudfoot NJ, (2014). R-loops induce repressive chromatin marks over mammalian gene terminators. Nature 516, 436–439. [PubMed: 25296254]

49. Huppert JL, (2008). Thermodynamic prediction of RNA–DNA duplex-forming regions in the human genome. Mol. BioSyst 4, 686–691. [PubMed: 18493667]

50. Core LJ, Waterfall JJ, Gilchrist DA, Fargo DC, Kwak H, Adelman K, Lis JT, (2012). Defining the Status of RNA Polymerase at Promoters. Cell Rep. 2, 1025–1035. [PubMed: 23062713]

51. Saunders A, Core LJ, Sutcliffe C, Lis JT, Ashe HL, (2013). Extensive polymerase pausing during Drosophila axis patterning enables high-level and pliable transcription. Gene Dev 27, 1146–1158. [PubMed: 23699410]

52. Lesch BJ, Page DC, (2014). Poised chromatin in the mammalian germ line. Development 141, 3619–3626. [PubMed: 25249456]

53. Blythe SA, Wieschaus EF, (2015). Zygotic Genome Activation Triggers the DNA Replication Checkpoint at the Midblastula Transition. Cell 160, 1169–1181. [PubMed: 25748651]

54. Sibon OCM, Laurençon A, Hawley RS, Theurkauf WE, (1999). The Drosophila ATM homologue Mei-41 has an essential checkpoint function at the midblastula transition. Curr. Biol 9, 302–312. [PubMed: 10209095]

55. Shafiq S, Chen C, Yang J, Cheng L, Ma F, Widemann E, Sun Q, (2017). DNA Topoisomerase 1 Prevents R-loop Accumulation to Modulate Auxin-Regulated Root Development in Rice. Mol. Plant 10, 821–833. [PubMed: 28412545]

56. Lee C-Y, McNerney C, Ma K, Zhao W, Wang A, Myong S, (2020). R-loop induced G-quadruplex in non-template promotes transcription by successive R-loop formation. Nature Commun. 11, 3392. [PubMed: 32636376]

57. Capuano F, Mülleder M, Kok R, Blom HJ, Ralser M, (2014). Cytosine DNA Methylation Is Found in Drosophila melanogaster but Absent in Saccharomyces cerevisiae, *Schizosaccharomyces pombe*, and Other Yeast Species. Anal. Chem 86, 3697–3702. [PubMed: 24640988]

58. Pinter S, Knodel F, Choudalakis M, Schnee P, Kroll C, Fuchs M, Broehm A, Weirich S, Roth M, Eisler SA, et al. , (2021). A functional LSD1 coregulator screen reveals a novel transcriptional regulatory cascade connecting R-loop homeostasis with epigenetic regulation. Nucleic Acids Res. 49, 4350–4370. [PubMed: 33823549]

59. Herrera-Moyano E, Mergui X, García-Rubio ML, Barroso S, Aguilera A, (2014). The yeast and human FACT chromatin-reorganizing complexes solve R-loop-mediated transcription–replication conflicts. Genes Dev. 28, 735–748. [PubMed: 24636987]

60. Sun Q, Csorba T, Skourti-Stathaki K, Proudfoot NJ,Dean C, (2013). R-Loop Stabilization Represses Antisense Transcription at the Arabidopsis FLC Locus. Science 340, 619–621. [PubMed: 23641115]

61. Cózar JMGD, Gerards M, Teeri E, George J, Dufour E, Jacobs HT, Jõers P, (2019). RNase H1 promotes replication fork progression through oppositely transcribed regions of Drosophila mitochondrial DNA. J. Biol. Chem 294 jbc.RA118.007015.

62. Schneider CA, Rasband WS, Eliceiri KW, (2012). NIH Image to ImageJ: 25 years of image analysis. Nature Methods 9, 671–675. [PubMed: 22930834]

63. Ramirez P, Crouch RJ, Cheung VG, Grunseich C, (2021). R-Loop Analysis by Dot-Blot. J. Vis. Exp. 10.3791/62069.

64. Bolger AM, Lohse M, Usadel B, (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. [PubMed: 24695404]

65. Langmead B, Salzberg SL, (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. [PubMed: 22388286]

66. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup, 1000 Genome Project Data Processing, (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. [PubMed: 19505943]

67. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, et al. , (2008). Model-based Analysis of ChIP-Seq (MACS). Genome Biol. 9 R137–R137. [PubMed: 18798982]

68. Quinlan AR, Hall IM, (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. [PubMed: 20110278]

69. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T, (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 42, W187–W191. [PubMed: 24799436]

70. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. , (2000). Gene Ontology: tool for the unification of biology. Nature Genet. 25, 25–29. [PubMed: 10802651]

71. Consortium TGO, Carbon S, Douglass E, Good BM, Unni DR, Harris NL, Mungall CJ, Basu S, et al. , (2020). The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. 49, 325–334.

72. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD, (2019). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. 47, 419–426.

73. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, et al. , (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol. Cell 38, 576–589. [PubMed: 20513432]

74. Contrino S, Smith RN, Butano D, Carr A, Hu F, Lyne R, Rutherford K, Kalderimis A, Sullivan J, Carbon S, et al. , (2012). modMine: flexible access to modENCODE data. Nucleic Acids Res. 40, D1082–D1088. [PubMed: 22080565]

75. Amemiya HM, Kundaje A, Boyle AP, (2019). The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Sci. Rep.-uk 9, 9354.

76. McLean MJ, Wolfe KH, Devine KM, (1998). Base Composition Skews, Replication Orientation, and Gene Orientation in 12 Prokaryote Genomes. J. Mol. Evol 47, 691–696. [PubMed: 9847411]
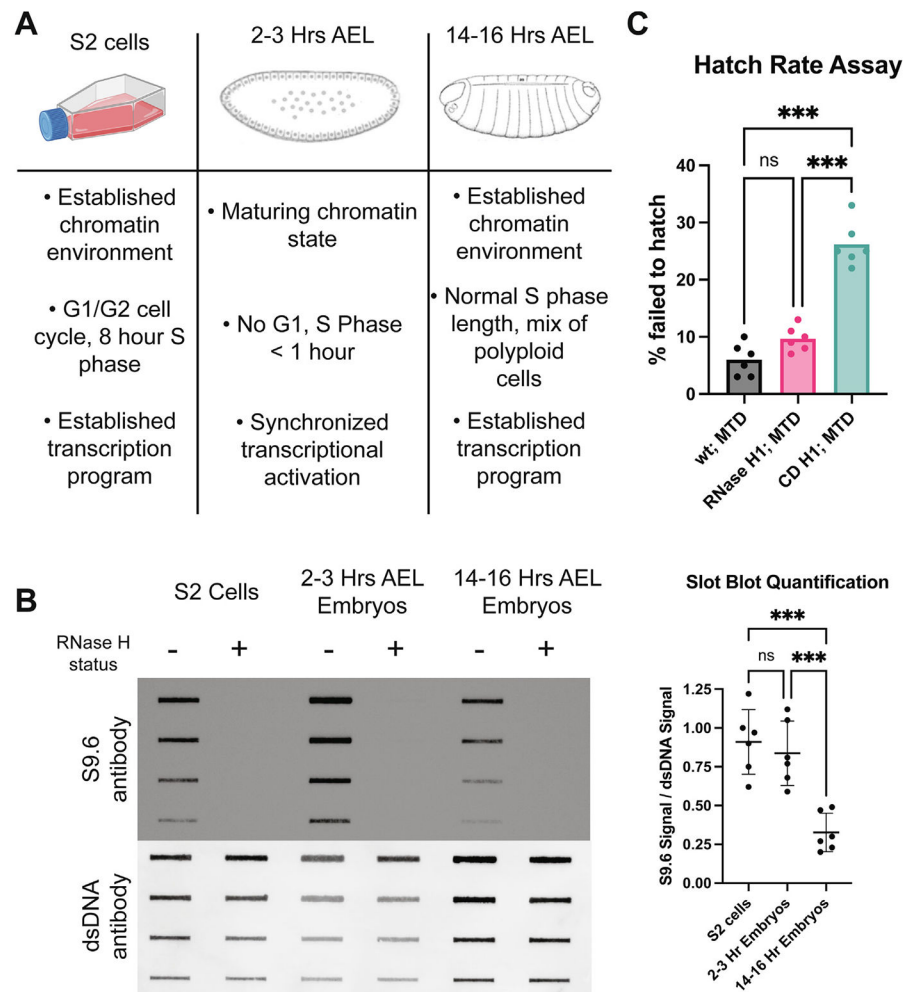
Figure 1.
R-loop abundance is developmentally regulated and R-loop homeostasis is necessary for development. (A) Schematic summarizing how the chromatin environment, developmental stage, and replication program vary among the developmental samples used. (B) Representative slot blot of RNA:DNA hybrid levels, measured by S9.6 antibody intensity, across samples. RNase H1 treatment verifies specificity of antibody, and antibody specific for double-stranded DNA is used as a loading control. Quantification of signal for six biological replicates is to the right. *** < 0.05, one-way ANOVA with Tukey's multiple comparisons test. (C) Hatch rate among embryos that overexpress RNase H1 (H1) or a catalytic dead RNase H1 (CD). 6 biological replicates from 2 independent crosses, counting 100 embryos in each replicate. *** < 0.05, one-way ANOVA with Tukey's multiple comparisons test.
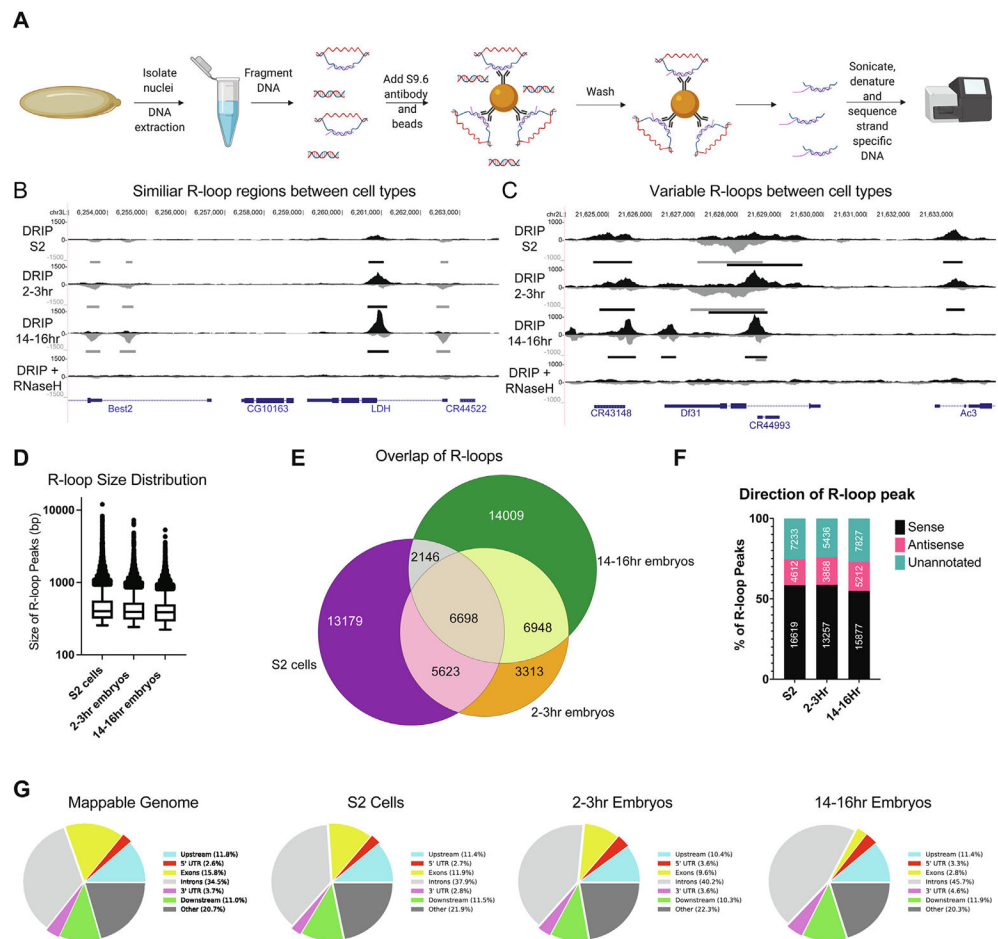
Figure 2.
The R-loop landscape changes as a function of development. (A) Diagram of the ssDRIP-seq mapping strategy. (B) ssDRIP-seq snapshot of a 10 kb region on chromosome *3L* where R-loop distribution is similar between samples. Black and grey bars below each track represent peak calls for forward and reverse strands, respectively. (C) ssDRIP-seq snapshot of a 10 kb region on chromosome *2L* where R-loop distribution varies between samples. Note the reverse strand coming from a lncRNA in the middle of the Df31 gene. (D) The distribution of R-loop sizes for each developmental sample. (E) Overlap of R-loops between developmental samples. (F) Quantification of the percent of R-loops mapping to sense, antisense and untranscribed regions of the genome. Numbers represent absolute R-loop peaks in each category. (G) R-loop enrichment relative to the expected distribution for common genomic features.
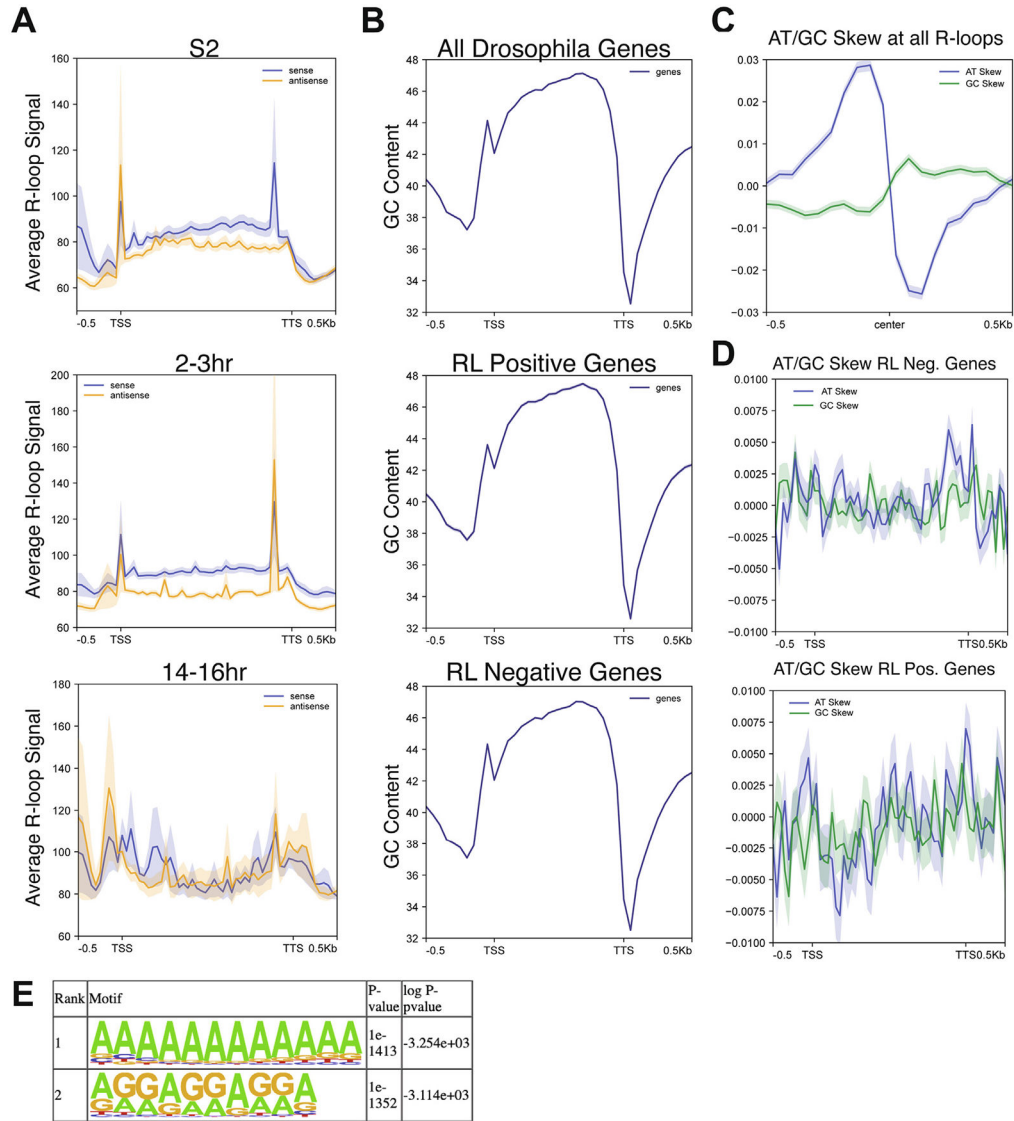
Figure 3.

R-loop signal as a function of transcription unit and sequence composition. (A) Metaplots of ssDRIP-seq signal for all samples relative to the gene body. Each plot represents the signal derived from sense R-loops in blue and antisense R-loops in orange. Shaded region represents the standard error of the mean (SEM). (B) The GC composition of all Drosophila genes, genes that have an R-loop in one of the developmental samples and genes that lack any R-loop signal. Shaded region represents the SEM. (C) Metaplot of GC and AT skew across all identified R-loops. Shaded region represents the SEM. (D) Metaplot of GC and AT skew across the gene body of genes that lack R-loops (top) and genes that form an R-loop. Shaded region represents the SEM. (E) DNA sequence motifs in the peaks of all R-loops identified by HOMER. Motif analysis was not strand specific.
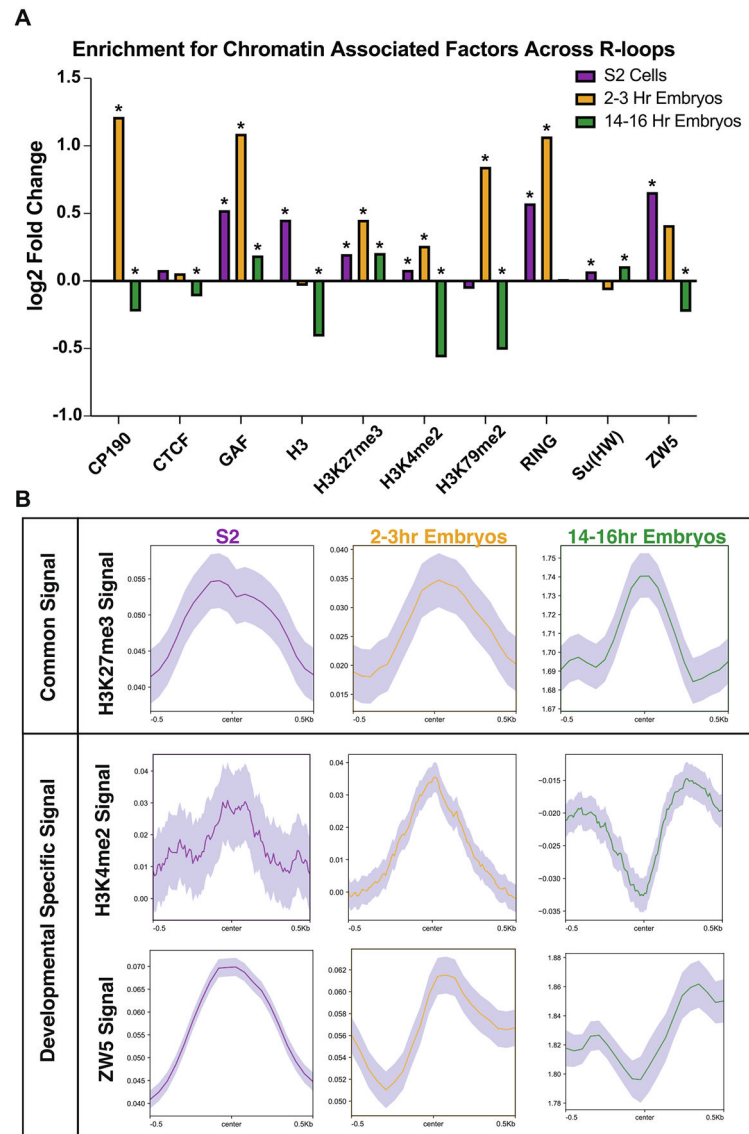
Figure 4.
Common chromatin features associated with R-loops. (A) Log2 fold enrichments of chromatin-associated factors within R-loop regions in common for S2 cells, 2–3 hour embryos and 14–16 hour embryos. * < 0.05 with Bonferroni correction for multiple testing. (B) Metaplots of H3K27me3, H3K4me2, and ZW5 ChIP-chip (S2 and 2–4 hour embryos) and ChIP-seq (14–16 hour embryos) confirming common and developmental-specific enrichment of chromatin factors at R-loops. Shaded region represents the standard error of the mean (SEM).

Figure 5.
R-loop formation as a function of transcription. (A) GRO-seq values for genes that contain strand-specific R-loops (RL Pos), genes that do not contain strand-specific R-loops (RL Neg) in S2 cells, every transcript in S2 cells, and transcripts that only form R-loops in S2 cells. (B) Transcripts were sorted into quartiles based upon GRO-seq expression, and R-loop forming genes were assigned to their respective quartile. (C) Same as A, except for 2–3 h embryos. (D) Same as B, except for 2–3 h embryos. (E) The average number of R-loops peaks detected for each gene in each of the expression quartiles is graphed for S2 cells and 2–3 h embryos. (F) The difference in GRO-seq values between S2 cell and 2–3 h embryos were queried for genes that showed developmental-specific R-loop formation. (G) Log2 fold enrichments of chromatin-associated factors within R-loop regions in the highest or lowest expression quartiles in S2 cells. ns > 0.05 with Bonferroni correction for multiple testing. (H) Log2 fold enrichments of chromatin-associated factors within R-loop regions

in the highest or lowest expression quartiles in 2–3 h embryos. ns > 0.05 with Bonferroni correction for multiple testing.
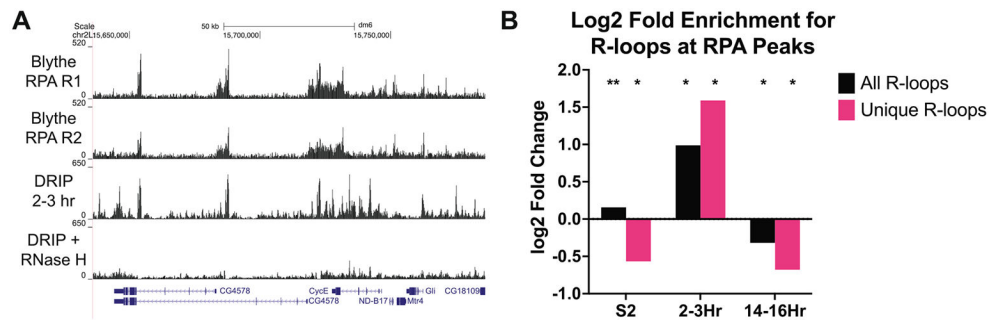
Figure 6.
R-loops have the potential to trigger ATR activation at the MZT. (A) Overlap of RPA ChIP-seq profiles from cycle 13 embryos (Blythe and Wieschaus et al. 2015) and ssDRIP-seq profiles from 2-3 h embryos. (B) Log2-fold enrichment of RPA at R-loop peaks for all samples. Each sample was separated into total R-loops or R-loops unique to that sample type. P values were generated with Bonferroni correction for multiple testing. * = P value < 0.01 and ** = P value < 0.001.

Table 1

List of available ChIP-chip and ChIP-seq from modENCODE.

Assay	Time	Mark
ChIP-chip	2–4 hr	BEAF-32, CP-190, CTCF, RING, SFMBT, GAF, H2Av, H2Bubi, H3, H3K18ac, H3K23ac H3K27ac, H3K27me3, H3K36me1, H3K36me3, H3K4me1, H3K4me2, H3K4me3, H3K79me1, H3K79me2, H3K79me3, H3K9ac, H3K9me2, H3K9me3, H4, H4K20me1, HP1a, HP1c, HP2, Polycomb, POF, Su(HW), ZW5
	S2 cells	ACF1, ASH1, BEAF-70, BEAF-HB, CG10630, Chriz-WR, CP190, CTCF, Mi–2, TopoII, RING, SFMBT, E(z), GAF, H1, H2Av, H2BK5ac, H2Bubi, H3, H3K18ac, H3K23ac, H3K27ac, H3K27me1, H3K27me2, H3K27me3, H3K36me1, H3K36me3, H3K4me1, H3K4me2, H3K4me3, H3K79me1, H3K79me2, H3K79me3, H3K9ac, H3K9acS10P, H3K9me1, H3K9me2, H3K9me3, H4, H4acTetra, H4K12ac, H4K16ac, H4K20me1, H4K5ac, H4K8ac, HP1a, HP1b, HP1c, HP2, HP4, ISWI, JHDMI, JIL–2, JMJD2A, LSD1, MBD-R2, MLE, mod(mdg4), MOF, MRG15, MSL–1, NURF301, ORC2, Polycomb, PCL, Pho, Pof, PR-Set7, Psc, Rhino, RNAPolII, RPD3, Smc3, Spt16, Su(HW), Su(var)3–7, Su(var)3–9, WDS, ZW5
ChIP-seq	14–16 hr	Beaf-HB, Chriz, CP190, CTCF, Mi–2, RING, GAF, H1, H2Av, H2B-ubi, H3, H3K18ac, H3K23ac, H3K27ac, H3K27me2, H3K27me3, H3K36me1, H3K36me2, H3K36me3, H3K4me1, H3K4me3, H3K79me1, H3K79me2, H3K79me3, H3K9acS10P, H3K9me1, H3K9me2, H3K9me3, H4, H4K16ac, H4K20me1, HP1a, HP1b, HP1c, HP2, HP4, JHDMI, LSD1, MBD-R2, MOF, NURF301, POF, Psc, RNAPolII, RPD3, Su(HW), Su (var)3–7, ZW5