

Review

Nascent Connections:
R-Loops and Chromatin
PatterningFrédéric Chédin^{1,*}

RNA molecules, such as long noncoding RNAs (lncRNAs), have critical roles in regulating gene expression, chromosome architecture, and the modification states of chromatin. Recent developments suggest that RNA also influences gene expression and chromatin patterns through the interaction of nascent transcripts with their DNA template via the formation of co-transcriptional R-loop structures. R-loop formation over specific, conserved, hotspots occurs at thousands of genes in mammalian genomes and represents an important and dynamic feature of mammalian chromatin. Here, focusing primarily on mammalian systems, I describe the accumulating connections and possible mechanisms linking R-loop formation and chromatin patterning. The possible contribution of aberrant R-loops to pathological conditions is also discussed.

Basic Determinants of Co-Transcriptional R-Loop Formation

During transcription, one strand of the DNA double helix is copied into a complementary RNA transcript. RNA synthesis occurs within the context of a moving transcription bubble where the two strands of DNA are physically separated and the nascent transcript is held through a transient 8-base pair (bp) RNA:DNA hybrid. The transcription elongation complex architecture ensures that the outgoing nascent RNA does not become entangled with the DNA helix as the RNA polymerase (RNAP) moves forward [1] (Figure 1). For this reason, **R-loop structures** (see Glossary), which can form upon reannealing of the nascent RNA to the template DNA strand, have long been considered rare and accidental byproducts of transcription.

In vitro transcription studies of mammalian switch regions first revealed that R-loop formation could be highly efficient [2,3]. Additional studies demonstrated that the efficiency of R-loop formation related directly to the sequence of the transcribed region. Repetitive switch regions characterized by strong G clustering in close proximity to the transcription start site (TSS) efficiently form R-loops *in vitro* when transcription results in a G-rich RNA [4,5]. Similarly, *in vitro* transcription through unique, nonrepetitive, human and mouse **CpG island** (CGI) promoter sequences leads to efficient R-loop formation when the transcribed regions are GC rich, have strong **GC skew**, and transcription results in a G-rich nascent transcript [6]. These observations are consistent with the fact that RNA:DNA hybrids comprising G-rich RNA strands have a higher thermodynamic stability compared with DNA–DNA duplexes [7]. Thus, R-loop formation can be understood as a competition between the nascent RNA and the nontemplate DNA strand for reannealing with the template DNA strand behind the advancing RNAP [5,8]. Factors that are conducive to reannealing of the RNA strand, including favorable RNA:DNA base-pair energetics and negative **supercoiling**, facilitate R-loop formation [9–11] (Figure 1).

Trends

R-loop structures are a type of non-B DNA structure that can form during transcription upon reannealing of the nascent RNA to the template DNA template. R-loops can now be profiled globally at high resolution in any genome.

R-loops in mammalian systems are abundant, covering over 100 Mb of DNA sequence, and form dynamically over conserved regions.

R-loops have important roles in chromosome dynamics, in particular because it concerns transcription termination and chromatin patterning.

Deregulation of R-loop metabolism is linked to an increasing number of human diseases, including cancers and several neurodegenerative disorders.

¹Department of Molecular and Cellular Biology and Genome Center, One Shields Avenue, University of California, Davis, CA 95616, USA

*Correspondence: fichedin@ucdavis.edu (F. Chédin).

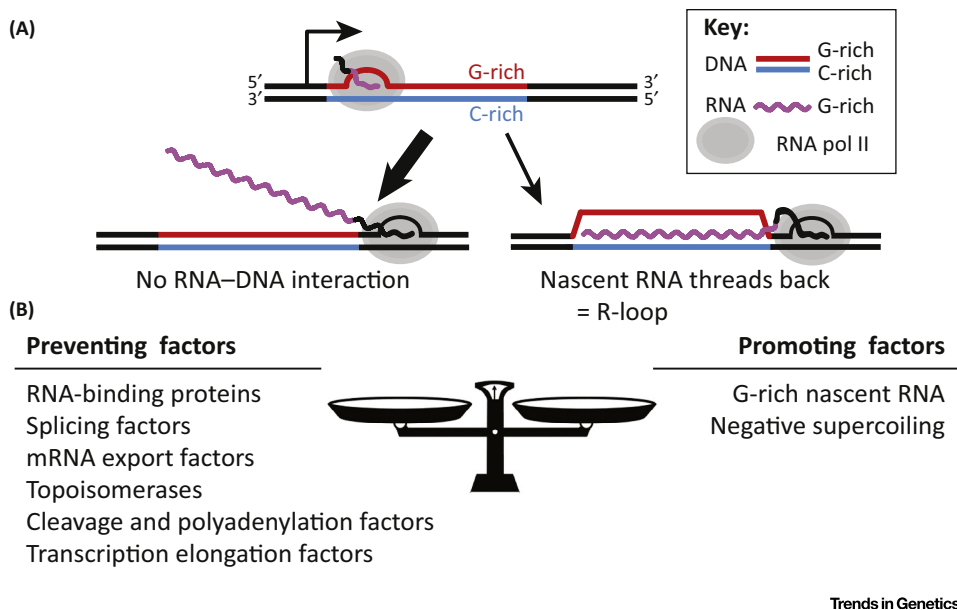


Figure 1. Schematic Description of Co-transcriptional R-loop Formation. (A) Schematic description of co-transcriptional R-loop formation. In most cases, transcription proceeds without R-loop formation (left; thick arrow). However, in some cases, in particular over DNA regions characterized by GC strand asymmetries, the nascent RNA can reanneal and lead to R-loop formation. (B) A brief list of factors known to prevent or facilitate R-loop formation. Abbreviation: RNA Pol, RNA polymerase.

Glossary

CpG island: a stretch of DNA sequence characterized by high GC content and high CG dinucleotide density compared with the rest of the human genome. CpG islands often map to the beginning of mammalian genes, where they serve as promoters. Nearly 60% of human genes have a CpG island promoter.

GC skew: a sequence property that describes the strand asymmetry in the distribution of guanine versus cytosine residues.

R-loop structure: a type of three-stranded non-B DNA structure in which the nontemplate DNA strand exists in a looped out single-stranded state owing to the presence of the RNA:DNA hybrid (Figure 1, main text).

Supercoiling: the torsional stress that is exerted on the DNA double helix. In negatively supercoiled DNA, the two DNA strands are slightly underwound compared with B DNA.

Trends in Genetics

Box 1. Methods for R-Loop Mapping

Nondenaturing (native) sodium bisulfite treatment allows R-loop footprinting by triggering the deamination of intrinsically unpaired cytosines, such as those on the looped-out strand of an R-loop, to uracil (U). After PCR amplification and sequencing of individual DNA molecules, R-loops can be detected by the presence of long, strand-specific, C to T conversion ‘footprints’. Importantly, these footprints are sensitive to pretreatment with Ribonuclease H, an enzyme that degrades RNA strands specifically in the context of RNA:DNA hybrids [84]. The S9.6 antibody and its sequence-independent subnanomolar affinity for RNA:DNA hybrids [13,14] enabled R-loop immunoprecipitation from genomic material [30,92]. This reagent also allowed R-loop profiling at the genome scale through DNA:RNA Immunoprecipitation coupled to high-throughput DNA sequencing (DRIP-seq) [6]. Variations of the DRIP-seq method have now enabled the profiling of R-loops at higher resolution and in a strand-specific manner, either through sequencing of RNA strands involved in R-loop formation after cDNA synthesis (DRIP-RNA-seq [22] or DRIPc-seq [19]), or through sequencing of the template DNA strand (RDIP-seq [18] and S1-DRIP-seq [93]).

Prevalent and Conserved R-Loop Formation in Mammalian Genomes

R-loops can now be effectively mapped at endogenous loci in genomic DNA using a variety of methods, including low-throughput, single-molecule approaches based on nondenaturing sodium bisulfite footprinting [12] and high-throughput, population average techniques based on the S9.6 anti-RNA:DNA hybrid antibody [13,14] (Box 1). Initial evidence for endogenous R-loop formation came from footprinting analysis of murine switch regions, where R-loops were detected upon induction of the corresponding upstream promoter in primary mouse B cells [12]. Consistent with thermodynamic predictions, these R-loops often initiated in proximity to G clusters and terminated over regions of lower G density [15–17]. However, R-loop initiation and termination sites were heterogeneous, resulting in structures of variable lengths ranging from a few hundred base pairs up to over 1 kb. Similar results were obtained upon footprinting R-loops at three different CGI promoters in the human and mouse genomes [6]. Therefore, R-loops form in mammalian chromosomes and individual R-loops can be long, covering stretches of DNA corresponding to multiple nucleosomes.

High-throughput R-loop mapping studies revealed that R-loop formation in mammalian genomes is more prevalent than initially thought, occurring over tens of thousands of genomic loci [6,18–21] and covering up to 150 Mb of sequence space or nearly 5% of the human genome [19]. Consistent with results from *in vitro* transcription and R-loop footprinting, CGI promoters are major R-loop hotspots in the human genome [6,18,20,22]. R-loops formed at CGI promoters were preceded by significant GC skew. The observation that GC skew is a conserved characteristic of mammalian CGI promoters suggests that these loci are prone to R-loop formation in a range of organisms [23]. Promoter R-loops tend to peak 1–2 kb downstream of the TSS and are distinct from the short (40–60 bp) nascent transcripts held within the transcription complex at sites of promoter-proximal pausing [24]. The 3' end of genes surrounding the polyadenylation site (PAS) is another major hotspot of R-loop formation [19,20,22]. Terminal R-loops often do not associate with positive GC skew [19], and terminal GC skew, contrary to promoter GC skew, is conserved only in a small subset of vertebrate genes [23]. Thus, terminal R-loops may not be driven by the energetic favorability of the RNA:DNA hybrid. In addition, the ends of genes, in particular convergent genes that are notable R-loop hotspots [19], are thought to associate with positive supercoiling during transcription elongation [25]. The lack of favorable sequence and topological characteristics at gene ends raises questions regarding the mechanism(s) underlying terminal R-loop formation. Interestingly, these structures were only observed for transcripts undergoing PAS-dependent cleavage and polyadenylation [19]. In budding yeast, mutations in cleavage and polyadenylation factors, as well as other 3' end-processing factors, lead to R-loop-mediated genomic instability [26–28]. R-loop structures were further proposed to play a key role in the transcription termination process [19,29–31]. It is possible that terminal R-loops are formed as an intermediate in transcription termination and 3' processing and that their formation involves the activity of as yet unknown specialized complexes. While gene ends are clearly R-loop hotspots, most genic R-loops form within gene bodies [18,19]. Gene-body R-loops typically occupy only a small portion of the sequence space of a gene. This indicates that R-loop formation is not only a simple consequence of transcription, but also results from interplay between transcription, DNA sequence, topology, and possibly other characteristics of the chromatin template.

R-loops have now been profiled in multiple human and mouse cell lines [18,19,32], including primary human fibroblasts [21] and murine embryonic stem cells (mESCs) [19,22]. The promoter and terminator regions of genes are conserved R-loop formation hotspots both within cell lines [18,19,32] and across species [22], including over orthologous gene pairs [19]. Gene-body R-loop signal is also often conserved, albeit slightly less so than the signal at gene extremities. In addition to genic R-loop formation, repetitive regions of mammalian genomes are also capable of structure formation (Box 2). Collectively, these studies highlight that R-loop formation occurs at specific, conserved loci associated with transcription initiation, elongation, and termination.

Box 2. R-Loop Formation in Intergenic and/or Repetitive Regions

While genic R-loops have been the best studied so far, evidence suggests that these structures also form over intergenic and/or repetitive regions. For instance, rDNA arrays are known R-loop hotspots in bacteria [83], yeast [92], and mammals [18]. Bright S9.6 staining over mammalian nucleoli [6,87] supports the notion that nucleoli are a strong R-loop hotspots, although the extent to which this signal is solely representative of RNA:DNA hybrids or also caused by abundant double-stranded rRNAs remains unclear. Recent mapping data suggest that RNA:DNA hybrids form over exons of mammalian rDNA units and over the intergenic spacer region [18]. In addition, evidence exists that R-loops can form over telomeric regions in yeast [94] and mammals. In this latter case, transcription initiation upstream of the G-rich telomeric repeats (5'-TTAGGG-3') through the CGI-like TERRA promoter, gives rise to R-loop formation [95]. RNA:DNA hybrids were also suggested to form over specific human pericentromeric regions [18] and have been long described to form over mitochondrial DNA from yeast to mouse, where they are involved in priming DNA replication [96–98].

Dynamic Turnover of R-Loop Structures

While genomic studies provide a wealth of information on the range of sequences that can form R-loops, they provide no information on two key parameters: the frequency at which R-loops form and the turnover rate of an R-loop once formed. DRIP-qPCR approaches suggest that R-loop formation frequencies range from 1% to 10% of input genomic DNA, depending on the locus [19,29,32]. By contrast, negative loci (untranscribed and/or intergenic regions) range from 0.01% to 0.1% of input. Thus, while R-loops can form over a surprisingly large fraction of the genome, the frequencies at which they are formed remain relatively modest. Given that most genic R-loops form co-transcriptionally, R-loop turnover rates can be measured after blocking *de novo* R-loop formation by inhibiting transcription. Treatment with transcription-blocking drugs combined with a kinetic analysis of R-loop frequencies showed that R-loops are dynamic structures. At promoters, which are easiest to analyze, R-loops are rapidly turned over, showing a half-life of 10 min [19]. This implies that R-loops are continuously formed and resolved and that the retention of nascent transcripts at their site of transcription is a dynamic feature of mammalian chromatin.

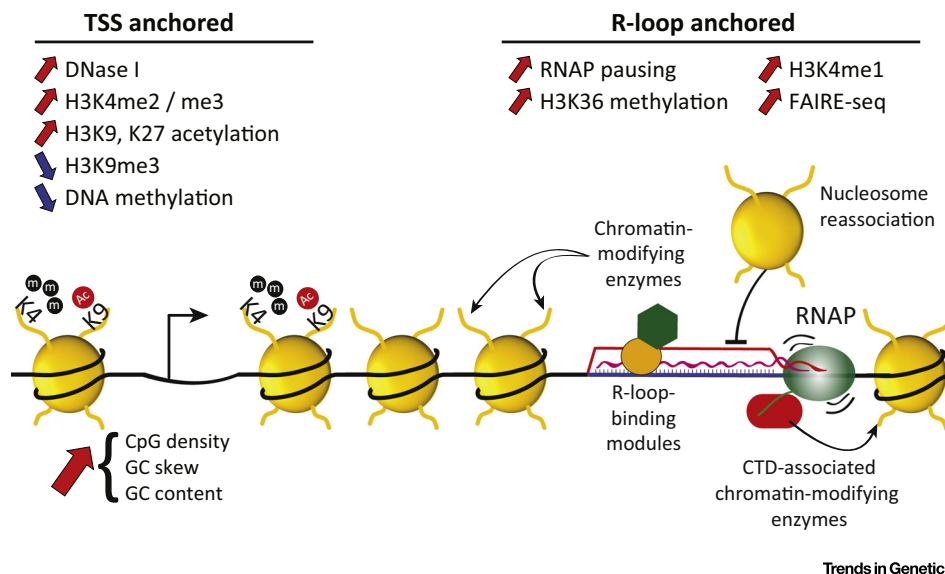
Heightened Accessibility and RNAP Stalling Are Common Features of R-Loop Chromatin

The formation of long, stable R-loops likely alters local chromatin, and numerous studies have focused on defining these effects. Consistent with observations that RNA:DNA hybrids adopt a rigid A form-like conformation [33] and prevent nucleosome wrapping *in vitro* [34], R-loops associate with DNase I hyperaccessibility at all genic positions, as per DRIPc-seq [19]. This association could be explained by interference with nucleosome redeposition behind the advancing RNAP [35]. In support of this, levels of H3.3, a histone variant that dynamically replaces H3 lost owing to nucleosome disruption [36], are elevated over both promoter and terminal R-loop regions [19]. R-loops further correlate with chromatin decondensation and lower nucleosome occupancy [37] and, conversely, R-loop destabilization causes chromatin compaction [38]. Finally, loss of the FACT histone chaperone causes R-loop accumulation, suggesting that proper transcription-coupled nucleosome redeposition prevents R-loop formation [39]. The association of R-loops with hyperaccessible chromatin is generally consistent with long-standing observations that nuclear-retained RNAs associate with the maintenance of an open, active, chromatin state and that treatment of cells with ribonucleases or transcription inhibitors leads to the collapse of chromatin architecture [40,41]. Altogether, these observations suggest that R-loops locally open the chromatin structure, possibly by regulating nucleosome occupancy, positioning, and/or turnover. The proposed association between 'aberrant' R-loops and chromatin condensation is specifically discussed below.

Another strong hallmark of R-loop-forming [R-loop(+)] regions is a local increase in the density of RNAP [19]. This suggests that R-loop formation causes transient RNAP stalling, as observed *in vitro* [11]. The alternative possibility that RNAP stalling promotes R-loop formation cannot be ruled out, although a recent study of elongation-defective yeast RNAP mutants did not reveal an increase in R-loop formation [42]. Interestingly, promoter distal RNAP-pausing sites were described at CGI promoters over R-loop-prone regions characterized by high GC skew [43]. These distal pause sites mediate contacts with long-range enhancers, suggesting that R-loop-associated pausing participates in the control of gene expression [43]. Interestingly, switch regions, one of the best-documented R-loop regions (see above), are also characterized by chromatin accessibility and RNAP stalling [44,45], suggesting that these two features are consistent hallmarks of R-loop chromatin.

Defining Histone Modification Signatures of R-Loop Chromatin

In addition to chromatin accessibility and RNAP stalling, high-resolution R-loop maps have enabled the identification of a set of histone modifications that associate with R-loop chromatin under normal conditions [19,22]. Histone modifications that characterize 'aberrant' R-loops



Trends in Genetics

Figure 2. Summary of Chromatin Features Associated with R-Loop Formation at Promoter Regions. The features were broken down depending on their location. Transcription start site (TSS)-anchored features typically occur in the immediate surrounding of the promoter region, often at CGI promoters marked by increased CpG density, GC skew, and GC content. By contrast, R-loop-anchored features were found to physically coincide with the location of R-loops, typically 1–2 kb downstream of the TSS. Multiple mechanisms are envisioned for coupling R-loop formation to chromatin patterning, including the recruitment of chromatin-modifying complexes through R-loop binding modules, the prevention of nucleosome redeposition behind the RNA polymerase (RNAP) machinery, and the co-transcriptional recruitment of chromatin-modifying enzymes to transiently stalled RNAP complexes. Abbreviation: CTD, C-terminal domain.

formed under altered conditions are described separately below. The existence of such signatures suggests that the transient formation of RNA:DNA hybrids is sensed by chromatin-modifying enzymes and translated into specific modification states. Thus, R-loop formation, which occurs at thousands of loci of varied DNA sequence in a transient and reversible manner, may represent an additional layer of epigenetic information.

At promoter regions, R-loops associate with higher levels of histone marks characteristic of active transcription (H3K4me1, H3K4me3, and H3 acetylation) and transcription elongation (H3K36me3) (Figure 2) [19,22]. Analysis of the chromatin of activated switch regions in B cells showed that H3K4me3 and H3 acetylation are recruited upon transcriptional induction [46–48]. Furthermore, genomic off-targets of the activated B cell-specific AID cytidine deaminase, an enzyme that acts on single-stranded (ss)DNA, including that found at R-loop structures [49], are characterized by high-transcription, RNAP stalling, and the combined R-loop-enriched marks H3K4me1, H3K27Ac, and H3K36me3 [50]. Terminal R-loops were predominantly marked by increased H3K4me1 levels [19], suggesting that H3K4me1 is a common mark of R-loop regions. Terminal R-loops also showed increased recruitment of the p300 acetyltransferase and of the CTCF-cohesin complex, conferring them an enhancer-like and insulator-like chromatin state [19]. Interestingly, these observations may account for properties often linked to efficient transcription terminators, including the production of antisense noncoding RNAs and the formation of gene loops [29,51,52]. Indeed, evidence suggests that terminal R-loops represent a broad class of efficient transcription terminators [19], as previously suggested [30].

Mechanisms of R-Loop-Mediated Chromatin Patterning

Now that a coherent set of R-loop chromatin signatures is emerging, the task is to decipher which of these marks can be causally linked to R-loop formation and to provide a mechanistic understanding of these relations. One way that R-loops may affect the chromatin landscape is by

favoring the recruitment of chromatin-modifying complexes that are normally recruited co-transcriptionally. Trimethylation of histone H3 at lysine 36 (H3K36me3) is mediated by the SETD2 histone methyltransferase upon interaction with the C-terminal domain (CTD) of RNAP [53,54]. Genes with R-loop(+) promoters show heightened H3K36me3 recruitment precisely over the R-loop formation peak and the associated increased RNAP density [19] (Figure 2). This suggests that R-loops facilitate recruitment of SETD2 to RNAP complexes by transiently stalling the transcription machinery. Additional histone modifications that are deposited co-transcriptionally, most significantly H3K4 methylation [55,56], are also enriched over R-loop(+) promoters. However, in this case, it is worth distinguishing between H3K4 trimethylation, which shows a TSS-anchored distribution, from H3K4 monomethylation, which shows an R-loop-anchored distribution 1–2 kb downstream of TSS (Figure 2) [19]. In the first case, it is possible that the higher CpG densities and GC contents of R-loop(+) promoters [19] are driving increased H3K4me3. Indeed, CpG density favors the recruitment of the unmethylated CpG-binding protein CFP1 [57] and its associated H3K4 methyltransferase SET1/COMPASS complex [55,58]. In the latter case, it is possible that R-loops have a more direct role in recruiting H3K4me1. While the exact mechanism linking R-loop formation and H3K4me1 deposition remains to be elucidated, several SET1/MLL H3K4 methyltransferase family members have been shown to bind ssDNA [59]. Furthermore, the PAF1 complex, which facilitates both H3K4 and H3K36 methylation [60], is enriched over R-loop regions [19] and may itself be able to sense ssDNA [61]. Interestingly, increased PAF1 recruitment over terminal R-loops may also account for H3K4me1 deposition over these regions and for efficient transcription termination, given the ability of the PAF1 complex to recruit termination factors [60]. R-loops may also favor the recruitment of the XRN2 ‘torpedo’ ribonuclease at terminal pause sites, triggering the release of the RNAP machinery upon degradation of the residual transcript downstream of the PAS [30,31,62].

Histone acetylation, as discussed above, is a TSS-anchored mark of R-loop(+) promoters (Figure 2). Recent data show that recruitment of the Tip60-p400 chromatin-remodeling and acetylase complex is substantially reduced in mESCs expressing RNASEH1 [22], an enzyme that resolves R-loops. This, together with evidence that many Tip60-p400 targets form R-loops in mESCs, suggests that R-loops contribute to recruiting the Tip60-p400 complex to chromatin [22] through an as-yet-unidentified R-loop-binding subunit. Interestingly, RNASEH1 overexpression also led to increased recruitment of the Polycomb Responsive Complex 2 (PRC2) and increased deposition of the PRC2-catalyzed H3K27me3-silencing mark to a subset of genomic targets in mESCs [22]. This suggests that R-loops prevent PRC2 chromatin loading. Notably, R-loop(+) loci in human carcinoma cells showed significantly higher H3K27me3 levels compared with matched R-loop(-) loci, especially for poorly expressed genes [19]. One possibility to account for this discrepancy is that pluripotent ESCs have distinct regulatory mechanisms controlling H3K27me3 deposition compared with differentiated cells. However, successfully preventing the recruitment of a complex to chromatin requires an efficient and long-lived inhibitory signal when R-loops form transiently and at modest frequencies.

Altogether, evidence suggests that R-loops influence chromatin modification states by affecting nucleosome density, through co-transcriptional mechanisms, or even directly via recognition by chromatin regulators. Another speculative mechanism by which R-loops could contribute to chromatin patterning might involve the targeting of lncRNAs (Box 3) [63].

Aberrant R-Loops and Chromatin Condensation

Defects in the THO mRNA export complex in yeast triggers marked genomic instability and increased R-loop formation [64]. This increased R-loop load is accompanied by higher levels of H3S10P [65], a mark typically associated with condensed chromosomes during mitosis [66]. Increased H3S10P and condensation were observed outside of mitosis and, while most visible

Box 3. *Cis* or *Trans* R-Loops as a Mechanism for Targeting Noncoding RNAs?

The study of genic R-loops provides strong support for a co-transcriptional mode of formation in *cis*: (i) in 90% of cases, R-loop are formed codirectionally with transcription [19]; (ii) R-loop levels correlate with expression levels [18,19,74]; (iii) R-loop levels respond to transcriptional changes [19,74]; (iv) DNA-based (DRIP-seq) and RNA-based (DRIPc-seq) R-loop maps agree well with each other [19]; and (v) genes showing allele-specific expression show allele-specific R-loops [6]. However, it remains possible that R-loops also form in *trans*. In this case, an RNA strand transcribed from a locus could conceivably hybridize onto a distant locus through regions of RNA–DNA complementarity (Figure 1). This mechanism likely requires proteins to catalyze strand invasion and assist in homology searching, as is the case for the CRISPR-Cas9 complex [99,100]. Evidence for a *trans* R-loop mechanism dependent on the homologous recombination machinery was obtained in budding yeast [101,102]. Similarly, *trans* R-loops dependent on the RecA DNA strand exchange protein have been described in *Escherichia coli* and reconstituted biochemically [103–105]. *Trans* R-loop formation provides an attractive mechanism for directing lncRNAs to potential targets. Recent evidence suggests that the GAL lncRNAs in budding yeast regulate transcriptional induction via *trans* R-loop-mediated gene looping [106]. In addition to these two mechanisms, a mixed *cis/trans* model is also possible (Figure 1). In this case, co-transcriptional R-loops might provide a ready-made ssDNA landing pad on the nontemplate strand for lncRNAs to interact with in *trans*. Such *trans* RNA:DNA hybrid formation might be easier to achieve mechanistically and only require the activity of a protein with strand-annealing capacity. The conserved Rad52 protein, which mediates strand annealing upon binding to ssDNA and displacement of the ssDNA-binding protein RPA [107], is a good candidate for such an activity. Recent evidence shows that Rad52 can catalyze the annealing of an RNA strand to a DNA strand, creating RNA:DNA hybrids [108]. Interestingly, this mixed *cis/trans* model predicts that transcription itself, through the formation of *cis* R-loops, might regulate the potential repertoire of some lncRNAs *trans* targets, as observed [109].

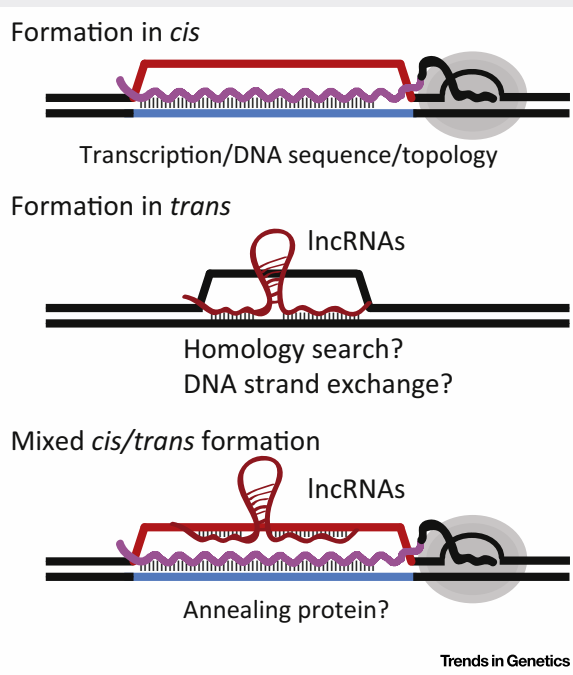


Figure 1. Possible Mechanisms for R-Loop Formation in *cis* and in *trans*, Including a Mixed *cis/trans* Mechanism. Factors that may contribute to these mechanisms are also listed.

over pericentromeric regions, they were also observed over coding genic regions. Since H3S10P levels could be suppressed by RNASEH1 expression [65], a link with R-loops was suggested. Accumulation of R-loops and H3S10P in THO-defective conditions was demonstrated in *Caenorhabditis elegans* and human cells, further arguing that this link is conserved [65]. Thus, under conditions associated with ‘aberrant’ R-loop formation (see below), R-loops appear to associate with H3S10P and chromatin condensation, a startling conclusion given the significant evidence linking R-loops and active, hyperaccessible chromatin under normal conditions. Interestingly, prior studies of the chromatin state of R-loop-prone murine switch regions showed that H3S10P is increased upon transcriptional induction [48]. However, the mark was

found in the context of active, acetylated, and H3K4 methylated chromatin [48]. Indeed, H3S10P in interphase cells has been associated with transcription and can coexist with active histone modifications [67–69]. In *Drosophila*, H3S10 phosphorylation was suggested to permit transcription by counteracting H3K9 dimethylation, a mark of gene silencing [70]. Thus, while evidence suggests that H3S10P associates with R-loop formation, additional work is required to delineate the mechanism underlying this association and to clarify the relations between H3S10P, R-loops, and the physical state of chromatin.

Recently, R-loop formation has been linked to increased deposition of H3K9me2/me3 and heterochromatin formation in the context of triplet expansion diseases, such as Friedrich's ataxia and Fragile X syndrome [71–73]. This association between R-loops and a silencing histone modification could reflect the fact that R-loops formed over expanded repeat arrays are unusually complex, including formation of hairpins on the displaced ssDNA strand [74], and, therefore, are hard to resolve. Complex, persistent, R-loops might in turn be more prone to double-strand break (DSB) formation, a hallmark of R-loops [75]. Given that DSB repair involves the transient repression of chromatin via H3K9me2/3 deposition [76,77], the association of R-loops with H3K9 methylated chromatin observed at specific loci may reflect R-loop-induced fragility. Recently, H3K9me2 deposition was also observed broadly over R-loop(+) terminator regions of murine genes and proposed to participate in the transcription termination process [29]. However, re-examination of global data sets did not support this broad association [19]; instead, terminators were shown to adopt an open, insulator-like state. Nevertheless, it remains possible that specific terminator regions, such as those of highly expressed genes, may carry high loads of persistent R-loops that predispose them to H3K9me2/3 deposition via a mechanism similar to that delineated above for triplet-repeat expansion loci.

Defining Aberrant R-Loop Formation under Pathological Conditions

R-loops are well known for their association with genomic instability, a topic extensively covered by several excellent recent reviews [75,78–82]. This association is clearest in the context of defects in a variety of factors involved in co-transcriptional processes, such as splicing, mRNA export, 3'-RNA processing, and transcription elongation, among others. Defects in other factors, such as topoisomerases, which are thought to prevent R-loop formation [83], or in enzymes such as Ribonucleases H [84] and RNA/DNA helicases, which are thought to mediate R-loop resolution, also associate with DSBs and genomic instability [62,85–89]. Not surprisingly, dysfunctional R-loop metabolism has been implicated in multiple human disorders, particularly cancers and neurological diseases (reviewed in [71,90,91]). Despite these clear links, the genomic profiles of R-loops generated under pathological conditions have seldom been examined, raising questions as to what 'aberrant', 'increased', or 'unscheduled' R-loop formation truly entails. In a recent study, estrogen signaling in breast cells was shown to lead to a burst of R-loop formation at thousands of loci, including estrogen-responsive genes, and to R-loop- and replication-dependent DSBs [32]. Thus, upregulation of transcriptional programs can lead to elevated genic R-loop loads and fuel cancer-relevant instability and rearrangements. However, R-loop profiling in cells of patients with Aicardi-Goutières syndrome harboring partial RNASEH2 loss of function did not reveal any significant changes in genic R-loop distribution [21]. Instead, RNA:DNA hybrids were unexpectedly found to accumulate over late-replicating, DNA hypomethylated, intergenic regions characterized by facultative heterochromatin [21]. Thus, assumptions about increased genic R-loop levels in pathological situations may not necessarily be correct. Instead, more subtle alterations of R-loop turnover rates caused by mutations in specific factors or by sequence changes at specific R-loop forming loci may be sufficient to lead to persistent R-loops and threaten genome stability. One could even envision that a loss of R-loop formation either globally or locally could lead to detrimental consequences by preventing R-loops from participating in their physiological functions. For instance, R-loop loss could compromise transcription termination mechanisms, thus increasing transcriptional interference and

the likelihood of runaway transcription-replication encounters. Additional studies are required to carefully define the distribution, molecular nature, and turnover rates of the aberrant R-loops caused upon alteration of the many factors to which they have been tied.

Concluding Remarks

The field of R-loop biology has made rapid progress in recent years thanks in part to technological advances in the detection of endogenous R-loops. While we now understand that R-loops are more abundant than previously thought, many questions remain (see Outstanding Questions). Efforts should focus in particular on identifying the protein complexes that maintain R-loop homeostasis in normal cells and how breakdown of this equilibrium is linked to disease. These studies will undoubtedly provide new insights into the mechanisms by which nascent RNAs regulate important aspects of chromosome dynamics, such as transcription termination, chromatin patterning, replication licensing, and gene looping.

Acknowledgments

Research in my laboratory is supported by NIH grants GM094299, GM113929, and GM120607.

References

- Westover, K.D. *et al.* (2004) Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science* 303, 1014–1016
- Daniels, G.A. and Lieber, M.R. (1995) RNA:DNA complex formation upon transcription of immunoglobulin switch regions: implications for the mechanism and regulation of class switch recombination. *Nucleic Acids Res.* 23, 5006–5011
- Reaban, M.E. *et al.* (1994) Transcription induces the formation of a stable RNA:DNA hybrid in the immunoglobulin alpha switch region. *J. Biol. Chem.* 269, 21850–21857
- Roy, D. and Lieber, M.R. (2009) G clustering is important for the initiation of transcription-induced R-loops in vitro, whereas high G density without clustering is sufficient thereafter. *Mol. Cell Biol.* 29, 3124–3133
- Roy, D. *et al.* (2010) Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation in vitro: a nick can serve as a strong R-loop initiation site. *Mol. Cell Biol.* 30, 146–159
- Ginno, P.A. *et al.* (2012) R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell* 45, 814–825
- Ratmeyer, L. *et al.* (1994) Sequence specific thermodynamic and structural properties for DNA:RNA duplexes. *Biochemistry* 33, 5298–5304
- Roy, D. *et al.* (2008) Mechanism of R-loop formation at immunoglobulin class switch sequences. *Mol. Cell Biol.* 28, 50–60
- Drolet, M. *et al.* (1994) Hypernegative supercoiling of the DNA template during transcription elongation in vitro. *J. Biol. Chem.* 269, 2068–2074
- Phoenix, P. *et al.* (1997) Roles of DNA topoisomerases in the regulation of R-loop formation in vitro. *J. Biol. Chem.* 272, 1473–1479
- Belotserkovskii, B.P. *et al.* (2010) Mechanisms and implications of transcription blockage by guanine-rich DNA sequences. *Proc. Natl. Acad. Sci. U.S.A.* 107, 12816–12821
- Yu, K. *et al.* (2003) R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat. Immunol.* 4, 442–451
- Boguslawski, S.J. *et al.* (1986) Characterization of monoclonal antibody to DNA:RNA and its application to immunodetection of hybrids. *J. Immunol. Methods* 89, 123–130
- Phillips, D.D. *et al.* (2013) The sub-nanomolar binding of DNA:RNA hybrids by the single-chain Fv fragment of antibody S9.6. *J. Mol. Recognit.* 26, 376–381
- Huang, F.T. *et al.* (2007) Sequence dependence of chromosomal R-loops at the immunoglobulin heavy-chain S μ class switch region. *Mol. Cell Biol.* 27, 5921–5932
- Huang, F.T. *et al.* (2006) Downstream boundary of chromosomal R-loops at murine switch regions: implications for the mechanism of class switch recombination. *Proc. Natl. Acad. Sci. U.S.A.* 103, 5030–5035
- Kao, Y.P. *et al.* (2013) Detection and characterization of R-loops at the murine immunoglobulin S α region. *Mol. Immunol.* 54, 208–216
- Nadel, J. *et al.* (2015) RNA: DNA hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships. *Epigenetics Chromatin* 8, 46
- Sanz, L.A. *et al.* (2016) Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals. *Mol. Cell* 63, 167–178
- Ginno, P.A. *et al.* (2013) GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome Res.* 23, 1590–1600
- Lim, Y.W. *et al.* (2015) Genome-wide DNA hypomethylation and RNA:DNA hybrid accumulation in Aicardi-Goutieres syndrome. *Elife* 4
- Chen, P.B. *et al.* (2015) R loops regulate promoter-proximal chromatin architecture and cellular differentiation. *Nat. Struct. Mol. Biol.* 22, 999–1007
- Hartono, S.R. *et al.* (2015) GC skew is a conserved property of unmethylated CpG island promoters across vertebrates. *Nucleic Acids Res.* 43, 9729–9741
- Adelman, K. and Lis, J.T. (2012) Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* 13, 720–731
- Liu, C. *et al.* (2007) Proteomic patterns for classification of ovarian cancer and CTCL serum samples utilizing peak pairs indicative of post-translational modifications. *Proteomics* 7, 4045–4052
- Stirling, P.C. *et al.* (2012) R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. *Genes Dev.* 26, 163–175
- Gomez-Gonzalez, B. *et al.* (2011) Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. *EMBO J.* 30, 3106–3119
- Wahba, L. *et al.* (2011) RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Mol. Cell* 44, 978–988
- Skourti-Stathaki, K. *et al.* (2014) R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature* 516, 436–439
- Skourti-Stathaki, K. *et al.* (2011) Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell* 42, 794–805

Outstanding Questions

What enzymes and proteins complexes are involved in forming, stabilizing, sensing, and resolving R-loop structures? Our understanding of the metabolism of these structures is still limited. Filling this gap is critical to revealing the eventual cellular functions of R-loops.

What is the fate of the RNA transcripts involved in R-loop formation? Do they represent terminal intermediates that will be targeted for degradation or are they normal intermediates in the life of a transcript?

What are the distinguishing features of pathological and physiological R-loops? Methodologies to measure global R-loop distribution at high resolution, as well as dynamic R-loop turnover and individual R-loop footprints, will need to be leveraged to address this key question.

How do pathological R-loops trigger genomic instability? An abundant literature shows that conflicts between co-transcriptional R-loops and replication have a key role in driving instability. However, mechanistic insights into the process are still lacking, despite recent progress.

31. Proudfoot, N.J. (2016) Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science* 352, aad9926
32. Stork, C.T. *et al.* (2016) Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage. *Elife* 5, e17548
33. Noy, A. *et al.* (2005) Structure, recognition properties, and flexibility of the DNA:RNA hybrid. *J. Am. Chem. Soc.* 127, 4910–4920
34. Dunn, K. and Griffith, J.D. (1980) The presence of RNA in a double helix inhibits its interaction with histone protein. *Nucleic Acids Res.* 8, 555–566
35. Venkatesh, S. and Workman, J.L. (2015) Histone exchange, chromatin structure and the regulation of transcription. *Nat. Rev. Mol. Cell Biol.* 16, 178–189
36. Weber, C.M. and Henikoff, S. (2014) Histone variants: dynamic punctuation in transcription. *Genes Dev.* 28, 672–682
37. Powell, W.T. *et al.* (2013) R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation. *Proc. Natl. Acad. Sci. U.S.A.* 110, 13938–13943
38. Boque-Sastre, R. *et al.* (2015) Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. *Proc. Natl. Acad. Sci. U.S.A.* 112, 5785–6590
39. Herrera-Moyano, E. *et al.* (2014) The yeast and human FACT chromatin-reorganizing complexes solve R-loop-mediated transcription-replication conflicts. *Genes Dev.* 28, 735–748
40. Nickerson, J.A. *et al.* (1989) Chromatin architecture, nuclear RN. *A. Proc. Natl. Acad. Sci. U.S.A.* 86, 177–181
41. Caudron-Herger, M. and Rippe, K. (2012) Nuclear architecture by RNA. *Curr. Opin. Genet. Dev.* 22, 179–187
42. Felipe-Abrio, I. *et al.* (2015) RNA polymerase II contributes to preventing transcription-mediated replication fork stalls. *EMBO J.* 34, 236–250
43. Kellner, W.A. *et al.* (2015) GC skew defines distinct RNA polymerase pause sites in CpG island promoters. *Genome Res.* 25, 1600–1609
44. Wang, L. *et al.* (2009) S region sequence, RNA polymerase II, and histone modifications create chromatin accessibility during class switch recombination. *J. Exp. Med.* 206, 1817–1830
45. Pavri, R. *et al.* (2010) Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell* 143, 122–133
46. Balter, B.B. *et al.* (2012) Mice lacking Smu tandem repeats maintain RNA polymerase patterns but exhibit histone modification pattern shifts linked to class switch site locations. *Mol. Immunol.* 52, 1–8
47. Schotta, G. *et al.* (2008) A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. *Genes Dev.* 22, 2048–2061
48. Li, G. *et al.* (2013) Combinatorial H3K9acS10 ph histone modification in IgH locus S regions targets 14-3-3 adaptors and AID to specify antibody class-switch DNA recombination. *Cell Rep.* 5, 702–714
49. Yu, K. *et al.* (2005) Fine-structure analysis of activation-induced deaminase accessibility to class switch region R-loops. *Mol. Cell Biol.* 25, 1730–1736
50. Wang, I.X. *et al.* (2014) RNA-DNA differences are generated in human cells within seconds after RNA exits polymerase II. *Cell Rep.* 6, 906–915
51. Caminci, P. *et al.* (2005) The transcriptional landscape of the mammalian genome. *Science* 309, 1559–1563
52. Grzechnik, P. *et al.* (2014) Terminate and make a loop: regulation of transcriptional directionality. *Trends Biochem. Sci.* 39, 319–327
53. Strahl, B.D. *et al.* (2002) Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. *Mol. Cell Biol.* 22, 1298–1306
54. Kizer, K.O. *et al.* (2005) A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. *Mol. Cell Biol.* 25, 3305–3316
55. Shilatifard, A. (2012) The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annu. Rev. Biochem.* 81, 65–95
56. Krogan, N.J. *et al.* (2003) The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. *Mol. Cell* 11, 721–729
57. Thomson, J.P. *et al.* (2010) CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* 464, 1082–1086
58. Lee, J.H. and Skalknik, D.G. (2005) CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. *J. Biol. Chem.* 280, 41725–41731
59. Krajewski, W.A. *et al.* (2005) A motif within SET-domain proteins binds single-stranded nucleic acids and transcribed and supercoiled DNAs and can interfere with assembly of nucleosomes. *Mol. Cell Biol.* 25, 1891–1899
60. Tomson, B.N. and Arndt, K.M. (2013) The many roles of the conserved eukaryotic Paf1 complex in regulating transcription, histone modifications, and disease states. *Biochim. Biophys. Acta* 1829, 116–126
61. de Jong, R.N. *et al.* (2008) Structure and DNA binding of the human Rtf1 Plus3 domain. *Structure* 16, 149–159
62. Morales, J.C. *et al.* (2016) XRN2 Links transcription termination to DNA damage and replication stress. *PLoS Genet.* 12, e1006107
63. Rinn, J.L. and Chang, H.Y. (2012) Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* 81, 145–166
64. Huertas, P. and Aguilera, A. (2003) Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol. Cell* 12, 711–721
65. Castellano-Pozo, M. *et al.* (2013) R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. *Mol. Cell* 52, 583–590
66. Hsu, J.Y. *et al.* (2000) Mitotic phosphorylation of histone H3 is governed by Ipl1/aurora kinase and Glc7/PP1 phosphatase in budding yeast and nematodes. *Cell* 102, 279–291
67. Sharma, A.K. *et al.* (2015) Dynamic alteration in H3 serine 10 phosphorylation is G1-phase specific during ionizing radiation induced DNA damage response in human cells. *Mutat Res.* 773, 83–91
68. Ivaldi, M.S. *et al.* (2007) Phosphorylation of histone H3 at Ser10 facilitates RNA polymerase II release from promoter-proximal pausing in *Drosophila*. *Genes Dev.* 21, 2818–2831
69. Zippo, A. *et al.* (2009) Histone crosstalk between H3S10 ph and H4K16ac generates a histone code that mediates transcription elongation. *Cell* 138, 1122–1136
70. Cai, W. *et al.* (2014) Genome-wide analysis of regulation of gene expression and H3K9me2 distribution by JIL-1 kinase mediated histone H3S10 phosphorylation in *Drosophila*. *Nucleic Acids Res.* 42, 5456–5467
71. Groh, M. and Gromak, N. (2014) Out of Balance: R-loops in Human Disease. *PLoS Genet.* 10, e1004630
72. Groh, M. *et al.* (2014) R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. *PLoS Genet.* 10, e1004318
73. Colak, D. *et al.* (2014) Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. *Science* 343, 1002–1005
74. Loomis, E.W. *et al.* (2014) Transcription-associated R-loop formation across the human FMR1 CGG-repeat region. *PLoS Genet.* 10, e1004294
75. Sollier, J. and Cimprich, K.A. (2015) Breaking bad: R-loops and genome integrity. *Trends Cell Biol.* 25, 514–522
76. Ayrapetov, M.K. *et al.* (2014) DNA double-strand breaks promote methylation of histone H3 on lysine 9 and transient formation of repressive chromatin. *Proc. Natl. Acad. Sci. U.S.A.* 111, 9169–9174
77. Gursoy-Yuzugullu, O. *et al.* (2016) Patching broken DNA: nucleosome dynamics and the repair of DNA breaks. *J. Mol. Biol.* 428 (9 Pt B), 1846–1860

78. Hamperl, S. and Cimprich, K.A. (2014) The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability. *DNA Repair (Amst)* 19, 84–94
79. Aguilera, A. and Garcia-Muse, T. (2012) R loops: from transcription byproducts to threats to genome stability. *Mol. Cell* 46, 115–124
80. Santos-Pereira, J.M. and Aguilera, A. (2015) R loops: new modulators of genome dynamics and function. *Nat. Rev. Genet.* 16, 583–597
81. Costantino, L. and Koshland, D. (2015) The Yin and Yang of R-loop biology. *Curr. Opin. Cell Biol.* 34, 39–45
82. Chan, Y.A. *et al.* (2014) Mechanisms of genome instability induced by RNA-processing defects. *Trends Genet.* 30, 245–253
83. Masse, E. *et al.* (1997) DNA topoisomerases regulate R-loop formation during transcription of the *rrnB* operon in *Escherichia coli*. *J. Biol. Chem.* 272, 12816–12823
84. Cerritelli, S.M. and Crouch, R.J. (2009) Ribonuclease H: the enzymes in eukaryotes. *FEBS J.* 276, 1494–1505
85. Tuduri, S. *et al.* (2009) Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nat. Cell Biol.* 11, 131513–131524
86. Yang, Y. *et al.* (2014) Arginine methylation facilitates the recruitment of TOP3B to chromatin to prevent R loop accumulation. *Mol. Cell* 53, 484–497
87. Sollier, J. *et al.* (2014) Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol. Cell* 56, 777–785
88. Yuce, O. and West, S.C. (2013) Senataxin, defective in the neurodegenerative disorder ataxia with oculomotor apraxia 2, lies at the interface of transcription and the DNA damage response. *Mol. Cell Biol.* 33, 406–417
89. Mischo, H.E. *et al.* (2011) Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol. Cell* 41, 21–32
90. Richard, P. and Manley, J.L. (2016) R loops and links to human disease. *J. Mol. Biol.*
91. Salvi, J.S. and Mekhal, K. (2015) R-loops highlight the nucleus in ALS. *Nucleus* 6, 23–29
92. El Hage, A. *et al.* (2010) Loss of topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev.* 24, 1546–1558
93. Wahba, L. *et al.* (2016) S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes Dev.* 30, 1327–1338
94. Pfeiffer, V. *et al.* (2013) The THO complex component Thp2 counteracts telomeric R-loops and telomere shortening. *EMBO J.* 32, 2861–2871
95. Arora, R. *et al.* (2014) RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. *Nat. Commun.* 5, 5220
96. El Hage, A. *et al.* (2014) Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria. *PLoS Genet.* 10, e1004716
97. Brown, T.A. *et al.* (2008) Native R-loops persist throughout the mouse mitochondrial DNA genome. *J. Biol. Chem.* 283, 36743–36751
98. Xu, B. and Clayton, D.A. (1995) A persistent RNA-DNA hybrid is formed during transcription at a phylogenetically conserved mitochondrial DNA sequence. *Mol. Cell Biol.* 15, 580–589
99. Jiang, F. *et al.* (2016) Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* 351, 867–871
100. Szczelkun, M.D. *et al.* (2014) Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl. Acad. Sci. U.S.A.* 111, 9798–9803
101. Wahba, L. *et al.* (2013) The homologous recombination machinery modulates the formation of RNA-DNA hybrids and associated chromosome instability. *Elife* 2, e00505
102. Aronica, L. *et al.* (2016) The spliceosome-associated protein Nrl1 suppresses homologous recombination-dependent R-loop formation in fission yeast. *Nucleic Acids Res.* 44, 1703–1717
103. Kogoma, T. (1997) Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiol. Mol. Biol. Rev.* 61, 212–238
104. Kasahara, M. *et al.* (2000) RecA protein-dependent R-loop formation in vitro. *Genes Dev.* 14, 360–365
105. Zaitsev, E.N. and Kowalczykowski, S.C. (2000) A novel pairing process promoted by *Escherichia coli* RecA protein: inverse DNA and RNA strand exchange. *Genes Dev.* 14, 740–749
106. Cloutier, S.C. *et al.* (2016) Regulated formation of lncRNA-DNA hybrids enables faster transcriptional induction and environmental adaptation. *Mol. Cell* 61, 393–404
107. Sugiyama, T. *et al.* (1998) DNA annealing by RAD52 protein is stimulated by specific interaction with the complex of replication protein A and single-stranded DNA. *Proc. Natl. Acad. Sci. U.S.A.* 95, 6049–6054
108. Keskin, H. *et al.* (2014) Transcript-RNA-templated DNA recombination and repair. *Nature* 515, 436–439
109. West, J.A. *et al.* (2014) The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol. Cell* 55, 791–802