Cell

R-loops: the hidden face of transcription

--Manuscript Draft--

Manuscript Number:	CELL-D-19-01374R1
Full Title:	R-loops: the hidden face of transcription
Article Type:	Commissioned Review
Corresponding Author:	Andres Aguilera Universidad de Sevilla Sevilla, SPAIN
First Author:	Andres Aguilera
Order of Authors:	Andres Aguilera
	Tatiana Garcia-Muse
Abstract:	DNA-RNA hybrids play a physiological role in cellular processes, but often, they represent non-scheduled co-transcriptional structures with a negative impact on transcription, replication and DNA repair. Accumulating evidence suggests that they constitute a source of replication stress, DNA breaks and genome instability. Reciprocally, DNA breaks facilitate DNA-RNA hybrid formation by releasing the double helix torsional conformation. Cells avoid DNA-RNA accumulation by either preventing or removing hybrids directly or by DNA repair-coupled mechanisms. Given the R-loop impact on chromatin and genome organization and its potential relation with genetic diseases, we review R-loop homeostasis as well as their physiological and pathological roles.
Opposed Reviewers:	
Suggested Reviewers:	Marco Foiani
	Philippe Pasero
	Nicholas Proudfoot
	Frederic Chedin

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

R-loops: the hidden face of transcription

Tatiana García-Muse and Andrés Aguilera

Centro Andaluz de Biología Molecular y Medicina Regenerativa (CABIMER), Universidad de Sevilla-CSIC, 41092 SEVILLE, Spain

Correspondence to: aguilo@us.es

Abstract

DNA-RNA hybrids play a physiological role in cellular processes, but often, they represent non-scheduled co-transcriptional structures with a negative impact on transcription, replication and DNA repair. Accumulating evidence suggests that they constitute a source of replication stress, DNA breaks and genome instability. Reciprocally, DNA breaks facilitate DNA-RNA hybrid formation by releasing the double helix torsional conformation. Cells avoid DNA-RNA accumulation by either preventing or removing hybrids directly or by DNA repair-coupled mechanisms. Given the R-loop impact on chromatin and genome organization and its potential relation with genetic diseases, we review R-loop homeostasis as well as their physiological and pathological roles.

R-loops are non-B DNA structures that form during transcription when the nascent RNA anneals with the template DNA strand, thus displacing the non-template DNA strand. The term R-loop refers to the three-stranded structure formed by the DNA-RNA hybrid and the displaced single-stranded DNA (ssDNA), whereas DNA-RNA hybrid only refers to the double-stranded structure, even though in many reports both terms are used indistinctly. R-loops form generally behind elongating RNA polymerases as structures that can be longer than 1 kb (Garcia-Pichardo et al., 2017; Li and Manley, 2005; Yu et al., 2003). There are two types of R loops, physiological and pathological. The first ones normally rely on a programmed process requiring specific factors that warrant their formation, the second ones occur accidentally in a non-scheduled manner. DNA-RNA hybrid formation is specifically enhanced at certain regions where they play a physiological function. These include Immunoglobulin class switching recombination of B cells in vertebrates, mitochondrial DNA replication, CRISPR-Cas9 gene editing, specific regulatory steps in transcription initiation and termination, telomere homeostasis, and some cases of bacterial plasmid replication, among others (see Aguilera and Garcia-Muse, 2012; Crossley et al., 2019). However, hybrids have also been detected all over the genome of bacterial and eukaryotic cells. The frequency at which these occur at every transcribed unit has not been established yet, but it is likely too low to represent a physiological structure with a global functional role, but rather an accidental structure with a negative impact in the cell.

In general, R-loops may interfere with DNA replication, repair and transcription, thus compromising genome integrity and function. Therefore, cells have developed different mechanisms to either prevent or resolve such DNA-RNA hybrids. When any of these mechanisms fail, R-loops become a threat to genome integrity and cell proliferation, becoming a potential source of cellular pathologies. Apart from the fact that a ssDNA may be highly accessible to metabolites, reactive oxygen species (ROS), DNA modifying enzymes or nucleases that would increase the incidence of DNA damage (Lindahl, 1993) features that can apply to the displaced ssDNA of an R-loop, accumulated evidence suggests that R-loop-mediated replication fork (RF) stalling is a major feature of transcription-replication conflicts and R-loop-induced DNA damage (Boubakri et al., 2010; Tuduri et al., 2009; Wellinger et al., 2006). In order to understand the impact of R-loops in cell fate and proliferation, we need to answer questions such as what differentiates physiological from unscheduled pathological R-loops, how often unscheduled R-loops form along genomes, how cells protect themselves from pathological R-loops, or how R-loops can affect chromatin structure and compromise genome integrity. Here we review the relevant data that have

garnered attention for R-loops in the contexts of transcription, replication and repair, with the aim at providing a current view of the molecular mechanisms controlling R-loop homeostasis and how this might relate to disease.

Distribution, length and frequency of R-loops

Research on R-loop characteristics and distribution along the genome depends on the tools available to detect them *in vivo*. Different methodologies have been developed to detect DNA-RNA hybrids *in vivo* that include the isolation and physical analysis of nucleic acids resistant to RNase A and sensitive to RNase H (Huertas and Aguilera, 2003); electron microscopy (EM) (Backert, 2002) or chromatin immunoprecipitation (ChIP) and/or immunofluorescence (IF) using an inactivated RNase H (Chen et al., 2017; Ginno et al., 2012) and the RNase H1 hybrid-binding domain (HB) fused to the green fluorescent protein (GFP) (Bhatia et al., 2014) (**Figure 1A**). Nevertheless, the most commonly used methods rely on the S9.6 anti-10 nucleotide DNA-RNA hybrid monoclonal antibody (Boguslawski et al., 1986) (**Figure 1A**). This antibody has been extensively used for DNA-RNA hybrid immunoprecipitation (DRIP) followed by either qPCR at specific DNA regions or by sequencing for genome-wide studies, as well as for IF. These analyses have enabled a relatively precise idea on the distribution of hybrids along the different regions of eukaryotic genomes, mainly from *Saccahromyces cerevisae* and human cells (El Hage et al., 2010; Ginno et al., 2012; Wahba et al., 2011; Sanz et al., 2016). In both DRIP and IF assays, treatment with RNase H1, which specifically degrades the RNA moiety of the hybrid, is required, since the S9.6 antibody is also able to recognize double-stranded RNAs (dsRNAs) present at high concentration at specific sites (Hartono et al., 2018; Silva et al., 2018).

In general, the initial genome-wide DRIP analyses have been confirmed and extended by other approaches. These include using strand specific high-throughput sequencing of either DNA libraries, marked with dUTP or with adaptors (Nadel et al., 2015; Xu et al., 2017), or cDNA libraries obtained by RT-PCR from the RNA moiety of hybrids after DNase I treatment (Sanz et al., 2016). Genome-wide analysis have unveiled R-loops to be present in normal/wild-type cells at more DNA regions than anticipated. Indeed, genome coverage data oscillates from 8% in yeast to 10% in Arabidopsis or 5% in human cells (Wahba et al., 2016; Xu et al., 2017; Sanz et al., 2016). Altogether the data indicate that DNA-RNA hybrids accumulate preferentially at highly transcribed genes, including the rRNA and tRNA loci, as well as in some transposable elements (Ty in yeast), centromeres and telomeres, antisense-RNAs or ncRNAs regions (Chan et al., 2014; Chen et al., 2017; El Hage et al., 2014; Ginno et al., 2012; Wahba et al., 2016). In yeast the R-loop-prone open reading frames (ORFs)

are generally highly transcribed and with high GC content, but the proportion of ORFs observed fluctuates from 20% to more that 65% depending on the study (Wahba et al., 2016; El Hage et al., 2014; Chan et al., 2014). This variability may be due to technical differences in the DNA-RNA immunoprecipitation protocols (see Halasz et al, 2017). In the case of mammalian cells R-loops are also detected mainly at active genes and they also accumulate at promoter and terminator regions in which they play a role in gene expression regulation (Ginno et al, 2012; Sanz et al, 2016). In addition, these studies have permitted the identification of sequences prone to DNA-RNA hybridization such as GC-rich sequences, CpG islands, transcription termination sites with high GC-skew or G-quadruplex (G4)-containing sequences (Ginno et al., 2012; Sanz et al., 2016; Chen et al., 2017; Stork et al., 2016). Despite a general agreement about such features, certain regions were observed using specific protocols. For example, R-loop-prone regions are preferentially AT-rich when detected by DRIP analysis combined with S1 nuclease digestion in yeast (Wahba et al., 2016). However, in this case it would be important to exclude the possibility that S1-induced nicks do not favor DNA-RNA annealing during the DRIP protocol.

In contrast to DNA-RNA hybrids, R-loop detection relies on the identification of the displaced ssDNA. Several reports have shown that the ssDNA-binding protein RPA binds the displaced ssDNA (Chaudhuri et al., 2004; Nguyen et al., 2018), suggesting that one could use anti-RPA antibodies as a way to detect R-loops. However, since RPA binds to ssDNA generated at resected double-strand breaks (DSBs) and at the lagging strand of RFs, it is unlikely that an RPA signal can be unambiguously attributed to R-loops. Indirectly, *in vivo* R-loops can be inferred with high precision by the mutation profile generated by sodium bisulfite in single DNA molecules purified from cell cultures (Yu et al., 2003) or by the *in vivo* action of the human activation-induced cytidine deaminase (AID) (Gomez-Gonzalez and Aguilera, 2007) (**Figure 1A**). Whereas both have proved to be reliable to detect R-loops *in vivo* (Garcia-Pichardo et al., 2017) only the bisulfite mutagenesis can provide a measurement of the length of DNA-RNA hybrids, since the *in vitro* treatment by bisulfite induces deamination of all cytidines within DNA whereas *in vivo* AID-induced mutation is not as processive. With this method, the mutation profile of the non-template ssDNA region allows the inference of the average length of R-loops, provided that such a profile is abolished by RNase H1 treatment. Some estimates have been obtained in human cells and yeast reporting sizes ranging from less than 100 bp up to 2 kb (Garcia-Pichardo et al., 2017; Li and Manley, 2005; Yu et al., 2003). Comparison of the length of the DNA regions identified as R-loop-prone in the available DRIP-seq experiments shows two different behaviors, a narrow-sized distribution (180-2350 bp) or a wide-sized distribution (180-

22500 bp) depending on the methodology used to break the DNA preparation prior to immunoprecipitation. Sonication yields shorter R-loop-prone DNA segments, while restriction enzymes yields long segments likely due to the more spaced distribution of the DNA restriction sites (Halasz et al., 2017).

Importantly, with the described methodology we are yet unable to determine the frequency at which R-loops originate, so we do not have a measure of how often R-loops occur in cells or at specific DNA regions. However, we can affirm that R-loop occurs co-transcriptionally over conserved genomic regions, presumably facing a dynamic turnover. Thus, to understand the physiological meaning and causes of R-loop distribution we must consider the different transcription profiles, the signals and environmental conditions of growth, the cell cycle stage or the specific epigenetic features of each cell or DNA region among other features.

DNA-RNA hybrids with a physiological role

R-loops are intermediates required for several processes regulating genome dynamics such as immunoglobulin (Ig) class-switch recombination (CSR) (Yu et al., 2003) or CRISPR-Cas9 activity, in which the guide-RNA forms a DNA-RNA hybrid as an intermediate to identify the target for Cas9-mediated cleavage (Jinek et al., 2012), or the initiation of DNA replication in mitochondrial DNA, bacterial plasmids, and the bacteriophages ColE1 and T4 (see Aguilera and Garcia-Muse, 2012) (**Figure 1B**). Also, increasing evidence indicates that R-loops play roles in gene expression, as suggested by the fact that such structures are enriched in gene promoters and termination regions. DNA-RNA hybrids have been linked to the unmethylated state of CpG islands, present at promoters of many genes in mammalian cells. Computational and experimental data suggest that R-loops formed at CpG islands protect these regions against DNA methylation, ultimately avoiding transcription silencing (Ginno et al., 2012; Grunseich et al., 2018). This protection could be explained by the observation that DNA methyl-transferase 1 preferentially binds and methylates double-stranded DNA over the DNA-RNA hybrid (Grunseich et al., 2018) (**Figure 1C**).

In addition, R-loops can guide the binding of transcription regulation factors, as described for the regulation driven by long non-coding RNAs (lncRNAs) (**Figure 1C**). Thus, in *Arabidopsis* the expression of a set of lncRNAs called COOLAIR leads to silencing of the *FLC* gene. The homeodomain protein AtNDX, inhibits the expression of COOLAIR by binding to the ssDNA of an R-loop formed at its promoter ultimately allowing transcription of the *FLC* gene (Sun et al., 2013). Interestingly, yeast cells also respond to environmental changes with the formation of R-loops by lncRNAs, but in this case to activate expression (Cloutier et al., 2016). In human cells, the vimentin

(*VIM*) gene involved in cell and tissue integrity is upregulated in different types of cancer due to antisense transcription of a lncRNA, *VIM-AS1*, that forms an R-loop at the VIM promoter region and transcription start site (TSS) favoring the recruitment of transcription factors. Depletion of the *VIM-AS1* or R-loop disruption by RNase H overexpression inactivates *VIM* expression (Boque-Sastre et al., 2015). Similarly, the antisense lncRNA TARID also forms an R-loop that regulates gene expression. In this case, TARID is located at the promoter region of the tumor suppressor gene *TCF21*, and the stress response protein GADD45A binds to this R-loop and recruits methylcytosine dioxygenase TET1 leading to local demethylation and gene transcription (Arab et al., 2019).

In addition to localization at promoters in human genes, R-loops are enriched over G-rich terminator elements. These would facilitate RNA Pol II pausing before efficient termination (**Figure 1C**). The SETX helicase could act on these R-loops positioned downstream of gene poly(A) signals (Skourti-Stathaki et al., 2011), where SETX seems to act together with the Tudor-domain protein SMN, which is recruited to the di-methylated CTD of RNAPII (Zhao et al., 2016). Removal of these R-loops allows the access of the Xrn2 exoribonuclease and termination factors (Morales et al., 2016; Skourti-Stathaki et al., 2011). It is unclear how general this mechanism of transcription termination is since this system has been shown for few transcribed loci. Interestingly, R-loops at G-rich terminator elements also induce antisense transcription, which in turn leads to the generation of double-stranded RNA and the recruitment of the silencing machinery including DICER, AGO1, and AGO2 as shown by ChIP (Skourti-Stathaki et al., 2014). In *Schizosaccharomyces pombe*, Dicer favors the release of RNAPII at termination sites previously identified as replication stress hotspots. Loss of Dicer seems to cause R-loop accumulation at such sites (Castel et al., 2014), but this needs confirmation. These results fit with genome-wide analyses in mammals showing a correlation between R-loop location and transcription termination sites (Sanz et al., 2016). Therefore, R-loops may contribute to transcription termination in some cases.

The existence of R-loops with a physiological role, even though limited to specific cases, raises the question of whether they occur spontaneously or if their formation is a regulated process mediated by the action of specific protein factors. Indeed, several proteins exhibit RNA binding activity and the capability to hybridize it with the complementary strand of the DNA duplex. *In vitro,* the mammalian capping enzyme bound to phosphorylated RNAPII is able to promote formation of R-loops in transcription assays (Kaneko et al., 2007), and the virus-encoded ssDNA-binding protein ICP8 promotes invasion of an RNA molecule into a homologous acceptor plasmid to generate an R-loop (Boehmer, 2004). Perhaps the best example for the

need of a regulated process of DNA-RNA hybridization *in vivo* is provided by CRISPR-Cas9 (Jinek et al., 2012).

The need for the regulation of DNA-RNA hybrids with a functional role is also exemplified by the existence of specific protein factors that favor DNA-RNA hybrid accumulation. Thus, the DDX1 helicase disentangles the G4 structures found in the IgH S-region transcript allowing the RNA to hybridize with the DNA and thus providing the ssDNA substrate needed for AID action and thereafter class switch recombination (Ribeiro de Almeida et al., 2018). Similarly, the DHX9 RNA helicase promotes R-loop accumulation in spliceosome-defective conditions (Chakraborty et al., 2018). Given that knockdown of DHX9 suppressed the accumulation of physiological R-loops at centromeric regions, DHX9 was suggested to be involved in the formation of these structures even in normal conditions, probably through unwinding RNA secondary structures (Chakraborty et al., 2018). Other protein activities that can promote R-loop accumulation may rely on sequestering the non-template ssDNA via ssDNA-binding proteins. Thus, a recent observation has shown that mitochondrial ssDNA-binding proteins stabilize the formation of R-loops (Posse et al., 2019).

In summary, the prevailing experimental evidence indicates that DNA-RNA hybrids can provide a specific mechanism for the control of gene expression likely promoted by specific protein machineries, which in most cases need yet to be deciphered. This contrasts with spontaneous and non-scheduled R-loops for which cells have developed machineries to limit their formation and harmful consequences. The mechanisms to prevent the deleterious effects on unscheduled R-loops are three-fold: i) prevention, ii) resolution, and iii) DNA replication/repair-coupled removal. A discussion of these activities is provided in the next three sections.

Factors preventing non-scheduled spontaneous R-loop accumulation

From the original observations in yeast THO-complex mutants (Huertas and Aguilera, 2003), supported by subsequent reports implicating other RNA processing/export factors in DNA-RNA hybrid prevention (Li and Manley, 2005) (**Table 1**) it was proposed that in eukaryotic cells, DNA-RNA hybrids are prevented by the coating of the nascent RNA molecule with proteins involved in processing and export (**Figure 2A**). The accumulation of non-scheduled DNA-RNA hybrids in mutant strains is associated with an increase in genome instability as determined by either hyper-mutation, hyper-recombination, gross-chromosomal rearrangements or different forms of replication stress (Aguilera and Garcia-Muse, 2012). The protective role against hybrid accumulation is not a property of all RNA binding and processing factors, but only of a

subset of factors that, presumably, function in the assembly of mRNA-protein particles during transcription elongation.

A second key factor conditioning the ability of the nascent RNA to hybridize back with the DNA template is DNA topology (**Figure 2A**). Thus, the negative supercoiling transiently generated behind the transcribing RNA polymerase makes the double-stranded DNA prone to open, favoring DNA-RNA hybrid formation as observed *in vitro* and *in vivo* in topoisomerase I-depleted bacteria, yeast and human cells (Drolet et al., 1995; El Hage et al., 2010; Tuduri et al., 2009). In agreement, most DNA-RNA hybrids occur co-transcriptionally.

Transcription is an obligatory prerequisite for R-loop formation and, in addition, R-loop levels correlate with high transcriptional levels in normal cells (Chan et al., 2014; Stork et al., 2016; Wahba et al., 2016). However, high transcription levels do not ensure DNA-RNA hybrid formation in all cases. Several R-loop accumulating mutants show severe transcriptional defects (Hraiky et al., 2000; Huertas and Aguilera, 2003), some of which might be exacerbated by the DNA-RNA hybrid itself. Indeed, DNA-RNA hybrids negatively affect transcription elongation (Tous and Aguilera, 2007) and R-loop induction correlates with transcriptional pausing (Chen et al., 2017). In addition, accumulation of RNA in the nucleus caused by deficient RNA export is not sufficient to induce DNA-RNA hybrid accumulation (Garcia-Rubio et al., 2018). Therefore, even though high transcription increases the chances of hybrid formation, R-loop accumulation is not merely a consequence of high transcription.

Emerging data also supports that chromatin is one additional factor protecting the genome from R-loop accumulation (**Figure 2A**). Experimental evidence that chromatin controls R-loop homeostasis was provided in the yeast *S. cerevisiae* in which a screening of a non-lethal histone H3 and H4 mutant library identified a number of specific histone H3 and H4 mutations that strongly increased R-loop accumulation (Garcia-Pichardo et al., 2017). Comparison of R-loop maps in mammalian genomes with maps of DNAase I hypersensitivity revealed that R-loop positive regions correlate with “open chromatin” better than R-loop negative regions (Chen et al., 2017; Sanz et al., 2016). In agreement with a more accessible DNA favoring DNA-RNA hybridization, hybrids accumulate in human cells treated with histone deacetylase inhibitors or depleted of the SIN3A histone deacetylase (HDAC) complex (Salas-Armenteros et al., 2017) as well as in yeast *sin3*Δ mutants (Wahba et al., 2011). The specific role of the SIN3A histone deacetylase versus other deacetylases remains to be seen but it is interesting that SIN3A interacts with the RNA processing/export factor THO complex, suggesting that THO prevents R-loops not only by facilitating the formation of an optimal mRNP, but also by promoting local co-transcriptional histone deacetylation

(Salas-Armenteros et al., 2017). This result opens the possibility that there is a safeguard co-transcriptional mechanism by which specific proteins like the THO complex bind to the nascent RNA as it emerges from RNAP and talk to Sin3A to transiently deacetylate histones thus closing chromatin. On the contrary, the less accessible heterochromatin emerges as an R-loop opponent. In this regard, heterochromatic marks such as histone H3 lysine 9 di- and tri-methylations (H3K9me2/3) have been related to a low incidence of DNA-RNA hybrids in *C. elegans* (Zeller et al., 2016). Moreover, increased R-loop levels have been reported in *Drosophila* cells depleted of histone H1, which show defective heterochromatin (Bayona-Feliu et al., 2017), in mouse cells with de-condensed chromatin induced by topoisomerase inhibitors (Powell et al., 2013), or in *S. pombe* cells with a deletion of the SNF2-like nucleosome-remodeling factor Fft3 that alters the maintenance of heterochromatin and confers major histone turnover (Taneja et al., 2017). It is important to differentiate this protecting role of chromatin against R-loops from the association between R-loops and specific chromatin markers seen in *S. cerevisiae*, *S. pombe*, *C. elegans* and human cells. Indeed, different studies indicate that R-loops can indirectly affect chromatin by regulating the binding of chromatin-regulatory complexes (Arab et al., 2019; Chen et al., 2015). Thus, the way R-loops and chromatin structure affect or relate to each other will depend on their physiological or pathological impact.

Factors resolving R-loops

Despite all the direct mechanisms that prevent the co-transcriptional formation of R-loops, they sometimes fail to prevent R-loops albeit with low frequency. Consequently, cells have developed backup mechanisms to resolve R-loops and avoid their accumulation and potential threat to transcription and genome integrity. Probably the most relevant and well-known factor with this function is RNase H, a conserved ribonuclease from bacteria to humans that specifically degrades the RNA moiety of the RNA:DNA hybrids, as mentioned before (**Figure 2B**). In principle both types of RNase H, RNase H1 and RNase H2, are able to resolve hybrids, the latter being also involved in Ribonucleotide Excision Repair (RER). However, the key role of RNase H is to remove the RNA primer of the DNA-RNA hybrid in Okazaki fragments. This is a highly efficient back-up mechanism to resolve R-loops, which is supported by the observations that bacteria, yeast and human cells depleted of or inactivated for RNase H1 accumulate R-loops. Moreover, RNase H1 overexpression efficiently suppresses specific phenotypes associated with R-loops or eliminates *in vivo* the DNA-RNA hybrid signal detected by the S9.6 antibody (Drolet et al., 1995; Huertas and Aguilera, 2003; Li and Manley, 2005; Wahba et al., 2011). However, overexpression of RNase H is an

artificial condition and it cannot be excluded that an excess of RNH could act on certain hybrids that are not necessarily its main natural target. Given that non-scheduled R-loops may be formed during transcription, it seems very costly to resolve them by degrading the nascent RNA, in particular for human genes that can reach lengths beyond 600 kb.

In recent years numerous reports have identified an increasing number of RNA-dependent ATPases that in some cases have been shown to have *in vitro* DNA-RNA unwinding activity and when depleted from cells lead to an accumulation of DNA-RNA hybrids. These include human SETX and FANCM and their yeast homologs Sen1 and Mph1, AQR, DDX19, yeast Pif1, DDX23, DDX1, Dbp2 (human DDX5), Sgs1 (human BLM) and others (Chang et al., 2017; Cristini et al., 2018; Hodroj et al., 2017; Lafuente-Barquero et al., 2017; Li et al., 2016a; Mischo et al., 2011; Ribeiro de Almeida et al., 2018; Skourti-Stathaki et al., 2011; Sollier et al., 2014; Sridhara et al., 2017; Schwab et al., 2015; Tedeschi et al., 2018; Tran et al., 2017; Wang et al., 2018b) (**Table 1**). These results suggest that DNA-RNA helicases constitute a second type of factors to remove R-loops (**Figure 2B**). SETX has been implicated in the removal of R-loops located at terminators (Skourti-Stathaki et al., 2011). Knockdown of AQR, a DEAxQ-like putative RNA/DNA helicase, leads to R-loop accumulation as well as R-loop-dependent DNA damage (Sollier et al., 2014). DDX19 is a nucleopore-associated mRNA export factor that is able to unwind DNA-RNA hybrids *in vitro*. Based on the observation that proliferating cells lacking DDX19 show an increase in R-loops and DSBs it has been proposed that DDX19 acts on R-loops formed upon replication stress or DNA damage (Hodroj et al., 2017). In addition, it has been shown that pausing of RNAPII at R-loop accumulating sites initiates a signaling cascade that ends in the phosphorylation of the DDX23 RNA helicase, its recruitment to pause sites, and the dissolution of R-loops (Sridhara et al., 2017). As in other cases such us Pif1, BLM and Sgs1, DDX9 or FANCM an *in vitro* DNA-RNA unwinding activity was also shown. Other helicases, such as RECQL5 seems to act differently. RECQL5 prevents transcription-associated genome instability (Saponaro et al., 2014). R-loops accumulate in RECQL5-depleted cells and the expression of RNase H reduced the number of DNA breaks. However, the helicase domain, but not the helicase activity, is required for R-loop suppression and it has been proposed that RECQL5 promotes TOP1 activity (Li et al, 2015).

The role of RNA helicases in resolving R-loops is conceptually appealing because, in contrast to RNases H, helicases would resolve the hybrid without the costly action of degrading the nascent RNA. However, most of these enzymes have not been shown to specifically unwind DNA-RNA hybrids better than dsRNAs. Most importantly, it has not been determined whether all these helicases unwind hybrids *in*

vivo by, for example, overexpressing them and showing a reduction of hybrids and of any R-loop-dependent phenotype. Why cells would need so many different non-redundant DNA-RNA helicases? One possible explanation is that each helicase acts on a specific subset of DNA-RNA hybrids. Alternatively, most of these proteins could act *in vivo* not as DNA-RNA helicases but as RNA chaperones that prevent R-loop formation by promoting an optimal mRNP incapable of hybridizing with its DNA template, as is the case for THO.

Interestingly, two recent proteomic analyses of DNA-RNA hybrid pull-downs have reported more than 350 factors enriched in the DNA-RNA hybrid-binding fraction. Among these, there were more than 25 DEAD-box helicases including those discussed above (Cristini et al., 2018; Wang et al., 2018b). Both studies identified factors with a role in R-loop metabolism *in vivo*, but since the degree of coincidence between both studies is scarce and RNase H was not detected in the S9.6 pull down (Cristini et al., 2018) further validation analyses are still required. Indeed, there are several limitations to making further conclusions on the role of these proteins. For instance, one pull-down was performed on linear DNA-RNA substrates containing free ends that might stimulate DNA-RNA hybrids; thus, it is possible that at least a number of the identified proteins binds to such ends, as it is the case of ligase LIG3. In addition, many of the proteins identified were not related to RNA metabolism, such as DNA helicase MCM5, histone methyltransferase NSD2, nuclear matrix protein MATR3 or myosin MYO1, making their roles, if real, in DNA-RNA hybrid metabolism difficult to interpret. In any case, these hybrid interactomes provide an appealing tool to explore R-loop biology.

Removal of R-loops by DNA replication and repair-associated mechanisms

Persistent R-loops can constitute a road-block for transcription, which has been proposed to explain the negative impact on transcription of mutations in bacterial *rho* or the yeast THO complex (Hraiky et al., 2000; Huertas and Aguilera, 2003). Thus, mechanisms must exist to remove such roadblocks. Since RNA polymerases stacked in front of a DNA damage site (such as a cyclo-pyrimidine-dimer (CPD)) trigger the process of transcription-coupled repair (TCR), one possible mechanism to resolve R-loop-transcription conflicts is that a TCR-like mechanism could facilitate the removal of the R-loop to allow transcription resumption. However, so far there has not been any report indicating the existence of such kind of functions or TCR-like mechanism.

R-loops are also obstacles to RF progression, which is the main cause of R-loop-associated genome instability in S-G2 phase of cycling cells (Crossley et al., 2019; Gómez-González and Aguilera, 2019). Stalled RFs need to be protected to avoid their collapse and the resulting DNA breaks. As a consequence, different repair factors

and pathways function to allow RF restart. One mechanism relies on members of the Fanconi Anemia (FA) pathway, composed of a complex array of factors that have been mainly studied in the context of inter-strand crosslinks (ICLs) repair but that is active on any type of DNA lesion that blocks replication. Notably, cells depleted of FA factors such as FANCD2, FANCA and FANCM and BRCA2 among others, accumulate R-loops and R-loop-dependent DNA breaks (Bhatia et al., 2014; Garcia-Rubio et al., 2015; Hatchi et al., 2015; Liang et al., 2019; Madireddy et al., 2016; Okamoto et al., 2019; Schwab et al., 2015). Importantly, inhibition of transcription or RNase H1 overexpression, both diminishing R-loop accumulation, suppresses RF arrest and DNA damage in FANCD2-depleted cells (Garcia-Rubio et al., 2015; Schwab et al., 2015). FANCD2 foci formation, which is the first step in the activation of this repair pathway, is largely R-loop-dependent even in the absence of DNA damage. *in vitro* assays have shown the FANCI-FANCD2 complex binds the displaced ssDNA and ssRNA tail and this substrate stimulates FANCD2 activation via monoubiquitination, supporting the view that the FA pathway functions at R-loop containing sites (Garcia-Rubio et al., 2015; Liang et al., 2019). It was proposed that FANCM, which belongs to the DEAD family of DNA-RNA helicases, could remove R-loops, since it unwinds DNA-RNA molecules *in vitro* (Schwab et al., 2015). However, it has also been recently shown that FANCD2 interacts with RNA processing factors including DEAD-box RNA helicase DDX47 that could facilitate R-loop removal (Okamoto et al., 2019). How each of these proteins operates to facilitate removal of R-loops is unclear.

We have yet to elucidate the mechanisms by which the BRCA factors and the FA pathway dissolve R-loops that cause DNA damage. In principle, the FA pathway may cleave and eliminate the hybridized DNA strand, as it does with ICLs. In the case of the FA and DSB repair factor BRCA2, it has been proposed to help remove hybrids by acting directly on the R-loop provided that this partially resembles a blocked RF (Bhatia et al., 2014). Interestingly, BRCA2 binds to DSS1 and PCID2, whose yeast orthologs belong to the TREX-2/THSC complex involved in mRNP biogenesis and export, which might suggest a role not necessarily associated with replication. It is possible that BRCA2 also functions to resolve R-loops or to deal with the lesions originated by them (Bhatia et al., 2014). Additionally, BRCA1, another FA factor, is enriched at a subset of transcription termination regions, which accumulate R-loops, and mediates the recruitment of helicase SETX (Hatchi et al., 2015). The mechanisms by which BRCA2 and BRCA1 facilitate R-loop removal, therefore, are not necessarily related to each other. Given their roles as FA and DSB repair factors, and in protecting stalled forks from collapse, it is also possible that they act on hybrids accumulated at DSBs (see below). In this capacity, by facilitating DSB repair they would help remove

the hybrid as well (see below). Indeed, the importance of DNA-RNA hybrids as a background spontaneous source of replication stress and genome instability that can be indirectly eliminated via DNA repair is supported by the observation that silencing of DNA damage response (DDR) and DNA repair genes (ATR, ATM, CHK1, CHK2, UBE1, etc) cause DNA-RNA hybrid accumulation (Barroso et al., 2019). It would not be surprising that many genome instability conditions, whether or not caused by failures in DDR or repair, were associated with the accumulation of DNA-RNA hybrids that preceded spontaneous DNA breaks.

In conclusion, despite our poor knowledge on how DNA replication and repair functions help to remove R-loops and prevent R-loop-mediated genome instability, it seems that in addition to cellular functions directly involved in R-loop prevention (RNA biogenesis factors, chromatin) and resolution (RNAse H, DNA-RNA helicases), specific DNA repair pathways might constitute a back-up mechanism to dissolve R-loops accumulated in S-G2 at either RF stalls or, as we discuss below, DSBs. Whether this is in concert with other proteins such as DDX47, SETX, PCID2 or others, as it has been proposed for FANCD2, BRCA1, BRCA2 and other DNA repair factors is yet unknown (**Table 1**).

R-loops as a source of replication stress and genome instability

Although regulated R-loops are relevant in several cellular processes, unscheduled R-loops give rise to DNA damage and ultimately genome instability. This has been first shown in mutations of genes involved in the biogenesis and export of mRNP and in pre-mRNA splicing, which exhibit increased R-loops that correlate with increased DNA damage and transcription-associated recombination (Huertas and Aguilera, 2003; Li and Manley, 2005; Paulsen et al, 2009). One reason behind this may rely on the ssDNA fiber from the R-loop, which would be more susceptible to the action of nucleases and genotoxics (**Figure 3A**). Nevertheless, the most relevant mechanism by which an R-loop causes genome instability in cycling S-G2 cells is its capacity to stall RF progression and potentially cause fork breakage (**Figure 3B**). This is strongly supported by many different studies showing replication impairment through R-loop-enriched regions or in bacteria and yeast R-loop accumulating mutants (Gan et al., 2011; Wellinger et al., 2006; Gómez-González et al., 2011) or by the asymmetry increase in RFs as detected by DNA combing in human cells (Salas-Armenteros et al., 2017; Tuduri et al., 2009).

The difficulties of replication progression through transcribed chromatin are a long-known source of genome instability. It is possible that R-loops themselves or in association with a paused RNA polymerase contribute to the RF blockage. Strong

support for this model is provided by the observation that cells depleted of FA factors also accumulate R-loops, as discussed. Or, that depletion of the FACT chromatin-remodeling complex causes accumulation of R-loops and R-loop-dependent DNA damage. FACT exchanges nucleosomes around the RNA polymerase during transcription elongation and is important during RF progression to avoid R-loop-mediated transcription–replication conflicts (Herrera-Moyano et al., 2014). Also, analysis of separation-of-function □*C*□□allele, which cannot activate the checkpoint under replication stress but accomplishes normal replication, showed that yeast cells carrying such checkpoint defective allele of □*C*□□also accumulates DNA-RNA hybrids that correlate with an increase in damage, suggesting a role for the MCM replicative helicase in preventing transcription-replication conflicts (Vijayraghavan et al., 2016). Finally, PrimPol activity has been described to re-establish replication at sites where R-loops have impaired replication (Svikovic et al., 2019). We do not know exactly how or at which frequency the conflict occurs in normal cells, but certainly it can be highly stimulated in cells lacking R-loop-protecting factors (Dominguez-Sanchez et al., 2011; Stirling et al., 2012; Tuduri et al., 2009).

The conflicts between the transcription and replication can occur in two modes since the RNA polymerase either moves head-on or co-directionally with respect to the movement of the replication fork and both have different outcomes. Two reports in bacteria and yeast have proposed that R-loops are generated as a result of the transcription-replication conflicts (Hamperl et al., 2017; Lang et al., 2017), provided that they are mainly observed in systems in which conflicts occur head-on *vers*□*s* co-directional. The fact that head-on but not co-directional conflicts are responsible for transcription-mediated instability was shown from yeast and bacteria (Prado and Aguilera 2005; Mirkin and Mirkin, 2005), and recent results in yeast have shown that R-loop formation is independent of the orientation of the collision (Garcia-Rubio et al., 2018). However, only those present in head-on collision cause stalling and damage, whereas those in co-directional orientation do not. If a DNA-RNA binding protein stabilizes such R-loops, these would constitute a block to replication regardless of the collision orientation (**Figure 3B**). It is likely that the RF is able to dissolve the R-loop without undergoing a permanent stalling, unless the R-loop is stabilized or blocked, as in yeast with excess of Yra1 (Garcia-Rubio et al., 2018), or in human cells expressing an DNA-RNA hybrid domain-GFP (HB-GFP) fusion protein (Bhatia et al., 2014).

Recent genome-wide analysis in yeast has demonstrated that persistent R-loops are behind gross chromosomal rearrangements (Costantino and Koshland, 2018). This finding may be particularly relevant in the context of fragile sites that are genomic regions, often consisting of repetitive DNA sequences, which are difficult to

replicate and undergo DNA breaks under conditions of replication stress. Supporting that R-loops may be one factor influencing transcription-replication conflicts as a source of fragility, several groups have reported accumulation of R-loops at fragile sites (Groh et al., 2014; Helmrich et al., 2011). In agreement, it is known that transcription stimulates the genetic instability of trinucleotide repeat sequences and several studies in bacteria and eukaryotes have shown that persistent R-loops at trinucleotide repeats promote such instability (Grabczyk et al, 2007; Kim and Jinks-Robertson, 2011; Reddy et al, 2011). Thus, *in vitro* transcribed CG-rich repeating sequences form RNase A-resistant and RNase H-sensitive structures, supporting that they form R-loops since the RNase A degrades double stranded RNA while the RNase H removes the RNA from a DNA-RNA hybrid. Furthermore, instability of CTG·CAG repeats is stimulated by the reduced activity of RNase H1 in *E. coli* and by siRNA knockdown of RNase H1 and H2 in human cells (Lin et al., 2010).

Even though not all CFSs are explained by R-loops, in agreement with transcription influencing fragility, it has been shown that fragile sites are cell type-specific, as some of the sites are fragile in lymphocytes but not in fibroblasts and *vice versa* and, in each cell line, fragile sites correlate with chromatin status and gene expression (Letessier et al., 2011). Importantly, cells lacking proteins associated with the FA pathway are prone to chromosome breaks at common fragile sites (CFSs) (Madireddy et al., 2016; Wang et al., 2018a). In this case, analysis of FANCD2-/- lymphoblast cells revealed that DNA-RNA hybrids accumulate at the CFS-FRA16D. Removal of R-loops by overexpressing RNase H suppresses the replication perturbations observed at this CFS as shown by single-molecule analysis of replicated DNA (SMARD) (Madireddy et al., 2016). This can be extended to other FA factors. Thus, analysis with the AT-rich sequence Flex1 derived from CFS-FRA16D, which induces HR-mediated mitotic recombination, showed that FANCM together with Rad52 are vital in the protection of CFSs (Wang et al., 2018a). Therefore, under conditions of replication stress there are regions prone to induce the formation of DSBs and chromosomal translocations and in many cases these regions correlate with sites of transcription-replication conflicts at R-loop-enriched sites.

Apart from the effect on RFs, specific reports have raised the possibility that DNA breaks and genome instability may also occur in the absence of DNA replication. Thus, it has been proposed that DNA-RNA hybrids accumulated by the lack of RNA processing factors are actively processed into DSBs by the NER endonucleases XPF and XPG (Sollier et al., 2014; Brustel et al., 2018). However, new reports suggest that R-loop-dependent DSBs require replication. Thus, it has been shown at *Igγ* locus during class switch recombination (CSR) that DNA replication origin activation relies on

transcription-dependent R-loop formation and it is required for efficient CSR (Wiedemann et al., 2016) and that XPG has a nuclease-independent role in replication (Trego et al., 2016). Therefore, it remains to be seen if NER endonucleases coordinate DNA incision with DNA refilling.

DNA-RNA hybrids at DNA breaks: positive or negative effect on repair?

Several reports have uncovered that DNA breakage, whether single or double-stranded, is another driving force for DNA-RNA hybrid formation, as recently reviewed (**Figure 4**) (Aguilera and Gomez-Gonzalez, 2017). The use of an *in vitro* system, constituted by a T7 RNA polymerase upstream of defined Ig class switch substrates to elucidate the DNA features that contribute to R-loop formation, led to the observation that an ssDNA nick on the non-template strand strongly stimulates DNA-RNA hybrid formation (Roy et al., 2010). Similarly, analysis of the role of ribonucleoprotein SAF-A/hnRNP U during DNA damage response (DDR) suggested a transient transcription-dependent accumulation of DNA-RNA hybrids at sites of breaks in transcribed genes (Britton et al., 2014). Furthermore, in mammalian cells the RNA helicase DEAD box 1 (DDX1) involved in DDR forms foci after IR; importantly, those foci are R-loop-dependent as they can be reduced with RNase H or treatments with transcription inhibitor. Indeed, R-loops could be detected by DRIP around a DSB upon DDX1 depletion (Li et al., 2016a). These data support the conclusion that nicks and DSBs favor hybridization between the RNA and the template strand.

A free DNA end relieves the torsional stress imposed by a closed molecule, thus facilitating the entanglement of the invading RNA with its ssDNA template (**Figures 4B and C**). Topoisomerase I (Topo I) is required to relax DNA supercoils generated during transcription by the RNA polymerase. It is inhibited with high specificity by camptothecin (CPT) generating a Topo I-DNA cleavage complex (Top1cc) in which the ssDNA break persists with its 3' end covalently bound to Topo I. Several studies have shown that CPT treatment stimulates DNA-RNA hybrid formation (Groh et al., 2014; Marinello et al., 2013; Sordet et al., 2009). Analysis of post-mitotic primary neurons and lymphocytes show that DDR foci induced by CPT were reduced by inhibiting transcription or by overexpressing RNase H1, both limiting co-transcriptional R-loops. This could either be explained if Topo I inactivation stalled transcription promoting R-loop formation that would lead to DSBs, or the other way around, if the broken DNA resulting from Top1cc facilitated R-loop formation (Sordet et al., 2009). Supporting the first possibility, CPT triggers the accumulation of antisense RNAPII transcripts specifically at active divergent CpG-island promoters causing a transient stalling of RNAPII at the promoter and an increase of R-loops at highly-

transcribed regions including rDNA (Marinello et al., 2013). In the same line, analysis of endogenous □□□ and □□ □□ genes, which are altered in Friedreich ataxia (FRDA) and Fragile X syndromes, show R-loop accumulation at the CGG-repeats that lead to gene silencing (Colak et al., 2014; Groh et al., 2014). Importantly, treatment with CPT leads to increased levels of such R-loops at the □□□ resulting in strong transcriptional silencing (Groh et al., 2014). Altogether, these data reinforce the idea that DNA cleavage at actively transcribed regions favors R-loop formation.

The question emerging is if a DNA-RNA hybrid at a DNA break could impact on its repair. Several reports have led to the proposal that R-loops could mediate DSB repair. It has been shown that *S. pombe* RNase H-deficient cells are sensitive to genotoxic agents that generate DSBs, whereas RNase H overexpression delayed repair of DSBs artificially generated by an endonuclease. In addition, RNAPII was observed at higher levels at such DSBs, in which DNA-RNA hybrids are accumulated (Ohle et al., 2016). Taken together, these data suggested that efficient DSB repair requires transcription at resected DSBs followed by DNA-RNA hybrid formation and subsequent removal by RNase H. Nevertheless, a positive ChIP signal can result from □*e novo* protein recruitment or from accumulation of a pre-existing protein at the site analyzed; so a major occupancy of RNAPII at breaks could also result from a potential stalling of an elongating RNAP at the DNA region undergoing a break. On the other hand, a different study revealed that the *S. pombe* RNase H double mutant strain have neither defect in DSB repair nor enrichment of R-loops at the □ □□ programmed DSB. This RNase H double mutant is proficient for ionizing radiation-DSB repair and for survival to CPT treatment and did not show R-loop accumulation at the □ □□ locus by DRIP analysis in contrast to other regions (Zhao et al., 2018). Since RNase H1 and RNase H2 remove either RNA primers of Okazaki fragments during replication or ribonucleotides inserted into the DNA, sensitivity to specific genotoxic agents may also be due to an indirect effect on replication or repair.

Accumulation of human factors involved in homologous recombination (HR) at DSBs, such as RAD52, RAD51, BRCA1 and BRCA2, has been shown to be reduced by RNase H overexpression at active transcription regions or through specific reporter systems (D'Alessandro et al., 2018; Yasuhara et al., 2018). Nevertheless, genome-wide analysis of RAD51 accumulation at DSBs induced by the □*si*SI restriction endonuclease *in vivo* in human cells revealed that HR factors are enriched preferentially at DSBs located at transcriptionally active chromatin (Aymard et al., 2014). Whether this is enhanced by the existence of R-loops has not been determined. However, it has recently been shown that the transcription inhibitor DRB reduces RPA and Rad51 foci after IR exposure. Critical for this response was Rad52, whose

recruitment to laser tracked-DSBs was lessened in cells overexpressing RNase H; supporting this, DNA-RNA hybrid detection at such IR tracks was increased in RAD52 KO cells (Yasuhara et al., 2018). These results suggest the possibility that DNA-RNA hybrids at active transcription regions may direct DSB repair via HR. However, evidence that hybrids enhance HR is lacking. It is possible that HR factors are necessary to remove the hybrids to permit DSB repair. Indeed, as the first step for HR is resection of the 5' end, a hybrid at that strand will protect this end from resection, whereas a hybrid at the 3' end may facilitate resection of the 5' end, but will impair the ability of the 3' end ssDNA to invade the homologous DNA and initiate DSB repair by HR (Aguilera and Gomez-Gonzalez, 2017) (**Figure 4C**).

Consistent with the idea that hybrids would interfere with HR, genome-wide mapping of DNA-RNA hybrids in human cells carrying the DSB inducible system based on □*siS*□expression to cause DSBs all over the genome, revealed a significant R-loop enrichment on DSB-flanking chromatin. Interestingly, SETX was found to be recruited at such DSB ends. Moreover, SETX depletion impaired RAD51 recruitment and favored 53BP1 accumulation, a key DDR factor in NHEJ (Cohen et al., 2018). These data suggest that DNA-RNA hybrids may favor HR factor accumulation to potentially facilitate the elimination of the hybrids so that HR could occur, likely counteracting NHEJ at DSBs within transcribed genes. Whether DNA-RNA hybrids formed after a DSB could facilitate DNA-end resection as a first step of HR repair is unclear. However, DNA-RNA hybrids are likely a structure required to be removed to allow repair. Indeed, a number of reports favor a model that DNA-RNA hybrids would interfere with DSB repair rather than enhance it. Thus, in a RNase H double mutant in budding yeast, the absence of Top1 leads to increased lethality due to defective break-induced replication (BIR) within the repetitive rDNA locus (Amon and Koshland, 2016) and that persistent R-loops hinder DSB resection thus blocking proper repair, as suggested from the analysis of a yeast triple mutant strain in RNase H1, RNAse H2 and Sen1 (Costantino and Koshland, 2018). Similarly, in mammalian cells depletion of AQR leads to a reduced DNA damage-induced recruitment of HR proteins such as CtIP (Sakasai et al., 2017). On the contrary, it has recently been shown that human CtIP/yeast Sae2 involved in DSB resection also has a role in R-loop prevention in yeast and human cells. Although the study proposes an endonucleolitic role of CtIP/Sae2 in cleaving R-loops regardless of their known functions on DSB end resection (Makharashvili et al., 2018), it would be interesting to see whether CtIP action is more prominent on hybrids at DSBs. Indeed, we cannot discard the possibility that the role of BRCA2 and other putative DSB repair factors in protecting from R-loops were in part a consequence of the hybrids accumulated in unrepaired DSBs in repair-deficient cells. The recent

observation that depletion of the ATM/CHK2 DSB signaling pathway leads to R-loop accumulation would be consistent with this possibility (Barroso et al., 2019).

In conclusion, while it is clear that both ssDNA and dsDNA breaks favor R-loop formation, extensive research is still required to know whether hybrids are just a secondary consequence of breaks occurring at DNA undergoing transcription or are intermediates with a function in repair.

R-loops and disease

The overabundance of DNA-RNA hybrids has been documented in a number of syndromes, human neurological disorders and cancer cells (Perego et al., 2019; Richard and Manley, 2017). The latter may not be surprising since R-loops lead to genome instability and replication stress, which are hallmarks of tumor cells (Halazonetis et al., 2008). Apart from the involvement of the BRCA1 and BRCA2 tumor-suppressors and FA factors in R-loop prevention, a recent screening for cancer-related genome caretakers revealed components of mRNA biogenesis previously shown to be involved in R-loop prevention (Teloni et al., 2019). Regarding this, analysis of human breast cancer cell lines showed that translocations are located within genes induced by estrogen signaling, and that R-loops accumulate at such estrogen-induced genes (Stork et al., 2016), suggesting that DNA damage may be mediated by R-loops in active genes. This is supported by the mechanism to induce genome instability proposed for Kaposi sarcoma-associated herpesvirus (KSHV), for which the sequestration of the THO complex by the KSHV ORF57 protein leads to R-loop-dependent DNA damage and genome instability (Jackson et al., 2014). Also in connection with cancer, a particularly relevant recent observation is that in Ewing sarcoma expression of the EWS-FLI1 fusion enhances transcription leading to R-loop accumulation and at the same time impairs HR by interacting with BRCA1, thus causing the genome instability found in cancer cells (Gorthi et al., 2018). Therefore, it is alluring to place R-loops as the driving force during oncogenesis either as a source of replication stress or by preventing proper activity of the DNA repair pathways, however as yet there is no direct evidence for this.

Autoimmune diseases have also been related to R-loops that might compromise genome integrity by impacting DNA replication and/or repair. This is the case of the Wiskott-Aldrich syndrome (WAS), which has a high incidence of childhood leukaemia. Recently, it has been shown that deficiency of the WAS protein, WASp, causes R-loop accumulation, defects in splicing of specific genes and R-loop-mediated DSBs in T helper lymphocytes (Sarkar et al., 2018). Interestingly, the observation that WASp, a cytoplasmic protein, enters the nucleus and together with the actin-nucleating

ARP2/3 complex is recruited to damaged chromatin undergoing HR repair of DSBs (Schrank et al., 2018), raises the question of whether the increase in R-loops is preferentially occurring at DSBs or at replication stress sites. A link has also been described in the case of Aicardi-Goutieres syndrome (AGS). Mutations in either of the three subunits of the human RNase H2 are connected with AGS, for which R-loops have been shown to accumulate along the genome of patient cells (Lim et al., 2015). Interestingly, AGS is also linked to mutations in the dNTPase SAMHD1, a protein shown to promote ssDNA degradation at stalled RFs as a step required for replication restart (Coquel et al., 2018). We need further research to know whether it is the RER defect or if hybrid accumulation is the main cause behind this syndrome.

Finally, several diseases emerging as potentially related to R-loop accumulation are neurogenerative disorders. The key difference with cancer and previously discussed syndromes is that neurodegenerative disorders affect non-dividing neuronal cells, therefore R-loops accumulated in neuronal cells would not be compromising DNA replication. Repeat expansions are at the core of several neurogenerative disorders including ataxias, amyotrophic lateral sclerosis and nucleotide expansion disorders. In Friedreich's ataxia patient cell lines, which carry GAA repeat expansions, R-loops form at such trinucleotide repeats and trigger transcriptional silencing of the genes associated with the disease (Groh et al., 2014). Interestingly, introduction of antisense oligonucleotide and duplex RNAs to recognize the expanded GAA repeats in Friedreich's ataxia cell lines restore expression levels by blocking R-loop formation (Li et al., 2016b). This may be explained if the antisense RNA engages the RNA from the R-loop disrupting the hybrid and/or if the silencing machinery is yet another mechanism to solve R-loops. On the other hand, it has been proposed that R-loops promote repeat expansion by BIR, which has led to speculation that R-loops might be at the origin of such expansions (Neil et al., 2018). In the case of neurological diseases, we cannot forget SETX whose mutation is linked to ataxia with oculomotor apraxia type 2 (AOA2) and amyotrophic lateral sclerosis type 4 (ALS4), although is still unclear how SETX mutations cause neurodegeneration in humans. AOA2 cells have altered expression of neuronal genes, as determined by genome-wide studies, and increased R-loop levels (Becherel et al., 2015). In addition, amyotrophic lateral sclerosis (ALS) have been related to genes whose mutations also lead to R-loop accumulation in motor neurons (Perego et al., 2019). However, a gain-of-function mutation of SETX has been recently described that shows decreased R-loop formation in ALS4 patients implying that low levels of DNA-RNA hybrids do not always correlate with the disease (Grunseich et al., 2018). A similar situation is been shown in two autism-spectrum disorders with known genetic basis. Prader-Willi syndrome (PWS) and Angelman syndrome (AS) are caused

by deletion or mutation of the maternal allele of the ubiquitin protein ligase E3A (UBE3A). In neurons, the paternal allele of UBE3A is intact but epigenetically silenced because transcription continues through the convergent Snord116 gene to generate an antisense Ube3a transcript (Ub3a-ATS). The stabilization of R-loops at the GC-rich repetitive intronic segments within Snord116 locus with topotecan, a topoisomerase I inhibitor, leads to chromatin decondensation and impairs Ube3a-ATS transcription thus avoiding the silencing of Ube3a (Powell et al., 2013). Therefore, despite R-loop-dependent genomic instability appearing to be a common feature in neurological disorders, the correlation between neurological diseases and R-loops may not necessarily reflect a cause-effect relationship, but rather a collateral consequence of a defect in transcription caused by specific mutations.

Concluding remarks and perspectives

Evidence has accumulated the past two decades indicating that DNA-RNA hybrids occur all over the genome of cells and play specific functions in a number of regulated cellular processes but that, in most cases, they constitute a source of DNA damage and genome instability. DNA-RNA hybrids are found in protein-coding genes, tRNAs, rRNAs and non-coding RNAs, and they have a relevant impact on genome function and integrity. Three different types of factors protect genomes from unscheduled or pathological R-loop accumulation: i) factors preventing their formation, in most cases RNA binding and processing factors and chromatin remodelers; ii) factors resolving them, such as ribonucleases and DNA-RNA unwinding proteins, and iii) DNA repair factors that directly or indirectly help remove R-loops during repair or replication. The increasing identification of factors functioning in R-loop homeostasis is consistent with the relevance of these ubiquitous structures along genomes and their potential impact on cell proliferation and physiology. However, many questions emerge about the role of these factors on R-loop homeostasis.

Thus, we know much about the distribution of DNA-RNA hybrids in normal cells, but we do not know precisely: i) R-loop size and frequency along the genome; ii) what distinguishes a physiological R-loop from a pathological one; iii) the accumulation and impact of R-loops during different stages of the cell cycle; iv) why so many non-redundant RNA helicases with an *in vitro* DNA-RNA unwinding activity protect genomes from R-loop accumulation; v) how R-loops affect transcription-replication conflicts and *vice versa*; vi) whether hybrids at DNA breaks are made from de novo-synthesized RNA or from the nascent RNA generated by a previously engaged RNA polymerase during elongation, or vii) whether DNA-RNA hybrids have any role in DSB repair.

Current research points to a scenario in which R-loops represent a mark linked to genomic stress and dysfunction. The same way that any condition associated with genomic stress can be identified by the accumulation of DNA breaks, replication fork stalling or mitotic mis-segregation, among other features, we are starting to realize that they can also be identified by R-loop accumulation. High R-loop levels may just be a signature of cellular (nuclear) stress, thus the increasing identification of numerous factors involved in replication, repair, transcription, RNA processing, RNA export, chromatin remodeling, whose depletion is accompanied by R-loop accumulation, may not be identifying direct regulators of R-loop homeostasis. Even though unscheduled R-loops associate with genome instability and replication stress, hallmarks of cancer and a number of genetic diseases, and multiple neurodegenerative diseases are associated with R-loop accumulation, we have yet to establish a causal connection between R-loops and disease. Understanding the molecular basis of R-loop homeostasis, and the mechanisms that protect cells from unscheduled R-loops, is therefore required to comprehend the role of R-loops as the hidden face of transcription.

Acknowledgments

We would like to thank B. Gómez-González for her critical reading of the manuscript. We apologize to those whose work could not be cited due to space limitation.

Research in AA's lab is funded by the European Research Council, the Spanish Ministry of Science and Education and the Junta de Andalucía.

FIGURE LEGENDS

Figure 1. R-loop recognition and its involvement in cellular processes. A. R-loop detection systems. The DNA-RNA hybrid can be detected by the S9.6 antibody or GFP fused to an inactive RNase H or the RNA binging domains (HB) of the RNase H. The displaced ssDNA can be detected by induced-mutagenesis with bisulfite *in vitro*, which converts cytosines in the displaced ssDNA of the R-loop to uracil, or with activation-induced cytidine deaminase (AID) *in vivo*. **B.** R-loops as intermediaries of cellular processes. These include mammalian mitochondrial light DNA strand replication (top), Ig class-switch recombination (middle) or identification of a cognate target DNA site by a guide RNA in the CRISP-Cas9 system (bottom). **C.** R-loop contribution to gene regulation. This can be achieved at promoters, in transcription silencing by impeding transcription-factor binding (top) or in transcription activation by promoting transcription-factor binding or by blocking binding of transcriptional repressors or DNA methylases (middle), and at terminators, by facilitating transcription termination (bottom).

Figure 2. Avoiding R-loop accumulation. A. R-loop formation is prevented by specific RNA-binding proteins that are involved in RNA biogenesis (hnRNP), by topoisomerase 1 (Topo I) that resolves the negative supercoiling behind the elongating RNA polymerase II (RNAP), and by chromatin. **B.** R-loop removal is achieved by RNase H enzymes that degrade the RNA moiety of the hybrid and by DNA-RNA helicases that unwind the hybrid.

Figure 3. R-loops as a source of DNA damage. A. In non-cycling cells or in G_0/G_1 cells DNA damage and breaks might be originated from an R-loop because the ssDNA is a target of nucleases and genotoxics. **B.** In S/G_2 phases of cycling cells the most likely source of DNA damage derives from transcription-replication conflicts. These can be head-on (top), in which the RNA polymerase (RNAP) and RF (RF) travel on opposite direction, or co-directional (bottom), in which both RNAP and RF travel in the same direction. In the latter case, the R-loop must be stabilized in order to generate DNA breaks.

Figure 4. DNA break-dependent R-loops. A. A nick in the template DNA strand behind the RNA polymerase (RNAP) would free a DNA end favoring DNA-RNA hybrid formation. **B.** A break covering both DNA strands (DSB) will also favor DNA-RNA hybrid formation. In addition to the more likely scenario of the RNA hybridizing with its

proximal ssDNA template (continuous red line), it would be also possible the hybridization with the ssDNA at the other side of the break (orange dashed line).

REFERENCES

Aguilera, A. and Garcia-Muse, T. (2012) R loops: from transcription byproducts to threats to genome stability. Mol Cell □□, 115-124.

Aguilera, A., and Gomez-Gonzalez, B. (2017) DNA-RNA hybrids: the risks of DNA breakage during transcription.Nat Struct Mol Biol □□, 439-443.

Amon, J.D., and Koshland, D. (2016) RNase H enables efficient repair of R-loop induced DNA damage. eLife □.

Arab, K., Karaulanov, E., Musheev, M., Trnka, P., Schafer, A., Grummt, I., and Niehrs, C. (2019) GADD45A binds R-loops and recruits TET1 to CpG island promoters. Nat Genet □□, 217-223.

Aymard, F., Bugler, B., Schmidt, C.K., Guillou, E., Caron, P., Briois, S., Iacovoni, J.S., Daburon, V., Miller, K.M., Jackson, S.P.□□*t* □□(2014) Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. Nat Struct Mol Biol □□, 366-374.

Backert, S. (2002) R-loop-dependent rolling-circle replication and a new model for DNA concatemer resolution by mitochondrial plasmid mp1. EMBO J □□, 3128-3136.

Barroso, S., Herrera-Moyano, E., Muñoz, S., Garcia-Rubio, M., Gomez-Gonzalez, B., Aguilera, A. (2019) The DNA damage response acts as a safeguard against harmful DNA-RNA hybrids of different origins. EMBO Rep. □□, e47250.

Bayona-Feliu, A., Casas-Lamesa, A., Reina, O., Bernues, J., and Azorin, F. (2017) Linker histone H1 prevents R-loop accumulation and genome instability in heterochromatin. Nat Commun □, 283.

Becherel, O.J., Sun, J., Yeo, A.J., Nayler, S., Fogel, B.L., Gao, F., Coppola, G., Criscuolo, C., De Michele, G., Wolvetang, E.□□*t* □□(2015) A new model to study neurodegeneration in ataxia oculomotor apraxia type 2. Hum Mol Genet □□, 5759-5774.

Bhatia, V., Barroso, S.I., Garcia-Rubio, M.L., Tumini, E., Herrera-Moyano, E., and Aguilera, A. (2014) BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. Nature □□□, 362-365.

Boehmer, P.E. (2004) RNA binding and R-loop formation by the herpes simplex virus type-1 single-stranded DNA-binding protein (ICP8). Nucleic Acids Res □□, 4576-4584.

Boguslawski, S.J., Smith, D.E., Michalak, M.A., Mickelson, K.E., Yehle, C.O., Patterson, W.L., and Carrico, R.J. (1986) Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. J Immunol Methods □□, 123-130.

Boque-Sastre, R., Soler, M., Oliveira-Mateos, C., Portela, A., Moutinho, C., Sayols, S., Villanueva, A., Esteller, M., and Guil, S. (2015) Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. Proc Natl Acad Sci USA □□□, 5785-5790.

Boubakri, H., de Septenville, A.L., Viguera, E., and Michel, B. (2010) The helicases DinG, Rep and UvrD cooperate to promote replication across transcription units in vivo. The EMBO journal □□, 145-157.

Britton, S., Dernoncourt, E., Delteil, C., Froment, C., Schiltz, O., Salles, B., Frit, P., and Calsou, P. (2014) DNA damage triggers SAF-A and RNA biogenesis factors exclusion from chromatin coupled to R-loops removal. Nucleic Acids res □□, 9047-9062.

Brustel, J., Kozik, Z., Gromak, N., Savic, V., Sweet, S.M.M. (2018) Large XPF-dependent deletions following misrepair of a DNA double strand break are prevented by the RNA:DNA helicase Senataxin. Sci Rep □, 3850.

Castel, S.E., Ren, J., Bhattacharjee, S., Chang, A.Y., Sanchez, M., Valbuena, A., Antequera, F., and Martienssen, R.A. (2014) Dicer promotes transcription termination at sites of replication stress to maintain genome stability. Cell □□□, 572-583.

Chakraborty, P., Huang, J.T.J., and Hiom, K. (2018) DHX9 helicase promotes R-loop formation in cells with impaired RNA splicing. Nat Commun □, 4346.

Chan, Y.A., Aristizabal, M.J., Lu, P.Y., Luo, Z., Hamza, A., Kobor, M.S., Stirling, P.C., and Hieter, P. (2014) Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-chip. PLoS Genet □□, e1004288.

Chang, E.Y., Novoa, C.A., Aristizabal, M.J., Coulombe, Y., Segovia, R., Chaturvedi, R., Shen, Y., Keong, C., Tam, A.S., Jones, S.J.M.□□*t* □□(2017) RECQ-like helicases Sgs1 and BLM regulate R-loop-associated genome instability. j Cell Biol □□□, 3991-4005.

Chaudhuri, J., Khuong, C., and Alt, F.W. (2004) Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. Nature □□□, 992-998.

Chen PB, Chen HV, Acharya D, Rando OJ, Fazzio TG. (2015) R loops regulate promoter-proximal chromatin architecture and cellular differentiation. Nat. Struct. Mol. Biol. □□, 999-1007.
Chen, L., Chen, J.Y., Zhang, X., Gu, Y., Xiao, R., Shao, C., Tang, P., Qian, H., Luo, D., Li, H.□□*t* □□(2017) R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. Mol Cell □□, 745-757 e745.
Cloutier, S.C., Wang, S., Ma, W.K., Al Husini, N., Dhoondia, Z., Ansari, A., Pascuzzi, P.E., and Tran, E.J. (2016) Regulated Formation of lncRNA-DNA Hybrids Enables Faster Transcriptional Induction and Environmental Adaptation. Mol Cell □□, 393-404.
Cohen, S., Puget, N., Lin, Y.L., Clouaire, T., Aguirrebengoa, M., Rocher, V., Pasero, P., Canitrot, Y., and Legube, G. (2018) Senataxin resolves RNA:DNA hybrids forming at DNA double-strand breaks to prevent translocations. Nat Commun □, 533.
Colak, D., Zaninovic, N., Cohen, M.S., Rosenwaks, Z., Yang, W.Y., Gerhardt, J., Disney, M.D., and Jaffrey, S.R. (2014) Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. Science □□□, 1002-1005.
Coquel, F., Silva, M.J., Techer, H., Zadorozhny, K., Sharma, S., Nieminuszczy, J., Mettling, C., Dardillac, E., Barthe, A., Schmitz, A.L.□□*t* □□(2018) SAMHD1 acts at stalled replication forks to prevent interferon induction. Nature □□□, 57-61.
Costantino, L., and Koshland, D. (2018) Genome-wide Map of R-Loop-Induced Damage Reveals How a Subset of R-Loops Contributes to Genomic Instability. Mol Cell □□, 487-497 e483.
Cristini, A., Groh, M., Kristiansen, M.S., and Gromak, N. (2018) RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage. Cell Rep □□, 1891-1905.
Crossley, M.P., Bocek, M., and Cimprich, K.A. (2019) R-Loops as Cellular Regulators and Genomic Threats. Mol Cell □□, 398-411.
D'Alessandro, G., Whelan, D.R., Howard, S.M., Vitelli, V., Renaudin, X., Adamowicz, M., Iannelli, F., Jones-Weinert, C.W., Lee, M., Matti, V.□□*t* □□(2018) BRCA2 controls DNA:RNA hybrid level at DSBs by mediating RNase H2 recruitment. Nat Commun □, 5376.
Dominguez-Sanchez, M.S., Barroso, S., Gomez-Gonzalez, B., Luna, R., and Aguilera, A. (2011) Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. PLoS Genet □, e1002386.
Drolet, M., Phoenix, P., Menzel, R., Masse, E., Liu, L.F., and Crouch, R.J. (1995) Overexpression of RNase H partially complements the growth defect of an Escherichia coli delta topA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I.Proc Natl Acad Sci USA □□, 3526-3530.
El Hage, A., French, S.L., Beyer, A.L., and Tollervey, D. (2010) Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. Genes Dev □□, 1546-1558.
Gan, W., Guan, Z., Liu, J., Gui, T., Shen, K., Manley, J.L., and Li, X. (2011) R-loop-mediated genomic instability is caused by impairment of replication fork progression. Genes Dev □□, 2041-2056.
Garcia-Pichardo, D., Canas, J.C., Garcia-Rubio, M.L., Gomez-Gonzalez, B., Rondon, A.G., and Aguilera, A. (2017). Histone Mutants Separate R Loop Formation from Genome Instability Induction. Mol Cell □□, 597-609 e595.
Garcia-Rubio, M., Aguilera, P., Lafuente-Barquero, J., Ruiz, J.F., Simon, M.N., Geli, V., Rondon, A.G., and Aguilera, A. (2018) Yra1-bound RNA-DNA hybrids cause orientation-independent transcription-replication collisions and telomere instability. Genes Dev □□, 965-977.
Garcia-Rubio, M.L., Perez-Calero, C., Barroso, S.I., Tumini, E., Herrera-Moyano, E., Rosado, I.V., and Aguilera, A. (2015) The Fanconi Anemia Pathway Protects Genome Integrity from R-loops. PLoS Genet □□, e1005674.
Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I., and Chedin, F. (2012) R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol Cell □□, 814-825.
Gomez-Gonzalez, B., and Aguilera, A. (2007) Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. Proc Natl Acad Sci USA □□□, 8409-8414.

Gomez-Gonzalez, B., Garcia-Rubio, M.L., Bermejo, R., Gaillard, H., Shirahige, K., Marin, A., Foiani, M., and Aguilera, A. (2011) Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. EMBO J. □□, 3106-3119.
Gomez-Gonzalez, B., and Aguilera, A. (2019) Transcription-mediated replication hindrance: a major driver of genome instability. Genes Dev. □□, 1008-1026.
Gorthi, A., Romero, J.C., Loranc, E., Cao, L., Lawrence, L.A., Goodale, E., Iniguez, A.B., Bernard, X., Masamsetti, V.P., Roston, S.□□*t* □□(2018) EWS-FLI1 increases transcription to cause R-loops and block BRCA1 repair in Ewing sarcoma. Nature □□□, 387-391.
Grabczyk, E., Mancuso, M., and Sammarco, M.C. (2007) A persistent RNA.DNA hybrid formed by transcription of the Friedreich ataxia triplet repeat in live bacteria, and by T7 RNAP in vitro. Nucleic Acids Res □□, 5351–5359.
Groh, M., Lufino, M.M., Wade-Martins, R., and Gromak, N. (2014) R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. PLoS Genet □□, e1004318.
Grunseich, C., Wang, I.X., Watts, J.A., Burdick, J.T., Guber, R.D., Zhu, Z., Bruzel, A., Lanman, T., Chen, K., Schindler, A.B.□□*t* □□(2018) Senataxin Mutation Reveals How R-Loops Promote Transcription by Blocking DNA Methylation at Gene Promoters. Molecular cell □□, 426-437 e427.
Halász, L., Karányi, Z., Boros-Oláh, B., Kuik-Rózsa, T., Sipos, É., Nagy, É., Mosolygó-L, Á., Mázló, A., Rajnavölgyi, É., Halmos, G., Székvölgyi, L. (2017) RNA-DNA hybrid (R-loop) immunoprecipitation mapping: an analytical workflow to evaluate inherent biases. Genome Res □□, 1063-1073.
Halazonetis, T.D., Gorgoulis, V.G., Bartek, J. (2008) An oncogene-induced DNA damage model for cancer development. Science □□□, 1352-1355.
Hamperl, S., Bocek, M.J., Saldivar, J.C., Swigut, T., and Cimprich, K.A. (2017) Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. Cell □□□, 774-786 e719.
Hartono, S.R., Malapert, A., Legros, P., Bernard, P., Chedin, F., and 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 □□□, 272-284.
Hatchi, E., Skourti-Stathaki, K., Ventz, S., Pinello, L., Yen, A., Kamieniarz-Gdula, K., Dimitrov, S., Pathania, S., McKinney, K.M., Eaton, M.L.□□*t* □□(2015) BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. Mol Cell □□, 636-647.
Helmrich, A., Ballarino, M., and Tora, L. (2011) Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. MolCell □□, 966-977.
Herrera-Moyano, E., Mergui, X., Garcia-Rubio, M.L., Barroso, S., and Aguilera, A. (2014) The yeast and human FACT chromatin-reorganizing complexes solve R-loop-mediated transcription-replication conflicts. Genes Dev □□, 735-748.
Hodroj, D., Recolin, B., Serhal, K., Martinez, S., Tsanov, N., Abou Merhi, R., and Maiorano, D. (2017) An ATR-dependent function for the Ddx19 RNA helicase in nuclear R-loop metabolism. EMBO J □□, 1182-1198.
Hraiky, C., Raymond, M.A., and Drolet, M. (2000) RNase H overproduction corrects a defect at the level of transcription elongation during rRNA synthesis in the absence of DNA topoisomerase I in Escherichia coli. J Biol Chem □□□, 11257-11263.
Huertas, P., and Aguilera, A. (2003) Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. Mol Cell □□, 711-721.
Jackson, B.R., Noerenberg, M., and Whitehouse, A. (2014) A novel mechanism inducing genome instability in Kaposi's sarcoma-associated herpesvirus infected cells. PLoS Pathog □□, e1004098.
Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science □□□, 816-821.
Kaneko, S., Chu, C., Shatkin, A.J., and Manley, J.L. (2007) Human capping enzyme promotes formation of transcriptional R loops in vitro. Proc Natl Acad Sc i USA □□□, 17620-17625.
Kim, N., and Jinks-Robertson, S. (2011) Guanine repeat-containing sequences confer transcription-dependent instability in an orientation-specific manner in yeast. DNA Repair (Amst.) □□, 953–960.

Lafuente-Barquero, J., Luke-Glaser, S., Graf, M., Silva, S., Gomez-Gonzalez, B., Lockhart, A., Lisby, M., Aguilera, A., and Luke, B. (2017) The Smc5/6 complex regulates the yeast Mph1 helicase at RNA-DNA hybrid-mediated DNA damage. PLoS Genet □□, e1007136.
Lang, K.S., Hall, A.N., Merrikh, C.N., Ragheb, M., Tabakh, H., Pollock, A.J., Woodward, J.J., Dreifus, J.E., and Merrikh, H. (2017) Replication-Transcription Conflicts Generate R-Loops that Orchestrate Bacterial Stress Survival and Pathogenesis. Cell □□□, 787-799 e718.
Letessier, A., Millot, G.A., Koundrioukoff, S., Lachages, A.M., Vogt, N., Hansen, R.S., Malfoy, B., Brison, O., and Debatisse, M. (2011) Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. Nature □□□, 120-123.
Li, L., Germain, D.R., Poon, H.Y., Hildebrandt, M.R., Monckton, E.A., McDonald, D., Hendzel, M.J., and Godbout, R. (2016a) DEAD Box 1 Facilitates Removal of RNA and Homologous Recombination at DNA Double-Strand Breaks. Mol Cell Biol □□, 2794-2810.
Li, L., Matsui, M., and Corey, D.R. (2016b) Activating frataxin expression by repeat-targeted nucleic acids. Nat Commun □, 10606.
Li, M., Pokharel, S., Wang, J.T., Xu, X., Liu, Y. (2015) RECQ5-dependent SUMOylation of DNA topoisomerase I prevents transcription-associated genome instability. Nat Commun □, 6720.
Li, X., and Manley, J.L. (2005) Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. Cell □□□, 365-378.
Liang, Z., Liang, F., Teng, Y., Chen, X., Liu, J., Longerich, S., Rao, T., Green, A.M., Collins, N.B., Xiong, Y.□□*t* □□(2019) Binding of FANCI-FANCD2 Complex to RNA and R-Loops Stimulates Robust FANCD2 Monoubiquitination. Cell Rep □□, 564-572 e565.
Lim, Y.W., Sanz, L.A., Xu, X., Hartono, S.R., and Chedin, F. (2015) Genome-wide DNA hypomethylation and RNA:DNA hybrid accumulation in Aicardi-Goutieres syndrome. Elife □.
Lin, Y., Dent, S.Y., Wilson, J.H., Wells, R.D., and Napierala, M. (2010) R loops stimulate genetic instability of CTG.CAG repeats. Proc Natl Acad Sci USA □□□, 692-697.
Lindahl, T. (1993) Instability and decay of the primary structure of DNA. Nature □□□, 709-715.
Madireddy, A., Kosiyatrakul, S.T., Boisvert, R.A., Herrera-Moyano, E., Garcia-Rubio, M.L., Gerhardt, J., Vuono, E.A., Owen, N., Yan, Z., Olson, S.□□*t* □□(2016) FANCD2 Facilitates Replication through Common Fragile Sites. Mol Cell □□, 388-404.
Makharashvili, N., Arora, S., Yin, Y., Fu, Q., Wen, X., Lee, J.H., Kao, C.H., Leung, J.W., Miller, K.M., and Paull, T.T. (2018) Sae2/CtIP prevents R-loop accumulation in eukaryotic cells. eLife □.
Marinello, J., Chillemi, G., Bueno, S., Manzo, S.G., and Capranico, G. (2013) Antisense transcripts enhanced by camptothecin at divergent CpG-island promoters associated with bursts of topoisomerase I-DNA cleavage complex and R-loop formation. Nucleic Acids Res □□, 10110-10123.
Mirkin, E.V., and Mirkin, S.M. (2005) Mechanisms of transcription-replication collisions in bacteria. Mol. Cell Biol. □□, 888-895.
Mischo, H.E., Gomez-Gonzalez, B., Grzechnik, P., Rondon, A.G., Wei, W., Steinmetz, L., Aguilera, A., and Proudfoot, N.J. (2011) Yeast Sen1 helicase protects the genome from transcription-associated instability. Molecular Cell □□, 21-32.
Morales, J.C., Richard, P., Patidar, P.L., Motea, E.A., Dang, T.T., Manley, J.L., and Boothman, D.A. (2016) XRN2 Links Transcription Termination to DNA Damage and Replication Stress. PLoS Genet □□, e1006107.
Nadel, J., Athanasiadou, R., Lemetre, C., Wijetunga, N.A., P, O.B., Sato, H., Zhang, Z., Jeddeloh, J., Montagna, C., Golden, A.□□*t* □□(2015) RNA:DNA hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships. Epigenetics Chromatin □, 46.
Neil, A.J., Liang, M.U., Khristich, A.N., Shah, K.A., and Mirkin, S.M. (2018) RNA-DNA hybrids promote the expansion of Friedreich's ataxia (GAA)n repeats via break-induced replication. Nucleic Acids Res □□, 3487-3497.
Nguyen, H.D., Leong, W.Y., Li, W., Reddy, P.N.G., Sullivan, J.D., Walter, M.J., Zou, L., and Graubert, T.A. (2018) Spliceosome Mutations Induce R Loop-Associated Sensitivity to ATR Inhibition in Myelodysplastic Syndromes. Cancer Res □□, 5363-5374.
Ohle, C., Tesorero, R., Schermann, G., Dobrev, N., Sinning, I., and Fischer, T. (2016) Transient RNA-DNA Hybrids Are Required for Efficient Double-Strand Break Repair. Cell □□□, 1001-1013 e1007.
Okamoto, Y., Abe, M., Itaya, A., Tomida, J., Ishiai, M., Takaori-Kondo, A., Taoka, M., Isobe, T., and Takata, M. (2019) FANCD2 protects genome stability by recruiting RNA processing enzymes to resolve R-loops during mild replication stress. FEBS J □□□, 139-150.

Paulsen, R.D., Soni, D.V., Wollman, R., Hahn, A.T., Yee, M.C., Guan, A., Hesley, J.A., Miller, S.C., Cromwell, E.F., Solow-Cordero, D.E., et al. (2009) A genome-wide siRNA screen reveals diverse cellular processes and path- ways that mediate genome stability. Mol. Cell □□, 228–239.
Perego, M.G.L., Taiana, M., Bresolin, N., Comi, G.P., and Corti, S. (2019) R-Loops in Motor Neuron Diseases. Mol Neurobiol □□, 2579-2589.
Posse, V., Al-Behadili, A., Uhler, J.P., Clausen, A.R., Reyes, A., Zeviani, M., Falkenberg, M., and Gustafsson, C.M. (2019) RNase H1 directs origin-specific initiation of DNA replication in human mitochondria. PLoS Genet □□, e1007781.
Powell, W.T., Coulson, R.L., Gonzales, M.L., Crary, F.K., Wong, S.S., Adams, S., Ach, R.A., Tsang, P., Yamada, N.A., Yasui, D.H.□□*t* □□(2013) R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation. Proc Natl Acad Sci USA □□□, 13938-13943.
Prado, F., and Aguilera, A. (2005) Impairment of replication fork progression mediates RNA polII transcription-associated recombination. EMBO J. □□, 1267-1276.
Reddy, K., Tam, M., Bowater, R.P., Barber, M., Tomlinson, M., Nichol Edamura, K., Wang, Y.H., and Pearson, C.E. (2011) Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. Nucleic Acids Res □□, 1749–1762.
Ribeiro de Almeida, C., Dhir, S., Dhir, A., Moghaddam, A.E., Sattentau, Q., Meinhart, A., and Proudfoot, N.J. (2018) RNA Helicase DDX1 Converts RNA G-Quadruplex Structures into R-Loops to Promote IgH Class Switch Recombination. Mol Cell □□, 650-662 e658.
Richard, P., and Manley, J.L. (2017) R Loops and Links to Human Disease. Journal of molecular biology □□□, 3168-3180.
Roy, D., Zhang, Z., Lu, Z., Hsieh, C.L., and Lieber, M.R. (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 □□, 146-159.
Sakasai, R., Isono, M., Wakasugi, M., Hashimoto, M., Sunatani, Y., Matsui, T., Shibata, A., Matsunaga, T., and Iwabuchi, K. (2017) Aquarius is required for proper CtIP expression and homologous recombination repair. Sci Rep □, 13808.
Salas-Armenteros, I., Perez-Calero, C., Bayona-Feliu, A., Tumini, E., Luna, R., and Aguilera, A. (2017) Human THO-Sin3A interaction reveals new mechanisms to prevent R-loops that cause genome instability. EMBO J □□, 3532-3547.
Sanz, L.A., Hartono, S.R., Lim, Y.W., Steyaert, S., Rajpurkar, A., Ginno, P.A., Xu, X., and Chedin, F. (2016) Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. Mol Cell □□, 167-178.
Saponaro, M., Kantidakis, T., Mitter, R., Kelly, G.P., Heron, M., Williams, H., Söding, J., Stewart, A., Svejstrup, J.Q. (2014) RECQL5 controls transcript elongation and suppresses genome instability associated with transcription stress. Cell □□□, 1037-49.
Sarkar, K., Han, S.S., Wen, K.K., Ochs, H.D., Dupre, L., Seidman, M.M., and Vyas, Y.M. (2018) R-loops cause genomic instability in T helper lymphocytes from patients with Wiskott-Aldrich syndrome. J Allergy Clin Immunol □□□, 219-234.
Schrank, B.R., Aparicio, T., Li, Y., Chang, W., Chait, B.T., Gundersen, G.G., Gottesman, M.E., and Gautier, J. (2018) Nuclear ARP2/3 drives DNA break clustering for homology-directed repair. Nature □□□, 61-66.
Schwab, R.A., Nieminuszczy, J., Shah, F., Langton, J., Lopez Martinez, D., Liang, C.C., Cohn, M.A., Gibbons, R.J., Deans, A.J., and Niedzwiedz, W. (2015) The Fanconi Anemia Pathway Maintains Genome Stability by Coordinating Replication and Transcription. Mol Cell □□, 351-361.
Silva, S., Camino, L.P., and Aguilera, A. (2018) Human mitochondrial degradosome prevents harmful mitochondrial R loops and mitochondrial genome instability. Proc Natl Acad Sci USA □□□, 11024-11029.
Skourti-Stathaki, K., Kamieniarz-Gdula, K., and Proudfoot, N.J. (2014) R-loops induce repressive chromatin marks over mammalian gene terminators. Nature □□□, 436-439.
Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. (2011) Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Mol Cell □□, 794-805.
Sollier, J., Stork, C.T., Garcia-Rubio, M.L., Paulsen, R.D., Aguilera, A., and Cimprich, K.A. (2014) Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. Mol Cell □□, 777-785.

Sordet, O., Redon, C.E., Guirouilh-Barbat, J., Smith, S., Solier, S., Douarre, C., Conti, C., Nakamura, A.J., Das, B.B., Nicolas, E. *et al.* (2009) Ataxia telangiectasia mutated activation by transcription- and topoisomerase I-induced DNA double-strand breaks. EMBO Rep □□, 887-893.
Sridhara, S.C., Carvalho, S., Grosso, A.R., Gallego-Paez, L.M., Carmo-Fonseca, M., and de Almeida, S.F. (2017) Transcription Dynamics Prevent RNA-Mediated Genomic Instability through SRPK2-Dependent DDX23 Phosphorylation. Cell Rep □□, 334-343.
Stirling, P.C., Chan, Y.A., Minaker, S.W., Aristizabal, M.J., Barrett, I., Sipahimalani, P., Kobor, M.S., and Hieter, P. (2012) R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. Genes Dev □□, 163-175.
Stork, C.T., Bocek, M., Crossley, M.P., Sollier, J., Sanz, L.A., Chedin, F., Swigut, T., and Cimprich, K.A. (2016) Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage. Elife □.
Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N.J., and Dean, C. (2013) R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. Science □□□, 619-621.
Svikovic, S., Crisp, A., Tan-Wong, S.M., Guilliam, T.A., Doherty, A.J., Proudfoot, N.J., Guilbaud, G., and Sale, J.E. (2019) R-loop formation during S phase is restricted by PrimPol-mediated repriming. EMBO J □□.
Taneja, N., Zofall, M., Balachandran, V., Thillainadesan, G., Sugiyama, T., Wheeler, D., Zhou, M., and Grewal, S.I. (2017) SNF2 Family Protein Fft3 Suppresses Nucleosome Turnover to Promote Epigenetic Inheritance and Proper Replication. Mol Cell □□, 50-62 e56.
Tedeschi, F.A., Cloutier, S.C., Tran, E.J., and Jankowsky, E. (2018) The DEAD-box protein Dbp2p is linked to noncoding RNAs, the helicase Sen1p, and R-loops. RNA □□, 1693-1705.
Teloni, F., Michelena, J., Lezaja, A., Kilic, S., Ambrosi, C., Menon, S., Dobrovolna, J., Imhof, R., Janscak, P., Baubec, T. *et al.* (2019) Efficient Pre-mRNA Cleavage Prevents Replication-Stress-Associated Genome Instability. Mol Cell □□, 670-683 e612.
Tous, C., and Aguilera, A. (2007) Impairment of transcription elongation by R-loops in vitro. Biochem Biophys Res Commun □□□, 428-432.
Tran, P.L.T., Pohl, T.J., Chen, C.F., Chan, A., Pott, S., and Zakian, V.A. (2017) PIF1 family DNA helicases suppress R-loop mediated genome instability at tRNA genes. Nat Commun □, 15025.
Trego, K.S., Groesser, T., Davalos, A.R., Parplys, A.C., Zhao, W., Nelson, M.R., Hlaing, A., Shih, B., Rydberg, B., Pluth, J.M. *et al.* (2016) Non-catalytic Roles for XPG with BRCA1 and BRCA2 in Homologous Recombination and Genome Stability. Mol Cell □□, 535-546.
Tuduri, S., Crabbe, L., Conti, C., Tourriere, H., Holtgreve-Grez, H., Jauch, A., Pantesco, V., De Vos, J., Thomas, A., Theillet, C. *et al.* (2009) Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. Nat Cell Biol □□, 1315-1324.
Vijayraghavan, S., Tsai, F.L., and Schwacha, A. (2016) A Checkpoint-Related Function of the MCM Replicative Helicase Is Required to Avert Accumulation of RNA:DNA Hybrids during S-phase and Ensuing DSBs during G2/M. PLoS Genet □□, e1006277.
Wahba, L., Amon, J.D., Koshland, D., and Vuica-Ross, M. (2011) RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. Mol Cell □□, 978-988.
Wahba, L., Costantino, L., Tan, F.J., Zimmer, A., and Koshland, D. (2016) S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. Genes Dev □□, 1327-1338.
Wang, H., Li, S., Oaks, J., Ren, J., Li, L., and Wu, X. (2018a) The concerted roles of FANCM and Rad52 in the protection of common fragile sites. Nat Commun □, 2791.
Wang, I.X., Grunseich, C., Fox, J., Burdick, J., Zhu, Z., Ravazian, N., Hafner, M., and Cheung, V.G. (2018b) Human proteins that interact with RNA/DNA hybrids. Genome Res □□, 1405-1414.
Wellinger, R.E., Prado, F., and Aguilera, A. (2006) Replication fork progression is impaired by transcription in hyperrecombinant yeast cells lacking a functional THO complex. Mol Cell Biol □□, 3327-3334.
Wiedemann, E.M., Peycheva, M., and Pavri, R. (2016) DNA Replication Origins in Immunoglobulin Switch Regions Regulate Class Switch Recombination in an R-Loop-Dependent Manner. Cell Rep □□, 2927-2942.

Xu, W., Xu, H., Li, K., Fan, Y., Liu, Y., Yang, X., and Sun, Q. (2017) The R-loop is a common chromatin feature of the Arabidopsis genome. Nat Plants □, 704-714.
Yasuhara, T., Kato, R., Hagiwara, Y., Shiotani, B., Yamauchi, M., Nakada, S., Shibata, A., and Miyagawa, K. (2018) Human Rad52 Promotes XPG-Mediated R-loop Processing to Initiate Transcription-Associated Homologous Recombination Repair. Cell □□□, 558-570 e511.
Yu, K., Chedin, F., Hsieh, C.L., Wilson, T.E., and Lieber, M.R. (2003) R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nat Immunol □, 442-451.
Zeller, P., Padeken, J., van Schendel, R., Kalck, V., Tijsterman, M., and Gasser, S.M. (2016) Histone H3K9 methylation is dispensable for Caenorhabditis elegans development but suppresses RNA:DNA hybrid-associated repeat instability. Nat Genet □□, 1385-1395.
Zhao, D.Y., Gish, G., Braunschweig, U., Li, Y., Ni, Z., Schmitges, F.W., Zhong, G., Liu, K., Li, W., Moffat, J.□□*t* □□(2016) SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. Nature □□□, 48-53.
Zhao, H., Zhu, M., Limbo, O., and Russell, P. (2018) RNase H eliminates R-loops that disrupt DNA replication but is nonessential for efficient DSB repair. EMBO Rep □□, e45335

Table1. Selected factors regulating DNA-RNA hybrid homeostasis

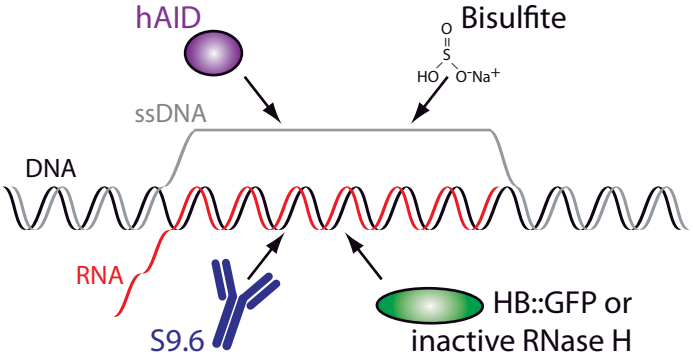
Biological activity	Mammals / yeast factors
Transcription and mRNA processing	THO complex; SPT6; SRFR1; CE (capping enzyme); DSS1/yDss1; yMex67; yRna14; SETX/ySen1; WDR33; XRN2; yNup64; yNup84
RNA surveillance	TRAMP complex; SUV3/PNPase
Ribonucleases	RNase H1; RNase H2
Helicases	SETX/ySen1; DDX23; DDX9; BLM/ySgs1; RECQL5; FANCM/yMph1; AQR; DDX19; DDX1; DDX21; DDX43; yPif1; yDbp2; ySub2
Topoisomerases	TOP1/yTop1; yTop2
Chromatin remodelers	FACT complex; SIN3A/ySin3; SNF2
DNA repair	BRCA1; BRCA2; FA (FANCM, FANCD2); XPG/XPF; CtIP/ySae2; PrimPol

For references see text.

Figure 1

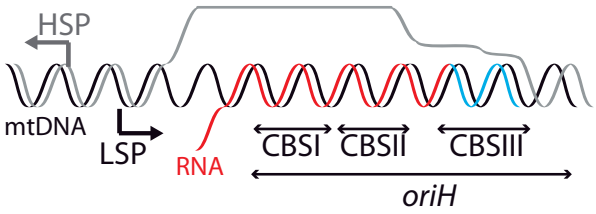
Click here to access/download;Figure;Garcia-Muse&Aguilera_Figure1.ai

A

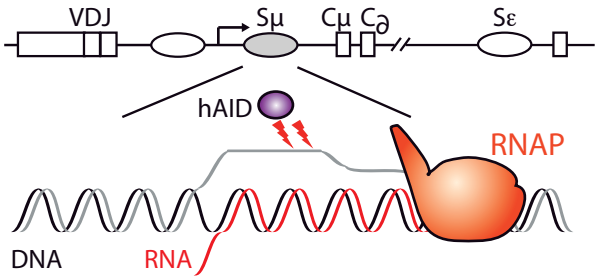


B

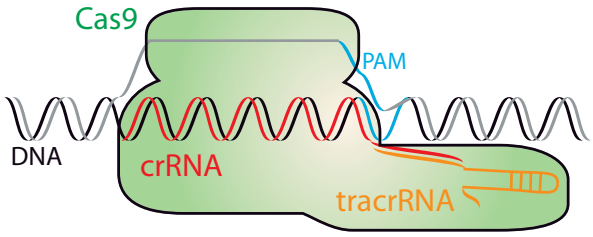
mtDNA REPLICATION



Ig CLASS-SWITCH RECOMBINATION

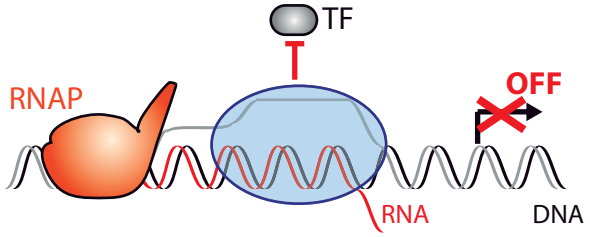


DNA EDITING

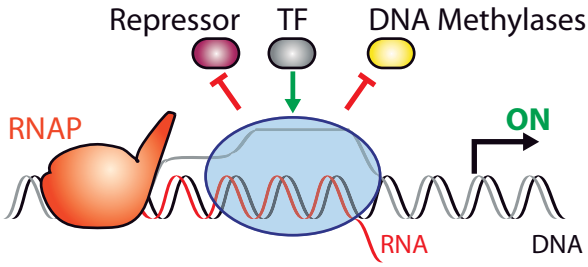


C

TRANSCRIPTION SILENCING



TRANSCRIPTION ACTIVATION



TRANSCRIPTION TERMINATION

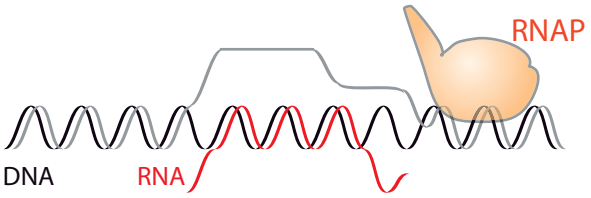
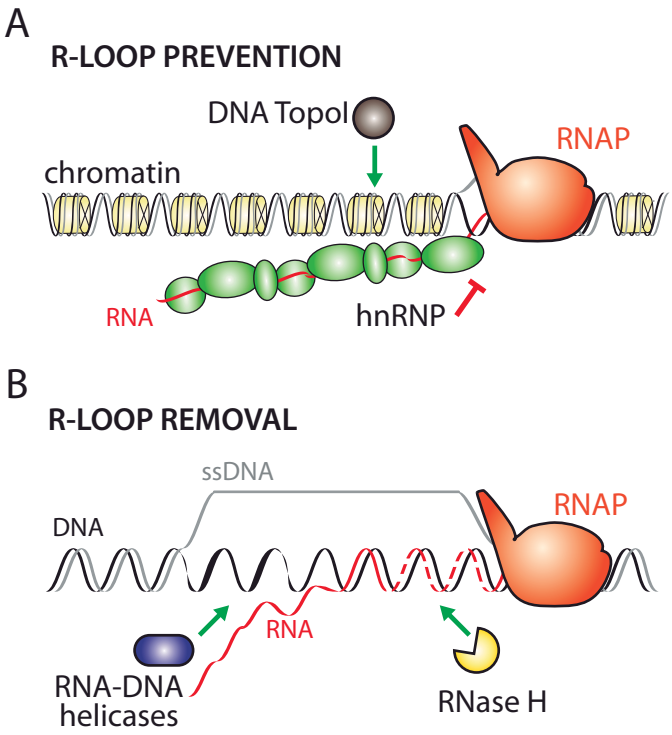


Figure 2

Click here to access/download;Figure;Garcia-Muse&Aguilera_Figure2.ai



Click here to access/download;Figure;Garcia-Muse&Aguilera_Figure3.ai

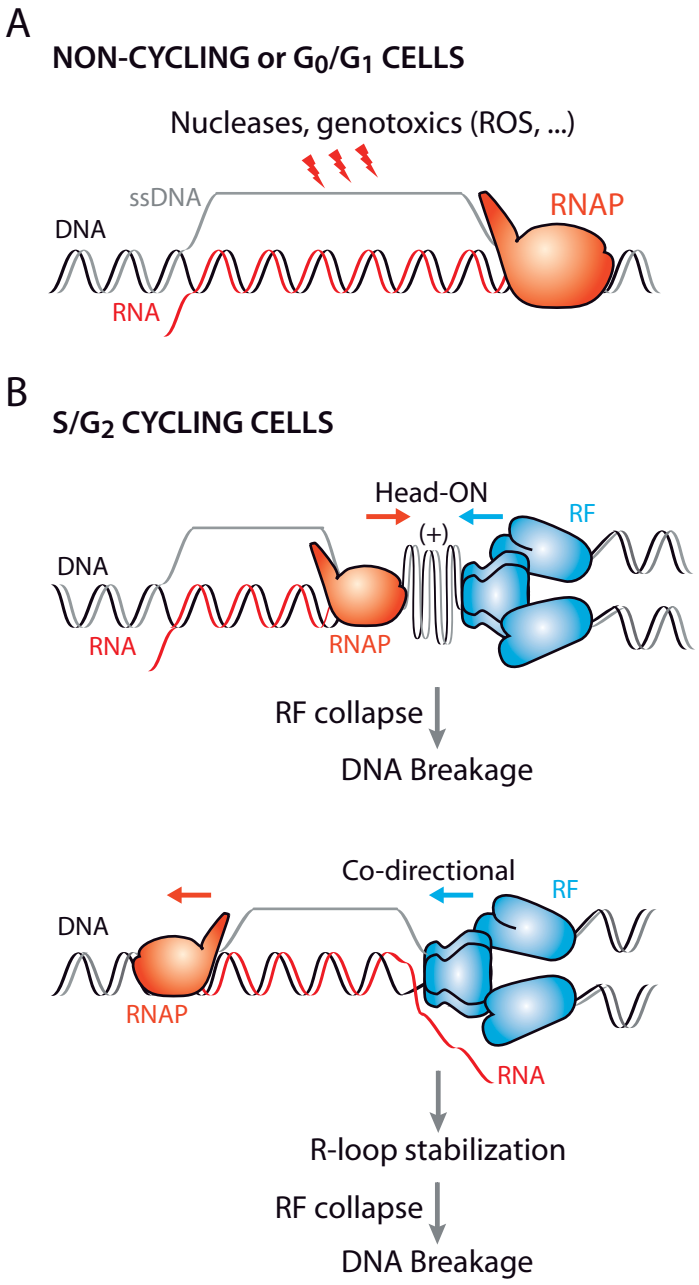


Figure 4

Click here to access/download;Figure;Garcia-Muse&Aguilera_Figure4.ai