Published in final edited form as:

DNA Repair (Amst). 2014 July ; 19: 84–94. doi:10.1016/j.dnarep.2014.03.023.

The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability

Stephan Hamperl and Karlene A. Cimprich*

Department of Chemical and Systems Biology, Stanford University School of Medicine, 318 Campus Drive, Stanford, CA 94305-5441, USA

Abstract

Accurate DNA replication and DNA repair are crucial for the maintenance of genome stability, and it is generally accepted that failure of these processes is a major source of DNA damage in cells. Intriguingly, recent evidence suggests that DNA damage is more likely to occur at genomic loci with high transcriptional activity. Furthermore, loss of certain RNA processing factors in eukaryotic cells is associated with increased formation of co-transcriptional RNA:DNA hybrid structures known as R-loops, resulting in double-strand breaks (DSBs) and DNA damage. However, the molecular mechanisms by which R-loop structures ultimately lead to DNA breaks and genome instability is not well understood. In this review, we summarize the current knowledge about the formation, recognition and processing of RNA:DNA hybrids, and discuss possible mechanisms by which these structures contribute to DNA damage and genome instability in the cell.

Keywords

R-loops; Genome instability; Double-strand breaks; G-quadruplex; Topoisomerase; THO/TREX complex; THSC/TREX-2; ASF splicing factor; mRNP biogenesis; RNA processing factors; RNase H; RNA:DNA helicases; Senataxin; Activation-induced deaminase (AID); APOBEC family; Transcription-replication conflicts

1. Introduction

RNA synthesis is one of the central processes by which other cellular machineries access the genetic information encoded by genomic DNA. Therefore, transcription represents an essential process of DNA metabolism. However, actively transcribing RNA polymerases (RNAPs) inevitably induce fundamental alterations in the underlying chromatin template in eukaryotic cells. The tight association of nucleosomes with the template DNA must be disrupted during transcription elongation, leading to partial or complete loss or exchange of

© 2014 Elsevier B.V. All rights reserved.

* Corresponding author, Tel.: +1 650 498 4720, Fax: +1 650 725 4665, cimprich@stanford.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

histone molecules (Dion et al., 2007; Jamai et al., 2007; Kimura and Cook, 2001; Schwabish and Struhl, 2006; Thiriet and Hayes, 2005, 2006). Moreover, when unwinding the DNA double helix at the site of active transcription, negative and positive supercoils arise behind and in front of the advancing RNAP, respectively (Liu and Wang, 1987; Wu et al., 1988). These rearrangements have potentially destabilizing consequences for the underlying DNA molecule. Not surprisingly, highly transcribed genes exhibit increased mutation and recombination rates compared to genomic regions with low transcriptional activity (Datta and Jinks-Robertson, 1995; Gottipati et al., 2008; Kim et al., 2007; Nickoloff, 1992). These transcription-associated mutagenesis (TAM) and transcription-associated recombination (TAR) events are conserved from bacteria to mammalian cells, and extensive research in the past few years has led to great progress in our understanding of the molecular mechanisms leading to TAR and TAM. A large body of evidence suggests that conflicts between the transcription and replication machineries are a major source of the observed genomic instabilities, and we refer the reader to several excellent reviews covering this topic (Aguilera and Gómez-González, 2008; Bermejo et al., 2012; Gottipati and Helleday, 2009; Kim and Jinks-Robertson, 2012; Li and Manley, 2006; Saxowsky and Doetsch, 2006; Svejstrup, 2010). However, work in the last few years suggests that formation of cotranscriptional RNA:DNA hybrid structures, known as R-loops, may significantly contribute to the above phenomena. In an R-loop structure, the nascent RNA strand invades the DNA duplex to hybridize with the complementary template strand after it exits RNAP. This results in a nucleic acid structure containing an RNA:DNA hybrid and a displaced tract of single-stranded DNA (ssDNA) (Aguilera and Gómez-González, 2008; Aguilera and García-Muse, 2012, 2013) (**Figure 1**).

RNA:DNA hybrids are implicated in a multitude of biological processes. Besides the short RNA primers of ~7-12 bp generated by DNA primase during replication of the lagging strand (Pellegrini, 2012) and transient formation at the center of the transcription bubble (~8 bp) during transcription by RNAPII (Westover et al., 2004), longer and more stable stretches of hybrids are key intermediates in replication and recombination. For example in *E.coli*, replication initiation of *ColEI*-type plasmids requires transcription-dependent formation of a stable RNA:DNA hybrid that extends past the origin of replication. Cleavage of the hybrid by RNase H, which specifically degrades the RNA in an RNA:DNA-hybrid, leaves a 3'-OH end that is extended by the replication machinery (Itoh and Tomizawa, 1980). A similar mechanism has been proposed for replication of yeast and mammalian mitochondrial DNA, with DNA synthesis being primed by a mitochondrial RNAP transcript also processed by RNase H (Baldacci et al., 1984; Pohjoismäki et al., 2010; Xu and Clayton, 1996). Finally, the formation of RNA:DNA hybrid structures plays a role in the generation of antibody diversity during class switch recombination (CSR) in activated B cells (reviewed in Chaudhuri and Alt, 2004; Manis et al., 2002). In this process, transcripts derived from the repetitive switch (S) sequences of IgH genes form an R-loop with the template strand *in vitro* and *in vivo* (Daniels and Lieber, 1995a, 1995b; Mizuta et al., 2003; Reaban et al., 1994; Tian and Alt, 2000; Yu et al., 2003). The ssDNA of these R-loops can then be targeted by Activation-Induced Deaminase (AID) (Muramatsu et al., 2000), and the resulting deoxyuridine residues are processed by components of the base-excision or mismatch repair machineries to single-strand breaks (SSBs) (Guikema et al., 2007; Masani et al., 2013;

Petersen-Mahrt et al., 2002; Rada et al., 2004). These DNA lesions are finally converted to a DSB, a necessary intermediate for recombination at the S sequences, in a process that involves non-homologous end joining factors (Petersen et al., 2001; Stavnezer et al., 2008).

The events occurring during CSR clearly highlight the potential for co-transcriptional formation of RNA:DNA hybrid structures to induce DNA breaks and recombination. They also raise the possibility that CSR-related mechanisms could contribute to R-loop mediated strand-break formation and chromosomal instability at other genomic regions and in other cell types. In this regard, numerous recent studies suggest that R-loops may form with higher frequency in eukaryotic genomes than previously anticipated. Immunofluorescence experiments performed using an antibody which detects RNA:DNA hybrids in a sequence-independent manner (Boguslawski et al., 1986) showed abundant signals distributed throughout the nucleoplasm in human H1 ESC and mouse NPC cells (Ginno et al., 2012; Powell et al., 2013). Moreover, DNA:RNA immunoprecipitation (DRIP) combined with high-throughput sequencing (DRIP-seq) detected putative RNA:DNA hybrids at more than 20,000 peak regions in human Ntera2 cells (Ginno et al., 2012). A recent bioinformatic study corroborated these results by creating a computational algorithm to identify potential R-loop forming sequences (RLFS) in the human genome. Strikingly, almost 60% of transcribed sequences contained at least 1 RLFS (Wongsurawat et al., 2012). Thus, R-loops may be abundant cellular intermediates. This finding suggests that cells may also have systems that prevent the processing of these R-loops into DNA breaks. Here, we summarize the current knowledge of the factors and cellular pathways implicated in the formation, recognition and processing of R-loop structures *in vivo*. Finally, we discuss possible mechanisms for how these aberrant nucleic acid structures may lead to DNA breaks and compromise genome integrity.

2. Formation of R-loops is dependent on the sequence and transcription-dependent topological state of the DNA molecule

2.1 Clusters of G-rich sequences and G-quadruplex structures

The molecular mechanism of R-loop formation has primarily been elucidated from *in vitro* transcription experiments which utilized prokaryotic or phage RNA polymerases and purified plasmid DNA coding for mammalian class-switch regions (see Introduction) (Daniels and Lieber, 1995a; Duquette et al., 2004; Reaban and Griffin, 1990; Reaban et al., 1994; Roy et al., 2008; Tian and Alt, 2000). In an elegant set of experiments, Roy and colleagues showed that the nascent RNA strand must pass through the exit pore of RNA polymerase before threading back to anneal with the template DNA (thread-back model), and that therefore the R-loop is not just an extension of the ~8 bp hybrid formed in the transcription bubble (extended hybrid model) (Roy et al., 2008). This is consistent with the conserved architecture of all cellular RNA polymerases, which requires that RNA and DNA strands exit at different sites from the enzyme (Cramer et al., 2001, 2008; Engel et al., 2013; Hirata et al., 2008; Korkhin et al., 2009; Zhang et al., 1999).

The process of R-loop formation necessitates a competition between the nascent RNA and the non-template DNA strand to hybridize with the template strand. Therefore, hybrid

formation should be thermodynamically favorable when compared to reannealing of the DNA duplex in the R-loop forming region. Indeed, synthetic RNA:DNA hybrid structures with a high RNA-purine/DNA-pyrimidine ratio were shown to be more stable than a DNA:DNA duplex of the same sequence composition (Hall and McLaughlin, 1991; Ratmeyer et al., 1994; Roberts and Crothers, 1992). Thus, a high guanine (G) density in the non-template DNA strand promotes R-loop formation *in vitro* and *in vivo* (Roy et al., 2008; Yu et al., 2003). More precisely, one or two clusters of consecutive (3 or more) G residues in the R-loop initiating zone efficiently nucleates hybrid formation, whereas a high G density (but not G clustering) is sufficient for elongation of the R-loop (**Figure 2A**, Roy and Lieber, 2009).

Other factors on the non-template strand may also drive R-loop formation. A recent study showed that nicks in the non-template DNA strand reduce its ability to reanneal to the template strand, thereby favoring hybridization of the RNA and R-loop formation, even in the absence of G clusters (Roy et al., 2010). As nicks are created frequently in the genome by exogenous or other endogenous sources of DNA damage, this finding may further expand the possible R-loop forming regions *in vivo*. Also on the non-template strand, clusters of G-rich sequences have the potential to fold into a secondary, non-B-form DNA structure referred to as G-quadruplex or G4 DNA (reviewed in Bochman et al., 2012). G4 DNA is characterized by the association of four guanines bound through Hoogsteen base pairing and variable stacks of guanine quartet planes (**Figure 2B**, Burge et al., 2006). Intriguingly, G-quadruplex structures have been directly observed during *in vitro* transcription of S regions by electron microscopy (Duquette et al., 2004), and were recently shown to be stable structures, detectable by immunostaining and sequencing analysis, in human genomic DNA (Henderson et al., 2013; Lam et al., 2013). The high stability of G4 DNA may help stabilize the single-stranded tract of DNA in an R-loop structure, making it tempting to speculate that formation of G4 DNA on the non-template strand may facilitate or contribute to RNA:DNA hybrid formation during transcription *in vivo* (**Figure 2B**).

2.2 Torsional stress and transcription-dependent supercoiling

The inherent conformation of the DNA double-helix also influences the propensity of the nascent RNA strand to invade duplex DNA. According to the twin-domain model, the elongating RNAP complex creates positive supercoiling ahead of, and negative supercoiling behind the enzyme (Liu and Wang, 1987; Wu et al., 1988). Positive supercoils can impede further DNA unwinding and block transcription (Roca, 2011), whereas excessive negative supercoiling imparts single-stranded character to the DNA duplex and promotes melting of susceptible DNA sequences (Vologodskii and Cozzarelli, 1994). In *E. coli*, the topology of its circular DNA is regulated by the combined activities of DNA gyrase and topoisomerase I (TopA) which introduce or resolve negative supercoils, respectively (reviewed in Drolet, 2006). TopA null mutants accumulate excessive negative supercoiling that leads to defects in full length RNA synthesis and growth inhibition under certain conditions (Baaklini et al., 2004, 2008; Drolet et al., 1994). Importantly, these phenotypes can be partially rescued by overexpression of RNase H (Baaklini et al., 2004; Drolet et al., 1995), supporting the notion that negative supercoiling facilitates the annealing of the RNA strand to the DNA. A connection between R-loop formation and hypernegative DNA supercoiling can also be

found in the highly transcribed ribosomal DNA (rDNA) cluster of *S. cerevisiae*. Similar to the specialized roles of TopA and DNA gyrase in *E. coli*, yeast Top1 mainly resolves the negative torsional stress behind RNAPI, whereas Top2 resolves positive supercoiling in front of it (French et al., 2011). Interestingly, chromatin immunoprecipitation (ChIP) experiments showed an increase in RNA:DNA hybrids along the rDNA cluster in mutant yeast cells lacking Top1 and/or Top2. Strikingly, R-loops were also detectable in wild-type yeast, with a peak in the 5'-region of the 18S rDNA (El Hage et al., 2010). Thus, torsional stress may transiently increase, especially in genes with high transcription initiation rates, such that it cannot be completely relieved even in the presence of active topoisomerases. This idea is supported by single-molecule experiments showing that topoisomerases are not processive enough to keep pace with transcription-coupled supercoiling (Darzacq et al., 2007; Koster et al., 2005). Therefore, the transient torsional force created in highly transcribed genes may result in partial melting of recently transcribed sequences, allowing RNA access to its DNA template (**Figure 2B**). Moreover, it is possible that R-loops arising from torsional stress may significantly contribute to the increased TAR and TAM rates observed in highly transcribed genes.

2.3 R-loop formation in cis or in trans?

Thus far, much of the work in this field has focused on RNA:DNA hybrids that form in *cis* by cotranscriptional re-annealing of the nascent transcript near the transcribing RNAP (Aguilera and García-Muse, 2012; Huertas and Aguilera, 2003). However, *in vitro* studies on bacterial RecA, a DNA recombinase that promotes strand-exchange during homologous recombination (Bell, 2005), showed that RecA can also promote invasion of a homologous RNA strand into the DNA duplex to form an R-loop in the absence of transcription (Kasahara et al., 2000; Zaitsev and Kowalczykowski, 2000). The possibility that the yeast ortholog of RecA, Rad51, could promote R-loop formation post-transcriptionally (in *trans*) was recently investigated by Wahba and colleagues. The authors inserted the same DNA sequence into a genomic locus and a yeast artificial chromosome (YAC). Surprisingly, upon induction of the chromosomal transcript, Rad51-dependent RNA:DNA hybrids were detected at the YAC repeat, distant from the original site of transcription. Importantly, this apparent trans-induced R-loop also affected stability of the YAC DNA (Wahba et al., 2013). If this alternative mechanism of R-loop formation also occurs in higher eukaryotes, this could be deleterious for the cells. For example, transcripts derived from repetitive elements could form R-loops and induce DNA breaks at multiple target sites in the genome. Further research will be necessary to assess the genome-wide frequency of Rad51-dependent post-transcriptional R-loop formation, compared to the previously postulated cotranscriptional mechanism.

3. RNA processing and metabolizing factors that prevent and resolve R-loops

3.1 RNA processing factors

As R-loop formation depends on the availability of the nascent RNA strand to re-hybridize with the template DNA strand, it follows that limiting the amount of naked RNA in the cell could help preventing the accumulation of R-loops. In bacteria, the absence of a nuclear

membrane facilitates the coupling of transcription and translation by allowing ribosomes to be loaded onto the nascent mRNA molecule as it emerges from the RNAP (Gowrishankar and Harinarayanan, 2004). This coupling prevents the nascent RNA molecule from annealing back to the bacterial chromosome (**Figure 3A**). In eukaryotes, transcription and translation of the mRNA are uncoupled due to the separation of the nucleus and cytoplasm. However, the nascent mRNA is co-transcriptionally assembled into ribonucleoprotein (RNP) complexes that, after further processing, result in export-competent mRNP particles that can be transported to the cytoplasm (Han et al., 2011; Luna et al., 2008; Perales and Bentley, 2009). Over the past several years, an increasing number of factors have been discovered that function at the interface of transcription and mRNP processing. Absence or mutation of these factors results in a wide variety of phenotypes including thermosensitivity, hyperrecombination, transcriptional impairment, mRNA export defects and genome instability. The emerging view is that co-transcriptional association of these factors may restrict interactions between the nascent RNA and template DNA in order to prevent the accumulation of R-loops and associated genomic instability in prokaryotic and eukaryotic genomes (**Figure 3A**) (Luna et al., 2012; Paulsen et al., 2009; Rodríguez-Navarro and Hurt, 2011; Rondón et al., 2010; Stirling et al., 2012; Wahba et al., 2011).

THO/TREX and THSC/TREX-2 complexes—The first evidence for a connection between R-loops and mRNA processing was derived from studies of the hyperrecombination and transcription elongation phenotypes observed in yeast *hpr1* mutants (Aguilera and Klein, 1990; Chávez and Aguilera, 1997). Hpr1 is a component of the conserved THO complex consisting of Hpr1, Tho2, Mft1 and Thp2 in yeast (Chávez et al., 2000). Under mild conditions, THO components co-purify with the ATP-dependent RNA helicase Sub2/UAP56 and the mRNA export adaptor Yra1/Aly, forming the TREX (transcription/export) complex (Peña et al., 2012; Strässer et al., 2002). It is thought that THO is recruited to active sites of transcription (Strässer et al., 2002; Zenklusen et al., 2002), a conclusion that is supported by a recent genome-wide study in yeast showing that Hpr1 and Sub2 co-localize to the majority of RNAPII-transcribed genes (Gómez-González et al., 2011). Interestingly, Hpr1 and Sub2 mutants show decreased expression of certain long, GC-rich and/or highly transcribed genes, which are preferential substrates for R-loop formation (see 2.1 and 2.2) (Gómez-González et al., 2011). Importantly, overexpression of RNase H can partially rescue the phenotypes in THO- and TREX-complex mutant cells (Huertas and Aguilera, 2003), indicating that transcription arrest and hyperrecombination in the absence of these RNA-processing factors is partially R-loop dependent.

Another functional module connecting transcriptional elongation with mRNP export is the conserved THSC or TREX-2 complex consisting of Thp1, Sac3, Sus1, Cdc31 and Sem1 (Faza et al., 2009; Fischer et al., 2004; Gallardo et al., 2003; Rodríguez-Navarro et al., 2004). This complex is located at the inner face of the nuclear pore, where it interacts with the nucleoporins Nup1 and Nup60 (Fischer et al., 2002). Despite acting at a different step in the mRNP maturation process than the THO/TREX complex, mutants of the TREX-2 complex share similar transcription defects and hyper-recombination phenotypes as THO complex mutants, and these phenotypes can be rescued by RNase H treatment (Gallardo et al., 2003; González-Aguilera et al., 2008). A potential bridge between transcriptional

regulation and the nuclear pore is the TREX-2 protein Sus1, which is also a component of the SAGA transcription initiation complex (Galán and Rodríguez-Navarro, 2012; Köhler et al., 2006). This could indicate that activated genes may be physically coupled to the nuclear pore, facilitating the export and reducing the half-life of the newly synthesized RNA in the nucleus. Thus, two strategies may be used to prevent the nuclear accumulation of R-loops in a temporal and spatial manner. Cells may protect the nascent RNA by co-transcriptional association of ribonucleoprotein factors, and/or they may limit the nuclear presence of RNA by coupling transcription to RNA export. These two methods are not mutually exclusive.

The splicing factor ASF1/SF2—In vertebrate cells, R-loop mediated genome instability was first documented in a chicken DT40 B cell line depleted of the splicing factor ASF1/SF2 (alternative splicing factor 1/splicing factor 2, hereafter ASF). Inactivation of ASF, a member of the SR (serine/arginine rich) protein family (reviewed in Sanford et al., 2003), resulted in G2 cell-cycle arrest, DSB formation and fragmentation of chromosomal DNA (Li and Manley, 2005; Li et al., 2005). Strikingly, a specific DSB was detected upon depletion of ASF in a locus prone to R-loop formation. Moreover, expression of RNase H suppressed DSB formation and hypermutation phenotypes observed in the absence of ASF, suggesting that formation of R-loops might be the source of the DNA damage (Li and Manley, 2005). Interestingly, ASF and other SR family proteins are regulated by the kinase activity of human topoisomerase I (Rossi et al., 1996), which also resolves torsional stress during transcription. Both of these functions may contribute to the ability of topoisomerase I to suppress R-loop formation and resolve conflicts between transcription and replication complexes (see 4.2) (Tuduri et al., 2009).

Other RNA processing factors—In the last few years, several unbiased large-scale genetic and proteomic screens have greatly extended the list of RNA processing factors involved in preventing genomic instability with a potential link to R-loops. Strikingly, a genome-wide siRNA screen that utilized the phosphorylation of histone H2AX (γ -H2AX) as a marker of DNA damage showed that more than 80 genes coding for mRNP processing factors suppressed DNA damage accumulation in human cells. Importantly, for a subset of the hits, the accumulation of γ -H2AX foci could be suppressed by overexpression of RNase H (Paulsen et al., 2009), including the putative RNA helicase AQR (Sam et al., 1998), Cdc40/Prp17 and Skiip/Snw1, two factors involved in splicing (Folk et al., 2004; Ben Yehuda et al., 1998), and several snRNP proteins including Snrpa1, Snrpd1 and Snrpd3. A second genetic screen, which monitored chromosome stability at both an artificial and endogenous locus in yeast, also linked RNA:DNA hybrid formation to genome instability. In this case, the mutants found to suppress hybrid formation were involved not only in mRNA processing, but in various stages of RNA metabolism from transcription initiation to RNA degradation and export. The authors also reported that the *sin3* mutant specifically enhanced the formation of R-loops and genetic instability at the rDNA repeats, which may reflect the importance of Sin3 for transcriptional regulation at this locus in yeast (Wahba et al., 2011). Notably, Sin3 is a component of the Rpb3 histone deacetylase complex involved in transcriptional regulation (Silverstein and Ekwall, 2005), highlighting the interconnection between (de)regulation of transcription and R-loop formation. A third screen, which monitored the formation of Rad52 positive recombination centers in yeast, led to the

identification of factors involved in mRNA cleavage and polyadenylation that suppress RNA:DNA hybrid formation. One of these factors has a human ortholog, FIP1L1, that has also been shown to prevent DNA damage and chromosome breakage (Stirling et al., 2012). Together, these studies implicate multiple aspects of mRNP biogenesis in the suppression of RNA:DNA hybrid formation, including transcription, splicing, modification and export of the nascent RNAs. Thus, proper maturation of the RNP particles and their delivery to the cytoplasm may be crucial to prevent genomic instability (**Figure 3A**). Interestingly, proteomic screens for substrates of DNA damage response kinases like ATM (ataxia telangiectasia mutated) and ATR (ATM and Rad3-related) have also identified proteins involved in RNA processing (Matsuoka et al., 2007; Smolka et al., 2007). Although not tied to R-loop formation directly, these findings may indicate that the DNA damage response regulates R-loop processing potentially by modification of RNA processing factors in the cell. In the future, it will be important to determine if the genome instability associated with these factors stems from R-loops themselves, or if other effects or molecular intermediates like DNA supercoiling, arrest of RNAP, chromatin structure or nuclear redistribution contribute to the genomic instability observed in the absence of these proteins. Moreover, it will be interesting to analyze if different mRNA biogenesis factors act in specific genomic regions to prevent R-loop formation.

3.2 RNase H enzymes

If the mechanisms to prevent R-loop formation fail, there are additional mechanisms to resolve or remove these structures that can be utilized. The RNase H family of enzymes, which degrades the RNA moiety of an RNA:DNA hybrid in a sequence-independent manner, plays a crucial role in this process (**Figure 3B**). Most organisms encode two types of RNase H enzymes that are classified according to sequence conservation and substrate specificity. Eukaryotic RNase H1 is similar to prokaryotic RNase HI, whereas eukaryotic RNase H2 is a heterotrimeric complex, unlike its counterpart in prokaryotes which is monomeric. Both eukaryotic RNase H machineries can cleave the RNA in the extended RNA:DNA hybrid structure found within R-loops, albeit with different efficiencies (reviewed in Cerritelli and Crouch, 2009). However, both the prokaryotic RNase HII and eukaryotic RNase H2 can also remove single ribonucleotides misincorporated during DNA replication. Intriguingly, deletion of one or both RNase H enzymes in yeast increases the rate of YAC loss and chromosome instability, and 97% of the nuclei in the double mutant show a significant increase in hybrid formation (Wahba et al., 2011). Moreover, overexpression of RNase H enzymes has been shown to suppress phenotypes arising from R-loop formation in yeast and higher eukaryotes, suggesting that the nuclear activity of these enzymes is required to safeguard the genome from the deleterious consequences of R-loop formation. Recently, loss of RNase H2 in yeast was shown to alter the expression level of 349 genes (Arana et al., 2012) and R-loops uniquely processed by RNase H2 were revealed with a mutant that can resolve R-loops but cannot remove single misincorporated ribonucleotides in DNA (Chon et al., 2013). Future research will show if the two different RNase H complexes remove R-loops from overlapping or distinct sets of target genes. In the latter case, the specificity of RNase H1 and RNase H2 activity at a given gene will be important to pursue.

3.3 RNA:DNA helicases

In addition to nucleolytic cleavage of the RNA strand by RNase H, R-loops can also be resolved through unwinding of the RNA:DNA hybrid by a helicase (**Figure 3C**). In fact, a number of helicases in prokaryotes have been shown to unwind RNA:DNA hybrids *in vitro*. The DNA translocase RecG from *E. coli*, for example, targets a variety of branched substrates including Holliday junctions, three-strand junctions, D-loops and R-loops (reviewed in Rudolph et al., 2010). Interestingly, yeast lacking RNase H alone are viable, whereas strains devoid of both RNase H and the helicase RecG are not (Hong et al., 1995). This supports the notion that RecG is involved in R-loop removal *in vivo*, and suggests that this function becomes essential in the absence of RNase H. One of the major transcription termination factors in bacteria, known as Rho, also has RNA:DNA helicase activity. This homohexameric ring protein terminates the synthesis of transcripts by promoting ATP-dependent dissociation of the RNAP from the RNA (Peters et al., 2011; Richardson, 2003). Intriguingly, a mutant defective in Rho-dependent termination showed a genome-wide increase in RNA:DNA hybrids, as inferred from the higher rate of C-T conversions after bisulfite treatment (Leela et al., 2013). This result suggests that during transcription termination, the 5' to 3' RNA translocase activity of Rho may also function as a surveillance mechanism to protect the genome from R-loop formation.

Eukaryotes also express numerous DNA and RNA helicases that are able to unwind RNA:DNA hybrids, including the family of Pif1 helicases (Boulé and Zakian, 2007), human DHX9/RHA (RNA helicase A) (Chakraborty and Grosse, 2011) and Senataxin (Sen1 in yeast) (Kim et al., 1999). Senataxin was initially identified as a RNAPII transcription termination factor of mostly non-polyadenylated small RNAs, in complex with the two RNA-binding proteins Nrd1 and Nab3 (Hazelbaker et al., 2013; Steinmetz et al., 2001; Ursic et al., 1997). The N-terminus of yeast Sen1 shows additional interactions with the C-terminal domain (CTD) of RNAPII, ribonuclease III and the NER factor Rad2/XPG (Ursic et al., 2004), whereas the C-terminus contains the essential helicase domain. Interestingly, the distribution of RNAPII complexes was altered genome-wide in a helicase mutant of Sen1, suggesting that Sen1 regulates transcription (Steinmetz et al., 2006). Recently, Mischo and colleagues demonstrated that Sen1 plays a key role in resolving RNA:DNA hybrids and preventing transcription-associated instability (Mischo et al., 2011). Moreover, a fraction of Sen1 associates with replication forks, likely protecting the integrity of forks that encounter highly expressed RNAPII genes (Alzu et al., 2012). These additional functions of Sen1 seem to be conserved, as human Senataxin (SETX) is also needed for resolution of RNA:DNA hybrids at G-rich pause sites downstream of the polyadenylation signal, thereby promoting cleavage and degradation of the RNA by the 5' to 3' exoribonuclease Xrn2 (Skourti-Stathaki et al., 2011). SETX also localizes to collision sites of the replication and transcription machinery, as indicated by co-localization with the DNA damage marker 53BP1 in response to replication blockage (Yüce and West, 2013). Interestingly, Setx knockout mice exhibit defects in spermatogenesis, meiotic homologous recombination and sex chromosome inactivation (Becherel et al., 2013), indicating widespread functions for this protein. Finally, mutations in human SETX are strongly linked to the neurodegenerative disorders ataxia with oculomotor apraxia 2 (AOA2) and amyotrophic lateral sclerosis type 4 (ALS4) (Chen et al., 2004; Moreira et al., 2004), although future research will be necessary to establish whether

there is a molecular link between Senataxin, R-loop formation and the disease-related phenotypes.

4. Possible mechanisms for R-loop-mediated genome instability

As outlined above, cells use a variety of co- and post-transcriptional processes to prevent the increased mutation and recombination rates associated with R-loop formation. However, reannealing of the nascent RNA transcript to the DNA template strand does not damage or mutate the DNA template per se. Rather, these structures are likely processed into DSBs or other unusual intermediates, giving rise to the observed genomic instability. Research in the last few years has begun to shed light on the different molecular mechanisms that may underlie or contribute to the DNA damage associated with these RNA:DNA hybrid structures.

4.1 The ssDNA of the R-loop as a source of single-strand breaks

Single-stranded DNA (ssDNA) is chemically more unstable and susceptible to DNA-damaging agents than duplex DNA (Lindahl, 1993). This is consistent with earlier studies showing that transcription-associated mutations primarily affect the non-template strand (Beletskii and Bhagwat, 1996; Kim and Jinks-Robertson, 2012). Extensive R-loop formation may considerably increase the amount of ssDNA in the genome by exposing the non-template DNA strand as a tract of vulnerable ssDNA. Thus, one reasonable hypothesis for the generation of mutagenic/recombinogenic lesions by R-loops is that they are a product of the exposed ssDNA portion of the non-template strand. There are several ways in which ssDNA can be damaged. One possibility is that specific protein factors recognize and modify the ssDNA (**Figure 4A**). One candidate factor is AID, a DNA-specific cytidine deaminase that converts dC to dU residues during CSR (Muramatsu et al., 2000; Petersen-Mahrt et al., 2002; Sohail et al., 2003). AID has high sequence homology to a class of DNA- and RNA-editing enzymes of the Apolipoprotein B mRNA-Editing Catalytic Polypeptide (APOBEC) family. Human cells encode AID and 10 other APOBECs, which are expressed in a large variety of cell types and are effective mutators of DNA and RNA *in vitro* (Beale et al., 2004; Conticello, 2008). Thus, AID/APOBEC enzymes could mutate dC to dU residues in the R-loop, in analogy to the normal mechanism of CSR. The BER enzyme uracil DNA glycosylase could then excise the uracil base to create an abasic site and generate the initiating DNA lesion (**Figure 4A**, left panel).

A second possibility is that Top1 activity may be a source of TAM in eukaryotic cells. Using a reporter sequence under control of a strong promoter, it was shown that characteristic 2-3bp deletions accumulate at discrete hotspots in a Top1-dependent manner (Lippert et al., 2011; Takahashi et al., 2011). Presumably, Top1 is recruited to regions of high transcriptional activity in order to resolve superhelical stress in the DNA template. However, it can also be irreversibly trapped during its cleavage-ligation cycle, giving rise to a covalent Top1-DNA complex attached to the 5' end of the nicked DNA. This irreversible complex is likely processed into a gap by specific endonucleases like Rad1/Rad10 or Mus81/Mms4 (Takahashi et al., 2011). Thus, the Top1-DNA complex could give rise to a break in the ssDNA portion of the R-loop (**Figure 4A**, middle panel). If the complementary DNA strand is also trapped in the RNA:DNA hybrid, these breaks may be difficult to repair.

Alternatively, if the transiently displaced ssDNA of the R-loop forms a G-quadruplex structure (see 2.1, **Figure 4A**), this secondary structure element could allow targeting of the R-loop by other specific factors in the cell. Indeed, a human G4-specific endonuclease activity that cleaves within the single-stranded region 5' of the stacked G quartets has been characterized (Sun et al., 2001). In this context, it is unlikely that the displaced ssDNA of an R-loop stays unprotected or “naked” in the cell. Due to the abundance and high binding affinity of replication protein A (RPA) for ssDNA (Oakley and Patrick, 2010), it seems likely that this ssDNA is also coated with RPA (**Figure 2C**). Interestingly, RPA was shown to interact with AID to promote deamination of somatic hypermutation targets (Chaudhuri et al., 2004), and a recent study showed that RPA accumulates during S-G2/M phases of the cell cycle at AID target regions (Yamane et al., 2013). Another interesting finding is that RPA couples incision to DNA repair synthesis during transcription-coupled nucleotide excision repair (TC-NER), thereby preventing further generation of DNA strand breaks that could lead to mutagenic and recombinogenic events (Overmeer et al., 2011). Future research will be necessary to determine the role of RPA in the processing of R-loops to single- or double-strand breaks.

In addition to ssDNA regions, other structural features of R loops may also contribute to making these structures targets for certain nucleases. Each R loop contains two duplex–single strand junctions, and several structure-specific endonucleases are able to recognize and cleave such loop-duplex junctions (Hanawalt and Spivak, 2008; Tian and Alt, 2000). Thus, flap-endonucleases could be an alternative source for ssDNA breaks in the R-loop structure (**Figure 4A**, right panel).

4.2 Converting R-loops into DSBs – conflicts between replication and transcription machineries

Several pieces of data suggest that aberrant processing or a failure to resolve R-loop structures gives rise to DSBs. As outlined above, there are multiple plausible mechanisms to create a SSB in the R-loop structure. However, an intriguing question remains -- how are R-loop structures processed into deleterious DSBs *in vivo*? One possibility is derived from the current model of DSB formation during CSR. It has been suggested that two closely spaced AID-mediated SSBs on opposite strands result in the necessary DSB intermediate to induce recombination at the S regions (Xu et al., 2012). Accordingly, it can be speculated that AID/APOBEC-mediated deamination and BER, trapping of Top1-DNA complexes and/or the activity of different endonucleases (see 4.1) work together to create two adjacent DNA breaks, inducing the DNA DSB and resulting chromosomal fragility (**Figure 4B**, 1 and 2).

A large body of evidence, however, suggests that R-loop-mediated DNA damage involves the process of DNA replication. In yeast THO mutants, for example, transcription-dependent hyperrecombination only occurs in DNA sequences transcribed in S-phase but not in G2-phase (Wellinger et al., 2006). Thus, it was postulated that R-loop formation leads to RNA polymerase pausing, which interferes with replication fork progression through collision of the replication fork and the transcription machinery (**Figure 4B**). Numerous recent findings support this concept. A genome-wide study in yeast showed that replication forks pause more frequently in highly transcribed RNAPII genes (Azvolinsky et al., 2009). Importantly,

Rrm3, a helicase which helps to resolve replication obstacles, is also enriched at highly transcribed genes, and this enrichment is reduced by expression of RNase H (Gómez-González et al., 2011). Moreover, Top1-deficient human cells accumulate stalled replication forks and chromosome breaks in S phase, and these phenotypes can be suppressed by overexpression of RNase H. In these cells, the DSB marker, γ H2AX, is also observed predominantly in transcribed regions of the genome (Tuduri et al., 2009), suggesting that breaks may arise from transcription-replication collisions. Along these lines, the longest human genes need more than one cell cycle to be transcribed, making the collision with a replication fork inevitable. Interestingly, a small subset of these sequences are common fragile sites (CFS), whose breakage can be partially rescued by overexpression of RNase H1 (Helmrich et al., 2011). This indicates that R-loops may contribute to the formation of certain fragile replication sites, although other mechanisms also appear to underlie the instability of these and other fragile sites (Debatisse et al., 2012; Le Tallec et al., 2013). Another link between R-loop structures and DNA replication is provided by the RNA:DNA helicase Senataxin (see 3.3), which co-localizes to replication forks where it may assist fork progression through RNAPII-transcribed genes (Alzu et al., 2012). Finally, another study showed that R-loops formed in the absence of ASF in human and chicken cells result in increased replication fork asymmetry, indicating increased levels of stalled replication forks (Gan et al., 2011). Together, these results clearly link R-loop processing and replication fork impairment, but the mechanism behind how R-loop structures interfere with DNA replication to induce DSBs is currently unclear.

The replisome and the transcription machinery can collide when polymerases move in the same direction but with different velocities (co-directional collision) or when they converge upon each other (head-on collision). If a persistent R-loop has already formed a SSB in the non-transcribed strand (see 4.1, **Figure 4A**), continued separation of the parental strands would result in a DSB on the lagging (co-directional collision) or leading strand (head-on collision) (**Figure 4B**, 3-8). Alternatively, RNA:DNA hybrids could directly impede the progression of the replisome and lead to DSBs without the requirement of an initiating lesion. Experimental evidence suggests that impairment of replication fork progression is mainly induced by head-on collisions (Prado and Aguilera, 2005). Since positive DNA supercoiling accumulates in front of both RNA and DNA polymerases (Wang, 2002), the enzymes would create strong torsional stress when they converge. The RNA:DNA hybrid may aggravate progression of the fork or interfere with fork restart, finally leading to DSBs by fork stalling, reversal, and/or collapse (**Figure 4B**, 10). However, co-directional collisions between the replisome and the R-loop structure (**Figure 4B**, 9) may also contribute to the observed genomic instability. Moreover, it cannot be excluded that the RNA strand of the R-loop provokes aberrant DNA replication by providing the stalled replication machinery a 3'-OH end to initiate replication. Future research will be necessary to answer these questions.

5. Concluding remarks and future perspectives

Unscheduled formation of RNA:DNA hybrids in the genome creates harmful intermediates, and a failure to resolve or to remove them can have deleterious consequences for genome stability. Not surprisingly, cells have evolved multiple strategies and pathways to prevent

their formation and to resolve or process these structures *in vivo*. In this review, we have summarized the current knowledge on the occurrence of R-loops, and their potential consequences for genome stability. R-loop formation is prevented by several independent factors and pathways in the cell, including the RNase H enzymes and RNA:DNA helicases which are thought to cleave or unwind the RNA:DNA hybrid, respectively. Studies in the last few years have also yielded a large list of factors and protein complexes involved in mRNP biogenesis that act at the intersection of transcription and mRNP processing to prevent R-loop formation. Thus, proper packaging and maturation of the RNA into functional mRNP particles seems to be a common strategy to prevent re-annealing of the nascent RNA strand with the DNA template. If an R-loop still forms, despite the overlapping mechanisms in place to prevent this, nicking of the vulnerable exposed ssDNA, processing of the R-loop by structure-specific nucleases or conflicts of the stalled RNAP with the replication machinery may all contribute to the generation of DNA breaks and ultimately chromosomal rearrangements and genomic instability.

There are still many unresolved questions in R-loop biology that await future research. For example, are all R-loops considered equal? It is possible that different types of R-loops in different genomic loci are processed by different pathways in the cell. R-loops containing G-quadruplex or other secondary structures in the ssDNA region may be targeted by different factors than R-loops without such additional elements. How can the cell distinguish between R-loops with a beneficial role for the cell compared to those which form in an aberrant manner and have potentially genome destabilizing consequences? It is also unclear how the local chromatin structure influences the formation and processing of R-loops in a transcription unit. Highly transcribed genes are largely devoid of nucleosomes (Kristjuhan and Svejstrup, 2004; Merz et al., 2008), and the resulting open conformation of chromatin may facilitate the formation of R-loops. Moreover, efficient histone deposition and chromatin reassembly behind the elongating RNAP may be another strategy for eukaryotic cells to avoid extensive R-loop formation (Schwabish and Struhl, 2004). Thus, the deregulation of transcription and the resulting changes of the chromatin structure may also play a role in R-loop formation and genome instability.

Finally, it will be interesting to determine how the R-loop containing DNA recruits processing factors like RNase H or RNA:DNA helicases. In human cells, R-loops accumulate downstream of unmethylated CpG-island promoters, and it has been proposed that the ssDNA attracts histones enriched in H3K4me3 marks and prevents genomic DNA methylation (Ginno et al., 2012). Interestingly, a recent study from the Aguilera lab showed that yeast mutants with high levels of R-loops accumulate histone H3S10 phosphorylation, resulting in chromatin compaction of the R-loop containing region (Castellano-Pozo et al., 2013). Therefore, it will be of great interest to explore the regulatory role of chromatin on R-loop formation and genomic instability. Unfortunately, the transient nature of RNA:DNA hybrids complicates their detection and biochemical analysis in a chromatin context, and it is likely that only a fraction of the participating proteins and machineries have been identified or characterized so far. Thus, a major challenge in the future will be to develop new methods to monitor the fate of specific R-loops under defined conditions *in vivo*, and gain further mechanistic insight into the role of R-loops in cells.

Acknowledgments

We would like to thank all the scientists whose results we discussed in this review, and we apologize to those whose work we could not cite due to space limitations. We are especially grateful to all members of the Cimprich lab for all their inspiring input, helpful comments on the manuscript and their constant support. This work was supported by a fellowship from the German Research Foundation (DFG) to S.H. and by awards from the Komen Foundation (IIR 12222368) and NIH (GM100489) to K.A.C.

Abbreviations

DSB(s)	Double-Strand Break(s)
RNAP(s)	RNA Polymerase(s)
TAM	Transcription-Associated Mutagenesis
TAR	Transcription-Associated Recombination
ssDNA	single-stranded DNA
CSR	Class Switch Recombination
AID	Activation Induced Deaminase
SSB(s)	Single-Strand Break(s)
rDNA	ribosomal DNA
ChIP	Chromatin Immuno-Precipitation
RNP	Ribonucleoprotein
CIN	Chromosome Instability
CTD	Carboxy-Terminal Domain
CFS	Common Fragile Site
APOBEC	Apolipoprotein B mRNA-Editing Catalytic poly-peptide

References

- Aguilera A, García-Muse T. R loops: from transcription byproducts to threats to genome stability. *Mol. Cell.* 2012; 46:115–124. [PubMed: 22541554]
- Aguilera A, García-Muse T. Causes of genome instability. *Annu. Rev. Genet.* 2013; 47:1–32. [PubMed: 23909437]
- Aguilera A, Gómez-González B. Genome instability: a mechanistic view of its causes and consequences. *Nat. Rev. Genet.* 2008; 9:204–217. [PubMed: 18227811]
- Aguilera A, Klein HL. HPR1, a novel yeast gene that prevents intrachromosomal excision recombination, shows carboxy-terminal homology to the *Saccharomyces cerevisiae* TOP1 gene. *Mol. Cell. Biol.* 1990; 10:1439–1451. [PubMed: 2181275]
- Alzu A, Bermejo R, Begnis M, Lucca C, Piccini D, Carotenuto W, Saponaro M, Brambati A, Cocito A, Foiani M, et al. Senataxin associates with replication forks to protect fork integrity across RNA-polymerase-II-transcribed genes. *Cell.* 2012; 151:835–846. [PubMed: 23141540]
- Arana ME, Kerns RT, Wharey L, Gerrish KE, Bushel PR, Kunkel TA. Transcriptional responses to loss of RNase H2 in *Saccharomyces cerevisiae*. *DNA Repair.* 2012; 11:933–941. [PubMed: 23079308]

- Azvolinsky A, Giresi PG, Lieb JD, Zakian VA. Highly transcribed RNA polymerase II genes are impediments to replication fork progression in *Saccharomyces cerevisiae*. *Mol. Cell.* 2009; 34:722–734. [PubMed: 19560424]
- Baaklini I, Hraiky C, Rallu F, Tse-Dinh Y-C, Drolet M. RNase HI overproduction is required for efficient full-length RNA synthesis in the absence of topoisomerase I in *Escherichia coli*. *Mol. Microbiol.* 2004; 54:198–211. [PubMed: 15458416]
- Baaklini I, Usongo V, Nolent F, Sanscartier P, Hraiky C, Drlica K, Drolet M. Hypernegative supercoiling inhibits growth by causing RNA degradation. *J. Bacteriol.* 2008; 190:7346–7356. [PubMed: 18790862]
- Baldacci G, Chérif-Zahar B, Bernardi G. The initiation of DNA replication in the mitochondrial genome of yeast. *EMBO J.* 1984; 3:2115–2120. [PubMed: 6092055]
- Beale RCL, Petersen-Mahrt SK, Watt IN, Harris RS, Rada C, Neuberger MS. Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. *J. Mol. Biol.* 2004; 337:585–596. [PubMed: 15019779]
- Becherel OJ, Yeo AJ, Stellati A, Heng EYH, Luff J, Suraweera AM, Woods R, Fleming J, Carrie D, McKinney K, et al. Senataxin plays an essential role with DNA damage response proteins in meiotic recombination and gene silencing. *PLoS Genet.* 2013; 9:e1003435. [PubMed: 23593030]
- Beletskii A, Bhagwat AS. Transcription-induced mutations: increase in C to T mutations in the nontranscribed strand during transcription in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 1996; 93:13919–13924. [PubMed: 8943036]
- Bell CE. Structure and mechanism of *Escherichia coli* RecA ATPase. *Mol. Microbiol.* 2005; 58:358–366. [PubMed: 16194225]
- Bermejo R, Kumar A, Foiani M. Preserving the genome by regulating chromatin association with the nuclear envelope. *Trends Cell Biol.* 2012; 22:465–473. [PubMed: 22771046]
- Bochman ML, Paeschke K, Zakian VA. DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.* 2012; 13:770–780. [PubMed: 23032257]
- Boguslawski SJ, Smith DE, Michalak MA, Mickelson KE, Yehle CO, Patterson WL, Carrico RJ. Characterization of monoclonal antibody to DNA:RNA and its application to immunodetection of hybrids. *J. Immunol. Methods.* 1986; 89:123–130. [PubMed: 2422282]
- Boulé J-B, Zakian VA. The yeast Pif1p DNA helicase preferentially unwinds RNA DNA substrates. *Nucleic Acids Res.* 2007; 35:5809–5818. [PubMed: 17720711]
- Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* 2006; 34:5402–5415. [PubMed: 17012276]
- Castellano-Pozo M, Santos-Pereira JM, Rondón AG, Barroso S, Andújar E, Pérez-Alegre M, García-Muse T, Aguilera A. R Loops Are Linked to Histone H3 S10 Phosphorylation and Chromatin Condensation. *Mol. Cell.* 2013
- Cerritelli SM, Crouch RJ. Ribonuclease H: the enzymes in eukaryotes. *FEBS J.* 2009; 276:1494–1505. [PubMed: 19228196]
- Chakraborty P, Grosse F. Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA Repair.* 2011; 10:654–665. [PubMed: 21561811]
- Chaudhuri J, Alt FW. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nat. Rev. Immunol.* 2004; 4:541–552. [PubMed: 15229473]
- Chaudhuri J, Khuong C, Alt F. Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. *Nature.* 2004; 430:992–998. [PubMed: 15273694]
- Chávez S, Aguilera A. The yeast HPR1 gene has a functional role in transcriptional elongation that uncovers a novel source of genome instability. *Genes Dev.* 1997; 11:3459–3470. [PubMed: 9407037]
- Chávez S, Beilharz T, Rondón AG, Erdjument-Bromage H, Tempst P, Svejstrup JQ, Lithgow T, Aguilera A. A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in *Saccharomyces cerevisiae*. *EMBO J.* 2000; 19:5824–5834. [PubMed: 11060033]

- Chen Y-Z, Bennett CL, Huynh HM, Blair IP, Puls I, Irobi J, Dierick I, Abel A, Kennerson ML, Rabin BA, et al. DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). *Am. J. Hum. Genet.* 2004; 74:1128–1135. [PubMed: 15106121]
- Chon H, Sparks JL, Rychlik M, Nowotny M, Burgers PM, Crouch RJ, Cerritelli SM. RNase H2 roles in genome integrity revealed by unlinking its activities. *Nucleic Acids Res.* 2013; 41:3130–3143. [PubMed: 23355612]
- Conticello SG. The AID/APOBEC family of nucleic acid mutators. *Genome Biol.* 2008; 9:229. [PubMed: 18598372]
- Cramer P, Bushnell DA, Kornberg RD. Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution. *Science.* 2001; 292:1863–1876. [PubMed: 11313498]
- Cramer P, Armache K-J, Baumli S, Benkert S, Brueckner F, Buchen C, Damsma GE, Dengl S, Geiger SR, Jasiak AJ, et al. Structure of Eukaryotic RNA Polymerases. *Annu. Rev. Biophys.* 2008; 37:337–352. [PubMed: 18573085]
- Daniels GA, Lieber MR. RNA:DNA complex formation upon transcription of immunoglobulin switch regions: implications for the mechanism and regulation of class switch recombination. *Nucleic Acids Res.* 1995a; 23:5006–5011. [PubMed: 8559658]
- Daniels GA, Lieber MR. Strand specificity in the transcriptional targeting of recombination at immunoglobulin switch sequences. *Proc. Natl. Acad. Sci. U. S. A.* 1995b; 92:5625–5629. [PubMed: 7777560]
- Darzacq X, Shav-Tal Y, de Turris V, Brody Y, Shenoy SM, Phair RD, Singer RH. In vivo dynamics of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* 2007; 14:796–806. [PubMed: 17676063]
- Datta A, Jinks-Robertson S. Association of increased spontaneous mutation rates with high levels of transcription in yeast. *Science.* 1995; 268:1616–1619. [PubMed: 7777859]
- Debatisse M, Le Tallec B, Letessier A, Dutrillaux B, Brison O. Common fragile sites: mechanisms of instability revisited. *Trends Genet.* 2012; 28:22–32. [PubMed: 22094264]
- Dion MF, Kaplan T, Kim M, Buratowski S, Friedman N, Rando OJ. Dynamics of replication-independent histone turnover in budding yeast. *Science.* 2007; 315:1405–1408. [PubMed: 17347438]
- Drolet M. Growth inhibition mediated by excess negative supercoiling: the interplay between transcription elongation, R-loop formation and DNA topology. *Mol. Microbiol.* 2006; 59:723–730. [PubMed: 16420346]
- Drolet M, Bi X, Liu LF. Hypernegative supercoiling of the DNA template during transcription elongation in vitro. *J. Biol. Chem.* 1994; 269:2068–2074. [PubMed: 8294458]
- Drolet M, Phoenix P, Menzel R, Massé E, Liu LF, Crouch RJ. 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. U. S. A.* 1995; 92:3526–3530. [PubMed: 7536935]
- Duquette ML, Handa P, Vincent JA, Taylor AF, Maizels N. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev.* 2004; 18:1618–1629. [PubMed: 15231739]
- El Hage A, French SL, Beyer AL, Tollervey D. Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev.* 2010; 24:1546–1558. [PubMed: 20634320]
- Engel C, Sainsbury S, Cheung AC, Kostrewa D, Cramer P. RNA polymerase I structure and transcription regulation. *Nature.* 2013
- Faza MB, Kemmler S, Jimeno S, González-Aguilera C, Aguilera A, Hurt E, Panse VG. Sem1 is a functional component of the nuclear pore complex-associated messenger RNA export machinery. *J. Cell Biol.* 2009; 184:833–846. [PubMed: 19289793]
- Fischer T, Strässer K, Rácz A, Rodríguez-Navarro S, Oppizzi M, Ihrig P, Lechner J, Hurt E. The mRNA export machinery requires the novel Sac3p-Thp1p complex to dock at the nucleoplasmic entrance of the nuclear pores. *EMBO J.* 2002; 21:5843–5852. [PubMed: 12411502]
- Fischer T, Rodríguez-Navarro S, Pereira G, Rácz A, Schiebel E, Hurt E. Yeast centrin Cdc31 is linked to the nuclear mRNA export machinery. *Nat. Cell Biol.* 2004; 6:840–848. [PubMed: 15311284]

- Folk P, Pta F, Skrzynny M. Transcriptional coregulator SNW/SKIP: the concealed tie of dissimilar pathways. *Cell. Mol. Life Sci. CMLS.* 2004; 61:629–640.
- French SL, Sikes ML, Hontz RD, Osheim YN, Lambert TE, El Hage A, Smith MM, Tollervey D, Smith JS, Beyer AL. Distinguishing the roles of Topoisomerases I and II in relief of transcription-induced torsional stress in yeast rRNA genes. *Mol. Cell. Biol.* 2011; 31:482–494. [PubMed: 21098118]
- Galán A, Rodríguez-Navarro S. Sus1/ENY2: a multitasking protein in eukaryotic gene expression. *Crit. Rev. Biochem. Mol. Biol.* 2012; 47:556–568. [PubMed: 23057668]
- Gallardo M, Luna R, Erdjument-Bromage H, Tempst P, Aguilera A. Nab2p and the Thp1p-Sac3p complex functionally interact at the interface between transcription and mRNA metabolism. *J. Biol. Chem.* 2003; 278:24225–24232. [PubMed: 12702719]
- Gan W, Guan Z, Liu J, Gui T, Shen K, Manley JL, Li X. R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev.* 2011; 25:2041–2056. [PubMed: 21979917]
- Ginno P, Lott P, Christensen H, Korf I, Chédin F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell.* 2012; 45:814–825. [PubMed: 22387027]
- Gómez-González B, García-Rubio M, Bermejo R, Gaillard H, Shirahige K, Marín A, Foiani M, Aguilera A. Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. *EMBO J.* 2011; 30:3106–3119. [PubMed: 21701562]
- González-Aguilera C, Tous C, Gómez-González B, Huertas P, Luna R, Aguilera A. The THP1-SAC3-SUS1-CDC31 complex works in transcription elongation-mRNA export preventing RNA-mediated genome instability. *Mol. Biol. Cell.* 2008; 19:4310–4318. [PubMed: 18667528]
- Gottipati P, Helleday T. Transcription-associated recombination in eukaryotes: link between transcription, replication and recombination. *Mutagenesis.* 2009; 24:203–210. [PubMed: 19139058]
- Gottipati P, Cassel T, Savolainen L, Helleday T. Transcription-associated recombination is dependent on replication in Mammalian cells. *Mol. Cell. Biol.* 2008; 28:154–164. [PubMed: 17967877]
- Gowrishankar J, Harinarayanan R. Why is transcription coupled to translation in bacteria? *Mol. Microbiol.* 2004; 54:598–603. [PubMed: 15491353]
- Guikema JEJ, Linehan EK, Tsuchimoto D, Nakabeppu Y, Strauss PR, Stavnezer J, Schrader CE. APE1- and APE2-dependent DNA breaks in immunoglobulin class switch recombination. *J. Exp. Med.* 2007; 204:3017–3026. [PubMed: 18025127]
- Hall KB, McLaughlin LW. Thermodynamic and structural properties of pentamer DNA:RNA, RNA:RNA, and DNA:RNA duplexes of identical sequence. *Biochemistry (Mosc.).* 1991; 30:10606–10613.
- Han J, Xiong J, Wang D, Fu X-D. Pre-mRNA splicing: where and when in the nucleus. *Trends Cell Biol.* 2011; 21:336–343. [PubMed: 21514162]
- Hanawalt PC, Spivak G. Transcription-coupled DNA repair: two decades of progress and surprises. *Nat. Rev. Mol. Cell Biol.* 2008; 9:958–970. [PubMed: 19023283]
- Hazelbaker DZ, Marquardt S, Wlotzka W, Buratowski S. Kinetic competition between RNA Polymerase II and Sen1-dependent transcription termination. *Mol. Cell.* 2013; 49:55–66. [PubMed: 23177741]
- Helmrich A, Ballarino M, Tora L. Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol. Cell.* 2011; 44:966–977. [PubMed: 22195969]
- Henderson A, Wu Y, Huang YC, Chavez EA, Platt J, Johnson FB, Brosh RM Jr, Sen D, Lansdorp PM. Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res.* 2013
- Hirata A, Klein BJ, Murakami KS. The X-ray crystal structure of RNA polymerase from Archaea. *Nature.* 2008; 451:851–854. [PubMed: 18235446]
- Hong X, Cadwell GW, Kogoma T. Escherichia coli RecG and RecA proteins in R-loop formation. *EMBO J.* 1995; 14:2385–2392. [PubMed: 7774596]
- Huertas P, Aguilera A. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol. Cell.* 2003; 12:711–721. [PubMed: 14527416]

- Itoh T, Tomizawa J. Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H. *Proc. Natl. Acad. Sci. U. S. A.* 1980; 77:2450–2454. [PubMed: 6156450]
- Jamai A, Imoberdorf RM, Strubin M. Continuous histone H2B and transcription-dependent histone H3 exchange in yeast cells outside of replication. *Mol. Cell.* 2007; 25:345–355. [PubMed: 17289583]
- Kasahara M, Clikeman JA, Bates DB, Kogoma T. RecA protein-dependent R-loop formation in vitro. *Genes Dev.* 2000; 14:360–365. [PubMed: 10673507]
- Kim N, Jinks-Robertson S. Transcription as a source of genome instability. *Nat. Rev. Genet.* 2012; 13:204–214. [PubMed: 22330764]
- Kim HD, Choe J, Seo YS. The *sen1(+)* gene of *Schizosaccharomyces pombe*, a homologue of budding yeast *SEN1*, encodes an RNA and DNA helicase. *Biochemistry (Mosc.)*. 1999; 38:14697–14710.
- Kim N, Abdulovic A, Gealy R, Lippert M, Jinks-Robertson S. Transcription-associated mutagenesis in yeast is directly proportional to the level of gene expression and influenced by the direction of DNA replication. *DNA Repair.* 2007; 6:1285–1296. [PubMed: 17398168]
- Kimura H, Cook PR. Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. *J. Cell Biol.* 2001; 153:1341–1353. [PubMed: 11425866]
- Köhler A, Pascual-García P, Llopis A, Zapater M, Posas F, Hurt E, Rodríguez-Navarro S. The mRNA export factor *Sus1* is involved in *Spt/Ada/Gcn5* acetyltransferase-mediated H2B deubiquitinylation through its interaction with *Ubp8* and *Sgf11*. *Mol. Biol. Cell.* 2006; 17:4228–4236. [PubMed: 16855026]
- Korkhin Y, Unligil UM, Littlefield O, Nelson PJ, Stuart DI, Sigler PB, Bell SD, Abrescia NGA. Evolution of Complex RNA Polymerases: The Complete Archaeal RNA Polymerase Structure. *PLoS Biol.* 2009; 7:e1000102. [PubMed: 19419240]
- Koster DA, Croquette V, Dekker C, Shuman S, Dekker NH. Friction and torque govern the relaxation of DNA supercoils by eukaryotic topoisomerase IB. *Nature.* 2005; 434:671–674. [PubMed: 15800630]
- Kristjuhan A, Svejstrup JQ. Evidence for distinct mechanisms facilitating transcript elongation through chromatin in vivo. *EMBO J.* 2004; 23:4243–4252. [PubMed: 15457216]
- Lam EYN, Beraldi D, Tannahill D, Balasubramanian S. G-quadruplex structures are stable and detectable in human genomic DNA. *Nat. Commun.* 2013; 4:1796. [PubMed: 23653208]
- Leela JK, Syeda AH, Anupama K, Gowrishankar J. Rho-dependent transcription termination is essential to prevent excessive genome-wide R-loops in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 2013; 110:258–263. [PubMed: 23251031]
- Li X, Manley JL. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell.* 2005; 122:365–378. [PubMed: 16096057]
- Li X, Manley JL. Cotranscriptional processes and their influence on genome stability. *Genes Dev.* 2006; 20:1838–1847. [PubMed: 16847344]
- Li X, Wang J, Manley JL. Loss of splicing factor ASF/SF2 induces G2 cell cycle arrest and apoptosis, but inhibits internucleosomal DNA fragmentation. *Genes Dev.* 2005; 19:2705–2714. [PubMed: 16260492]
- Lindahl T. Instability and decay of the primary structure of DNA. *Nature.* 1993; 362:709–715. [PubMed: 8469282]
- Lippert MJ, Kim N, Cho J-E, Larson RP, Schoenly NE, O'Shea SH, Jinks-Robertson S. Role for topoisomerase 1 in transcription-associated mutagenesis in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 2011; 108:698–703. [PubMed: 21177427]
- Liu LF, Wang JC. Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. U. S. A.* 1987; 84:7024–7027. [PubMed: 2823250]
- Luna R, Gaillard H, González-Aguilera C, Aguilera A. Biogenesis of mRNPs: integrating different processes in the eukaryotic nucleus. *Chromosoma.* 2008; 117:319–331. [PubMed: 18427828]
- Luna R, Rondón AG, Aguilera A. New clues to understand the role of THO and other functionally related factors in mRNP biogenesis. *Biochim. Biophys. Acta.* 2012; 1819:514–520. [PubMed: 22207203]
- Manis JP, Tian M, Alt FW. Mechanism and control of class-switch recombination. *Trends Immunol.* 2002; 23:31–39. [PubMed: 11801452]

- Masani S, Han L, Yu K. Apurinic/aprimidinic endonuclease 1 is the essential nuclease during immunoglobulin class switch recombination. *Mol. Cell. Biol.* 2013; 33:1468–1473. [PubMed: 23382073]
- Matsuoka S, Ballif BA, Smogorzewska A, McDonald ER 3rd, Hurov KE, Luo J, Bakalarski CE, Zhao Z, Solimini N, Lerenthal Y, et al. ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science*. 2007; 316:1160–1166. [PubMed: 17525332]
- Merz K, Hondele M, Goetze H, Gmelch K, Stoeckl U, Griesenbeck J. Actively transcribed rRNA genes in *S. cerevisiae* are organized in a specialized chromatin associated with the high-mobility group protein Hmo1 and are largely devoid of histone molecules. *Genes Dev.* 2008; 22:1190–1204. [PubMed: 18451108]
- Mischo HE, Gómez-González B, Grzechnik P, Rondón AG, Wei W, Steinmetz L, Aguilera A, Proudfoot NJ. Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol. Cell.* 2011; 41:21–32. [PubMed: 21211720]
- Mizuta R, Iwai K, Shigeno M, Mizuta M, Uemura T, Ushiki T, Kitamura D. Molecular visualization of immunoglobulin switch region RNA/DNA complex by atomic force microscope. *J. Biol. Chem.* 2003; 278:4431–4434. [PubMed: 12475983]
- Moreira M-C, Klur S, Watanabe M, Németh AH, Le Ber I, Moniz J-C, Tranchant C, Aubourg P, Tazir M, Schöls L, et al. Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. *Nat. Genet.* 2004; 36:225–227. [PubMed: 14770181]
- Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell*. 2000; 102:553–563. [PubMed: 11007474]
- Nickloff JA. Transcription enhances intrachromosomal homologous recombination in mammalian cells. *Mol. Cell. Biol.* 1992; 12:5311–5318. [PubMed: 1333040]
- Oakley GG, Patrick SM. Replication protein A: directing traffic at the intersection of replication and repair. *Front. Biosci. J. Virtual Libr.* 2010; 15:883–900.
- Overmeer RM, Moser J, Volker M, Kool H, Tomkinson AE, van Zeeland AA, Mullenders LHF, Foustier M. Replication protein A safeguards genome integrity by controlling NER incision events. *J. Cell Biol.* 2011; 192:401–415. [PubMed: 21282463]
- Paulsen R, Soni D, Wollman R, Hahn A, Yee M-C, Guan A, Hesley J, Miller S, Cromwell E, Solow-Cordero D, et al. A genome-wide siRNA screen reveals diverse cellular processes and pathways that mediate genome stability. *Mol. Cell.* 2009; 35:228–239. [PubMed: 19647519]
- Pellegrini L. The Pol α -Primase Complex. *Subcell. Biochem.* 2012; 62:157–169. [PubMed: 22918585]
- Peña A, Gewartowski K, Mroczek S, Cuéllar J, Szykowska A, Prokop A, Czarnocki-Cieciura M, Piwowarski J, Tous C, Aguilera A, et al. Architecture and nucleic acids recognition mechanism of the THO complex, an mRNP assembly factor. *EMBO J.* 2012; 31:1605–1616. [PubMed: 22314234]
- Perales R, Bentley D. “Cotranscriptionality”: the transcription elongation complex as a nexus for nuclear transactions. *Mol. Cell.* 2009; 36:178–191. [PubMed: 19854129]
- Peters JM, Vangeloff AD, Landick R. Bacterial transcription terminators: the RNA 3'-end chronicles. *J. Mol. Biol.* 2011; 412:793–813. [PubMed: 21439297]
- Petersen S, Casellas R, Reina-San-Martin B, Chen HT, Difilippantonio MJ, Wilson PC, Hanitsch L, Celeste A, Muramatsu M, Pilch DR, et al. AID is required to initiate Nbs1/gamma-H2AX focus formation and mutations at sites of class switching. *Nature*. 2001; 414:660–665. [PubMed: 11740565]
- Petersen-Mahrt S, Harris R, Neuberger M. AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature*. 2002; 418:99–103. [PubMed: 12097915]
- Pohjoismäki JLO, Holmes JB, Wood SR, Yang M-Y, Yasukawa T, Reyes A, Bailey LJ, Cluett TJ, Goffart S, Willcox S, et al. Mammalian mitochondrial DNA replication intermediates are essentially duplex but contain extensive tracts of RNA/DNA hybrid. *J. Mol. Biol.* 2010; 397:1144–1155. [PubMed: 20184890]
- Powell WT, Coulson RL, Gonzales ML, Cray FK, Wong SS, Adams S, Ach RA, Tsang P, Yamada NA, Yasui DH, et al. R-loop formation at Snrd116 mediates topotecan inhibition of Ube3a-

- antisense and allele-specific chromatin decondensation. *Proc. Natl. Acad. Sci. U. S. A.* 2013; 110:13938–13943. [PubMed: 23918391]
- Prado F, Aguilera A. Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *EMBO J.* 2005; 24:1267–1276. [PubMed: 15775982]
- Rada C, Di Noia JM, Neuberger MS. Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Mol. Cell.* 2004; 16:163–171. [PubMed: 15494304]
- Ratmeyer L, Vinayak R, Zhong YY, Zon G, Wilson WD. Sequence specific thermodynamic and structural properties for DNA:RNA duplexes. *Biochemistry (Mosc.)*. 1994; 33:5298–5304.
- Reaban M, Griffin J. Induction of RNA-stabilized DNA conformers by transcription of an immunoglobulin switch region. *Nature*. 1990; 348:342–344. [PubMed: 1701219]
- Reaban ME, Lebowitz J, Griffin JA. Transcription induces the formation of a stable RNA:DNA hybrid in the immunoglobulin alpha switch region. *J. Biol. Chem.* 1994; 269:21850–21857. [PubMed: 8063829]
- Richardson JP. Loading Rho to terminate transcription. *Cell*. 2003; 114:157–159. [PubMed: 12887917]
- Roberts RW, Crothers DM. Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. *Science*. 1992; 258:1463–1466. [PubMed: 1279808]
- Roca J. Transcriptional inhibition by DNA torsional stress. *Transcription*. 2011; 2:82–85. [PubMed: 21468234]
- Rodríguez-Navarro S, Hurt E. Linking gene regulation to mRNA production and export. *Curr. Opin. Cell Biol.* 2011; 23:302–309. [PubMed: 21227675]
- Rodríguez-Navarro S, Fischer T, Luo M-J, Antúnez O, Brettschneider S, Lechner J, Pérez-Ortín JE, Reed R, Hurt E. Sus1, a functional component of the SAGA histone acetylase complex and the nuclear pore-associated mRNA export machinery. *Cell*. 2004; 116:75–86. [PubMed: 14718168]
- Rondón AG, Jimeno S, Aguilera A. The interface between transcription and mRNP export: from THO to THSC/TREX-2. *Biochim. Biophys. Acta*. 2010; 1799:533–538. [PubMed: 20601280]
- Rossi F, Labourier E, Forné T, Divita G, Derancourt J, Riou JF, Antoine E, Cathala G, Brunel C, Tazi J. Specific phosphorylation of SR proteins by mammalian DNA topoisomerase I. *Nature*. 1996; 381:80–82. [PubMed: 8609994]
- Roy D, Lieber M. G clustering is important for the initiation of transcription-induced R-loops in vitro, whereas high G density without clustering is sufficient thereafter. *Mol. Cell. Biol.* 2009; 29:3124–3133. [PubMed: 19307304]
- Roy D, Yu K, Lieber M. Mechanism of R-loop formation at immunoglobulin class switch sequences. *Mol. Cell. Biol.* 2008; 28:50–60. [PubMed: 17954560]
- Roy D, Zhang Z, Lu Z, Hsieh C-L, Lieber MR. 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.* 2010; 30:146–159. [PubMed: 19841062]
- Rudolph CJ, Upton AL, Briggs GS, Lloyd RG. Is RecG a general guardian of the bacterial genome? *DNA Repair*. 2010; 9:210–223. [PubMed: 20093100]
- Sam M, Wurst W, Klüppel M, Jin O, Heng H, Bernstein A. Aquarius, a novel gene isolated by gene trapping with an RNA-dependent RNA polymerase motif. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* 1998; 212:304–317.
- Sanford JR, Longman D, Cáceres JF. Multiple roles of the SR protein family in splicing regulation. *Prog. Mol. Subcell. Biol.* 2003; 31:33–58. [PubMed: 12494762]
- Saxowsky T, Doetsch P. RNA polymerase encounters with DNA damage: transcription-coupled repair or transcriptional mutagenesis? *Chem. Rev.* 2006; 106:474–488. [PubMed: 16464015]
- Schwabish MA, Struhl K. Evidence for eviction and rapid deposition of histones upon transcriptional elongation by RNA polymerase II. *Mol. Cell. Biol.* 2004; 24:10111–10117. [PubMed: 15542822]
- Schwabish MA, Struhl K. Asf1 mediates histone eviction and deposition during elongation by RNA polymerase II. *Mol. Cell.* 2006; 22:415–422. [PubMed: 16678113]
- Silverstein RA, Ekwall K. Sin3: a flexible regulator of global gene expression and genome stability. *Curr. Genet.* 2005; 47:1–17. [PubMed: 15565322]

- Skourti-Stathaki K, Proudfoot NJ, Gromak N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell.* 2011; 42:794–805. [PubMed: 21700224]
- Smolka MB, Albuquerque CP, Chen S, Zhou H. Proteome-wide identification of in vivo targets of DNA damage checkpoint kinases. *Proc. Natl. Acad. Sci. U. S. A.* 2007; 104:10364–10369. [PubMed: 17563356]
- Sohail A, Klapacz J, Samaranayake M, Ullah A, Bhagwat AS. Human activation-induced cytidine deaminase causes transcription-dependent, strand-biased C to U deaminations. *Nucleic Acids Res.* 2003; 31:2990–2994. [PubMed: 12799424]
- Stavnezer J, Guikema JEJ, Schrader CE. Mechanism and regulation of class switch recombination. *Annu. Rev. Immunol.* 2008; 26:261–292. [PubMed: 18370922]
- Steinmetz EJ, Conrad NK, Brow DA, Corden JL. RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature.* 2001; 413:327–331. [PubMed: 11565036]
- Steinmetz EJ, Warren CL, Kuehner JN, Panbehi B, Ansari AZ, Brow DA. Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Mol. Cell.* 2006; 24:735–746. [PubMed: 17157256]
- Stirling P, Chan Y, Minaker S, Aristizabal M, Barrett I, Sipahimalani P, Kobor M, Hieter P. R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. *Genes Dev.* 2012; 26:163–175. [PubMed: 22279048]
- Strässer K, Masuda S, Mason P, Pfannstiel J, Oppizzi M, Rodriguez-Navarro S, Rondón AG, Aguilera A, Struhl K, Reed R, et al. TREX is a conserved complex coupling transcription with messenger RNA export. *Nature.* 2002; 417:304–308. [PubMed: 11979277]
- Sun H, Yabuki A, Maizels N. A human nuclease specific for G4 DNA. *Proc. Natl. Acad. Sci. U. S. A.* 2001; 98:12444–12449. [PubMed: 11675489]
- Svejstrup J. The interface between transcription and mechanisms maintaining genome integrity. *Trends Biochem. Sci.* 2010; 35:333–338. [PubMed: 20194025]
- Takahashi T, Burguiere-Slezak G, Van der Kemp PA, Boiteux S. Topoisomerase I provokes the formation of short deletions in repeated sequences upon high transcription in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* 2011; 108:692–697. [PubMed: 21177431]
- Le Tallec B, Millot GA, Blin ME, Brison O, Dutrillaux B, Debatisse M. Common fragile site profiling in epithelial and erythroid cells reveals that most recurrent cancer deletions lie in fragile sites hosting large genes. *Cell Rep.* 2013; 4:420–428. [PubMed: 23911288]
- Thiriet C, Hayes JJ. Replication-independent core histone dynamics at transcriptionally active loci in vivo. *Genes Dev.* 2005; 19:677–682. [PubMed: 15769942]
- Thiriet C, Hayes JJ. Histone dynamics during transcription: exchange of H2A/H2B dimers and H3/H4 tetramers during pol II elongation. *Results Probl. Cell Differ.* 2006; 41:77–90. [PubMed: 16909891]
- Tian M, Alt F. Transcription-induced cleavage of immunoglobulin switch regions by nucleotide excision repair nucleases in vitro. *J. Biol. Chem.* 2000; 275:24163–24172. [PubMed: 10811812]
- Tuduri S, Crabbé L, Conti C, Tourrière H, Holtgreve-Grez H, Jauch A, Pantesco V, De Vos J, Thomas A, Theillet C, et al. Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nat. Cell Biol.* 2009; 11:1315–1324. [PubMed: 19838172]
- Ursic D, Himmel KL, Gurley KA, Webb F, Culbertson MR. The yeast SEN1 gene is required for the processing of diverse RNA classes. *Nucleic Acids Res.* 1997; 25:4778–4785. [PubMed: 9365256]
- Ursic D, Chinchilla K, Finkel JS, Culbertson MR. Multiple protein/protein and protein/RNA interactions suggest roles for yeast DNA/RNA helicase Sen1p in transcription, transcription-coupled DNA repair and RNA processing. *Nucleic Acids Res.* 2004; 32:2441–2452. [PubMed: 15121901]
- Vologodskii AV, Cozzarelli NR. Conformational and thermodynamic properties of supercoiled DNA. *Annu. Rev. Biophys. Biomol. Struct.* 1994; 23:609–643. [PubMed: 7919794]

- Wahba L, Amon J, Koshland D, Vuica-Ross M. RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Mol. Cell.* 2011; 44:978–988. [PubMed: 22195970]
- Wahba L, Gore S, Koshland D. The homologous recombination machinery modulates the formation of RNA-DNA hybrids and associated chromosome instability. *eLife.* 2013; 2
- Wang JC. Cellular roles of DNA topoisomerases: a molecular perspective. *Nat. Rev. Mol. Cell Biol.* 2002; 3:430–440. [PubMed: 12042765]
- Wellinger RE, Prado F, Aguilera A. Replication fork progression is impaired by transcription in hyperrecombinant yeast cells lacking a functional THO complex. *Mol. Cell. Biol.* 2006; 26:3327–3334. [PubMed: 16581804]
- Westover KD, Bushnell DA, Kornberg RD. Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science.* 2004; 303:1014–1016. [PubMed: 14963331]
- Wongsurawat T, Jenjaroenpun P, Kwok CK, Kuznetsov V. Quantitative model of R-loop forming structures reveals a novel level of RNA-DNA interactome complexity. *Nucleic Acids Res.* 2012; 40:e16. [PubMed: 22121227]
- Wu HY, Shyy SH, Wang JC, Liu LF. Transcription generates positively and negatively supercoiled domains in the template. *Cell.* 1988; 53:433–440. [PubMed: 2835168]
- Xu B, Clayton DA. RNA-DNA hybrid formation at the human mitochondrial heavy-strand origin ceases at replication start sites: an implication for RNA-DNA hybrids serving as primers. *EMBO J.* 1996; 15:3135–3143. [PubMed: 8670814]
- Xu Z, Zan H, Pone EJ, Mai T, Casali P. Immunoglobulin class-switch DNA recombination: induction, targeting and beyond. *Nat. Rev. Immunol.* 2012; 12:517–531. [PubMed: 22728528]
- Yamane A, Robbiani DF, Resch W, Bothmer A, Nakahashi H, Oliveira T, Rommel PC, Brown EJ, Nussenzweig A, Nussenzweig MC, et al. RPA accumulation during class switch recombination represents 5'-3' DNA-end resection during the S-G2/M phase of the cell cycle. *Cell Rep.* 2013; 3:138–147. [PubMed: 23291097]
- Ben Yehuda S, Dix I, Russell CS, Levy S, Beggs JD, Kupiec M. Identification and functional analysis of hPRP17, the human homologue of the PRP17/CDC40 yeast gene involved in splicing and cell cycle control. *RNA N. Y. N.* 1998; 4:1304–1312.
- Yu K, Chedin F, Hsieh C-L, Wilson T, Lieber M. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat. Immunol.* 2003; 4:442–451. [PubMed: 12679812]
- Yüce Ö, West SC. Senataxin, defective in the neurodegenerative disorder ataxia with oculomotor apraxia 2, lies at the interface of transcription and the DNA damage response. *Mol. Cell. Biol.* 2013; 33:406–417. [PubMed: 23149945]
- Zaitsev EN, Kowalczykowski SC. A novel pairing process promoted by *Escherichia coli* RecA protein: inverse DNA and RNA strand exchange. *Genes Dev.* 2000; 14:740–749. [PubMed: 10733533]
- Zenklusen D, Vinciguerra P, Wyss J-C, Stutz F. Stable mRNP formation and export require cotranscriptional recruitment of the mRNA export factors Yra1p and Sub2p by Hpr1p. *Mol. Cell. Biol.* 2002; 22:8241–8253. [PubMed: 12417727]
- Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K, Darst SA. Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell.* 1999; 98:811–824. [PubMed: 10499798]

Figure 1. Schematic representation of an R-loop structure

The cartoon depicts the general structure of an R-loop. The nascent RNA strand (red) is synthesized by RNA polymerase (RNAP, red oval) and hybridizes with the complementary DNA template strand. The non-template strand is exposed as single-stranded DNA (ssDNA).

Figure 2. Formation of R-loops is facilitated by G-rich sequences and transcriptional supercoiling

A) R-loop formation is preferred for consecutive clusters of 3 or more G-residues on the non-template strand in the R-loop initiating zone (RIZ). G-rich sequences downstream in the elongation zone of the R-loop (REZ) facilitate extension of the RNA:DNA hybrid. B) Formation of RNA:DNA hybrids is also dependent on the stability of the resulting RNA:DNA hybrid and the exposed stretch of ssDNA. Clusters of G tracts on the non-template strand can fold into a stable G-quadruplex structure, which may help to stabilize the exposed ssDNA region of the R-loop. In addition, ssDNA binding proteins like RPA

(yellow ovals) may associate and contribute to the stability of the vulnerable DNA strand. C) Positive and negative supercoils of the DNA double helix arise ahead of and behind the elongating RNAP, respectively. Negative supercoiling favors partial unwinding of the DNA double helix and may result in transient accumulation of ssDNA, thereby facilitating intrusion of the nascent RNA strand to hybridize with the complementary template strand.

Figure 3. Prevention and resolution of R-loop structures in the genome by multiple conserved mechanisms

A) Co-transcriptional assembly of RNP particles on the nascent RNA is a conserved strategy to prevent formation of R-loops from bacteria to metazoans. In bacteria, ribosomes associate cotranscriptionally with the nascent RNA and couple the two processes of transcription and translation. In this way, the RNA strand is prevented from forming an R-loop on the bacterial DNA template. In yeast and metazoan genomes, the double-helix is wrapped around nucleosome core particles (brown ovals) that may help to prevent invasion of the RNA strand after passage of RNAP. In addition, numerous RNA processing and splicing factors assemble with the RNA strand and prevent the accumulation of R-loops.

Homologous subunits of the conserved protein complexes THO, TREX, and TREX-2/AMEX in yeast and metazoans are depicted in the same color. B) RNase H enzymes degrade the RNA moiety of the RNA:DNA hybrid. C) Specific helicases like Senataxin have been proposed to specifically unwind the RNA:DNA hybrid and allow re-annealing of the non-template strand to restore the DNA double helix.

A Single-strand break(s)

B Conflict with replication fork

Figure 4. Possible mechanisms for R-loop mediated formation of single-strand and double-strand breaks

A) R-loop mediated SSBs can be generated by a variety of factors/pathways. AID or other proteins of the APOBEC family were shown to deaminate cytosine residues preferentially on single-stranded DNA. The resulting uracil is recognized by proteins of the mismatch repair (MMR) or base excision repair (BER) machineries to create an abasic site, which is processed by apurinic-apyrimidinic endonucleases to generate a single-strand break. Alternatively, Topoisomerase 1 can be irreversibly trapped during its cleavage–ligation cycle, giving rise to a covalent Top1-DNA complex attached to the 5' end of the nicked DNA. This irreversible complex is likely to be processed into a small gap by specific

endonucleases like Rad1/Rad10 (yeast) or Mus81/Mms4 (human), giving rise to a break or a small gap in the template or non-template strand. Finally, G-quadruplex or flap endonucleases may recognize a secondary DNA structure on the ssDNA or other structural features like the loop-duplex junction generating a single-strand break in the DNA. B) Possible mechanisms that induce R-loop mediated DSBs or convert the initiating SSB into a DSB. Green ovals represent DNA polymerases. Green hexagons indicate the replicative helicase. (1) A second SSB generated by one of the mechanisms described in A) may arise in proximity to the first SSB, which would result in a DSB (2). A replication fork may encounter the R-loop containing the SSB in co-directional (3) or head-on (6) orientation. If the fork can progress over the RNA:DNA hybrid (4, 7), separation of the parental strands over the SSB would result in a DNA lesion on the lagging (5) or leading strand (8), respectively. In the absence of a SSB at the R-loop structure, the stalled RNAP may still collide with the replication machinery in a co-directional (9) or head-on (10) orientation. Head-on collisions may be especially deleterious, leading to fork stalling and collapse, ultimately resulting in a DSB.