

R-loops: formation, function, and relevance to cell stress

David F Allison^{1,2,*} and Gang Greg Wang^{1,2,*}

¹ Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27599, USA.

² Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27599, USA.

* Corresponding Authors:

David F Allison, E-mail: David_Allison@unc.edu;

Gang Greg Wang, E-mail: greg_wang@med.unc.edu

ABSTRACT Exposure of genomic, single-stranded DNA (ssDNA) during transcription and replication creates opportunities for the formation of inappropriate secondary structures. Cells manage this exposure by using topoisomerases and helicases to reduce the inherent topological stress that arises from unwinding the double helix and by coating ssDNA with protective protein complexes. Interestingly, specific DNA-RNA hybrids, known as R-loops, form during transcription and exist in homeostasis throughout the genomes of prokaryotes and eukaryotes. These hybrids nucleate from guanine rich clusters in the template strand and extend across GC rich spans of transcribed genes. *In vivo* regulatory functions have evolved from R-loops, including regulation of gene expression and telomere lengthening. However, they also exist as a form of stress, particularly when replication forks collide with the transcription machinery. New methodologies and models are being developed to delineate the biology of R-loops, including those related to cell stress-based diseases like cancer. As accumulation of R-loops is associated with disease, targeting molecular pathways that regulate their formation or removal could provide new avenues for therapeutic intervention. This review covers recent understandings of the molecular basis for R-loop formation, removal, and biological outcomes in the context of cellular stress.

Received originally: 15.10.2018

in revised form: 27.11.2018,

Accepted 30.11.2018,

Published 21.01.2019.

Keywords: R-loops, RNA-DNA hybrid, replication fork stalling, gene transcription, genome instability, mRNA splicing, DNA damage, chromatin modification, cancer, ataxia.

Abbreviations:

AID – activation-induced deaminase,

ds – double-stranded,

DSB – double strand break,

FA – Fanconi Anemia,

G-cluster – guanine-rich cluster,

mRNP – messenger ribonucleoprotein particle,

RF – replication fork,

ss – single-stranded,

TOP – topoisomerase,

TREX – THO and transcription export.

INTRODUCTION

A variety of topological, structural and hybridization events occur during DNA replication and gene transcription. Unwinding of the DNA double helix provides access for polymerase to a template strand, and creates torsional stress that can manifest anomalous formation of “non-traditional” moieties. One such structure, known as an R-loop [1, 2], occurs during transcription (**Figure 1**). As RNA polymerase progresses along the DNA double helix, newly transcribed RNA threads back to hybridize with the transiently accessible template strand and displace the non-template strand. Structurally, the hybrid adopts an intermediate conformation between B-form double-stranded (ds)DNA and A-form dsRNA. This form carries more stability than dsDNA and must be enzymatically resolved in order to restore the

native double helix. A homeostasis emerges from the constant formation and removal of R-loops throughout the genome. Moreover, their presence has biological relevance in regulating gene expression and specialized rearrangement events. Misregulation of R-loop homeostasis promotes genomic instability is associated with progressive neurodegenerative disorders or cancer. The purpose of this review is to detail the current understanding of R-loop contribution to cell stress and disease.

FORMATION AND REMOVAL OF R-LOOPS

R-loops canonically nucleate from guanine-rich clusters (G-clusters) during gene transcription [3, 4], and these clusters increase the favorability of nascent RNA to anneal with complementary single-stranded (ss)DNA (**Figure 1**). Re-

search demonstrates that RNA containing four or more consecutive guanine residues near the 5' end possesses significantly higher rates of R-loop formation due to this nucleation requirement. After initiation of the R-loop, RNA-DNA hybridization is elongated and stabilized by a subsequent stretch of guanine rich (G-Rich) sequence. Elongation becomes less favorable when G-rich content decreases and the structure terminates. Other factors also promote R-loop formation and reduce the requirement for G-Rich sequence. For example, increased negative supercoiling on the trailing fork of the transcription bubble extends the opportunity for potential interactions between newly transcribed RNA and the template strand (reviewed in [5]). Additionally, nicks in the non-template strand can promote DNA-RNA hybridization of nascent RNA to the template strand, even when the G-Rich sequence is distal to the initiating G-cluster [3, 4].

The build-up of R-loops is counteracted through a variety of mechanisms that maintain transcription bubble integrity, including nuclease activity and resolution of topological stress (Figure 1). These mechanisms are detailed in recent review articles [5-9]. Briefly, distinct enzymes carry out complementary actions to prevent accumulation of R-

loops. RNase H1 and RNase H2 utilize 5'–3' exonuclease activity to remove RNA from the loop. These enzymes are evolutionarily conserved in prokaryotes and eukaryotes and are the only RNA-specific ribonucleases known to digest hybridized RNA. Complementary to RNase H1/2, cells also possess helicases that “untangle” DNA-RNA lesions like mammalian DHX9 and aquarius (AQR) [10, 11]. Other helicases known to resolve R-loop structures include the yeast Sen1p orthologue, senataxin (SETX) [12] and an ATP-dependent DNA helicase PIF1 [13].

Preventative mechanisms exist as an additional measure of protection against R-loop accumulation (Figure 1). Topoisomerases work to decrease negative supercoiling that increase access of nascent mRNA to G-clusters on the template DNA. Topoisomerase 1 (TOP1) and Topoisomerase 2 (TOP2) have both been shown to relieve torsional stress and prevent R-loop accumulation at the rDNA locus [14]. Moreover, topoisomerase 3B (TOP3B) reduces R-loop formation at highly expressed genes like the proto-oncogene MYC [15]. Biogenesis of messenger ribonucleoprotein particles (mRNP) also prevents R-loop accumulation. During mRNP biogenesis, RNA-binding proteins (RBP) interact with nascent mRNA to promote mRNA processing

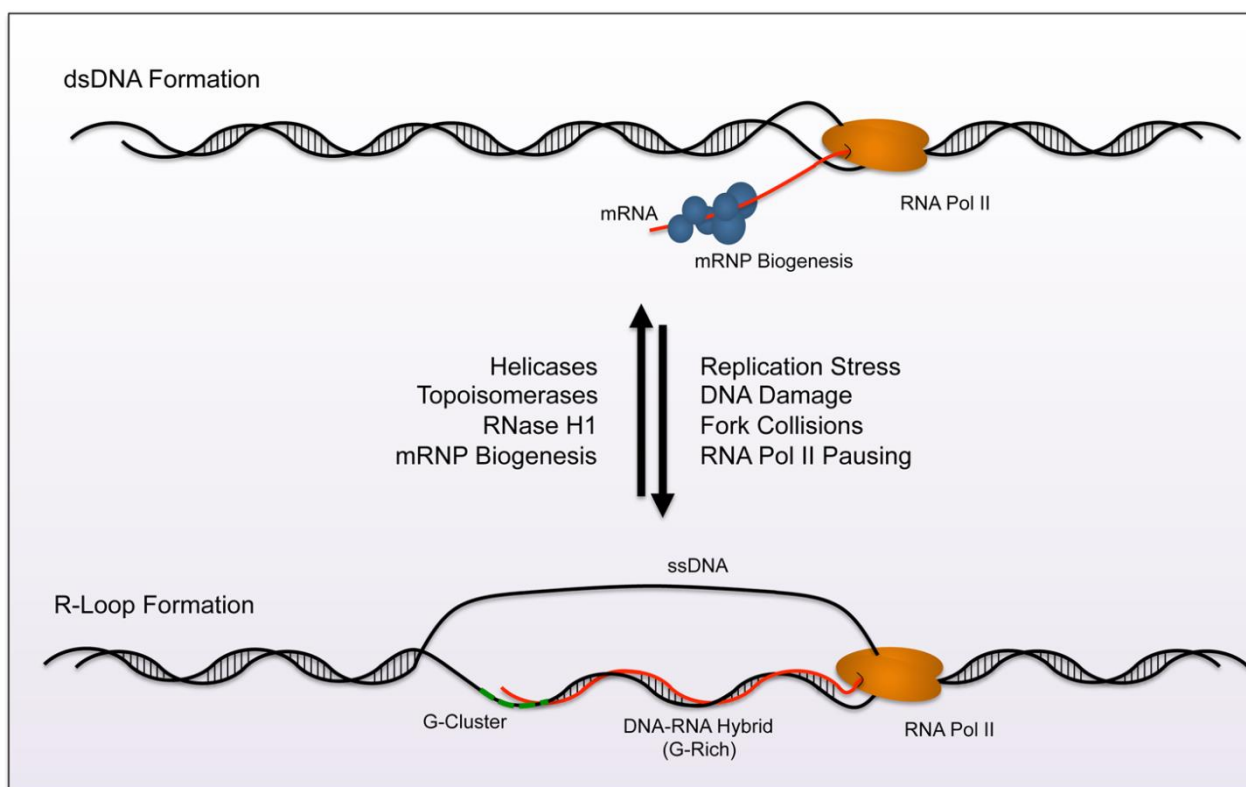


FIGURE 1: Model of R-loop homeostasis. A homeostasis exists between formation and removal of R-loops across the genome. Normally, mRNP biogenesis machinery targets nascent RNA and processes and prepares the strand for nuclear export (top). Topoisomerases and Helicases aid in preventing R-loop formation by reducing exposure times of ssDNA during transcription. RNase H removes R-loops by specifically digesting RNA in helical formation. Together these processes reduce R-loop accumulation during transcription. Factors that lead to pausing of RNA polymerase render conditions favorable for R-loop formation (replication stress, DNA damage, fork collisions, etc). R-loops nucleate from G-clusters (dashed green line). The loop then extends along GC rich sequences within the gene during elongation (bottom). The balance between removal and formation creates an equilibrium at nucleation sites across the genome.

and export from the nucleus. In eukaryotic cells, the THO/TREX complex, as well as Serine/Arginine Rich Splicing Factor (SRSF1), binds ssRNA to initiate mRNP maturation, and these interactions prohibit R-loop formation [16-18]. Additional mRNA splicing and surveillance machinery such as the exon junction complex (EJC) and the exosome also reduce

R-loop accumulation by sterically prohibiting interactions between the nascent mRNA and exposed ssDNA [18, 19].

Epigenetic mechanisms act upstream of topoisomerase and helicase function in R-loop homeostasis. For example, the MYST family acetyltransferase MOF aids in preventing R-loop accumulation that arises from replicative stress. Depletion of MOF decreases replication fork (RF) speed and potentiates fork stalling, factors that contribute to R-loop nucleation. Furthermore, recruitment of the repair machinery to sites of DNA damage occurs through MOF-mediated ubiquitination [20], while an interaction between THO and SIN3A has been described where histone deacetylation prevents R-loop formation [21]. How acetylation regulates R-loop formation is likely context dependent and further studies should delineate these differences.

Collectively, the cooperative activities of transcription bubble maintenance and mRNA processing machinery resolve and prevent R-loop accumulation. Each individual component contributes to the homeostasis of R-loops, and their respective regulation determines the local prevalence of these hybrid structures throughout the genome.

TECHNIQUES DEVELOPED TO STUDY R-LOOPS

Since being first described in 1976 [2], several techniques have been developed to study eukaryotic R-loop biology *in vitro* and *in vivo*. Various plasmid-based systems have been created to study the regulation of R-loop formation and removal. Known R-loop sequences are inserted into plasmids in order to examine the effect of these structures on replication or transcription *in vitro*.

Classically, a series of electrophoresis techniques have been developed to identify sequences that favorably form R-loops *in vitro*. The R-loop structure migrates more slowly in agarose gel electrophoresis and can be visualized with ethidium bromide (EtBr) after *in vitro* transcription [22, 23]. Moreover, the shifted band shows resistance to digestion by RNase A and sensitivity to digestion by RNase H1. These exonucleases target ssRNA and helical RNA, respectively. Using labeled uridine during the *in vitro* transcription step increases sensitivity of the assay. Moreover, it eliminates signal from dsDNA that appears when using a general nucleic acid stain like EtBr. As a method to identify perturbations in the dsDNA helix, bisulfite modification can be used to detect R-loops [22]. Briefly, unpaired cytosines on the displaced anti-sense ssDNA are converted to uracil through deamination by addition of bisulfite. The uracil subsequently converts CG basepairs to AT basepairs during PCR amplification and the change can be observed by sequencing. Because this technique targets accessible strands of ssDNA, R-loop formation should be verified through additional methodologies.

A powerful, genome-wide technique, known as DNA-RNA immunoprecipitation followed by sequencing (DRIP-seq), uses the S9.6 antibody to specifically pull down DNA-RNA hybrids that are subsequently identified by high throughput sequencing methods to map these structures to the genome [24-27]. While the use of DRIP-seq and its derivatives give an overall picture of R-loop distribution across the genome, called peaks should be validated as some inherent bias has been reported [28]. An alternative high throughput technique known as DNA-RNA *in vitro* enhancement followed by sequencing (DRIVE-seq) utilizes an epitope-tagged, catalytically dead RNase H1 and affinity pulldown to recover DNA-RNA hybrids for sequencing [29, 30]. Depending on experimental conditions, high throughput strategies identify between 1,000 and 20,000 R-loops across the genome. Interestingly, R-loops are largely mapped to gene promoters and terminator regions in human cell lines, showing their potential involvement in regulating RNA pol II-mediated transcription. However, several specialized regions also show enrichment for R-loops where they may act as a more structural component to the genome, including telomeres, ribosomal DNA (rDNA), and transposable elements.

Recently, IP-mass spectrometry techniques were used to identify R-loop interacting proteins in a high throughput format. The study identified the RNA helicase DHX9 and characterized its role in suppressing R-loop accumulation, reducing DNA damage, and promoting transcriptional termination [31]. R-loops can also be detected with microscopy to visually confirm their formation using *in vitro* transcription. For example, the S9.6 antibody, as well as a GFP-tagged DNA-RNA hybrid binding domain of RNase H1, allow for visualization of R-loop foci with fluorescence microscopy [26, 32] or fluorescence *in situ* hybridization (FISH) based techniques [33-35]. Electron microscopy provides another mechanism to study and visualize R-loop structures [36].

The continued development of techniques and methodologies to study R-loop dynamics and the downstream consequences of their accumulation will further detail the contribution that these structures have in gene regulation and genomic instability.

IN VIVO ROLES FOR R-LOOPS IN EUKARYOTIC CELLS

Predictably, formation of a DNA-RNA hybrid creates structural lesions in the genome that can affect replication, transcription, and recombination. Ongoing research continues to detail the *in vivo* role of R-loops as regulatory elements of specialized events in the eukaryotic cell. Previously, mitochondrial origins of replication contain R-loops that are used to facilitate replication of the mitochondrial DNA (mtDNA), a process likely conserved from prokaryotic ancestors [37]. Studies in activated B-lymphocytes also show that R-loops are used at the switch (S) region during immunoglobulin class switching recombination (CSR) to promote deamination of cytosine and generate a double-strand break (DSB) [22, 38]. In this case, activation-induced deaminase (AID) gains access to the WRC motifs (W=A/T,

R=A/G) within the repetitive S regions in the displaced, non-template strand downstream of the IgG promoter [39]. Recently, researchers describe a mechanism where the RNA helicase DDX1 resolves G-quadruplex structures into R-loops at the S region of immunoglobulin heavy chain locus (IgH) in order to recruit AID for CSR [40].

While R-loops contribute to specialized events or structures in the human genome, they also have distinct roles in regulating transcription. In one instance, R-loops regulate epigenetic mechanisms of gene repression. When R-loops form in downstream CpG island promoter regions, they inhibit the activity of DNA methyltransferases to promote expression [26]. In a separate study, R-loop formation can facilitate binding of transcription factors to regulatory elements on DNA. Invasion of anti-sense non-coding RNA (ncRNA) into the promoter region of the vimentin (VIM) gene reduces local nucleosome density. The formed R-loop subsequently increases gene expression by providing access of nuclear factor kappa B (NF- κ B) to its binding element [41]. Interestingly, this mechanism runs counter to previously described repressive activities of non-coding (nc)RNA, and provides new insight into how these moieties can positively regulate gene expression.

R-loop formation in G-Rich sequences downstream of the 3' polyadenylation (polyA) signal can promote RNA Pol II pausing [42]. The R-loop structure promotes antisense transcription and formation of dsRNA. As a result, RNAi machinery and associated histone modification machinery, including G9a, localize at the pause site. Subsequent deposition of the repressive H3K9me2 mark on chromatin by G9a leads to recruitment of heterochromatin protein 1y (HP1y), which reinforces the pausing. Transcriptional termination subsequently occurs and prevents run-through at regions of high gene density to maintain gene integrity.

MECHANISMS OF STRESS INDUCTION VIA R-LOOP ACCUMULATION

Accumulation of R-loops often results in increased cellular stress that leads to genomic instability. The current hypothesis states that stabilized, co-transcriptional R-loops disrupt RF progression during S-phase. This disruption causes stalling that can lead to fork collapse, DSBs, or incomplete replication before entry into mitosis. Recently, perturbation of the mRNA splicing, processing, and export factors have been shown to increase R-loop mediated genome instability. For example, disruption of the THO and Transcription Export (TREX) complex (discussed above) leads to a hyper-recombination phenotype in yeast and human cells [43, 44]. Interestingly, these mutants show genome-wide increases in histone H3 serine 10 phosphorylation (H3S10P) to promote chromatin condensation during mitosis. Evidence suggests that this modification, along with R-loop associated H3K9me2 deposition, impairs replication and leads to stress [45]. Independent studies demonstrate that dysfunction of the SRSF1 splicing complex increases R-loop formation detrimental to genomic stability [16]. In these cases, the observed phenotypes can be rescued through overexpression of RNase H, indicating

the importance of resolving DNA-RNA hybrids to prevent DSBs or recombination. Resolving topological stress remains an important aspect of genome maintenance where accumulation of R-loops after TOP1 inhibition increases genomic rearrangements [46]. These events result from replication fork collisions with a transcription bubble that was likely paused by the formation of an R-loop.

The deposition of R-loops at promoter and terminator regions has been shown to regulate initiation, pausing, and termination of transcription. Perturbation of these regulatory events could lead to inappropriate changes in gene expression or genetic stress. Studies indicate that mutation of the THO complex impairs elongation, possibly through the formation of R-loop intermediates that prevent RNA Pol II processivity [47]. Moreover, these impairments facilitate the hyper-recombination effects described above. A second form of genetic stress occurs from the exposed ssDNA segment in the R-loop. This feature is susceptible to alterations via DNA-modifying or repair enzymes, as well as chemical mutagens [48]. AID activity or other factors may inappropriately target the ssDNA strand of the R-loop and create lesions in the genome that lead to point mutations or other forms of deleterious repair. Unsurprisingly, stabilization of co-transcriptional R-loops can localize mutagenesis to gene regions and exacerbate their effect by targeting coding sequence [44].

DNA damage sensing and repair proteins have recently emerged as inhibitors of R-loop accumulation in eukaryotic cells (**Figure 2**). For example, the Breast Cancer Type 1 and 2 susceptibility genes (BRCA1/2) prevent potential harmful effects of R-loops through two distinct manners. BRCA1 recruits the helicase SETX to R-loops that form at transcription termination pause sites or sites of Negative Elongation Factor (NELF) mediated Pol-II pausing [49, 50]. BRCA1-dependent recruitment of SETX resolves R-loop structures at these sites and suppresses DNA damage. Interestingly, studies of breast cancer models with a BRCA1 mutation show increased insertion-deletion (in-del) mutations at BRCA1-bound termination sites known to inhabit R-loops.

Depletion of BRCA2 from cells also increases R-loop accumulation, suggesting an important role for this protein in R-loop homeostasis. However, dissecting the specific role (or roles) of BRCA2 in resolving R-loops has been difficult due to its contributions in DNA repair, RF stability, and other mechanisms associated with mitosis. Its inherent involvement in these processes may alone prevent R-loop formation by preventing RF collapse and recruiting the ssDNA binding protein, Rad51, to DSB sites [51]. However, recent evidence shows that BRCA2 associates with the TREX complex subunit, PCI-domain containing protein 2 (PCID2), which suggests a protective role for BRCA2 against the co-transcriptional formation of R-loops [33]. Separately, research shows that BRCA2 recruits RNA polymerase II associated factor-1 (PAF1) to promoter-bound Pol-II, potentiates pause release, and decreases R-loop formation [52].

Both BRCA1 (FANCS) and BRCA2 (FANCD1) are part of the Fanconi Anemia (FA) mediated DSB repair pathway. Therefore investigators elucidated that this complex may

regulate genome stability through R-loop resolution. In addition to the protective functions of BRCA1 and BRCA2, the disruption of critical FA complex members FANCD2, FANCA, and FANCM lead to genomic instability and DNA damage from R-loop mediated RF collapse [53, 54]. The activity of the FA complex during replication allows RFs to traverse intrastrand crosslinks (ICL), and promotes post-replicative repair of this lesion [55]. Briefly, the FA complex is recruited by ATR and subsequently activates FANCD2 through a mono-ubiquitination step. The complex then utilizes the DNA translocase activity of FANCM to traverse the lesion and prevent RF collapse. Emerging evidence demonstrates that the same FA complex mediated mechanisms facilitate RF progression through transcription-mediated conflicts, including R-loops. The FA complex is recruited to the stalled RFs due to collisions with the transcription machinery. It subsequently removes R-loops and restarts replication to prevent RF collapse. RNase H1 overexpression rescues the phenotypes observed during disruption of FA function and validates the role of R-loops in these events. Conversely, defects in the FA complex lead to increased R-loop formation upon treatment of ICL-inducing agents. This suggests that R-loops may be an outcome of genotoxic stress in FA patients. Interestingly, depletion of BRCA1 or BRCA2 individually shows that each gene regulates R-loop formation in distinct regions of the genome. Sequence or epigenetic differences that determine recruitment of BRCA1 or BRCA2 to R-loops remain an area of keen interest in the field.

The induction of stress due to R-loop accumulation largely results from collisions that occur between RFs and the transcriptional machinery, particularly in a head-on orientation [56]. Genomic instability and genome rearrangements occur if these structures cannot be resolved through the replication stress responses of the FA complex, DNA repair, and R-loop homeostasis. This creates a positive feedback model where genomic instability can potentiate additional R-loop accumulation. Buildup of R-loops during stress may identify new mechanisms that contribute to the pathology of these structures.

BIOLOGICAL OUTCOMES OF INAPPROPRIATE R-LOOP HOMEOSTASIS

R-loops can arise from a variety of cellular stresses, and lead to complications including replication defects, transcriptional irregularities, and genomic instability. The biological consequences of disrupted R-loop homeostasis has been described in multiple neurodegenerative diseases that largely associate with ataxia (Reviewed in [57]). GC-rich trinucleotide expansions of gene regions provide an extended landscape for R-loop formation that disrupt transcription and proper gene expression (**Figure 2A**). For example, Huntingtin (HTT; Huntington's Disease), Frataxin (FXN; Friedreich Ataxia), and Ataxin 1/2 (ATXN1/ATXN2; Spinocerebellar Ataxias) all have GC-rich trinucleotide expansion tracks that form R-loops *in vitro* and associate with disease progression [58, 59].

R-loop induced disruption of gene expression occurs through distinct mechanisms in trinucleotide repeat disorders. Expansion of the repeat region in the first intron of FXN leads to R-loop formation, increased histone H3K9 dimethyl (H3K9me2) chromatin deposition, and impeded Pol II mediated transcription [60]. Chemical inhibition of TOP1 in Friedreich Ataxia (FRDA) model cell lines with FXN expansions potentiates R-loop accumulation and H3K9me2 enrichment, providing further evidence that R-loop formation increases deposition of this repressive mark.

Fragile X syndrome harbors a mechanism where trinucleotide expansion in the 5' UTR of the FMR1 gene leads to DNA methylation mediated silencing of this locus [60, 61]. However, removal of DNA methylation from this locus reactivates expression to only twenty-five percent of wildtype cells. The reactivation of FMR1 expression is partial because of R-loop formation after depletion of DNA methylation. Examination of the relationship between DNA methylation and R-loop formation at this locus could reveal a distinct mechanism of gene silencing in this model. Separately, trinucleotide repeat expansions can form R-loops that induce RNA Pol II pausing and aborted transcripts, which subsequently sequester nucleolin (NCL) and cause nucleolar stress [62, 63].

Furthermore, mutations in the machinery that maintain R-loop homeostasis also associate with neurodegenerative disease. Dysfunctional TREX or RNaseH2 are linked to Aicardi-Goutières Syndrome (AGS) while SETX mutations contribute to ataxia with oculomotor apraxia (AOA2) [64, 65]. Mutations of SETX also occur in the motor neuron disease amyotrophic lateral sclerosis 4 (ALS4). In this case, the SETX mutation decreases R-loops in the promoter region of BMP and activin membrane bound inhibitor (BAMBI). Local R-loop loss increases DNMT-1 dependent methylation at the promoter, and downregulation of BAMBI leads to the activation of the TGF- β pathway, a process implicated in ALS [66].

As expected, the increased genome instability, chromosomal rearrangements, and DNA damage that occur during aberrant accumulation of R-loops is linked to oncogenesis [53, 67, 68] (**Figure 2B**). Loss of BRCA1 and BRCA2 function has already been well established in breast cancer. Cells show an increase in DSBs when lacking the protective effects that these genes carry against RF collapse [33, 49, 50, 69]. Moreover, the increased frequencies of in-dels in BRCA mutant cell lines localize these mutations to gene regions and can have deleterious effects on gene expression. The cancer-associated genotoxic stress that arises from mutations in BRCA1/2 or members of the FA complex can all be partially rescued through overexpression of RNase H1 in cancer cell lines, suggesting that the aberrant R-loop formation contributes to malignant progression.

The Tudor containing protein 3 (TDRD3) plays a critical role in maintaining proper R-loop homeostasis at the c-MYC locus [15]. TDRD3 recruits TOP3B to relieve negative super-coiling at c-MYC and prevents R-loop accumulation. Loss of TDRD3 function in B-lymphocytes leads genomic rearrangements between c-MYC and the Ig heavy chain (IGH) locus due to AID activity on R-loop associated ssDNA.

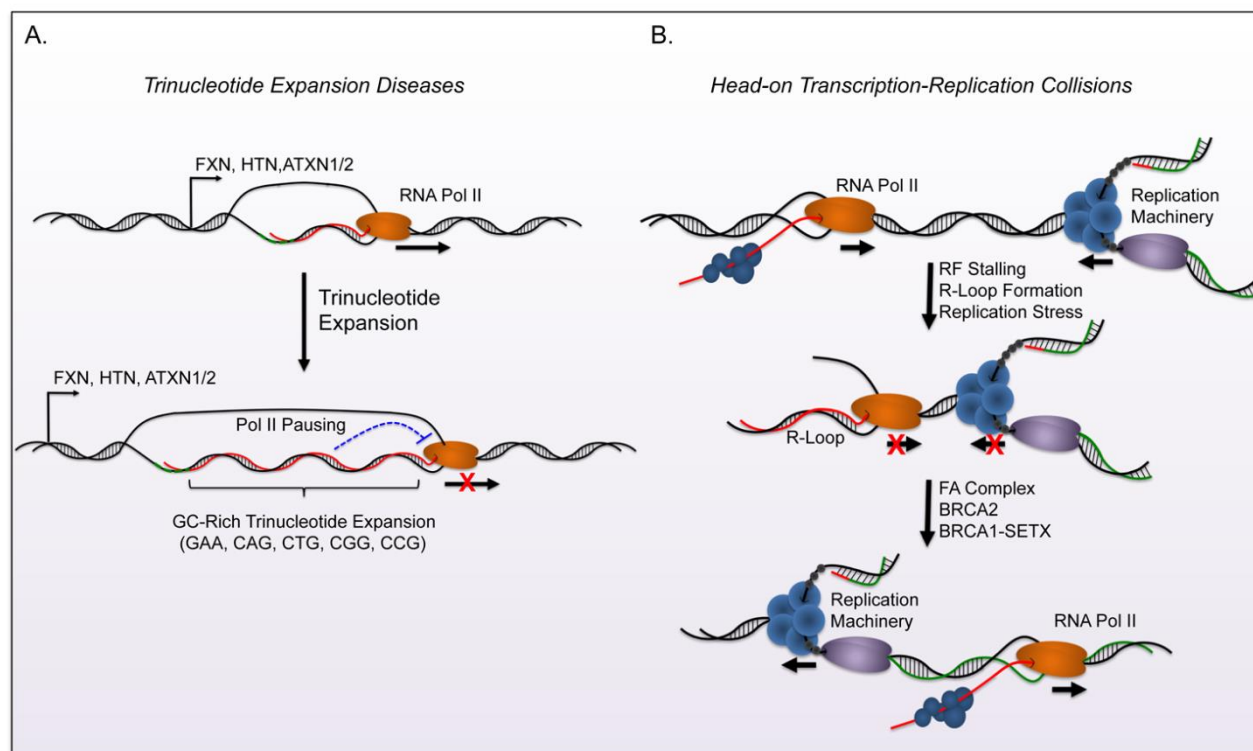


FIGURE 2: Biological consequences of R-loop formation. Formation of R-loops can result in defects in transcription and genomic instability. **(A)** Evidence indicates that R-loop formation contributes to trinucleotide expansion related diseases. The expansion of GC rich sequence in the gene body creates conditions that are favorable for R-loops and can subsequently cause reduced expression of specific genes (FXN, HTN, ATXN1/2). These diseases typically result in neurodegenerative disorders and ataxia. **(B)** Head-on collisions between the replication fork and transcription bubble create favorable conditions for R-loop formation. The FA complex, along with BRCA1, and other factors promote bypass and resolve R-loops during these collisions in order to complete replication and maintain genome stability. Mutation in FA complex subunits increase DNA damage associated with collisions and this damage is associated with cancer progression.

This translocation commonly occurs in Burkitt's lymphoma and multiple myeloma and causes oncogenic levels of *c-MYC* expression. Loss-of-function in mRNP biogenesis machinery also associates with known cancers. For example, sequestration of the THO/TREX complex through Kaposi's sarcoma herpesvirus expression of ORF57 leads to increased genomic instability and subsequent malignant progression in immunodeficiency-associated models [70].

In *Saccharomyces cerevisiae*, the formation of R-loops at the noncoding telomere repeat-containing RNA (TERRA) locus facilitates type II survival in telomerase mutants [71]. The survival mechanism depends on radiation sensitive (*RAD52*) dependent recombination events that result in long heterogenous tracts of telomeric repeats. Extended tracts of recombination dependent repeats appear in mammalian cells utilizing the alternative lengthening of telomere (ALT) pathway during bypass of senescence. Therefore, accumulation of R-loops at TERRA may also play a role in promoting ALT activation in some types of cancer.

Together, these models provide a wide variety of mechanisms where stress associates with increased R-loop formation and can lead to progressive neurodegenerative disease or cancer. The diversity of these mechanisms lends credence to the importance of proper R-loop homeostasis.

Redundancy in the machinery that resolves R-loops suggests these structures are dynamic and widespread (studies indicate that R-loops occupy up to 5% of the genome [25]). As we increase our understanding of the formation and resolution of these lesions *in vivo*, we may discover new disease associations of R-loops.

CONCLUDING REMARKS

Emerging evidence demonstrates the importance of proper R-loop homeostasis. These structures provide critical regulatory functions for gene expression and, in some cases, gene rearrangements. However, their formation and resolution must be tightly regulated to prevent topological events that disrupt replication through impaired RF progression. Such impairment provides an avenue for genome instability and contributes to the development of progressive diseases like cancer. Moreover, prolonged R-loop occupancy exposes ssDNA to genotoxic stress, which produces deleterious effects on gene expression.

Future examination of R-loops in cancer models should determine whether the seemingly independent functions of BRCA1 and BRCA2 are coordinated to prevent RF collapse through FANCD1-complex-mediated mechanisms. Specifically, research should elucidate whether BRCA1 recruits

SETX to unwind R-loops while BRCA2 acts to traverse the lesion in a manner to complete replication and allow DNA repair to occur at a later phase. Moreover, delineating the coordination between RNase H, helicases, and topoisomerases to prevent R-loop accumulation could provide insight into why specific cell types are susceptible to diseases associated with improper R-loop homeostasis.

Finally, the link between the epigenetics and sites of R-loop formation has been established, but further examination of the machinery that recruits chromatin-modifying enzymes is warranted. Discovery of mechanisms that link R-loop homeostasis to chromatin modification would lend further clues into the development of specific cancers or neurodegenerative diseases. Specifically, the field would benefit from identification of R-loop “readers” that allow cells to detect these structures and modify chromatin to repress gene expression or promote transcriptional termination.

R-loops provide an interesting new source of replicative and transcriptional stress that clearly have disease-causing consequences. As new methodologies and models are developed to study their regulation and homeostasis, re-

searches will continue to integrate this under-studied structure into cell biology and pathogenesis.

ACKNOWLEDGMENTS

This work was supported by the NIH (1R01CA218600 and 1R01CA211336 to G.G.W.). G.G.W. is an American Cancer Society (ACS) Research Scholar. D.F.A. was supported by the NIH (5T32CA009156) and the Debbie’s Dream Foundation – AACR Gastric Cancer Research Fellowship (14-40-41-ALLI).

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

COPYRIGHT

© 2019 Allison and Wang. This is an open-access article released under the terms of the Creative Commons Attribution (CC BY) license, which allows the unrestricted use, distribution, and reproduction in any medium, provided the original author and source are acknowledged.

Please cite this article as: David F Allison and Gang Greg Wang (2019). R-loops: formation, function, and relevance to cell stress. *Cell Stress*: in press.

REFERENCES

- White RL, Hogness DS (1977). R loop mapping of the 18S and 28S sequences in the long and short repeating units of *Drosophila melanogaster* rDNA. *Cell* 10(2): 177-192. doi: 10.1016/0092-8674(77)90213-6
- Thomas M, White RL, Davis RW (1976). Hybridization of RNA to double-stranded DNA: formation of R-loops. *Proc Natl Acad Sci U S A* 73(7): 2294-2298. doi: 10.1073/pnas.73.7.2294
- Roy D, Zhang Z, Lu Z, Hsieh CL, Lieber MR (2010). Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation in vitro: a nick can serve as a strong R-loop initiation site. *Mol Cell Biol* 30(1): 146-159. doi: 10.1128/MCB.00897-09
- Roy D, Lieber MR (2009). G clustering is important for the initiation of transcription-induced R-loops in vitro, whereas high G density without clustering is sufficient thereafter. *Mol Cell Biol* 29(11): 3124-3133. doi: 10.1128/MCB.00139-09
- Kuzminov A (2018). When DNA Topology Turns Deadly - RNA Polymerases Dig in Their R-Loops to Stand Their Ground: New Positive and Negative (Super)Twists in the Replication-Transcription Conflict. *Trends Genet* 34(2): 111-120. doi: 10.1016/j.tig.2017.10.007
- Santos-Pereira JM, Aguilera A (2015). R loops: new modulators of genome dynamics and function. *Nat Rev Genet* 16(10): 583-597. doi: 10.1038/nrg3961
- Freudenreich CH (2018). R-loops: targets for nuclease cleavage and repeat instability. *Curr Genet* 64(4): 789-794. doi: 10.1007/s00294-018-0806-z
- Sollier J, Cimprich KA (2015). Breaking bad: R-loops and genome integrity. *Trends Cell Biol* 25(9): 514-522. doi: 10.1016/j.tcb.2015.05.003
- Groh M, Gromak N (2014). Out of balance: R-loops in human disease. *PLoS Genet* 10(9): e1004630. doi: 10.1371/journal.pgen.1004630

- Sollier J, Stork CT, Garcia-Rubio ML, Paulsen RD, Aguilera A, Cimprich KA (2014). Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol Cell* 56(6): 777-785. doi: 10.1016/j.molcel.2014.10.020
- Chakraborty P, Grosse F (2011). Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA Repair* 10(6): 654-665. doi: 10.1016/j.dnarep.2011.04.013
- Groh M, Albulescu LO, Cristini A, Gromak N (2017). Senataxin: Genome Guardian at the Interface of Transcription and Neurodegeneration. *J Mol Biol* 429(21): 3181-3195. doi: 10.1016/j.jmb.2016.10.021
- Zhou R, Zhang J, Bochman ML, Zakian VA, Ha T (2014). Periodic DNA patrolling underlies diverse functions of Pif1 on R-loops and G-rich DNA. *Elife* 3: e02190. doi: 10.7554/eLife.02190
- El Hage A, French SL, Beyer AL, Tollervey D (2010). Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev* 24(14): 1546-1558. doi: 10.1101/gad.573310
- Yang Y, McBride KM, Hensley S, Lu Y, Chedin F, Bedford MT (2014). Arginine methylation facilitates the recruitment of TOP3B to chromatin to prevent R loop accumulation. *Mol Cell* 53(3): 484-497. doi: 10.1016/j.molcel.2014.01.011
- Li X, Manley JL (2005). Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* 122(3): 365-378. doi: 10.1016/j.cell.2005.06.008
- Gomez-Gonzalez B, Garcia-Rubio M, Bermejo R, Gaillard H, Shirahige K, Marin A, Foiani M, Aguilera A (2011). Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. *EMBO Journal* 30(15): 3106-3119. doi: 10.1038/emboj.2011.206

18. Li X, Niu T, Manley JL (2007). The RNA binding protein RNPS1 alleviates ASF/SF2 depletion-induced genomic instability. *RNA* 13(12): 2108-2115. doi: 10.1261/rna.734407
19. Pefanis E, Wang J, Rothschild G, Lim J, Kazadi D, Sun J, Federation A, Chao J, Elliott O, Liu ZP, Economides AN, Bradner JE, Rabadan R, Basu U (2015). RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. *Cell* 161(4): 774-789. doi: 10.1016/j.cell.2015.04.034
20. Singh DK, Pandita RK, Singh M, Chakraborty S, Hambarde S, Ramnarain D, Charaka V, Ahmed KM, Hunt CR, Pandita TK (2018). MOF Suppresses Replication Stress and Contributes to Resolution of Stalled Replication Forks. *Mol Cell Biol* 38(6). doi: 10.1128/MCB.00484-17
21. Salas-Armenteros I, Perez-Calero C, Bayona-Feliu A, Tumini E, Luna R, Aguilera A (2017). Human THO-Sin3A interaction reveals new mechanisms to prevent R-loops that cause genome instability. *EMBO J* 36(23): 3532-3547. doi: 10.15252/embj.201797208
22. Yu K, Chedin F, Hsieh CL, Wilson TE, Lieber MR (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat Immunol* 4(5): 442-451. doi: 10.1038/ni919
23. Daniels GA, Lieber MR (1995). RNA:DNA complex formation upon transcription of immunoglobulin switch regions: implications for the mechanism and regulation of class switch recombination. *Nucleic Acids Res* 23(24): 5006-5011. doi: 10.1093/nar/23.24.5006
24. Nadel J, Athanasiadou R, Lemetre C, Wijetunga NA, P OB, Sato H, Zhang Z, Jeddeloh J, Montagna C, Golden A, Seoighe C, Grealley JM (2015). RNA:DNA hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships. *Epigenetics Chromatin* 8: 46. doi: 10.1186/s13072-015-0040-6
25. Sanz LA, Hartono SR, Lim YW, Steyaert S, Rajpurkar A, Ginno PA, Xu X, Chedin F (2016). Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. *Mol Cell* 63(1): 167-178. doi: 10.1016/j.molcel.2016.05.032
26. Ginno PA, Lott PL, Christensen HC, Korf I, Chedin F (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell* 45(6): 814-825. doi: 10.1016/j.molcel.2012.01.017
27. Boguslawski SJ, Smith DE, Michalak MA, Mickelson KE, Yehle CO, Patterson WL, Carrico RJ (1986). Characterization of monoclonal antibody to DNA:RNA and its application to immunodetection of hybrids. *J Immunol Methods* 89(1): 123-130. doi: 10.1016/0022-1759(86)90040-2
28. Halasz L, Karanyi Z, Boros-Olah B, Kuik-Rozsa T, Sipos E, Nagy E, Mosolygo LA, Mazlo A, Rajnavolgyi E, Halmos G, Szekvolgyi L (2017). RNA-DNA hybrid (R-loop) immunoprecipitation mapping: an analytical workflow to evaluate inherent biases. *Genome Res* 27(6): 1063-1073. doi: 10.1101/gr.219394.116
29. Chen L, Chen JY, Zhang X, Gu Y, Xiao R, Shao C, Tang P, Qian H, Luo D, Li H, Zhou Y, Zhang DE, Fu XD (2017). R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. *Mol Cell* 68(4): 745-757 e745. doi: 10.1016/j.molcel.2017.10.008
30. Tresini M, Warmerdam DO, Kolovos P, Snijder L, Vrouwe MG, Demmers JA, van IWF, Grosveld FG, Medema RH, Hoeijmakers JH, Mullenders LH, Vermeulen W, Marteijn JA (2015). The core spliceosome as target and effector of non-canonical ATM signalling. *Nature* 523(7558): 53-58. doi: 10.1038/nature14512
31. Cristini A, Groh M, Kristiansen MS, Gromak N (2018). RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage. *Cell Rep* 23(6): 1891-1905. doi: 10.1016/j.celrep.2018.04.025
32. Hu Z, Zhang A, Storz G, Gottesman S, Leppla SH (2006). An antibody-based microarray assay for small RNA detection. *Nucleic Acids Res* 34(7): e52. doi: 10.1093/nar/gkl142
33. Bhatia V, Barroso SI, Garcia-Rubio ML, Tumini E, Herrera-Moyano E, Aguilera A (2014). BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature* 511(7509): 362-365. doi: 10.1038/nature13374
34. Powell WT, Coulson RL, Gonzales ML, Crary FK, Wong SS, Adams S, Ach RA, Tsang P, Yamada NA, Yasui DH, Chedin F, LaSalle JM (2013). R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation. *Proc Natl Acad Sci U S A* 110(34): 13938-13943. doi: 10.1073/pnas.1305426110
35. Soler M, Boque-Sastre R, Guil S (2017). RNA-FISH to Study Regulatory RNA at the Site of Transcription. *Methods Mol Biol* 1543: 221-229. doi: 10.1007/978-1-4939-6716-2_12
36. Kaback DB, Angerer LM, Davidson N (1979). Improved methods for the formation and stabilization of R-loops. *Nucleic Acids Res* 6(7): 2499-2517. doi: 10.1093/nar/6.7.2499
37. Xu B, Clayton DA (1996). 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* 15(12): 3135-3143. doi: 10.1002/j.1460-2075.1996.tb00676.x
38. Stavnezer J, Guikema JE, Schrader CE (2008). Mechanism and regulation of class switch recombination. *J Annu Rev Immunol* 26: 261-292. doi: 10.1146/annurev.immunol.26.021607.090248
39. Zarrin AA, Alt FW, Chaudhuri J, Stokes N, Kaushal D, Du Pasquier L, Tian M (2004). An evolutionarily conserved target motif for immunoglobulin class-switch recombination. *Nat Immunol* 5(12): 1275-1281. doi: 10.1038/ni1137
40. Ribeiro de Almeida C, Dhir S, Dhir A, Moghaddam AE, Sattentau Q, Meinhart A, Proudfoot NJ (2018). RNA Helicase DDX1 Converts RNA G-Quadruplex Structures into R-Loops to Promote IgH Class Switch Recombination. *Mol Cell* 70(4): 650-662 e658. doi: 10.1016/j.molcel.2018.04.001
41. Boque-Sastre R, Soler M, Oliveira-Mateos C, Portela A, Moutinho C, Sayols S, Villanueva A, Esteller M, Guil S (2015). Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. *Proc Natl Acad Sci U S A* 112(18): 5785-5790. doi: 10.1073/pnas.1421197112
42. Skourti-Stathaki K, Kamieniarz-Gdula K, Proudfoot NJ (2014). R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature* 516(7531): 436-439. doi: 10.1038/nature13787
43. Huertas P, Aguilera A (2003). Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol Cell* 12(3): 711-721. doi: 10.1016/j.molcel.2003.08.010
44. Gomez-Gonzalez B, Aguilera A (2007). Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. *Proc Natl Acad Sci U S A* 104(20): 8409-8414. doi: 10.1073/pnas.0702836104
45. Castellano-Pozo M, Santos-Pereira JM, Rondon AG, Barroso S, Andujar E, Perez-Alegre M, Garcia-Muse T, Aguilera A (2013). R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. *Mol Cell* 52(4): 583-590. doi: 10.1016/j.molcel.2013.10.006
46. Tuduri S, Crabbe L, Conti C, Tourriere H, Holtgreve-Grez H, Jauch A, Pantesco V, De Vos J, Thomas A, Theillet C, Pommier Y, Tazi J, Coquelle A, Pasero P (2009). Topoisomerase I suppresses genomic instability by

- preventing interference between replication and transcription. **Nat Cell Biol** 11(11): 1315-1324. doi: 10.1038/ncb1984
47. Dominguez-Sanchez MS, Barroso S, Gomez-Gonzalez B, Luna R, Aguilera A (2011). Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. **PLoS Genet** 7(12): e1002386. doi: 10.1371/journal.pgen.1002386
48. Polak P, Arndt PF (2008). Transcription induces strand-specific mutations at the 5' end of human genes. **Genome Res** 18(8): 1216-1223. doi: 10.1101/gr.076570.108
49. Hatchi E, Skourti-Stathaki K, Ventz S, Pinello L, Yen A, Kamieniarz-Gdula K, Dimitrov S, Pathania S, McKinney KM, Eaton ML, Kellis M, Hill SJ, Parmigiani G, Proudfoot NJ, Livingston DM (2015). BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. **Mol Cell** 57(4): 636-647. doi: 10.1016/j.molcel.2015.01.011
50. Zhang X, Chiang HC, Wang Y, Zhang C, Smith S, Zhao X, Nair SJ, Michalek J, Jatoi I, Lautner M, Oliver B, Wang H, Petit A, Soler T, Brunet J, Mateo F, Angel Pujana M, Poggi E, Chaldekias K, Isaacs C, Peshkin BN, Ochoa O, Chedin F, Theoharis C, Sun LZ, Curiel TJ, Elledge R, Jin VX, Hu Y, Li R (2017). Attenuation of RNA polymerase II pausing mitigates BRCA1-associated R-loop accumulation and tumorigenesis. **Nat Commun** 8: 15908. doi: 10.1038/ncomms15908
51. Schlacher K, Christ N, Siaud N, Egashira A, Wu H, Jasin M (2011). Double-strand break repair-independent role for BRCA2 in blocking stalled replication fork degradation by MRE11. **Cell** 145(4): 529-542. doi: 10.1016/j.cell.2011.03.041
52. Shivji MKK, Renaudin X, Williams CH, Venkitaraman AR (2018). BRCA2 Regulates Transcription Elongation by RNA Polymerase II to Prevent R-Loop Accumulation. **Cell Rep** 22(4): 1031-1039. doi: 10.1016/j.celrep.2017.12.086
53. Schwab RA, Nieminuszczy J, Shah F, Langton J, Lopez Martinez D, Liang CC, Cohn MA, Gibbons RJ, Deans AJ, Niedzwiedz W (2015). The Fanconi Anemia Pathway Maintains Genome Stability by Coordinating Replication and Transcription. **Mol Cell** 60(3): 351-361. doi: 10.1016/j.molcel.2015.09.012
54. Garcia-Rubio ML, Perez-Calero C, Barroso SI, Tumini E, Herrera-Moyano E, Rosado IV, Aguilera A (2015). The Fanconi Anemia Pathway Protects Genome Integrity from R-loops. **PLoS Genet** 11(11): e1005674. doi: 10.1371/journal.pgen.1005674
55. Ceccaldi R, Sarangi P, D'Andrea AD (2016). The Fanconi anaemia pathway: new players and new functions. **Nat Rev Mol Cell Biol** 17(6): 337-349. doi: 10.1038/nrm.2016.48
56. Hamperl S, Bocek MJ, Saldivar JC, Swigut T, Cimprich KA (2017). Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. **Cell** 170(4): 774-786 e719. doi: 10.1016/j.cell.2017.07.043
57. Richard P, Manley JL (2017). R Loops and Links to Human Disease. **J Mol Biol** 429(21): 3168-3180. doi: 10.1016/j.jmb.2016.08.031
58. Reddy K, Tam M, Bowater RP, Barber M, Tomlinson M, Nichol Edamura K, Wang YH, Pearson CE (2011). Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. **Nucleic Acids Res** 39(5): 1749-1762. doi: 10.1093/nar/gkq935
59. Loomis EW, Sanz LA, Chedin F, Hagerman PJ (2014). Transcription-associated R-loop formation across the human FMR1 CGG-repeat region. **PLoS Genet** 10(4): e1004294. doi: 10.1371/journal.pgen.1004294
60. Groh M, Lufino MM, Wade-Martins R, Gromak N (2014). R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. **PLoS Genet** 10(5): e1004318. doi: 10.1371/journal.pgen.1004318
61. Colak D, Zaninovic N, Cohen MS, Rosenwaks Z, Yang WY, Gerhardt J, Disney MD, Jaffrey SR (2014). Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. **Science** 343(6174): 1002-1005. doi: 10.1126/science.1245831
62. Tsoi H, Lau TC, Tsang SY, Lau KF, Chan HY (2012). CAG expansion induces nucleolar stress in polyglutamine diseases. **Proc Natl Acad Sci U S A** 109(33): 13428-13433. doi: 10.1073/pnas.1204089109
63. Haeusler AR, Donnelly CJ, Periz G, Simko EA, Shaw PG, Kim MS, Maragakis NJ, Troncoso JC, Pandey A, Sattler R, Rothstein JD, Wang J (2014). C9orf72 nucleotide repeat structures initiate molecular cascades of disease. **Nature** 507(7491): 195-200. doi: 10.1038/nature13124
64. Lim YW, Sanz LA, Xu X, Hartono SR, Chedin F (2015). Genome-wide DNA hypomethylation and RNA:DNA hybrid accumulation in Aicardi-Goutieres syndrome. **Elife** 4. doi: 10.7554/eLife.08007
65. Moreira MC, Klur S, Watanabe M, Nemeth AH, Le Ber I, Moniz JC, Tranchant C, Aubourg P, Tazir M, Schols L, Pandolfo M, Schulz JB, Pouget J, Calvas P, Shizuka-Ikeda M, Shoji M, Tanaka M, Izatt L, Shaw CE, M'Zahem A, Dunne E, Bomont P, Benhassine T, Bouslam N, Stevanin G, Brice A, Guimaraes J, Mendonca P, Barbot C, Coutinho P, et al. (2004). Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. **Nat Genet** 36(3): 225-227. doi: 10.1038/ng1303
66. Grunseich C, Wang IX, Watts JA, Burdick JT, Guber RD, Zhu Z, Bruzel A, Lanman T, Chen K, Schindler AB, Edwards N, Ray-Chaudhury A, Yao J, Lehky T, Piszczek G, Crain B, Fischbeck KH, Cheung VG (2018). Senataxin Mutation Reveals How R-Loops Promote Transcription by Blocking DNA Methylation at Gene Promoters. **Mol Cell** 69(3): 426-437 e427. doi: 10.1016/j.molcel.2017.12.030
67. Helmrich A, Ballarino M, Tora L (2011). Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. **Mol Cell** 44(6): 966-977. doi: 10.1016/j.molcel.2011.10.013
68. Kotsantis P, Silva LM, Irmischer S, Jones RM, Folkes L, Gromak N, Petermann E (2016). Increased global transcription activity as a mechanism of replication stress in cancer. **Nat Commun** 7: 13087. doi: 10.1038/ncomms13087
69. Hill SJ, Rolland T, Adelmant G, Xia X, Owen MS, Dricot A, Zack TI, Sahni N, Jacob Y, Hao T, McKinney KM, Clark AP, Reyon D, Tsai SQ, Joung JK, Beroukhir R, Marto JA, Vidal M, Gaudet S, Hill DE, Livingston DM (2014). Systematic screening reveals a role for BRCA1 in the response to transcription-associated DNA damage. **Genes Dev** 28(17): 1957-1975. doi: 10.1101/gad.241620.114
70. Jackson BR, Noerenberg M, Whitehouse A (2014). A novel mechanism inducing genome instability in Kaposi's sarcoma-associated herpesvirus infected cells. **PLoS Pathog** 10(5): e1004098. doi: 10.1371/journal.ppat.1004098
71. Yu TY, Kao YW, Lin JJ (2014). Telomeric transcripts stimulate telomere recombination to suppress senescence in cells lacking telomerase. **Proc Natl Acad Sci U S A** 111(9): 3377-3382. doi: 10.1073/pnas.1307415111