



Review

R-loop and its functions at the regulatory interfaces between transcription and (epi)genome

Arum Kim^{a,b}, Gang Greg Wang^{a,b,c,d,*}^a Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27599, USA^b Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27599, USA^c Curriculum in Genetics and Molecular Biology, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27599, USA^d Department of Pharmacology, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27599, USA

ARTICLE INFO

Keywords:

R loop
Chromatin
Transcription
DNA
RNA
RNA-DNA hybrid
Splicing
DNA damage
Cancer
N⁶-methyladenosine

ABSTRACT

R-loop represents a prevalent and specialized chromatin structure critically involved in a wide range of biological processes. In particular, co-transcriptional R-loops, produced often due to RNA polymerase pausing or RNA biogenesis malfunction, can initiate molecular events to context-dependently regulate local gene transcription and crosstalk with chromatin modifications. Cellular “readers” of R-loops are identified, exerting crucial impacts on R-loop homeostasis and gene regulation. Mounting evidence also supports R-loop deregulation as a frequent, sometimes initiating, event during the development of human pathologies, notably cancer and neurological disorder. The purpose of this review is to cover recent advances in understanding the fundamentals of R-loop biology, which have started to unveil complex interplays of R-loops with factors involved in various biological processes such as transcription, RNA processing and epitranscriptomic modification (such as N⁶-methyladenosine), DNA damage sensing and repair, and epigenetic regulation.

1. Introduction

When RNA hybridizes with its DNA template, the RNA-DNA hybrid duplex termed R-loop is formed. Such a heteronomous three-stranded structure includes an RNA-DNA hybrid and a displaced single-stranded DNA (ssDNA) [1–5]. R-loop represents an intermediate between the RNA (A-form) and DNA (B-form) conformations, often exhibiting a higher degree of stability than double-stranded DNAs (dsDNAs) [1–5]. Although R-loop was initially considered to be a by-product of transcription, R-loop can occur genome-wide and be detected among all organisms that cover bacteria, yeast, plant and animal. R-loops are known to be directly involved in a number of specialized cellular processes, such as replication of bacterial ColE1 DNA, bacteriophage T4 and mitochondrial DNA, clustered regularly interspaced short palindromic repeats (CRISPR) RNA (crRNA)-guided dsDNA recognition, RNAi-directed heterochromatin assembly in fission yeast, class switch recombination (CSR) of immunoglobulin genes, and formation and maintenance of **telomere**. More broadly, R-loops occur co-transcriptionally, often near gene promoters enriched with the C/G-high content, and are involved in regulation of RNA polymerase II (Pol-II) pausing, sense versus antisense transcription, and epigenetic

modifications [2–4]. R-loops generally form *in cis* near the sites of transcription but can also occur *in trans*. For the latter, Rad51p, an homologous recombination factor, was shown to promote formation of RNA-DNA hybrids *in trans* [6]; likewise, based on sequence complementarity, a long noncoding RNA (lncRNA) termed *APOLO* in *Arabidopsis* was reported to act *in trans* mediating R-loop formation at its target [7]. R-loops can also be detected among repeated elements such as LINE-1 [8].

It is generally viewed that excessive formation of R-loops induces the replication fork collision and Pol-II blocking, which are deleterious for productive transcriptional elongation and often lead to pathogenesis due to DNA double-strand breaks (DSBs) and genomic instability [2,9,10]. Therefore, understanding the homeostasis and regulation of R-loops shall have a far-reaching implication in biology (related to those termed “regulatory R-loops” [2,3]) and human diseases (which, in most cases, are related to excessive production of R-loops, also termed “unscheduled or unwanted R-loops” [2,3]). Here, we summarize the biochemical basis and molecular players that underlie the “creation”, “recognition”, “maintenance” and “resolution” of R-loops. In particular, this review will focus on the modulation of R-loops via epigenetic regulation of chromatin and RNA such as N⁶A methylation of DNA or RNA during the

* Corresponding author at: Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27599, USA.
E-mail address: greg.wang@med.unc.edu (G.G. Wang).

gene transcriptional and post-transcriptional processes, the newly unveiled RNA-DNA hybrid “readers”, and cancers resulting from the impaired R-loop balance.

1.1. Roles for R-loops in gene and epigenome regulation

1.1.1. Formation of R-loops is modulated by a number of factors, contributing to diverse cellular processes

In the “thread-back” model [10–12], co-transcriptional R-loops (also known as *cis* R-loops) can be established genome-wide when a nascent RNA transcript reanneals with its template DNA strand at the site of transcription. However, evidence exists showing that R-loops can also form *in trans* [6,11,12], in which an RNA transcript hybridizes to the complementary strand of DNA at a location distant from the original transcription site. R-loop formation is highly supported by conformational features (Fig. 1), such as negative supercoiling of DNA, the guanine (G) content and presence of G-quadruplex (G4) second structure or a single-strand DNA (ssDNA) nick [13–15]. Negative DNA superhelicity promotes R-loop formation and a recent study further suggested the important interplay between DNA base sequence and negative superhelicity in controlling R-loop stability [16]. The presence of G clusters on the RNA transcript facilitates the initiation of R-loop formation during transcription [14] due to the thermal stability of rG/dC base pairing. In accordance with thermal stability of G:C base pairing, R-loop hotspot regions were mapped to those with a feature of GC skew, including promoter and transcription termination regions exhibiting high G + C contents [17,18]. Once the formation of R-loops is initiated, the elongation of R-loops depends on the density of G on the non-template strand of DNA [14]. In addition, R loops are favoured by nicks on the non-template DNA strand downstream of the promoter [15]. Mechanistically, the transient removal of the non-template strand after the nicked position may increase the opportunity to form RNA-DNA hybrids [15]. Additionally, a set of biological processes, which at least include RNA Pol-II pausing, antisense transcription (often producing lncRNA), RNA and DNA modifications, and availability of a free RNA end (often increased due to RNA splicing or biogenesis malfunction), all enhance R-loop occurrence. Adenosine methylation of RNA (m^6A) or DNA (6mA) was also reported to modulate R-loops, potentially through recruitment of m^6A “readers” (Fig. 2). Conceivably, proteins with activity for binding ssDNA or R-loop may also influence R-loop formation/stabilization.

While regulatory R-loops actively contribute to diverse cellular processes, excessive R-loops (i.e., “unscheduled and unwanted R-loops” [2,3]) can be a source of a cell-intrinsic stress arising from the defects in transcription elongation, replication and DNA repair. Regarding the regulatory R-loops that are critically involved in normal physiological and biological processes, they can exert impact, not only on the RNA Pol-II behaviours but also on local chromatin structure or modifications

(notably, DNA demethylation and Polycomb-related modifications). In addition, the importance of regulatory R-loops has been also demonstrated in class switch recombination (CSR) of immunoglobulin genes [19,20]. It was reported that non-coding switch region transcripts promote formation of R-loops, which facilitate CSR, a process that also depends on DSBs, and the involvements of the origin recognition complex (ORC), MCM complex and the replicative helicase [21–23]. In the sections below, we discuss about recent advances in (epi)genomic regulation by R-loops.

1.1.2. Interplay between RNA Pol-II and co-transcriptional R-loop profoundly influences local transcription

It is generally viewed that, following RNA Pol-II pausing at transcription start sites (TSSs), R-loop accumulation often occurs (Fig. 3A). In support, Negative Elongation Factor Complex Member B (NELFB; also known as COBRA1), an integral subunit for mediating RNA Pol-II pausing, was reported to interact physically with and functionally antagonize the Breast cancer type 1 susceptibility protein (BRCA1) [24]; due to enhanced NELFB/COBRA1 activity and Pol-II pausing, the *BRCA1*-deficient tumor cells are featured by excessive accumulation of R-loops [24]. On the other hand, Kim et al. recently reported that the bromodomain-containing proteins BRD2 and BRD4, which are histone acetylation “readers” with transcriptional-elongation-facilitating activities, suppress R-loop formation [25]. While BRD4 recruits pTEFb for RNA Pol-II phosphorylation during transcriptional activation/elongation, a C-terminal region of BRD2 associates with DNA topoisomerase 1 (TOP1) inducing activation of the latter [25]. In consistence, a separate study also showed that depletion of BRD4 in cancer cells induces R-loop accumulation and transcription–replication collision events, leading to DNA damage and apoptosis of tumor cells [26] (Fig. 3A). Likewise, TFIIS recognizes RNA Pol-II backtracking, the reversible sliding of Pol-II along DNA and RNA, induces transcript cleavage and enables Pol-II to resume transcript elongation [27]; a recent work further suggested that expression of a transcriptional-elongation-defective mutant of TFIIS (which inhibits transcript cleavage) caused Pol-II pausing, extended backtracking, and triggered cellular accumulation of anterior R-loops, which are different from posterior R-loops [28] (Fig. 3A).

At promoter regions, GC skew-triggered formation of R-loop is often observed between TSS and the first exon-intron junction. Also, G-rich sequences potentially produce the G4 secondary structure on the non-template strand (Fig. 1) [29–31], which further contributes to R-loop stabilization and influences transcription. In agreement, the G4-binding proteins were reported to increase R-loop establishment in cancer cells, thereby inducing DNA damage [29–31]. Meanwhile, GC skew at transcription termination sites (TTSs) also causes R-loop, leading to RNA Pol-II pausing downstream from the poly(A) signals; here, it is viewed that R-loops enable efficient transcription termination of genes within

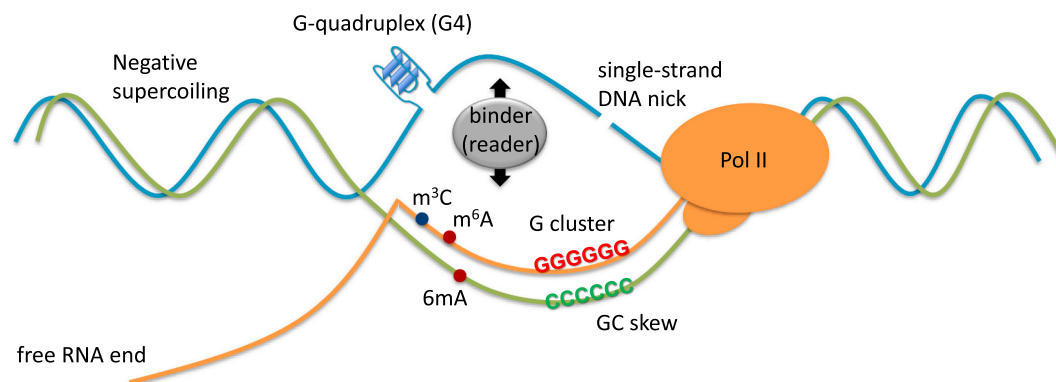


Fig. 1. Biophysical basis underlying the R-loop formation. R-loop formation is achieved by conformational features such as negative supercoiling, a nick and G-quadruplex on the displaced strand of DNA, G-cluster and GC skew on the RNA transcript, and free RNA end. RNA modifications such as m^3C and m^6A and/or DNA modification of 6mA may also potentially modulate homeostasis of R-loops.

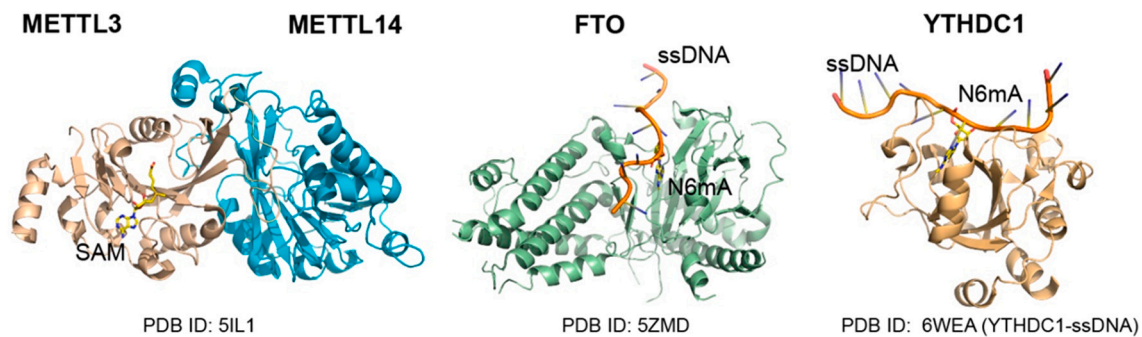


Fig. 2. Adenosine Methylation of RNA and DNA. RNA m⁶A is installed by writers such as the METTL3/METTL14 methyltransferase, removed by erasers include FTO and other demethylases, and can be recognized by readers such as YTH domain proteins. Please note that the above enzymes have promiscuous activity towards ssDNA.

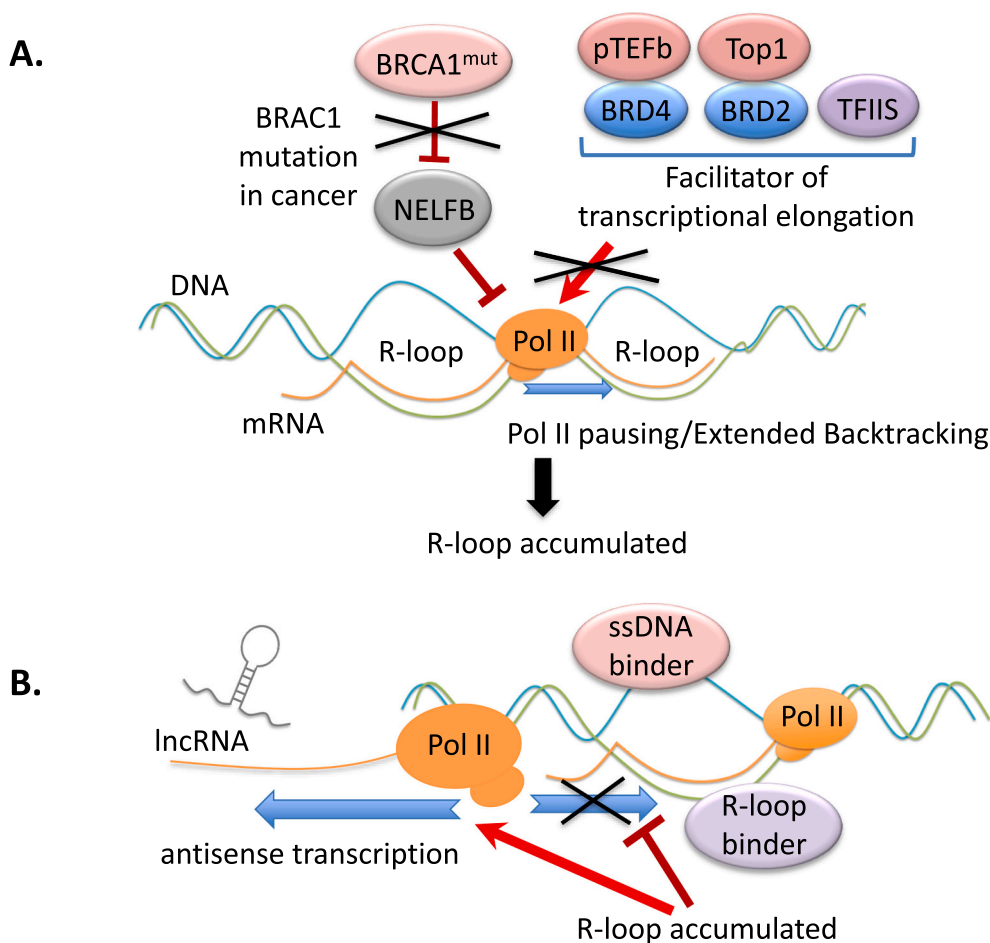


Fig. 3. Interplay between RNA Pol-II and co-transcriptional R-loop. (A) The impairment of transcriptional elongation causes RNA Pol-II pausing and promotes R-loop accumulation. Depletion of BRCA1 in cancer induces Pol-II pausing and triggers excessive R-loop formation, partly through the enhanced NELFB/COBRA1 activity. Along with TFIIS, BRD4 and BRD2 recruit pTEFb and TOP1, respectively, to facilitate transcriptional elongation. A transcriptional-elongation-defective mutant of TFIIS stimulates anterior R-loop formation. (B) Crosstalk among R-loops, Pol-II-mediated antisense transcription and various R-loop structure “sensors” such as ssDNA-specific binder and R-loop (RNA-DNA hybrid) “readers” is crucial for metabolism of R-loop and R-loop-related gene regulation. In the mammalian cells, accumulated R-loops promote antisense transcription of lncRNAs near the TSS of lncRNAs.

gene-dense regions [17,32,33], exhibiting a feature that resembles insulator or enhancer [18].

Furthermore, a free RNA end was recently appreciated to be an important initiating factor by allowing invasion of a nascent RNA into the DNA duplex [34]. Such a requirement for R-loop formation partially explains why diseased cells with mutation of RNA splicing factor or malfunction of premature transcriptional termination often display a phenotype of R-loop accumulation (will be discussed next).

Increasing evidence sheds light on reciprocal regulation involving R-loop, antisense transcription and lncRNA. Tan-Wong et al. recently demonstrated that the mammalian R-loops act as intrinsic promoters for promoting antisense transcription and thus generating lncRNAs

genome-wide [35] (Fig. 3B); here, R-loops are located near TSSs of lncRNAs and lncRNA expression was found to be sensitive to RNase H1, suggesting that R-loops act to enhance synthesis of antisense lncRNAs [35]. Previously, it was also reported that lncRNAs often induce R-loop formation, content-dependently regulating mRNA transcription of adjacent genes, as observed for *SPHK1* and *vimentin* in human cancer cells [36,37] and *FLOWERING LOCUS C (FLC)* in *Arabidopsis* [38]. For the latter, AtNDX, an ssDNA-binding homeodomain protein in *Arabidopsis*, was identified to stabilize R-loop at the promoter of *COOLAIR*, which is an antisense lncRNA at the *FLC* locus that functions to antagonize *FLC* expression through epigenetic silencing mechanisms [38].

Altogether, these observations highlight a general cell stress

produced by transcription and also unveil crucial crosstalk among RNA Pol-II, co-transcriptional R-loop and R-loop “sensor” (Fig. 3B; such as ssDNA binder and RNA-DNA hybrid “reader” to be discussed next). Importantly, R-loop can serve as a platform for recruiting or repelling DNA/RNA-binding factor and chromatin modifier, leading to context-dependent regulation of chromatin structure and transcription control, which is elaborated below.

1.1.3. RNA or DNA modifications modulate homeostasis of R-loops

In eukaryotes, methylation of adenosine within RNA (N^6 -methyladenosine or m^6A) emerges as the most prevalent post-transcriptional modification that modulates RNA metabolism [39,40]. RNA m^6A is installed by m^6A “writers” such as the METTL3/METTL14 methyltransferase and removed by m^6A “erasers” such as the m^6A demethylases, FTO and ALKBH5 [41] (Fig. 2). Additionally, a set of m^6A “readers”, including the YT521-B Homology (YTH) domain-containing proteins (such as YTHDC1, YTHDC2, YTHDF1, YTHDF2 and YTHDF3) (Fig. 2), heterogeneous nuclear ribonucleoproteins (hnRNPs; such as HNRNPC, HNRNPG and HNRNPA2B1), Insulin-like Growth Factor 2 mRNA Binding Proteins 1, 2, and 3 (IGF2BP1/2/3), and FMRP [41,42], selectively recognize and bind to m^6A -containing RNAs, mediating the biological consequence of m^6A [42].

A recent study uncovered that the METTL3/METTL14 complex deposits m^6A in nascent RNAs, which stabilizes R-loops around TTSs of m^6A^+ genes (defined as genes whose transcripts harbor m^6A), thereby facilitating transcription termination [43] (Fig. 4A, top). Such a cross-talk of METTL3/METTL14 with R-loop can be applied to cellular response to DNA double-strand break (DSB), in which METTL3 is

phosphorylated by the “sensor” of DSB, ATM, and localized to DSB sites for m^6A deposition [44]; in consistence, Kang et al. additionally reported that TonEBP recognizes R-loops generated by DNA damaging agents (see also next section of R-loop “readers”) and also subsequently recruits METTL3 to R-loop-associated damage sites for “writing” m^6A (Fig. 4A, top) [45]. Then, RNA m^6A recruits its “reader” YTHDC1 to enhance R-loop accumulation at DSB sites, leading to further recruitment of RAD51 and BRCA1 for homologous recombination (HR)-mediated repair and eventual resolution of R-loop [44] (Fig. 4A, bottom). “Reader” of m^6A can also act to suppress R-loop accumulation, thereby preventing R-loop-induced genomic instability [8]. Abakir et al. recently reported that, in human pluripotent stem cells, m^6A is prevalent in the RNA strands of RNA-DNA hybrids and such m^6A -containing R-loops undergo turnover in a cell cycle-dependent fashion— m^6A -marked R-loops are accumulated during S and G2/M phases, but depleted at G0/G1 phase [8]; here, YTHDF1, YTHDF2, HNRNPA2B1 and METTL3 all interact with R-loops, but in dividing cells, YTHDF2, an m^6A “reader” that regulates degradation of cytoplasmic mRNA, associates to mitotic chromatin and binds the R-loops within the LINE-1 repeated elements and intronic regions, leading to removal of R-loops [8]. It is important because YTHDF2 depletion leads to R-loop accumulation, cell growth retardation and DNA DSBs, supporting that YTHDF2 operates to remove excessive R-loops for maintenance of genomic stability [8] (Fig. 4A, bottom).

In human cells, METTL8, another methyltransferase-like family member, was shown to induce 3-methylcytidine (m^3C) within RNA, with implication in R-loop formation [46]. METTL8 interacts with various splicing and RNA-binding complexes and, in particular, a complex of

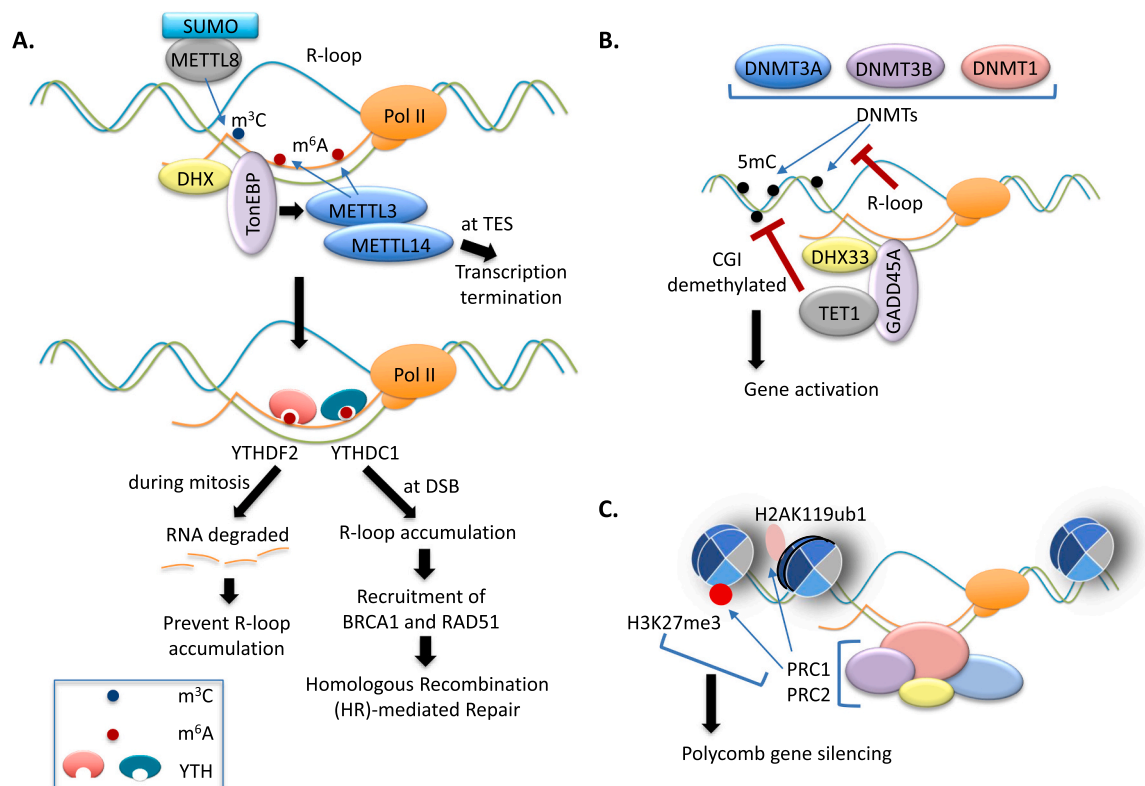


Fig. 4. RNA and DNA modifications regulate R-loop homeostasis during gene transcription and cellular response to DNA damage. (A) The m^6A methylation, installed by METTL3/METTL14 on nascent RNA, or the m^3C methylation, installed by SUMOylated METTL8, promotes R-loop formation, which leads to transcription termination at TTSs. The m^6A marks in the RNA strands of R-loops can be recognized by m^6A “readers” such as YTHDF2 and YTHDC1. In human pluripotent stem cells, YTHDF2 binds to R-loops leading to depletion of m^6A -marked R-loops during mitosis. At DSB sites, YTHDC1 increases accumulation of RNA-DNA hybrids and recruits BRCA1 and RAD51 for homologous recombination-mediated repair of DSB. (B) R-loops are poor substrates for methylation by mammalian DNA methyltransferases (DNMTs). GADD45A binds to R-loop, an event which can be further facilitated by the RNA helicase DHX33, and then recruits TET1 to induce CpG island (CGI) demethylation, leading to gene activation. (C) PcG proteins recognize R-loops and trigger local histone modifications such as H3K27me3 and H2AK119ub1, which provides a mechanism underlying PcG gene silencing.

METTL8-THOC2-RPA3 was suggested for R-loop association [46]. SUMOylation of METTL8 lysine 80 induces its nuclear localization and METTL8's methyltransferase activity of m³C was shown to be essential for the METTL8-related R-loop formation [46] (Fig. 4A, top).

In rice, m⁶A of DNA was reported to silence genes when present at promoters, but activate gene when present at gene bodies [47]. Furthermore, the rice genes containing R-loops contain more m⁶A within gene body and a short region downstream from TTS, and m⁶A+ genes are expressed at higher levels than those lacking m⁶A [48]. Though OsALKBH1 was suggested as m⁶A demethylase in rice [47], the causal relationship between DNA m⁶A and R-loop homeostasis remains to be fully determined.

Altogether, recent evidence starts to unveil context-dependent, dynamic modulation of R-loop through RNA or DNA modifications and their molecular players.

1.1.4. R-loop “readers” act to exert epigenetic impact for regulating gene expression

Once formed and stabilized, R-loops can induce subsequent change of local chromatin. Initially, R-loop was shown to be a poor substrate of mammalian DNA methyltransferases (DNMTs) [49], a phenomenon corroborated in vivo by genomic profiling [50] (Fig. 4B). Furthermore, Growth Arrest and DNA Damage-inducible Protein 45 alpha (GADD45A; Fig. 4B), a protein essential for cell cycle arrest and genomic stability, was recently identified as RNA-DNA hybrid “readers”, exhibiting a preferential affinity in vitro towards RNA-DNA hybrid over ssDNA/ssRNA or dsDNA/dsRNA [51]. We here use R-loop reader to refer to RNA-DNA hybrid reader. In mammalian cells, GADD45A binds R-loop generated by the antisense lncRNA TARID at the TCF21 promoter where GADD45A recruits TET1 to induce demethylation of CpG island (CGI), leading to activated expression of TCF21, a tumor suppressor [51]. Furthermore, R-loop-dependent recruitment of TET1 can be applied to thousands of TSS-associated CGIs in mouse embryonic stem cells (mESCs) as demonstrated by genomic profiling following RNase H1 treatment [51]. In agreement, Feng et al. reported that the RNA helicase DHX33 also recruits GADD45A leading to DNA demethylation and transcriptional activation [52] (Fig. 4B); DHX33 deficiency decreases R-loop formation in the promoter of Aurora kinase B, indicative of a role for DHX33 in R-loop stabilization and/or GADD45A:R-loop-mediated gene activation [52]. As GADD45A only binds to a subset of R-loops in cells [51], its binding specificity and structural basis remain to be characterized.

Furthermore, R-loops are frequently detected at Polycomb response elements (PREs) and involved in the chromatin recruitment of Polycomb Group (PcG) Proteins in *Drosophila* [53] (Fig. 4C). In the in vitro reconstitution system, PRC2 can mediate the formation of RNA-DNA hybrid and PcG proteins such as PRC2 and PRC1 also can recognize and bind the R-loops and open DNA bubbles, indicative of an interplay between R-loop and PcG gene silencing [53]. However, a separate study using mESCs did not support the promoting effect by PcG on R-loop, as PcG depletion did not affect R-loop formation in cells [54]. Nevertheless, the role for R-loop in enhancing PcG functionality appears to be conserved among *Drosophila* [53], animal [54] and plant [7]. In the mESCs and *Drosophila*, removal of R-loop by degrading RNAs with treatment of RNase H1 reduced the chromatin binding of PcG factors, leading to reduction in PcG-induced histone modifications (namely, H3K27me3 and H2AK119ub1) and derepression of the R-loop-associated target genes, but not those R-loop-negative ones [53,54]. At the genome-wide level, the silencing effects by PcG and R-loop are additive, suggesting their cooperation for optimal gene repression [54]. In *Arabidopsis*, the lncRNA APOLO acts in trans and mediates R-loop formation at its sequence-complementary targets, where R-loop subsequently recruits and decoys plant-specific Polycomb Repressive Complex 1 (PRC1) component LHP1 for controlling the local chromatin 3D conformation and gene transcription [7].

Additionally, AtALBA1 and AtALBA2, the two *Arabidopsis*-specific

Acetylation Lowers Binding Affinity (ALBA) family proteins, were reported to be RNA-DNA hybrid “readers” [55]. In vitro, AtALBA1 and AtALBA2 bind to the RNA-DNA hybrid and ssDNA in R-loops, respectively. Colocalized in the nucleus, AtALBA1 and AtALBA2 form a heterodimer or heteropolymer and bind R-loops at genic regions to maintain genome stability [55]. The molecular underpinning of ALBA1/2-mediated actions remains to be defined.

Thus, akin to binding of the displaced ssRNA by ssRNA-specific proteins such as RPA (Replication protein A), RNA-DNA hybrid is “sensed” and engaged by their specific “readers”, causing downstream changes of (epi)genome modification and chromatin structure.

1.2. Prevention and removal of excessive R-loops

R-loops are constantly removed/resolved, with an estimated 27,000 R-loops turned over every day and about 300 R-loops at the steady state in mammalian cells [56]. When excessively formed, R-loops induce DNA damage and replication stress, resulting in genomic instability. R-loops are prone to mutagenesis induced by activation-induced cytidine deaminase (AID) [57], the DNA replication fork stalling [58] and DSBs [8,59,60], all of which threaten genome integrity. In addition, spontaneous deamination of the displaced ssDNA in R-loops, replication fork collisions with blocked Pol II by transcriptional R-loops and R-loops-induced histone 3 Ser10 phosphorylation (H3S10P), a mark of chromatin compaction, eventually lead to DSBs and genomic instability [9]. Moreover, aberrant R-loop accumulation activates the replicative stress “sensor”, ATR, via replication fork reversal and the MUS81 nuclease [61]. ATR then represses transcription-replication collisions and promotes DNA synthesis and a G2/M cell-cycle arrest, thereby preventing genomic instability [61].

Given the importance of regulating R-loop abundance and distribution, a wealth of mechanisms evolves in controlling excessive R-loop formation, which at least include (i) R-loop formation prevention by TOP1 and (ii) RNA-binding proteins (RBPs), (iii) R-loop removal by RNase H1/2, (iv) R-loop prevention/surveillance by various proteins involving RNA biogenesis, degradation and metabolism, such as spliceosome, RNA exosome and export complexes, and (v) R-loops resolution by helicase.

While TOP1 inhibits R-loop formation by resolving negative supercoiling behind Pol-II, RBPs, involved in RNA surveillance, interact with nascent RNAs to prevent transcription-associated R-loop formation (Fig. 5A). For example, Npl3, an RNA-binding heterogeneous nuclear RNP (hnRNP), also prevents the R-loop-mediated genome instability [62]. RNases H is a class of ribonuclease that specifically degrades the RNA moiety in RNA-DNA hybrids. In eukaryotes, there are two types of RNases H, RNase H1 and RNase H2, and human RNase H1 has been adopted for genome-wide R-loop profiling. Loss of RNase H increases R-loop formation in cells. RNase H2 also binds with Breast cancer type 2 susceptibility protein (BRCA2) and manages R-loops at DSBs [63]. RNA-processing factors (Fig. 5A), such as Serine/arginine-rich splicing factor (SRSF1, also known as ASF1/SF2), the THO/TREX complex and Pcf11, have been reported to prevent R-loops by promoting mRNA processing and exporting [64–66]. The absence of Trf4, a poly(A) polymerase of TRAMP (Trf4–Air2–Mtr4p polyadenylation) complex, induces R-loop accumulation [67]. RNA exosome, including exosome component 3 (EXOSC3, Rrp40 in yeast) and exosome component 10 (EXOSC10, Rrp6 in yeast), mediated degradation of non-coding RNAs prevents R-loops to maintain genome integrity [68]. R loops can be resolved by helicases, with the reported activities reported for human senataxin (SETX) or its yeast homologue Sen1, Aquarius (AQR), DDX5, DDX23, DHX9, UAP56/DDX39B and the bacterial RNA helicase Rho (Fig. 5B). For instance, SETX resolves R-loops at transcriptional pause sites and enables access of XRN2, a 5′-3′ exoribonuclease that plays a role in preventing formation of R loops, at 3′ cleavage poly(A) sites, thereby promoting the termination of transcription [32,69,70]. AQR (also known as intron-binding protein 160) also acts as a R-loop removal factor. AQR is

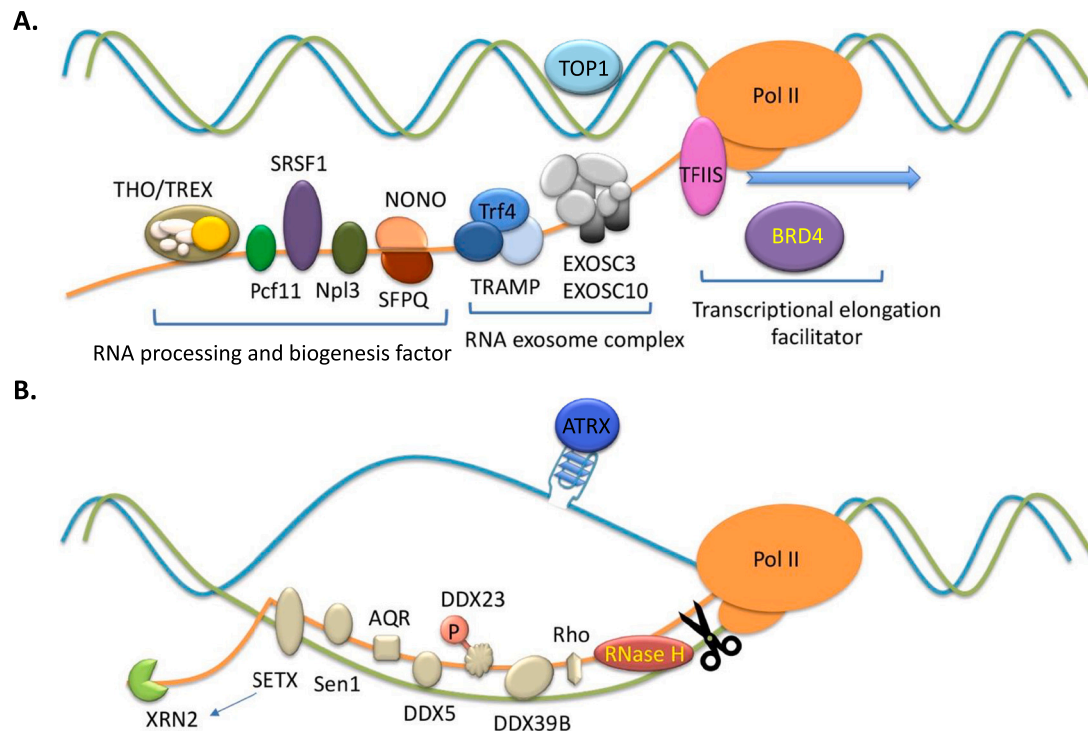


Fig. 5. Surveillance and prevention of excessive R-loop formation. (A) R-loop formation is prevented by topoisomerase 1 (TOP1) as well as specific RNA-binding proteins (RBPs) such as SRSF1, the THO/TREX complex, Pcf11, Npl3, non-POU domain-containing octamer-binding protein (NONO) and Splicing factor proline- and glutamine-rich (SFPQ). RNA exosome complexes such as Trf4, exosome component 3 (EXOSC3) and exosome component 10 (EXOSC10), as well as transcription factor TFIIS and epigenetic reader Bromodomain-containing protein 4 (BRD4) play roles in preventing R-loops. (B) The removal of R-loops is mediated by RNase H and helicases such as senataxin (SETX), Sen1, Aquarius (AQR), DDX5, DDX23, DDX39B and Rho. 5'-3' exoribonuclease XRN2 cooperates with SETX or DDX5 to remove R-loops. ATRX at telomere may be involved in resolving R-loops by suppressing deleterious DNA secondary structures.

integrated into the spliceosome in a position-specific manner, and loss of AQR results in DSB formation and R-loop accumulation [60,71]. This pathway depends on the transcription-coupled nucleotide excision repair (TC-NER) factor Cockayne syndrome group B (CSB), which recruits RAD52, then initiate transcription-coupled homologous recombination (TC-HR) by reactive oxygen species (ROS)-induced R-loops [60,72,73]. In brief, CSB recruits the nucleotide excision repair endonucleases XPF and XPG, and these endonucleases lead to the processing of R-loops into DSBs, indicating the role of AQR in R-loop repression [60,72]. Rho, the bacterial RNA helicase involved in transcription termination, is essential for bacteria survival and Rho-dependent transcription termination is essential for R-loop prevention [74].

Additionally, recent studies highlight important crosstalk and cooperation of helicase, RNA biogenesis/splicing factor and DNA damage repair protein in regulating R-loop homeostasis and safeguarding genome integrity. For instance, DDX5, a DEAD-box RNA helicase, reduces R-loop occurrence and, when it is methylated by PRMT5 at its C-terminal RGG/RG motif, DDX5 also interacts with XRN2 for the release of Pol-II at TTSs [75]. Likewise, UAP56/DDX39B was reported as a key RNA-DNA helicase for R-loops' resolution [76]. RNA Pol-II pausing induces the recruitment of serine/arginine protein kinase 2 (SRPK2) and DDX23 to specific R-loop-containing loci where SRPK2 phosphorylates DDX23, which then acts to prevent the R-loop-related genomic instability [77]. Mutations of DDX23 are observed in human cancer, with its homozygous deletion frequently observed in adenoid cystic carcinoma [77]. DHX9, a helicase carrying the *in vitro* R-loop-resolving activity, contributes to R-loop suppression and transcriptional termination in cells [78]. DHX9 interacts with poly(ADP-ribose) polymerase 1 (PARP1) and prevents R-loop-associated DNA damage in response to TOP1 inhibitor, camptothecin [78]. Another study demonstrated that, in the absence of RBPs and splicing factors, DHX9 enhances R-loop occurrence through unwinding nascent RNAs to generate the free RNA end [79],

which indicates a coordinated process for R-loop prevention and RNA biogenesis. The helicase senataxin (SETX), on the other hand, is associated with AOA2/ALS4 neurodegenerative disorders. SETX interacts with BRCA1 and recruitment of the BRCA1:SETX complex to the R-loop-associated TTSs helps to restrict R-loop-induced DNA damage [80]. Transcription-dependent recruitment of ATRX, a member of the SWI/SNF family of chromatin remodeling factors, to telomeric G-rich repeats (TTAGGG)_n, is also associated with the presence of R-loops [81]. ATRX has been shown to resolve R-loops or recruit other enzymes that degrade R-loops [81].

1.3. R-loops and DNA repair

RNA-DNA hybrid can also be formed upon DNA damage at transcribed DNA and regulate the DNA repair process. For example, a ssDNA nick downstream of the promoter stimulates R-loop formation; and laser irradiation or radiomimetics-induced nicks in DNA trigger the formation of R-loops and RNA-DNA hybrids at DSB sites [82–84]. In addition to R-loop-mediated recruitment of RAD51 or RAD52 during the HR-mediated DSB repair (Fig. 5), R-loops affect DSB repair process by HR or non-homologous end joining (NHEJ) pathway. In fission yeast, RNA-DNA hybrids in damaged regions recruit replication protein A (RPA) to ssDNA overhangs, but these hybrids should be degraded by RNase H to complete the DNA repair process [83]. Similarly, persistent R-loops in damaged sites block repair process in budding yeast [85]. On the contrary, the existence of RNA-DNA hybrids around DSB facilitates DNA repair by both HR and NHEJ pathways in human cells and these processes require Drosha, miRNA biogenesis enzyme, which manages the recruitment of repair factors [86].

1.4. Genome-wide R-loop profiling methodologies

Techniques for profiling R-loop distribution genome-wide were developed, including (i) bisulfite-based footprinting, (ii) S9.6 antibody-based DNA:RNA Immunoprecipitation coupled with sequencing (DRIP-seq) and (iii) RNase H-based approach (Table 1). Sodium bisulfite treatment induces the deamination of unmethylated cytosines to uracils in ssDNA [87]. Under non-denaturing conditions, bisulfite conversion can be used as marks for R-loop footprinting which detects RNA-DNA hybrid-dependent ssDNA footprints in genomic DNA [50]. The most widely adopted strategy for profiling R-loops has been based on the intrinsic specificity of the monoclonal S9.6 antibody for the RNA-DNA hybrid structure [88]. DNA:RNA Immunoprecipitation (DRIP), which displays a specificity for RNA-DNA hybrids, has been coupled with high-throughput sequencing, termed DRIP-seq, for mapping R-loops [50,89]. Based on the robustness of DRIP-seq, various modified methodologies have been established such as RDIP-seq (RNA:DNA immunoprecipitation, followed by sequencing) [90], S1-DRIP-seq (S1 nuclease DNA:RNA immunoprecipitation, followed by deep sequencing) [91], DRIPc-seq (DNA:RNA immunoprecipitation, followed by cDNA conversion coupled to high-throughput sequencing) [18,89], ssDRIP-seq (single-strand DNA ligation-based library construction from RNA-DNA hybrid immunoprecipitation, followed by sequencing) [92] and bisDRIP-seq (bisulfite-DNA:RNA immunoprecipitation, followed by sequencing) [93]. Recently, a modified DRIP-seq technology termed qDRIP-seq (quantitative differential DNA:RNA immunoprecipitation followed by sequencing) uses synthetic RNA-DNA hybrid internal standards as spike-in control, which enables absolute quantification of R-loop content with high resolution [56]. Alternatively, DRIP-chip (DNA:RNA immunoprecipitation followed by hybridization on tiling microarrays) [94] and S9.6 ChIP-seq (chromatin immunoprecipitation with the S9.6 antibody, followed by deep sequencing) [95] have been also proposed as S9.6-based methods for the identification of R-loop structures. Distinctive from S9.6 antibody-dependent R-loop detection, other mapping techniques rely on RNase H, an enzyme that degrades the RNA strand in RNA-DNA hybrids in the 3' → 5' direction. In DRIVE-seq (DNA:RNA in vitro Enrichment, followed by sequencing), a catalytically-dead human RNase H1, which binds to RNA-DNA hybrids but not resolve them, is used to affinity pull-down to capture in vitro R-loops [50]. R-ChIP (R-loop chromatin immunoprecipitation) is also based on a catalytically-dead human RNase H1, fused with a nuclear localization signal (NLS) at the N terminus and a V5 tag at the C terminus, for in vivo R-loop profiling [34,96]. A R-loop-capturing method, termed MapR, employs CUT&RUN strategy to overcome certain disadvantages of previous RNase H-based approaches [97,98]. In MapR, R-loop regions are recognized by a catalytically-dead RNase H1, cleaved by micrococcal nuclease (MNase) digestion, and released from chromatin for sequencing [97,98]. More recently, a R-loop imaging technique for detecting the RNA-DNA hybrids by immunofluorescence has been reported [99]. Using catalytically inactive human RNase H1 tagged with GFP (GFP-dRNH1), this method allows imaging and quantifying R-loops [99]. Please note that both S9.6 antibody- and RNase H1-based mapping techniques exhibit intrinsic target bias [56,100], off-target effects (e.g. S9.6 for dsRNA) or high background (for RNase H1). Further optimized strategies for accurately identifying R-loops genome-wide are required, due to recently appreciated relevance of R-loops in human diseases.

1.5. R-Loops and human disease

Previous evidence has demonstrated an intimate relationship between R-loop and pathogenesis of neurodegenerative disease, motor neuron disorder and autoimmune disease; for these topics, reviewers shall refer to other reviews [101–103]. We focus this section on the role for R-loops in oncogenesis. R-loop-induced DNA damage causes genomic instability, which can drive cancer initiation/progression [104]. In support, the tumor suppressors BRCA1 and BRCA2 participate in repair

of R-loop-induced DNA damage and depletion of BRCA1/2 increased R-loop accumulation [63,80,105]. In ER-responsive breast cancer, estrogen can stimulate R-loop formation leading to DNA damage and genome instability, suggesting that estrogen-induced R-loops to be tumorigenic [106].

1.5.1. R-loop accumulation underlying cancer with spliceosome mutation

Myelodysplastic syndromes (MDS) are clonal hematopoietic diseases, featured with ineffective hematopoiesis and morphological dysplasia leading to cytopenias and a higher risk of development to acute myeloid leukemia (AML) [107]. Somatic mutations of splicing factors, including SF3B1, U2AF1 (also known as U2AF35), SRSF2 and ZRSR2, are prevalent in MDS patients [107,108]. SF3B1, a component of the core spliceosome U2 complex, is the most commonly mutated spliceosomal gene in MDS [109]. SF3B1 mutations were typically found in MDS with increased ring sideroblasts, including refractory anaemia with ring sideroblasts (RARS) and refractory cytopenia with multilineage dysplasia with ring sideroblasts (RCMD-RS). In bone marrow stem/progenitor cells from MDS and AML patients with SF3B1 mutation, there is a significant increase of R-loops, which causes DNA damage and ATR-Chk1 activation [110] (Fig. 6A). Similar effects of U2AF1 and SRSF2 mutations were also observed [104,111,112]. Mechanistically, SRSF2 mutation interferes with p-TEFb translocation from the 7SK complex to Pol-II resulting in transcriptional elongation defection, which may explain the contribution of this splicing factor mutation to R-loop elevation [111]. Accumulating evidence also revealed that splicing factor mutations are mutually exclusive in MDS patients [113] and that each mutation exhibits own specific/selective effect on R-loop formation. Indeed, SRSF2 mutation induces R-loops at TSSs and U2AF1 mutation enhances R-loops in both promoter and non-promoter regions [111], indicating dynamic characteristics of R-loop formation. Importantly, splicing factor mutation-caused perturbations of RNA splicing and R-loop accumulation occur independently of one another [111,112], indicative of cooperation between the two changes during MDS pathogenesis (Fig. 6A).

1.5.2. R-loop accumulation underlying cancer with microRNA-processing mutation

R-loop is potentially involved in generating chromosomal breakpoints and rearrangement during tumorigenesis. Embryonal tumors with multilayered rosettes (ETMRs) are highly aggressive pediatric CNS malignancies, including embryonal tumor with abundant neuropil and true rosettes (ETANTR), ependymoblastoma (EBL) and medulloepithelioma (MEPL) [114]. A proposed driver mutation of ETMRs is genomic alteration of the chromosome 19 (19q13.42) microRNA cluster (C19MC) [114–116], either amplification of C19MC or aberrant fusion of TTYH1-C19MC, which leads to high expression of C19MC and an embryonic brain-specific DNMT3B isoform. However, the molecular mechanism underlying C19MC abnormality was previously unclear. Lambo et al. recently reported that R-loops accumulated in ETMRs are enriched around C19MC, which likely induced chromosomal instability and breakage leading to C19MC amplification and/or rearrangement in ETMRs [117]. Interestingly, a subset of ETMRs is featured with damaging lesion of DICER1, an miRNA-processing endoribonuclease, with somatic or germline mutations found clustered at its helicase and RNASE functional domains [117]. DICER1 malfunction, likely acting as a cancer susceptibility mutation, potentially interferes with biogenesis of microRNAs being transcribed at the C19MC locus, indicative of a casual mechanism underlying the R-loop-associated instability/breakage of C19MC [117]. In support, somatic amplification of the miR-17-92 miRNA cluster on chromosome 13 (also known as MIR17HG) also exists in ETMRs, regardless of C19MC amplification [117] (Fig. 6B).

1.5.3. RNA-binding factor aberration resulting in R-loop accumulation in sarcoma

Ewing sarcoma, a highly aggressive cancer that occurs primarily in

Table 1
Techniques for profiling genome-wide R-loop distribution.

	Technique	Innovation	Experimental workflow	Strengths	Weaknesses	Ref.
DRIP-seq and its derivatives	DRIP-seq	Using the intrinsic specificity of the S9.6 antibody for RNA-DNA hybrid	1. DNA extraction 2. Restriction enzyme digestion 3. Immunoprecipitation with S9.6 antibody 4. ds DNA sequencing	Convenient, Robust signal	Low resolution, Higher back-ground, Not strand specific, S9.6 antibody's bias in sequence recognition [100]	[50,89]
	RDIP-seq	Pre-treatment with RNase I, Use of sonication, Directional sequencing	1. Whole cell nucleic acid sonication 2. RNase I pre-treatment 3. Immunoprecipitation with S9.6 antibody 4. RNA sequencing	Reduced background, Strand-specific	A sonication step for nucleic acid fragmentation, which has been shown to reduce the number of genomic R-loops, Off target affinity of the S9.6 antibody for dsRNA	[90]
	S1-DRIP-seq	Using optimized levels of S1 nuclease to preserve RNA-DNA hybrid during sonication	1. DNA extraction 2. S1 nuclease digestion 3. Sonication 4. Immunoprecipitation with S9.6 antibody	Quantitative recovery of R loops, High-resolution	Not strand specific, S9.6 antibody's bias in sequence recognition	[91]
	DRIPc-seq	Further digestion by DNase I, cDNA conversion	5. dsDNA sequencing 1. DNA extraction 2. Restriction enzyme digestion 3. Immunoprecipitation with S9.6 antibody 4. DNase I treatment 5. RNA recovery and reverse transcription to cDNA 6. RNA sequencing	High resolution, Strand specific	Off target affinity of the S9.6 antibody for dsRNA	[18,89]
	ssDRIP-seq	Distinguish specific DNA strands with fewer steps for library construction than DRIPc-seq	1. DNA extraction 2. Sonication 3. Immunoprecipitation with S9.6 antibody 4. ssDNA sequencing	Strand specific	S9.6 antibody's bias in sequence recognition	[92]
	bisDRIP-seq	In vivo R-loop profiling, combining the use of the S9.6 antibody with sodium bisulfite treatment	1. Cell lysis in the presence of bisulfite 2. DNA extraction 3. Restriction enzyme digestion 4. Immunoprecipitation with S9.6 antibody 5. Bisulfite-modified dsDNA sequencing	Discriminate between the R-loop sequence and the surrounding non-R-loop sequence, High resolution, Strand specific	High sequencing depth is needed, Background conversions in ds DNA by bisulfite, S9.6 antibody's bias in sequence recognition	[93]
	qDRIP-seq	Combining synthetic RNA-DNA hybrid internal standards (spike-in)	1. Adding spike-in to cell lysate 2. Sonication 3. Immunoprecipitation with S9.6 antibody 4. ssDNA sequencing	Accurate cross-condition normalization, Absolute quantitation, Sensitive, high resolution, Strand-specific	S9.6 antibody's bias in sequence recognition	[56]
Other S9.6 based approaches	DRIP-chip	DRIP followed by hybridization on tiling microarray	1. Crosslinking with formaldehyde 2. Sonication 3. Immunoprecipitation with S9.6 antibody 4. T7 RNA polymerase amplification 5. Biotin labeling 6. Microarray		S9.6 antibody's bias in sequence recognition	[94]

(continued on next page)

Table 1 (continued)

	Technique	Innovation	Experimental workflow	Strengths	Weaknesses	Ref.
	S9.6 ChIP-seq (chromatin immunoprecipitation with antibody S9.6, followed by deep sequencing of immunopurified DNA fragments)	Application of ChIP-seq to mapping R loops	1. Crosslinking with formaldehyde 2. Sonication 3. Immunoprecipitation with S9.6 antibody 4. Reverse crosslinking 5. dsDNA sequencing	High-resolution	Formaldehyde crosslinking could affect results, Not strand specific	[95]
RNase H based approaches	DRIVE-seq (DNA:RNA in vitro Enrichment)	Using specificity of catalytically dead RNase H1 for RNA-DNA hybrid	1. Genomic DNA extraction 2. Restriction enzyme digestion 3. Catalytically dead RNase H1 incubation 4. Pull-down 5. dsDNA sequencing	Enable the specific and near quantitative recovery of R loop molecules in complex nucleic acid mixture Exquisite specificity of RNase H	Low capture efficiency, Not strand specific, Substantial RNase H-resistant regions on the genome	[50]
	R-ChIP	In vivo R-loop profiling using catalytically dead RNase H1	1. Introduce V5-tagged catalytically dead mutant RNase H1 into cells 2. Sonication 3. Immunoprecipitation with anti-V5 antibody 4. ssDNA sequencing	Exquisite specificity of RNase H, Strand-specific	Substantial RNase H-resistant regions on the genome, Require the generation of stable cell lines (time-consuming)	[34,96]
	MapR	Combining the specificity of RNase H for RNA-DNA hybrid with CUT&RUN approach	1. Immobilize cells on beads 2. Incubate with GST-RNaseHΔcat-MNase 3. R-loop recognition by RNaseHΔcat 4. MNase mediated R-loop cleavage 5. DNA sequencing	Antibody-independent, Does not require the generation of stable cell lines (fast and convenient), High sensitivity with low input material	substantial RNase H-resistant regions on the genome	[97,98]
	R-loop imaging	Imaging and quantifying R-loops using GFP-catalytically dead RNase H1 (dRNH1)	1. Expression and purification of GFP-dRNH1 2. Transfection in fixed cells 3. Imaging	Using purified GFP-dRNH1 protein, Bypassing the need for cell line engineering	GFP-dRNH1 has its own limitations (Non-specific binding, Binding preference to G-rich)	[99]

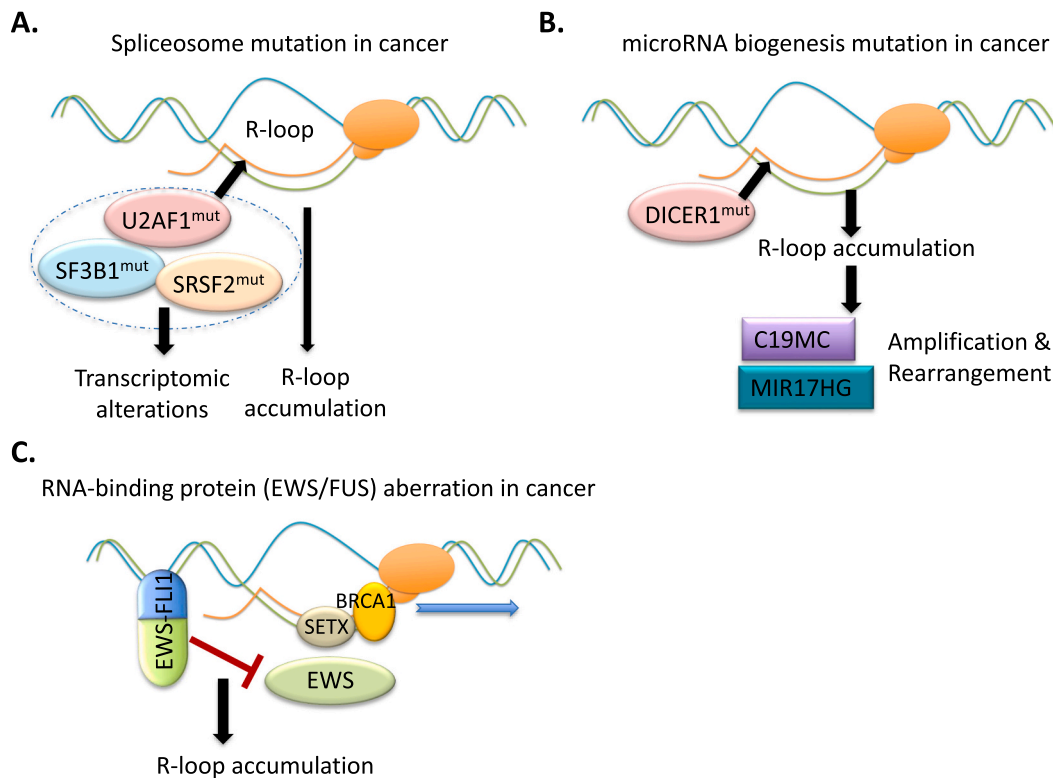


Fig. 6. R-loop aberration is prevalent, sometimes causal, during development of human cancer. (A) Splicing factor mutations induce Pol II pausing and increase R-loops, which lead to DNA damage and ATR-Chk1 activation, influencing clonal hematopoiesis and/or leukemic transformation. (B) During clonal evolution of ETMRs, R-loops are enriched around C19MC or MIR17HG, likely due to the loss-of-function mutation of the microRNA-processing enzyme DICER1, which induces the chromosomal breakage and rearrangement, such as the TTYH1-C19MC fusion and somatic amplification of C19MC or MIR17HG as recurrently detected in ETMRs. (C) In Ewing sarcoma, the oncogenic EWS-FLI1 chimera promotes the accumulation of R-loops. The N-terminal protein-protein interaction domain of EWS-FLI1 is essential for R-loop accumulation, which likely acts to interfere with functionality of wild-type EWS; and a functional sequestration and impairment of BRCA1, due to R-loop accumulation and DNA damage management, might underlie the genomic instability and therapeutic response to PARP inhibitors.

the bone or soft tissue, is characterized by chromosomal translocation of t(11;22)(q12;q24), which generates a chimeric oncoprotein EWS-FLI1. Recently, Gorthi et al. reported that wild-type EWS exhibits a multifaceted function in binding RNA, slowing transcriptional elongation and decreasing R-loop formation; on the other hand, EWS-FLI1 facilitates transcription and enhances the R-loop accumulation, resulting in the increased replication stress in Ewing sarcoma [118]. EWS-FLI1 forms interaction with EWS through its N-terminal region and the R-loop-inducing effect by EWS-FLI1 is likely due to EWS interference [118] (Fig. 6C). Interestingly, Ewing sarcoma also displays BRCA1-deficient tumor like phenotypes due to association/sequestration of BRCA1 (and associated helicase SETX [80]) with R-loop and stalled transcriptional machinery, suggesting a potential mechanism underlying oncogenesis [118]. Importantly, BRCA1 deficiency rendered a hypersensitivity of Ewing sarcoma to chemotherapy such as PARP1 inhibitor [118].

2. Concluding remarks

Intensive studies during recent years have allowed important insights into R-loop biology. R-loop is now appreciated to be relevant in a wide range of biological processes. As mentioned above, their roles can be beneficial (so-called “regulatory R-loops”) or detrimental (so-called “unscheduled or unwanted R-loops” when too many of them are excessively generated and not resolved), which depends on cellular context.

On one hand, regulatory roles of R-loops are important in normal physiology. Besides the above-mentioned functions of regulatory R-loops in chromatin regulation (DNA methylation and histone

modifications), transcription and a range of specialized processes such as CRISPR and class switch recombination, other documented examples include prevention of telomere shortening and premature cellular senescence by R-loops [119]; also, the oncogene *c-MYC* is activated by the interaction of Tudor domain containing protein 3 (TDRD3) and topoisomerase IIIB (TOP3B), which decreases negative supercoiled DNA and transcription-generated R-loops [120], indicating a regulatory role of R-loops. On the other hand, when R-loops cannot be efficiently resolved, unwanted, excessively accumulated R-loops elicit signals of DNA damage repair and responses, which can shape and/or initiate the development of human diseases.

R-loop's function and regulation are context-dependent and complicated, with questions and issues to be answered. First, further research needs to be directed at developing optimized techniques for more accurate genome-wide R-loop detection and mapping strategies, partly due to a prevalent involvement of R-loop in human pathologies such as cancer. Second, a set of outstanding questions related to R-loop biology and regulation remain unaddressed. For example, while most of the produced R-loops are either removed or resolved, the subset of R-loops are retained in the genome, so what determines the fate of R-loops? In other words, what differentiates a minority of ‘regulatory’ R-loops from a majority of those ‘unscheduled or unwanted’ ones to be removed? Also, once R-loop is formed and retained, what are the molecular determinants as of context-dependent functions of R-loops? Because it has been reported that some of R-loops are related to gene activation and some others associated with repression. Furthermore, m⁶A modifications occur frequently in the RNA strands of RNA-DNA hybrids, involving m⁶A “writer” (i.e., the METTL3-METTL14 heterodimer) and “reader” (YTH domain proteins such as YTHDF2 and

YTHDC1). Then, what is the causal and sequential relationship among these events, which include the R-loop formation and removal and m⁶A “writing” and “reading”? Does m⁶A serve as a cellular ‘sensing’ mechanism for R-loop detection and subsequent removal? Moreover, as de-regulations of RNA splicing and processing processes and alteration/mutation of RNA-binding factors are frequent in a wide range of human cancers. Does this mean that excessive R-loop accumulation so far observed in a few of cancer types can be generalized to many others tumor types.

Last, can the field leverage upon current understanding of R-loops for the development of new therapeutic strategies? One important therapeutic implication from studies of cancers exhibiting R-loop accumulation is that these cancers generally have a high demand for DNA damage management and repair, which thereby renders a relative hypersensitivity to therapy with PARP and/or TOP1 inhibitors as reported in ETMRs [117], Ewing sarcoma [118] and other tumor models [26]. Future investigation is warranted to expand such ideas to other tumor types exhibiting similar R-loop accumulation, since several PARP inhibitors are now FDA-approved for treatment of breast, ovarian and prostate cancer patients carrying the mutation of DNA damage repair gene (BRCA1/2, ATM or ATR). A better understanding of R-loop homeostasis and the underlying molecular basis shall continue to impact on biology and medicine in the years to come.

CRedit authorship contribution statement

Arum Kim: Writing, Reviewing and Editing

Gang Greg Wang: Writing, Reviewing and Editing

Declaration of competing interest

None.

Acknowledgement

We graciously thank S Zhao and the Wang Laboratory members for providing illustration and helpful discussion. We also apologize to authors whose works could not be covered due to the space limit.

Funding

This work was supported by NIH grants (R01-CA215284 and R01-CA218600 to G.G.W.). G.G.W. is an American Cancer Society (ACS) Research Scholar and a Leukemia and Lymphoma Society (LLS) Scholar.

References

- [1] R.W. Roberts, D.M. Crothers, Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition, *Science* 258 (1992) 1463–1466.
- [2] T. Garcia-Muse, A. Aguilera, R. Loops, From physiological to pathological roles, *Cell* 179 (2019) 604–618.
- [3] C. Niehrs, B. Luke, Regulatory R-loops as facilitators of gene expression and genome stability, *Nat. Rev. Mol. Cell. Biol.* 21 (2020) 167–178.
- [4] J. Sollier, K.A. Cimprich, Breaking bad: R-loops and genome integrity, *Trends Cell Biol.* 25 (2015) 514–522.
- [5] F. Chedin, Nascent connections: R-Loops and chromatin patterning, *Trends Genet.* 32 (2016) 828–838.
- [6] L. Wahba, S.K. Gore, D. Koshland, The homologous recombination machinery modulates the formation of RNA-DNA hybrids and associated chromosome instability, *Elife* 2 (2013), e00505.
- [7] F. Ariel, L. Lucero, A. Christ, M.F. Mammarella, T. Jegu, A. Veluchamy, K. Mariappan, D. Latrasse, T. Blein, C. Liu, M. Benhamed, M. Crespi, R-loop mediated transaction of the APOLO long noncoding RNA, *Mol. Cell* 77 (2020) 1055–1065 (e1054).
- [8] A. Abakir, T.C. Giles, A. Cristini, J.M. Foster, N. Dai, M. Starczak, A. Rubio-Roldan, M. Li, M. Eleftheriou, J. Crutchley, L. Flatt, L. Young, D.J. Gaffney, C. Denning, B. Dalhus, R.D. Ems, D. Gackowski, I.R. Correa Jr., J.L. Garcia-Perez, A. Klungland, N. Gromak, A. Ruzov, N(6)-methyladenosine regulates the stability of RNA:DNA hybrids in human cells, *Nat. Genet.* 52 (2020) 48–55.
- [9] K. Skourti-Stathaki, N.J. Proudfoot, A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression, *Genes Dev.* 28 (2014) 1384–1396.
- [10] A. Aguilera, T. Garcia-Muse, R loops: from transcription byproducts to threats to genome stability, *Mol. Cell* 46 (2012) 115–124.
- [11] M. Kasahara, J.A. Clikeman, D.B. Bates, T. Kogoma, RecA protein-dependent R-loop formation in vitro, *Genes Dev.* 14 (2000) 360–365.
- [12] E.N. Zaitsev, S.C. Kowalczykowski, A novel pairing process promoted by *Escherichia coli* RecA protein: inverse DNA and RNA strand exchange, *Genes Dev.* 14 (2000) 740–749.
- [13] S. Broccoli, F. Rallu, P. Sanscartier, S.M. Cerritelli, R.J. Crouch, M. Drolet, Effects of RNA polymerase modifications on transcription-induced negative supercoiling and associated R-loop formation, *Mol. Microbiol.* 52 (2004) 1769–1779.
- [14] D. Roy, M.R. Lieber, 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 (2009) 3124–3133.
- [15] D. Roy, Z. Zhang, Z. Lu, C.L. Hsieh, M.R. Lieber, 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 (2010) 146–159.
- [16] R. Stolz, S. Sulthana, S.R. Hartono, M. Malig, C.J. Benham, F. Chedin, Interplay between DNA sequence and negative superhelicity drives R-loop structures, *Proc. Natl. Acad. Sci. U. S. A.* 116 (2019) 6260–6269.
- [17] P.A. Ginno, Y.W. Lim, P.L. Lott, I. Korf, F. Chedin, GC skew at the 5′ and 3′ ends of human genes links R-loop formation to epigenetic regulation and transcription termination, *Genome Res.* 23 (2013) 1590–1600.
- [18] L.A. Sanz, S.R. Hartono, Y.W. Lim, S. Steyaert, A. Rajpurkar, P.A. Ginno, X. Xu, F. Chedin, Prevalent, dynamic, and conserved R-loop structures associate with specific epigenomic signatures in mammals, *Mol. Cell* 63 (2016) 167–178.
- [19] R. Pavri, R loops in the regulation of antibody gene diversification, *Genes (Basel)* 8 (2017).
- [20] K. Yu, M.R. Lieber, Current insights into the mechanism of mammalian immunoglobulin class switch recombination, *Crit. Rev. Biochem. Mol. Biol.* 54 (2019) 333–351.
- [21] D. Roy, K. Yu, M.R. Lieber, Mechanism of R-loop formation at immunoglobulin class switch sequences, *Mol. Cell. Biol.* 28 (2008) 50–60.
- [22] E.M. Wiedemann, M. Peycheva, R. Pavri, DNA replication origins in immunoglobulin switch regions regulate class switch recombination in an R-loop-dependent manner, *Cell Rep.* 17 (2016) 2927–2942.
- [23] Z.Z. Zhang, N.R. Pannunzio, L. Han, C.L. Hsieh, K. Yu, M.R. Lieber, The strength of an Ig switch region is determined by its ability to drive R loop formation and its number of WGCW sites, *Cell Rep.* 8 (2014) 557–569.
- [24] X. Zhang, H.C. Chiang, Y. Wang, C. Zhang, S. Smith, X. Zhao, S.J. Nair, J. Michalek, I. Jatoi, M. Lautner, B. Oliver, H. Wang, A. Petit, T. Soler, J. Brunet, F. Mateo, M. Angel Pujana, E. Poggi, K. Chaldekis, C. Isaacs, B.N. Peshkin, O. Ochoa, F. Chedin, C. Theoharis, L.Z. Sun, T.J. Curiel, R. Elledge, V.X. Jin, Y. Hu, R. Li, Attenuation of RNA polymerase II pausing mitigates BRCA1-associated R-loop accumulation and tumorigenesis, *Nat. Commun.* 8 (2017) 15908.
- [25] J.J. Kim, S.Y. Lee, F. Gong, A.M. Battenhouse, D.R. Boutz, A. Bashyal, S.T. Refvik, C.M. Chiang, B. Xhemalce, T.T. Paull, J.S. Brodbelt, E.M. Marcotte, K.M. Miller, Systematic bromodomain protein screens identify homologous recombination and R-loop suppression pathways involved in genome integrity, *Genes Dev.* 33 (2019) 1751–1774.
- [26] F.C. Lam, Y.W. Kong, Q. Huang, T.L. Vu Han, A.D. Maffa, E.M. Kasper, M.B. Yaffe, BRD4 prevents the accumulation of R-loops and protects against transcription-replication collision events and DNA damage, *Nat. Commun.* 11 (2020) 4083.
- [27] E. Nudler, RNA polymerase backtracking in gene regulation and genome instability, *Cell* 149 (2012) 1438–1445.
- [28] D. Zatreanu, Z. Han, R. Mitter, E. Tumini, H. Williams, L. Gregersen, A.B. Dirac-Svejstrup, S. Roma, A. Stewart, A. Aguilera, J.Q. Svejstrup, Elongation factor TFIIS prevents transcription stress and R-loop accumulation to maintain genome stability, *Mol. Cell* 76 (2019) 57–69 (e59).
- [29] A. De Magis, S.G. Manzo, M. Russo, J. Marinello, R. Morigi, O. Sordet, G. Capranico, DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells, *Proc. Natl. Acad. Sci. U. S. A.* 116 (2019) 816–825.
- [30] C.Y. Lee, C. McNeerney, K. Ma, W. Zhao, A. Wang, S. Myong, R-loop induced G-quadruplex in non-template promotes transcription by successive R-loop formation, *Nat. Commun.* 11 (2020) 3392.
- [31] G. Miglietta, M. Russo, G. Capranico, G-quadruplex-R-loop interactions and the mechanism of anticancer G-quadruplex binders, *Nucleic Acids Res.* 48 (2020) 11942–11957.
- [32] K. Skourti-Stathaki, N.J. Proudfoot, N. Gromak, Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination, *Mol. Cell* 42 (2011) 794–805.
- [33] D.Y. Zhao, G. Gish, U. Braunschweig, Y. Li, Z. Ni, F.W. Schmitges, G. Zhong, K. Liu, W. Li, J. Moffat, M. Vedadi, J. Min, T.J. Pawson, B.J. Blencowe, J. F. Greenblatt, SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination, *Nature* 529 (2016) 48–53.
- [34] L. Chen, J.Y. Chen, X. Zhang, Y. Gu, R. Xiao, C. Shao, P. Tang, H. Qian, D. Luo, H. Li, Y. Zhou, D.E. Zhang, X.D. Fu, R-ChIP using inactive RNase H reveals dynamic coupling of R-loops with transcriptional pausing at gene promoters, *Mol. Cell* 68 (2017) 745–757 (e745).
- [35] S.M. Tan-Wong, S. Dhir, N.J. Proudfoot, R-loops promote antisense transcription across the mammalian genome, *Mol. Cell* 76 (2019) 600–616 (e606).

- [36] R. Boque-Sastre, M. Soler, C. Oliveira-Mateos, A. Portela, C. Moutinho, S. Sayols, A. Villanueva, M. Esteller, S. Guil, Head-to-head antisense transcription and R-loop formation promotes transcriptional activation, *Proc. Natl. Acad. Sci. U. S. A.* 112 (2015) 5785–5790.
- [37] A. Postepska-Igielska, A. Giwojna, L. Gasri-Plotnitsky, N. Schmitt, A. Dold, D. Ginsberg, I. Grummt, LncRNA Khps1 regulates expression of the proto-oncogene SPHK1 via triplex-mediated changes in chromatin structure, *Mol. Cell* 60 (2015) 626–636.
- [38] Q. Sun, T. Csorba, K. Skourti-Stathaki, N.J. Proudfoot, C. Dean, R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus, *Science* 340 (2013) 619–621.
- [39] X. Wang, Z. Lu, A. Gomez, G.C. Hon, Y. Yue, D. Han, Y. Fu, M. Parisien, Q. Dai, G. Jia, B. Ren, T. Pan, C. He, N6-methyladenosine-dependent regulation of messenger RNA stability, *Nature* 505 (2014) 117–120.
- [40] I.A. Roundtree, M.E. Evans, T. Pan, C. He, Dynamic RNA modifications in gene expression regulation, *Cell* 169 (2017) 1187–1200.
- [41] S. Zaccara, R.J. Ries, S.R. Jaffrey, Reading, writing and erasing mRNA methylation, *Nat. Rev. Mol. Cell Biol.* 20 (2019) 608–624.
- [42] K.I. Zhou, T. Pan, An additional class of m(6)A readers, *Nat. Cell Biol.* 20 (2018) 230–232.
- [43] X. Yang, Q.L. Liu, W. Xu, Y.C. Zhang, Y. Yang, L.F. Ju, J. Chen, Y.S. Chen, K. Li, J. Ren, Q. Sun, Y.G. Yang, m(6)A promotes R-loop formation to facilitate transcription termination, *Cell Res.* 29 (2019) 1035–1038.
- [44] C. Zhang, L. Chen, D. Peng, A. Jiang, Y. He, Y. Zeng, C. Xie, H. Zhou, X. Luo, H. Liu, L. Chen, J. Ren, W. Wang, Y. Zhao, METTL3 and N6-methyladenosine promote homologous recombination-mediated repair of DSBs by modulating DNA-RNA hybrid accumulation, *Mol. Cell* 79 (2020) 425–442 (e427).
- [45] H.J. Kang, N.Y. Cheon, H. Park, G.W. Jeong, B.J. Ye, E.J. Yoo, J.H. Lee, J.H. Hur, E.A. Lee, H. Kim, K.Y. Lee, S.Y. Choi, W. Lee-Kwon, K. Myung, J.Y. Lee, H. M. Kwon, TonEBP recognizes R-loops and initiates m6A RNA methylation for R-loop resolution, *Nucleic Acids Res.* 49 (2021) 269–284.
- [46] L.H. Zhang, X.Y. Zhang, T. Hu, X.Y. Chen, J.J. Li, M. Rida, N. Sun, Y. Luo, X. Gao, The SUMOylated METTL8 induces R-loop and tumorigenesis via m3C, *iScience* 23 (2020), 100968.
- [47] C. Zhou, C. Wang, H. Liu, Q. Zhou, Q. Liu, Y. Guo, T. Peng, J. Song, J. Zhang, L. Chen, Y. Zhao, Z. Zeng, D.X. Zhou, Identification and analysis of adenine N(6)-methylation sites in the rice genome, *Nat. Plants* 4 (2018) 554–563.
- [48] Y. Fang, L. Chen, K. Lin, Y. Feng, P. Zhang, X. Pan, J. Sanders, Y. Wu, X.E. Wang, Z. Su, C. Chen, H. Wei, W. Zhang, Characterization of functional relationships of R-loops with gene transcription and epigenetic modifications in rice, *Genome Res.* 29 (2019) 1287–1297.
- [49] J.P. Ross, I. Suetake, S. Tajima, P.L. Molloy, Recombinant mammalian DNA methyltransferase activity on model transcriptional gene silencing short RNA-DNA heteroduplex substrates, *Biochem. J.* 432 (2010) 323–332.
- [50] P.A. Ginno, P.L. Lott, H.C. Christensen, I. Korf, F. Chedin, R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters, *Mol. Cell* 45 (2012) 814–825.
- [51] K. Arab, E. Karaulanov, M. Musheev, P. Trnka, A. Schafer, I. Grummt, C. Niehrs, GADD45A binds R-loops and recruits TET1 to CpG island promoters, *Nat. Genet.* 51 (2019) 217–223.
- [52] W. Feng, S. Chen, J. Wang, X. Wang, H. Chen, W. Ning, Y. Zhang, DHX33 recruits Gadd45a to cause DNA demethylation and regulates a subset of gene transcription, *Mol. Cell Biol.* 40 (2020).
- [53] C. Alecki, V. Chiwara, L.A. Sanz, D. Grau, O. Arias Perez, E.L. Boulter, K. J. Armache, F. Chedin, N.J. Francis, RNA-DNA strand exchange by the Drosophila Polycomb complex PRC2, *Nat. Commun.* 11 (2020) 1781.
- [54] K. Skourti-Stathaki, E. Torlai Triglia, M. Warburton, P. Voigt, A. Bird, A. Pombo, R-loops enhance polycomb repression at a subset of developmental regulator genes, *Mol. Cell* 73 (2019) 930–945 (e934).
- [55] W. Yuan, J. Zhou, J. Tong, W. Zhuo, L. Wang, Y. Li, Q. Sun, W. Qian, ALBA protein complex reads genetic R-loops to maintain genome stability in Arabidopsis, *Sci. Adv.* 5 (2019) eaav9040.
- [56] M.P. Crossley, M.J. Bocek, S. Hamperl, T. Swigut, K.A. Cimprich, qDRIP: a method to quantitatively assess RNA-DNA hybrid formation genome-wide, *Nucleic Acids Res.* 48 (2020), e84.
- [57] S.K. Petersen-Mahrt, R.S. Harris, M.S. Neuberger, AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification, *Nature* 418 (2002) 99–103.
- [58] W. Gan, Z. Guan, J. Liu, T. Gui, K. Shen, J.L. Manley, X. Li, R-loop-mediated genomic instability is caused by impairment of replication fork progression, *Genes Dev.* 25 (2011) 2041–2056.
- [59] A. Cristini, G. Ricci, S. Britton, S. Salimbeni, S.N. Huang, J. Marinello, P. Calsou, Y. Pommier, G. Favre, G. Capranico, N. Gromak, O. Sordet, Dual processing of R-loops and topoisomerase I induces transcription-dependent DNA double-strand breaks, *Cell Rep.* 28 (2019) 3167–3181 (e3166).
- [60] J. Sollier, C.T. Stork, M.L. Garcia-Rubio, R.D. Paulsen, A. Aguilera, K.A. Cimprich, Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability, *Mol. Cell* 56 (2014) 777–785.
- [61] D.A. Matos, J.M. Zhang, J. Ouyang, H.D. Nguyen, M.M. Genois, L. Zou, ATR protects the genome against R loops through a MUS81-triggered feedback loop, *Mol. Cell* 77 (2020) 514–527 (e514).
- [62] J.M. Santos-Pereira, A.B. Herrero, M.L. Garcia-Rubio, A. Marin, S. Moreno, A. Aguilera, The Npl3 hnRNP prevents R-loop-mediated transcription-replication conflicts and genome instability, *Genes Dev.* 27 (2013) 2445–2458.
- [63] G. D'Alessandro, D.R. Whelan, S.M. Howard, V. Vitelli, X. Renaudin, M. Adamowicz, F. Iannelli, C.W. Jones-Weinert, M. Lee, V. Matti, W.T.C. Lee, M. J. Morten, A.R. Venkitaraman, P. Pejka, E. Rothenberg, F. d'Adda di Fagnaga, BRCA2 controls DNA:RNA hybrid level at DSBs by mediating RNase H2 recruitment, *Nat. Commun.* 9 (2018) 5376.
- [64] X. Li, J.L. Manley, Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability, *Cell* 122 (2005) 365–378.
- [65] B. Gomez-Gonzalez, M. Garcia-Rubio, R. Bermejo, H. Gaillard, K. Shirahige, A. Marin, M. Foiani, A. Aguilera, Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles, *EMBO J.* 30 (2011) 3106–3119.
- [66] P. Grzechnik, M.R. Gdula, N.J. Proudfoot, Pcf11 orchestrates transcription termination pathways in yeast, *Genes Dev.* 29 (2015) 849–861.
- [67] S. Gavalda, M. Gallardo, R. Luna, A. Aguilera, R-loop mediated transcription-associated recombination in trf4Delta mutants reveals new links between RNA surveillance and genome integrity, *PLoS One* 8 (2013), e65541.
- [68] E. Pefanis, J. Wang, G. Rothschild, J. Lim, D. Kazadi, J. Sun, A. Federation, J. Chao, O. Elliott, Z.P. Liu, A.N. Economides, J.E. Bradner, R. Rabadan, U. Basu, RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity, *Cell* 161 (2015) 774–789.
- [69] S. Cohen, N. Puget, Y.L. Lin, T. Clouaire, M. Aguirrebengoa, V. Rocher, P. Pasero, Y. Canitrot, G. Legube, Senataxin resolves RNA:DNA hybrids forming at DNA double-strand breaks to prevent translocations, *Nat. Commun.* 9 (2018) 533.
- [70] J.C. Morales, P. Richard, P.L. Patidar, E.A. Motea, T.T. Dang, J.L. Manley, D. A. Boothman, XRN2 links transcription termination to DNA damage and replication stress, *PLoS Genet.* 12 (2016), e1006107.
- [71] I. De, S. Bessonov, R. Hofe, K. dos Santos, C.L. Will, H. Urlaub, R. Luhrmann, V. Pena, The RNA helicase Aquarius exhibits structural adaptations mediating its recruitment to spliceosomes, *Nat. Struct. Mol. Biol.* 22 (2015) 138–144.
- [72] R.D. Paulsen, D.V. Soni, R. Wollman, A.T. Hahn, M.C. Yee, A. Guan, J.A. Hesley, S.C. Miller, E.F. Cromwell, D.E. Solow-Cordero, T. Meyer, K.A. Cimprich, A genome-wide siRNA screen reveals diverse cellular processes and pathways that mediate genome stability, *Mol. Cell* 35 (2009) 228–239.
- [73] Y. Teng, T. Yadav, M. Duan, J. Tan, Y. Xiang, B. Gao, J. Xu, Z. Liang, Y. Liu, S. Nakajima, Y. Shi, A.S. Levine, L. Zou, L. Lan, ROS-induced R loops trigger a transcription-coupled but BRCA1/2-independent homologous recombination pathway through CSB, *Nat. Commun.* 9 (2018) 4115.
- [74] J.K. Leela, A.H. Syeda, K. Anupama, J. Gowrishankar, Rho-dependent transcription termination is essential to prevent excessive genome-wide R-loops in *Escherichia coli*, *Proc. Natl. Acad. Sci. U. S. A.* 110 (2013) 258–263.
- [75] S.Y. Mersaoui, Z. Yu, Y. Coulombe, M. Karam, F.F. Busatto, J.Y. Masson, S. Richard, Arginine methylation of the DDX5 helicase RGG/RG motif by PRMT5 regulates resolution of RNA:DNA hybrids, *EMBO J.* 38 (2019), e100986.
- [76] C. Perez-Calero, A. Bayona-Feliu, X. Xue, S.I. Barroso, S. Munoz, V.M. Gonzalez-Basallote, P. Sung, A. Aguilera, UAP56/DDX39B is a major cotranscriptional RNA-DNA helicase that unwinds harmful R loops genome-wide, *Genes Dev.* 34 (2020) 898–912.
- [77] S.C. Sridhara, S. Carvalho, A.R. Grosso, L.M. Gallego-Paez, M. Carmo-Fonseca, S. F. de Almeida, Transcription dynamics prevent RNA-mediated genomic instability through SRPK2-dependent DDX23 phosphorylation, *Cell Rep.* 18 (2017) 334–343.
- [78] A. Cristini, M. Groh, M.S. Kristiansen, N. Gromak, RNA/DNA hybrid Interactome identifies DHX9 as a molecular player in transcriptional termination and R-loop-associated DNA damage, *Cell Rep.* 23 (2018) 1891–1905.
- [79] P. Chakraborty, J.T.J. Huang, K. Hiom, DHX9 helicase promotes R-loop formation in cells with impaired RNA splicing, *Nat. Commun.* 9 (2018) 4346.
- [80] E. Hatchi, K. Skourti-Stathaki, S. Ventz, L. Pinello, A. Yen, K. Kamieniarz-Gdula, S. Dimitrov, S. Pathania, K.M. McKinney, M.L. Eaton, M. Kellis, S.J. Hill, G. Parmigiani, N.J. Proudfoot, D.M. Livingston, BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair, *Mol. Cell* 57 (2015) 636–647.
- [81] D.T. Nguyen, H.P.J. Voon, B. Xella, C. Scott, D. Clynes, C. Babbs, H. Ayyub, J. Kerry, J.A. Sharpe, J.A. Sloane-Stanley, S. Butler, C.A. Fisher, N.E. Gray, T. Jenuwein, D.R. Higgs, R.J. Gibbons, The chromatin remodelling factor ATRX suppresses R-loops in transcribed telomeric repeats, *EMBO Rep.* 18 (2017) 914–928.
- [82] A. Aguilera, B. Gomez-Gonzalez, DNA-RNA hybrids: the risks of DNA breakage during transcription, *Nat. Struct. Mol. Biol.* 24 (2017) 439–443.
- [83] C. Ohle, R. Tesorero, G. Schermann, N. Dobrev, I. Sinning, T. Fischer, Transient RNA-DNA hybrids are required for efficient double-strand break repair, *Cell* 167 (2016) 1001–1013 (e1007).
- [84] M.P. Crossley, M. Bocek, K.A. Cimprich, R-Loops as cellular regulators and genomic threats, *Mol. Cell* 73 (2019) 398–411.
- [85] L. Costantino, D. Koshland, Genome-wide map of R-loop-induced damage reveals how a subset of R-loops contributes to genomic instability, *Mol. Cell* 71 (2018) 487–497 (e483).
- [86] W.T. Lu, B.R. Hawley, G.L. Skalka, R.A. Baldock, E.M. Smith, A.S. Bader, M. Malewicz, F.Z. Watts, A. Wilczynska, M. Bushell, Drosha drives the formation of DNA:RNA hybrids around DNA break sites to facilitate DNA repair, *Nat. Commun.* 9 (2018) 532.
- [87] H. Hayatsu, Y. Wataya, K. Kazushige, The addition of sodium bisulfite to uracil and to cytosine, *J. Am. Chem. Soc.* 92 (1970) 724–726.
- [88] S.J. Boguslawski, D.E. Smith, M.A. Michalak, K.E. Mickelson, C.O. Yehle, W. L. Patterson, R.J. Carrico, Characterization of monoclonal antibody to DNA:RNA and its application to immunodetection of hybrids, *J. Immunol. Methods* 89 (1986) 123–130.

- [89] L.A. Sanz, F. Chedin, High-resolution, strand-specific R-loop mapping via S9.6-based DNA-RNA immunoprecipitation and high-throughput sequencing, *Nat. Protoc.* 14 (2019) 1734–1755.
- [90] J. Nadel, R. Athanasiadou, C. Lemetre, N.A. Wijetunga, O.B. P, H. Sato, Z. Zhang, J. Jeddell, C. Montagna, A. Golden, C. Seoighe, J.M. Greally, RNA:DNA hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships, *Epigenetics Chromatin* 8 (2015) 46.
- [91] L. Wahba, L. Costantino, F.J. Tan, A. Zimmer, D. Koshland, S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation, *Genes Dev.* 30 (2016) 1327–1338.
- [92] W. Xu, H. Xu, K. Li, Y. Fan, Y. Liu, X. Yang, Q. Sun, The R-loop is a common chromatin feature of the Arabidopsis genome, *Nat. Plants* 3 (2017) 704–714.
- [93] J.G. Dumelie, S.R. Jaffrey, Defining the location of promoter-associated R-loops at near-nucleotide resolution using bisDRIP-seq, *Elife* 6 (2017).
- [94] Y.A. Chan, M.J. Aristizabal, P.Y. Lu, Z. Luo, A. Hamza, M.S. Kobor, P.C. Stirling, P. Hieter, Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-chip, *PLoS Genet.* 10 (2014), e1004288.
- [95] A. El Hage, S. Webb, A. Kerr, D. Tollervy, Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria, *PLoS Genet.* 10 (2014), e1004716.
- [96] J.Y. Chen, X. Zhang, X.D. Fu, L. Chen, R-ChIP for genome-wide mapping of R-loops by using catalytically inactive RNASEH1, *Nat. Protoc.* 14 (2019) 1661–1685.
- [97] Q. Yan, K. Sarma, MapR: a method for identifying native R-loops genome wide, *Curr. Protoc. Mol. Biol.* 130 (2020), e113.
- [98] Q. Yan, E.J. Shields, R. Bonasio, K. Sarma, Mapping native R-loops genome-wide using a targeted nuclease approach, *Cell Rep.* 29 (2019) 1369–1380 (e1365).
- [99] M.P. Crossley, J.R. Brickner, C. Song, S.M.T. Zar, S.S. Maw, F. Chedin, M.S. Tsai, K.A. Cimprich, Catalytically inactive, purified RNase H1: a specific and sensitive probe for RNA-DNA hybrid imaging, *J. Cell Biol.* 220 (2021).
- [100] F. König, T. Schubert, G. Langst, The monoclonal S9.6 antibody exhibits highly variable binding affinities towards different R-loop sequences, *PLoS One* 12 (2017), e0178875.
- [101] P. Richard, J.L. Manley, R loops and links to human disease, *J. Mol. Biol.* 429 (2017) 3168–3180.
- [102] M. Groh, N. Gromak, Out of balance: R-loops in human disease, *PLoS Genet.* 10 (2014), e1004630.
- [103] M.G.L. Perego, M. Taiana, N. Bresolin, G.P. Comi, S. Corti, R-loops in motor neuron diseases, *Mol. Neurobiol.* 56 (2019) 2579–2589.
- [104] H. Gaillard, T. Garcia-Muse, A. Aguilera, Replication stress and cancer, *Nat. Rev. Cancer* 15 (2015) 276–289.
- [105] V. Bhatia, S.I. Barroso, M.L. Garcia-Rubio, E. Tumini, E. Herrera-Moyano, A. Aguilera, BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2, *Nature* 511 (2014) 362–365.
- [106] C.T. Stork, M. Bocek, M.P. Crossley, J. Sollier, L.A. Sanz, F. Chedin, T. Swigut, K. A. Cimprich, Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage, *Elife* 5 (2016).
- [107] A.S. Sperling, C.J. Gibson, B.L. Ebert, The genetics of myelodysplastic syndrome: from clonal haematopoiesis to secondary leukaemia, *Nat. Rev. Cancer* 17 (2017) 5–19.
- [108] K. Yoshida, M. Sanada, Y. Shiraishi, D. Nowak, Y. Nagata, R. Yamamoto, Y. Sato, A. Sato-Otsubo, A. Kon, M. Nagasaki, G. Chalkidis, Y. Suzuki, M. Shiosaka, R. Kawahata, T. Yamaguchi, M. Otsu, N. Obara, M. Sakata-Yanagimoto, K. Ishiyama, H. Mori, F. Nolte, W.K. Hofmann, S. Miyawaki, S. Sugano, C. Haferlach, H.P. Koeffler, L.Y. Shih, T. Haferlach, S. Chiba, H. Nakauchi, S. Miyano, S. Ogawa, Frequent pathway mutations of splicing machinery in myelodysplasia, *Nature* 478 (2011) 64–69.
- [109] L. Malcovati, K. Stevenson, E. Papaemmanuil, D. Neuberg, R. Bejar, J. Boulwood, D.T. Bowen, P.J. Campbell, B.L. Ebert, P. Fenaux, T. Haferlach, M. Heuser, J. H. Jansen, R.S. Komrokji, J.P. Maciejewski, M.J. Walter, M. Fontenay, G. Garcia-Manero, T.A. Graubert, A. Karsan, M. Meggendorfer, A. Pellagatti, D.A. Sallman, M.R. Savona, M.A. Sekeres, D.P. Steensma, S. Tauro, F. Thol, P. Vyas, A.A. Van de Loosdrecht, D. Haase, H. Tuchler, P.L. Greenberg, S. Ogawa, E. Hellstrom-Lindberg, M. Cazzola, SF3B1-mutant MDS as a distinct disease subtype: a proposal from the International Working Group for the Prognosis of MDS, *Blood* 136 (2020) 157–170.
- [110] S. Singh, D. Ahmed, H. Dolatshad, D. Tatwavedi, U. Schulze, A. Sanchi, S. Ryley, A. Dhir, L. Carpenter, S.M. Watt, D.J. Roberts, A.M. Abdel-Aal, S.K. Sayed, S. A. Mohamed, A. Schuh, P. Vyas, S. Killick, A.G. Kotini, E.P. Papapetrou, D. H. Wiseman, A. Pellagatti, J. Boulwood, SF3B1 mutations induce R-loop accumulation and DNA damage in MDS and leukemia cells with therapeutic implications, *Leukemia* 34 (2020) 2525–2530.
- [111] L. Chen, J.Y. Chen, Y.J. Huang, Y. Gu, J. Qiu, H. Qian, C. Shao, X. Zhang, J. Hu, H. Li, S. He, Y. Zhou, O. Abdel-Wahab, D.E. Zhang, X.D. Fu, The augmented R-loop is a unifying mechanism for myelodysplastic syndromes induced by high-risk splicing factor mutations, *Mol. Cell* 69 (2018) 412–425 (e416).
- [112] H.D. Nguyen, W.Y. Leong, W. Li, P.N.G. Reddy, J.D. Sullivan, M.J. Walter, L. Zou, T.A. Graubert, Spliceosome mutations induce R loop-associated sensitivity to ATR inhibition in myelodysplastic syndromes, *Cancer Res.* 78 (2018) 5363–5374.
- [113] H. Dvinge, E. Kim, O. Abdel-Wahab, R.K. Bradley, RNA splicing factors as oncoproteins and tumour suppressors, *Nat. Rev. Cancer* 16 (2016) 413–430.
- [114] A. Korshunov, D. Sturm, M. Ryzhova, V. Hovestadt, M. Gessi, D.T. Jones, M. Remke, P. Northcott, A. Perry, D. Picard, M. Rosenblum, M. Antonelli, E. Aronica, U. Schuller, M. Hasselblatt, A. Woehrer, O. Zheludkova, E. Kumirova, S. Puget, M.D. Taylor, F. Giangaspero, V. Peter Collins, A. von Deimling, P. Lichter, A. Huang, T. Pietsch, S.M. Pfister, M. Kool, Embryonal tumor with abundant neuropil and true rosettes (ETANTR), ependymoblastoma, and medulloepithelioma share molecular similarity and comprise a single clinicopathological entity, *Acta Neuropathol.* 128 (2014) 279–289.
- [115] S. Pfister, M. Remke, M. Castoldi, A.H. Bai, M.U. Muckenthaler, A. Kulozik, A. von Deimling, A. Pscherer, P. Lichter, A. Korshunov, Novel genomic amplification targeting the microRNA cluster at 19q13.42 in a pediatric embryonal tumor with abundant neuropil and true rosettes, *Acta Neuropathol.* 117 (2009) 457–464.
- [116] C.L. Kleinman, N. Gerges, S. Papillon-Cavanagh, P. Sin-Chan, A. Pramatarova, D. A. Quang, V. Adoue, S. Busche, M. Caron, H. Djambazian, A. Bemmo, A. M. Fontebasso, T. Spence, J. Schwartzentruber, S. Albrecht, P. Hauser, M. Garami, A. Klekner, L. Bognar, J.L. Montes, A. Staffa, A. Montpetit, P. Berube, M. Zakrzewska, K. Zakrzewski, P.P. Liberski, Z. Dong, P.M. Siegel, T. Duchaine, C. Perotti, A. Fleming, D. Faury, M. Remke, M. Gallo, P. Dirks, M.D. Taylor, R. Sladek, T. Pastinen, J.A. Chan, A. Huang, J. Majewski, N. Jabado, Fusion of TTYH1 with the C19MC microRNA cluster drives expression of a brain-specific DNMT3B isoform in the embryonal brain tumor ETMR, *Nat. Genet.* 46 (2014) 39–44.
- [117] S. Lambo, S.N. Grobner, T. Rausch, S.M. Waszak, C. Schmidt, A. Gorthi, J. C. Romero, M. Mauerer, S. Brabetz, S. Krausert, I. Buchhalter, J. Koster, D. A. Zwijnenburg, M. Sill, J.M. Hubner, N. Mack, B. Schwalm, M. Ryzhova, V. Hovestadt, S. Papillon-Cavanagh, J.A. Chan, P. Landgraf, B. Ho, T. Milde, O. Witt, J. Ecker, F. Sahm, D. Sumerauer, D.W. Ellison, B.A. Orr, A. Darabi, C. Haberler, D. Figarella-Branger, P. Wesseling, J. Schittenhelm, M. Remke, M. D. Taylor, M.J. Gil-da-Costa, M. Lastowska, W. Grajkowska, M. Hasselblatt, P. Hauser, T. Pietsch, E. Uro-Coste, F. Bourdeaut, J. Masliah-Planchon, V. Rigau, S. Alexandrescu, S. Wolf, X.N. Li, U. Schuller, M. Snuderl, M.A. Karajannis, F. Giangaspero, N. Jabado, A. von Deimling, D.T.W. Jones, J.O. Korbel, K. von Hoff, P. Lichter, A. Huang, A.J.R. Bishop, S.M. Pfister, A. Korshunov, M. Kool, The molecular landscape of ETMR at diagnosis and relapse, *Nature* 576 (2019) 274–280.
- [118] A. Gorthi, J.C. Romero, E. Loranc, L. Cao, L.A. Lawrence, E. Goodale, A.B. Iniguez, X. Bernard, V.P. Masamsetti, S. Roston, E.R. Lawlor, J.A. Toretsky, K. Stegmaier, S.L. Lessnick, Y. Chen, A.J.R. Bishop, EWS-FLI1 increases transcription to cause R-loops and block BRCA1 repair in Ewing sarcoma, *Nature* 555 (2018) 387–391.
- [119] L. Perez-Martinez, M. Ozturk, F. Butter, B. Luke, Npl3 stabilizes R-loops at telomeres to prevent accelerated replicative senescence, *EMBO Rep.* 21 (2020), e49087.
- [120] Y. Yang, K.M. McBride, S. Hensley, Y. Lu, F. Chedin, M.T. Bedford, Arginine methylation facilitates the recruitment of TOP3B to chromatin to prevent R loop accumulation, *Mol. Cell* 53 (2014) 484–497.