Acta Biochim Biophys Sin, 2016, 48(7), 623–631
doi: 10.1093/abbs/gmw052
Advance Access Publication Date: 1 June 2016
Review

Review

R-loop: an emerging regulator of chromatin dynamics

Qais Al-Hadid and Yanzhong Yang*

Department of Cancer Genetics and Epigenetics, Beckman Research Institute, City of Hope Cancer Center, Duarte, CA 91010, USA

*Correspondence address. Tel: +1-626-218-9859; Fax: +1-626-301-8892; E-mail: yyang@coh.org

Received 6 March 2016; Accepted 29 April 2016

Abstract

The dynamic structure of chromatin, which exists in two conformational states: heterochromatin and euchromatin, alters the accessibility of the DNA to regulatory factors during transcription, replication, recombination, and DNA damage repair. Chemical modifications of histones and DNA, as well as adenosine triphospahate-dependent nucleosome remodeling, have been the major focus of research on chromatin dynamics over the past two decades. However, recent studies using a DNA–RNA hybrid-specific antibody and next-generation sequencing approaches have revealed that the formation of R-loops, one of the most common non-canonical DNA structures, is an emerging regulator of chromatin states. This review focuses on recent insights into the interplay between R-loop formation and the epigenetic modifications of chromatin in normal and disease states.

Key words: DNA–RNA hybrid, R-loop, epigenetic modification, transcription

Introduction

DNA can adopt a number of alternative structures in addition to the right-handed double helical B-form, such as left-handed Z-DNA, the four-stranded G-quardruplex, D-loops, and R-loops. R-loops are three-stranded nucleic acid structures, consisting of a DNA–RNA hybrid and a displaced single-stranded DNA [1]. Davis and colleagues first reported R-loops 40 years ago [2]. They showed that RNA could hybridize with a complementary double-stranded DNA, resulting in a D-loop-like structure, which they called an R-loop. Twenty years after the discovery of R-loops, Crouch and colleagues showed that R-loops could form *in vivo* during transcription in bacterial cells, and were regarded as mere infrequent transcriptional by-products [3,4]. Since then and especially in the past decade, R-loop biology has become an increasingly expanded area of research, as these nucleic acid structures are conserved and found in a variety of organisms, ranging from bacteria to mammals, and are implicated in genomic instability and many genetic-based diseases, such as cancer and neurodegeneration diseases.

Precisely, how R-loops form *in vivo* is still unclear; however, three nonexclusive models have been presented. First, they are predominantly thought to form by the 'thread-back' model, wherein the newly synthesized RNA strand, exiting the RNA polymerase, re-anneals back with its homologous sequence in the duplex DNA (in *cis*), displacing the non-template strand. This is especially the case when RNA polymerase transcribes a C-rich DNA template, resulting in a G-rich mRNA that invades the homologous DNA sequence behind the elongating RNA polymerase, leaving the G-rich non-template strand single-stranded. Second, the normal 8-bp DNA–RNA hybrid in the transcription machinery can be extended, leaving the non-template strand single-stranded, leading to R-loop formation. However, this model is unlikely in light of structural studies showing that the nascent RNA transcript and template DNA exit through different channels during transcription [5]. Third, R-loops can also form in *trans*, post-transcriptionally, when an RNA strand anneals to a homologous DNA sequence at a different locus from where it was transcribed [6,7]. This model has gained more support from a recent R-loop genome-wide mapping study, in which R-loops were found in untranscribed regions of human genome [8]. Nevertheless, it has become increasingly apparent that R-loops occur much more frequently than previously imagined and genome-wide mapping

established R-loops to be the most abundant non-B DNA structures in mammalian genomes [8,9]. They are involved not only in transcriptional regulation and replication, but also in genomic instability, class-switch recombination in B cells, and DNA damage and repair. These areas of R-loop research have been extensively reviewed recently [1,10–13]. Our review focuses on the role of R-loops as an emerging regulator of chromatin dynamics, highlighting the epigenetic interplay of chromatin with this unique DNA structure. There have been considerable progresses in understanding these aspects of R-loop biology, although areas of controversy remain.

Factors Determining the Formation of R-Loops

DNA sequence composition

In silico analysis of the human genome has identified ~250,000 putative R-loop forming sequences [14]. These DNA sequences are rich in cytosine on the template strand and guanine on the non-template strand. *In vitro* transcription studies have shown that clusters of three or more consecutive guanines promote R-loop formation and high guanine density, without clustering, is important for maintaining R-loops [15]. In addition, during transcription, the guanine-rich non-template single-stranded DNA can form very stable secondary structures, known as G-quadruplexes [16], which may facilitate hybridization of the nascent mRNA with the template DNA via the thread-back model. The single-stranded portion of the R-loop is also susceptible to damage [17], which hinders its ability to re-anneal with the template strand, thereby promoting RNA binding to the template strand leading to R-loop formation.

RNA biogenesis

Currently, >50 yeast gene mutations have been found to lead to DNA–RNA hybrid accumulation [18,19]. These include mutants that cause defects in transcription elongation, termination, mRNA splicing, cleavage, polyadenylation, mRNA export, RNA degradation, and rDNA processing [7,18–28], suggesting that RNA biogenesis factors play a key role in mitigating or preventing R-loop formation. Three DNA topoisomerases (I, II, and IIIB), which unwind supercoiled DNA during transcription elongation, are also involved in preventing R-loop formation [25,27].

Ribonucleases and helicases

Once formed, R-loops are very stable and exhibit a structure that is intermediary between B-form DNA and A-form double-stranded RNA (dsRNA) that is thermodynamically more favorable than duplex DNA. It is critical to remove excessive R-loops in a time- and space-dependent manner to avoid its catastrophic outcomes [11,13]. This can be achieved by either ribonucleases or helicases. RNase H enzymes remove R-loops by degrading the RNA strand of the DNA–RNA hybrid [29]. Overexpression of RNase H has been shown to partially complement the growth defect of DNA topoisomerase I mutants in *Escherichia coli* by reducing R-loops formation [3,30]. This mechanism is evolutionary conserved and found in yeast and mammals [31].

Another mechanism cell utilizes to remove R-loops is through DNA–RNA helicases that unwind the hybrids. This group of enzymes belongs to either DNA or RNA helicases, but with a preference for DNA–RNA hybrids as substrates [32,33]. For example, the yeast Pif1p DNA helicase preferentially unwinds DNA–RNA substrates and is involved in mitochondrial DNA maintenance and telomeric DNA synthesis [34]. The human DEAH box protein DHX9 preferentially unwinds R-loops and G-quadruplexes for transcription activation and genome stability [35]. The human senataxin (SETX) is a DNA and RNA helicase that resolves R-loops formed at transcription termination sites [28,36]. Loss of SETX leads to aberrant R-loop accumulation and failure of meiotic recombination and infertility in mice [37].

DNA damage response pathways

In line with the increasing evidence showing that R-loops are threats to genome stability, proteins involved in DNA damage response (DDR) pathways can also process R-loops. For example, R-loops are accumulated at damaged transcribed sites in a transcription-dependent manner and are regulated by DDR pathway proteins such as SAF-A, FUS, and TAF15 [38]. It was also found that R-loops induced by loss of function mutations of SETX and AQR helicases or inhibition of topoisomerase 1 are processed into DNA double-strand breaks through transcription-coupled nucleotide excision repair pathway [17]. R-loops are believed to be a major source of spontaneous replication stress and that BRCA2 and Fanconi anaemia proteins contribute to the elimination of R-loops that block replication folk progression [39–41]. Furthermore, BRCA1 was found to interact with SETX at transcription termination regions to prevent R-loop formation, as deletion of either one of them increases R-loops [42]. The physical interaction between BRCA1 and SETX provides the first evidence for targeting the helicase activity for R-loop resolution.

Recombination-driven R-loop formation

The concept that R-loop formation invariably occurs at the site of transcription (in *cis*) was challenged by a series of recent findings showing that R-loops can also form by RNA invading the DNA duplex at a remote genomic location (in *trans*). The first evidence comes from the *E. coli* homologous recombination and DNA damage repair protein, RecA. It catalyzes the assimilation of complementary RNA into a homologous region of a DNA duplex to form R-loops [43,44]. Interestingly, mutation of eukaryotic Rad51 protein (the yeast homolog of RecA), which promotes strand exchange by forming nucleoprotein filaments during homolog recombination, was reported to result in reduced R-loop formation, suggesting that Rad51 likely promotes R-loop formation in *trans* [7]. However, further biochemistry assays are required to demonstrate the Rad51 activity in this regard. Recently, an *in vitro* study shows that transcripts harboring AGGAG repeats can form R-loops with CTCCT-repeat template in *trans* [6], suggesting that R-loops formed in *trans* may likely prefer some sequences that are prone to undergo strand separation under negative supercoiled condition, thus favoring spontaneous R-loop formation.

The Genomic Distribution of R-Loops

Detection of R-loops *in vivo* had been a challenge before the DNA–RNA hybrid-specific antibody (S9.6 monoclonal antibody) was developed [45]. This antibody is widely used to visualize R-loops by immunofluorescence and to isolate R-loops from cells by immunoprecipitation. In combination with microarray and next-generation sequencing approaches, several genome-wide mapping strategies were developed to detect R-loops *in vivo*. For example, Ginno *et al.* [46,47] developed the first DNA–RNA immunoprecipitation-sequencing (DRIP-Seq) technique to quantitatively recover R-loops

from complex nucleic acid mixtures of human genomic DNA. This approach identified thousands of R-loop peaks distributed along the human genome [46,47]. The majority of these R-loop forming regions showed enrichment at the promoters of highly transcribed genes (e.g. housekeeping genes). In addition, a significant amount of R-loops were also found at the 3′ end of genes, consistent with the role of R-loop formation at transcription termination sites [36,48,49]. Ginno *et al.* [46,47] also developed another R-loop enrichment method, called DNA–RNA *in vitro* enrichment (DRIVE) to complement with DRIP-Seq, which relies on the intrinsic specificity of the S9.6 monoclonal antibody. DRIVE makes use of a catalytically deficient but DNA–RNA hybrid binding competent human RNASEH1 protein for affinity pull down of R-loops from nucleic acid mixtures. Over 1000 R-loop forming regions were identified and they are significantly overlapping with core promoter regions.

Hage *et al.* [50] applied chromatin immunoprecipitation followed by sequencing (ChIP-Seq) to systematically identify R-loop forming sites throughout the nuclear and mitochondrial genomes in budding yeast. They detected R-loops in actively transcribed genes by all RNA polymerases (Pol I, II, III) and mitochondrial RNA polymerase (mtRNAP). These regions include rDNA genes transcribed by Pol I, tRNAs, and U6 snRNA transcribed by Pol III, transcription units of the mtDNA, and a subset of intron-containing genes transcribed by Pol II. Ty1 retrotransposons also form low levels of R-loops in wild-type strains, but notable accumulation was observed in RNase H mutants. Exon 2 of spliced protein-coding genes was found to be a favored site for R-loop formation [50]. It was speculated that R-loops over exon 2 decelerate Pol II elongation and create a chromatin environment favorable for Pol II pausing, thus promoting co-transcriptional mRNA splicing [51]. Although this hypothesis still needs to be tested, the observed correlation between transcription activity, R-loop formation, and GC content is consistent with the features of R-loop forming sequences in mammalian cells [46,47]. Similar but not identical observations have been made from a separate study mapping DNA–RNA hybrid sites in yeast using DRIP followed by hybridization on tiling microarrays (DRIP-Chip) [18]. In that study, Chan *et al.* [18] showed that R-loops not only preferentially accumulated at rDNA1, Ty2 retrotransposons, telomeric-repeat regions, and a subset of open reading frames (ORFs), but also significantly enriched at genes with antisense transcripts. The expression of these genes was shown to be sensitive to RNase H overexpression. Although the role of R-loop formation in antisense regulation is still unclear, available evidence shows that it may both up- and down-regulate the expression of certain antisense transcripts, and more interestingly, play a role in determining the direction of divergent gene transcription [52–54].

It is also notable that although strand asymmetry in the distribution of guanine (G) and cytosine (C) residues (as measured by GC skew) is a strong predictor of R-loop forming regions in human genome, in yeast, the levels of gene expression contribute more to R-loop forming potential than GC content. Additionally, DRIP-Seq and DRIVE-Seq did not enrich significant R-loops formed at rDNA, tRNAs, retrotransposons, and genes with antisense transcription. It is possible that these differential enrichments are due to different mechanisms that evolved from different species. However, different R-loop detection approaches used in these studies could also explain the variation. When mapping R-loops in mammals, genomic DNA was first extracted and then fragmented using a cocktail of restriction enzymes before R-loops were enriched by S9.6 monoclonal antibody; while in yeast studies, whole cells were cross-linked with formaldehyde and the chromatin was fragmented by sonication before S9.6 monoclonal antibody immunoprecipitation (similar to conventional ChIP). It is likely that R-loops formed at ORFs are more sensitive to formaldehyde cross-link and sonication, while repetitive regions are either digested too frequently or rarely digested by selected restriction enzymes.

Nadel *et al.* [8] recently modified the DRIP protocol and mapped R-loop forming regions in SV40 large T antigen transformed human embryonic kidney cells (HEK293T). These modifications include pretreatment of nucleic acid with RNase I, the use of sonication for genomic DNA fragmentation and addition of directional information for RNA strand in the hybrid. Although the new data set supports the correlation of R-loops with DNA hypomethylation at promoters to some extent, several distinct observations are unexpected. First, a large amount of R-loops were found in the intergenic regions, suggesting that R-loops formed in *trans* are more prevalent that previous thought [6,7]. Second, R-loops are rather relatively depleted than enriched at the end of RefSeq genes. Third, a significant amount of R-loops were detected at enhancers as evident by their association with H3K27ac and H3K4me1 histone modifications, indicating the involvement of R-loops in enhancer–promoter looping and high-order chromatin architecture. Although further validations of these bioinformatics analysis are required, this study sheds light on the importance of the formation and regulatory roles of R-loops in intergenic regions.

R-Loops as Emerging Regulator of Chromatin Dynamics

In this section, we discuss in detail about the interplay between R-loops formed at different genomic loci and changes in epigenetic modifications. A brief summary in this regard is listed in **Table 1**.

R-loops formed at the gene promoters

R-loops have been implicated in all stages of gene expression from transcription initiation to termination, as they modulate chromatin architecture and hence, the accessibility of the transcription machinery to the underlying DNA. R-loop forming sequences are enriched at the 5′ and 3′ end of genes, in the CpG promoter and termination regions, respectively. Recent studies showed that in the human genome, these promoter and termination regions are characterized by a significant asymmetry in the distribution of guanine and cytosine residues, called GC skew, that are highly prone to R-loop formation [46,69]. In CpG island-containing promoter regions, the formation of R-loops negatively correlates with DNA methylation, and is associated with activation of gene expression. However, the underlying mechanism by which this occurs is unknown. One possibility is that CpG islands associated with R-loops are poor substrates and sterically inaccessible for the *de novo* DNA methyltransferase, DNMT3B1, due to the presence of the transcription machinery, resulting in hypomethylated CpG promoters and transcriptional activation. In support of this, RNA polymerase presence at CpG island promoters has been shown to counteract DNA methylation, whereas, in the absence of RNA polymerase, CpG island methylation is enhanced [70,71]. A more intriguing possibility is that the R-loop structure acts as a binding site for factors that antagonize CpG promoter methylation to promote transcription activation or repels factors associated with silencing. In support of this model, histone marks associated with transcription activation including H3K4me3, H4K20me1, H3K36me3, and H3K79me2 were associated with

Table 1. R-loop associated epigenetic modifications

R-loop forming regions	Determining factors	Epigenetic modifications	References
Promoters	GC skew, transcription activity	Depletes DNA methylation, promotes active histone marks, e.g. H3K4me3, H3K79me2, H4K20me1, and H3K36me3, and prevents H3K27me3	[8,46,47,55]
	Trinucleotide repeats	Promotes repressive histone mark H3K9me2	[56,57]
Enhancers	Mediators and RNA exosome pathway	Long-range gene looping, higher-order chromatin structure, loss of repressive mark H3K9me2	[19,58,59]
Transcription termination sites	GC skew, antisense transcription	Promotes repressive histone marks, e.g. H3K9me2 for Pol II release	[28,36,47,49]
Centromeres	ncRNA of centromeric repeats, defects in mRNP biogenesis	Promotes repressive histone marks H3K9me and H3S10p	[60,61]
Telomeres	Telomeric repeat-containing lncRNA (TERRA)	Promotes repressive histone marks H3K9me2 and H3K9me3	[62–65]
rDNA repeats	Topoisomerase activity, pRNA and asRNA of rDNA	Promotes DNA methylation and repressive histone mark H4K20me3	[66–68]

highly GC skewed CpG island promoters that are prone to R-loop formation, whereas a specific type of CpG island characterized by 'reverse GC skew' was enriched for the repressive histone mark H3K27me3 [47]. The factors responsible for these transcription-activating modifications, SET-domain lysine methyltransferases, were shown to contain motifs that bind single-stranded DNA. These motifs potentially may recognize the displaced single-stranded DNA of the R-loop and target the methyltransferases to genomic loci to lay the active chromatin marks [72]. Regulation of the antagonistic relationship between R-loops and CpG promoter methylation is important because its deregulation can lead to the autoimmune disease, Aicardi–Goutieres syndrome (AGS), which is characterized by elevated levels of R-loops [55]. A recent study showed that patients with AGS had decreased DNA methylation levels, particularly at loci where R-loops accumulated, suggesting a potential role for R-loop in inhibiting DNA methylation. It is likely that this R-loop-driven inhibition of DNA methylation results in aberrant gene expression profiles in cells, which triggers an immune response leading to autoimmunity [55].

Although the consequence could be either gene activation or repression, transcription regulation by R-loops formed at promoter-proximal region seems to be more and more a general mechanism. A recent study, using embryonic stem cells (ESCs), showed that R-loops at the promoters of specific sets of transcribed genes recruit the transcriptional activating acetyltransferase, Tip60-p400, and simultaneously inhibit the binding of the polycomb repressive complex 2 (PRC2), which deposits the transcriptional repressive mark, H3K27me3, a process that is important for ESC differentiation [73]. However, it is unclear which portion of the R-loop, DNA–RNA hybrid or the single-stranded DNA, is important for this discriminatory binding mechanism. Fragile X syndrome, the most common genetic form of mental retardation, is characterized by a progressive CGG trinucleotide-repeat expansion adjacent to the fragile X mental retardation 1 (*FMR1*) gene promoter, which leads to epigenetic silencing of *FMR1* gene [74]. It was recently shown that during differentiation of ESCs into neurons, *FMR1* gene expression was silenced by R-loops formed between *FMR1* mRNA and *FMR1* DNA at CGG-repeat-containing promoter region, which was accompanied by a loss of the chromatin active mark, H3K4me2 and gain of the repressive mark, H3K9me2 [56,75]. Thus, the formation of R-loops at promoters can either promote the recruitment of transcription activator complex (Tip60-p400) for transcription activation or initiate the deposition of repressive histone mark (H3K9me2) for gene silencing. It is likely that this complex regulation mediated by R-loops is gene specific and the key factors that determine the specific outcome have yet to be discovered. Interestingly, although GC content does not contribute to R-loop formation potential in yeast, the trinucleotide-repeat induced R-loops seem to be specific to the length of the CGGs—gene silencing occurs at loci containing >200 repeats [56,74].

R-loops formed during transcription elongation

During transcription elongation, the negatively supercoiled DNA behind the elongating RNA polymerase is loose and susceptible to invasion by the mRNA transcript to form R-loops. Co-transcriptional R-loops impede the movement of RNA polymerase both *in vitro* and in the cells. For example, Aguilera and colleagues provide the first experimental evidence that stably formed DNA–RNA hybrids in a negatively supercoiled transcription template reduces the efficiency of transcription elongation *in vitro* [76]. Furthermore, transcription studies *in vitro* using an artificial DNA mimic, with greater nucleic acid affinity than natural nucleic acids, triggers R-loop formation and blocks the progression of RNA polymerase II [77]. Additionally, the GAA trinucleotide-repeat expansion in the first intron of the frataxin (*FXN*) gene, which leads to reduced *FXN* gene expression, is considered to be the molecular basis of Friedreich ataxia (FRDA), one of the most frequent autosomal recessive ataxia [78]. Studies from Gromak's group found that R-loops form in FRDA patient cells on expanded GAA repeats of endogenous *FXN* gene. These R-loops are stable and co-localize with repressive histone marks H3K9me2, which impedes RNA Pol II elongation [57]. To inhibit R-loop formation during transcription, cells employ three different enzymes: topoisomerase type 1A, RNase H, and DNA–RNA helicase. How these surveillance mechanisms discriminate and target these co-transcriptional R-loops, while avoiding the beneficial R-loops is still unclear. A recent study showed that methylation of histones at H4R3 and H3R17 at actively transcribing regions, by the arginine methyltransferases PRMT1 and CARM1, respectively, recruits the methylarginine reader protein TDRD3 in complex with topoisomerase IIIβ (TOP3B). TOP3B recruitment at the actively transcribing regions relaxes underwound DNA behind the elongating RNA polymerase and inhibits R-loop formation, resulting in transcription activation [27,79]. It is likely that transcribing genes are pre-loaded with the TDRD3–TOP3B complex prior to R-loop formation. In this case,

methylation of H4R3 and H3R17 precedes R-loop formation. This model is supported by further identification of arginine methylation of RNA Pol II C-terminal Domain (CTD) by CARM1 at R1810 [80]. This modification promotes the interaction of CTD with TDRD3–TOP3B complex, potentially resolves underwound DNA during transcription elongation and prevents R-loop formation.

On the contrary, R-loops formed in the gene body have beneficial effects in yeast cells, where its enrichment is found at the second exon of ORFs to facilitate efficient mRNA splicing [50]. This leads to an intriguing hypothesis that R-loop might be involved in the regulation of alternative splicing. R-loop is known to form at antisense transcribed genes [18] and antisense transcripts have been shown to be a conserved mechanism contributing to splicing regulation across multiple metazoan species [81,82].

R-loops formed at the transcription termination sites

Regulation of transcription termination was best illustrated at the human β-actin (*ACTB*) gene 3′ end [36]. It is characterized by the formation of R-loops over G-rich pause sites downstream of polyA signals. Consistent with this observation, genome-wide mapping of R-loop forming regions in mammalian cells also shows positive GC skew and enrichment of R-loops at 3′ end of genes [47]. Both in yeast and in mammalian cells, R-loops formed at 3′ end of genes have been reported to be critical for transcription termination and RNA Pol II release, a process involving the coordination of helicases and nucleases, including Senataxin and XRN2 [83]. The working model includes R-loops accumulation after RNAP II transcribes through G-rich pause site and resolve of R-loops by Senataxin to allow XRN2-mediated Pol II release [36,48]. Overexpression of RNASEH1 and depletion of Senataxin from cells cause transcription read-through, indicating that efficient R-loop-dependent termination mechanism prevents transcription read-through. In line with this hypothesis, genes with positive GC skew at 3′ end are localized in regions of high gene density, where transcription read-through tends to have more detrimental effects to the cell [47]. In this model, Senataxin is the major helicase to resolve R-loops, but how it is recruited to termination site is not clear. A recent study on RNAP II CTD arginine methylation sheds light on understanding the targeting mechanism for Senataxin. Protein arginine methyltransferase 5 (PRMT5) catalyzes the symmetrical dimethylation (SDMA) of CTD at arginine 1810 (R1810), which promotes the recruitment of the survival of motor neuron protein, SMN. SMN interacts with Senataxin and targets it to regions where R1810me2s-modified RNAP II is enriched. Knockdown of PRMT5 or SMN, as well as a R1810 methylation-deficient mutation in RNAP II CTD, all cause R-loop accumulation at *ACTB* termination sites, supporting the role of a PRMT5-mediated RNAP II methylation—SMN—Senataxin cascade in R-loop resolution at 3′ end of gene [28]. This provides an additional targeting mechanism for Senataxin helicase, in addition to BRCA1-mediated SETX recruitment [42].

Despite the appreciation of the role of R-loops in regulating transcription termination, how its formation disengages RNAP II from its DNA template is still unknown. In addition to acting as a roadblock for RNAP II, chromatin modifications induced by R-loop formation additionally contribute to its role in termination. For example, R-loops formed at the G-rich termination pause site of the *ACTB* gene induce antisense transcription and initiate the formation of dsRNA, which recruits the RNA interference (RNAi) machinery, including DICER1, Argonaute 1 and 2 (AGO1, AGO2), and chromatin modifying enzyme G9A. In this model, the methyltransferase G9A deposits H3K9me2 repressive mark [84] and recruits heterochromatin protein 1γ (HP1γ) to reinforce a condensed local chromatin status for impeding RNA Pol II progression and pausing, culminating in efficient termination [49]. Although these results indicate a connection between R-loops and chromatin condensation or focal heterochromatin formation at termination sites, others have found that transcription termination sites actually tend to be nucleosome free, especially in yeast genome [85–87]. Accordingly, R-loops at Snord116 locus in human cells correlate with increased chromatin decondensation [88]. However, it is still unclear why in some cases R-loops promote chromatin condensation, but in other cases R-loops are associated with open chromatin. Since the length of the R-loop varies from a few bp to a few kb, different lengths of R-loops may affect epigenome differently.

R-loops and antisense transcription

R-loops have also been shown to regulate antisense transcription of long noncoding RNAs (lncRNAs). In *Arabidopsis*, one of the ways that flowering is regulated is by an epigenetic switch called vernalization [89]. This epigenetic switch is triggered by prolonged exposure to cold in the winter and is mediated by PRC2, which marks the Flowering Locus C (FLC) with the silencing mark H3K27me3. PRC2 is recruited to the *FLC* gene by antisense noncoding RNA, such as COOLAIR, and therefore the relative levels of *COOLAIR* are crucial for the proper transition to flowering during vernalization. The mechanism that regulates these lncRNAs is unclear. However, a recent study showed that R-loops are important for COOLAIR antisense expression, in line with the observation in yeast that R-loops form at genes with antisense transcription [18]. These R-loops accumulate in the promoter region of the *COOLAIR* gene, which is in the terminator region of *FLC*, and are thought to stall the elongating RNA Pol II, leading to abortion of *COOLAIR* transcription [90]. Although the mechanism is not yet clear, a small patch of heterochromatin marked by H3K9me2 was identified immediately downstream of the *FLC* sense transcript polyadenylation site where R-loop forms [90], indicating that the formation of R-loops may promote a repressive chromatin status to suppress *COOLAIR* expression.

In supporting R-loop-mediated transcription repression of antisense genes in plants, R-loops formed at the human *Snord116* locus suppress the expression of its downstream antisense gene, *Ube3a-ATS* [88]. Expression of this antisense transcript leads to silencing of the *Ube3a* gene, whose loss-of-function has been characterized as the fundamental genetic basis for Angelman syndrome (AS) [91]. Topotecan was recently identified as an AS candidate drug for its inhibition of *Ube3a-ATS* and restoration of Ube3a expression in mouse neurons and brain [92]. Topotecan acts as a topoisomerase inhibitor and stabilizes R-loops at the *Snord116* locus, leading to excessive stalling of the transcription machinery, resulting in decreased *Ube3a-ATS* expression [88]. Interestingly, although no particular histone modification has been linked to Snord116-induced R-loop formation, more decondensed chromatin was observed at R-loop forming regions using combined immuno-FISH assays, possibly due to the depletion of nucleosome occupancy at these regions. This seems to be also the case for R-loops induced by antisense transcription at the promoter of the human vimentin (*VIM*) gene, which increases chromatin decondensation and facilitates *VIM* transcription [93]. VIM gene expression is positively correlated with its promoter antisense R-loop levels, as knockdown of

antisense expression or overexpression of RNASEH1, which reduces promoter R-loops, both lead to reduction in gene expression.

R-loops at telomeres and centromeres

Telomeres are important chromatin territories that protect chromosome ends from degradation and activation of DDR pathways [94]. Telomeric repeat-containing long noncoding RNA (TERRA) plays an important role in regulation of heterochromatin formation at chromosome ends, a process involving the formation of DNA–RNA hybrids between TERRA transcripts and telomere DNA [95]. Additionally, loss of function mutations of the THO complex, an evolutionary conserved mRNA processing component, causes accumulation of TERRA-dependent R-loops and telomere shortening [62]. In contrast, in telomerase-deficient cells, where telomere length is maintained through telomerase-independent alternative lengthening of telomere pathway, overexpression of RNASEH1, which disrupts TERRA R-loops, leads to telomere shortening [63]. These results suggest that TERRA transcript-induced R-loop levels need to be precisely controlled to maintain telomere integrity. R-loop-induced heterochromatin formation could play a role in this process, because TERRA transcripts have been reported to interact with a group of chromatin modifying enzymes to promote repressive chromatin states, including depositions of H3K9me2 and H3K9me3 and recruitment of heterochromatin protein HP1 [64,65].

Noncoding RNAs transcribed from centromeric repeats can also form R-loops, which mediates heterochromatin states of the centromere. The RNA-induced transcriptional silencing (RITS) complex was targeted to either chromatin-associated noncoding RNA or the single-stranded DNA of the R-loops to promote heterochromatin formation in yeast [60]. Similar mechanisms have also been proposed in the R-loop-induced repressive chromatin marks over mammalian gene terminators [49]. A recent study also showed a direct correlation between R-loops and the repressive chromatin compaction marks, phosphorylation at H3S10 (H3S10P) in both yeast and human cells at centromeres and pericentromeric regions [61]. Cells defective in mRNP biogenesis pathways, including loss-of-function mutations of THO complex and Senataxin, show accumulated R-loops and the H3S10P mark. Chromatin condensation caused by accumulated R-loops can act as a strong barrier to replication progression and hence, a major source of replication stress and genome instability [61]. For the establishment of H3S10P mark following R-loop formation, it is likely that enhanced recruitment of the responsive kinase or inhibition of phosphatase activity could be involved. However, their targeting mechanisms have not been explored.

R-loops regulate gene looping

Besides local epigenetic modification changes, R-loops formed by lncRNA are likely involved in gene looping and higher-order chromatin architecture, which promote interactions between distal gene elements for efficient transcription regulation and potential recombination [96]. The mediator complex has long been known as the major component of gene looping machinery, but the mechanisms are still elusive [97]. Recently, a group of mediator-related lncRNAs was discovered, which sheds new light into the mechanisms by which the mediator complexes promote chromatin architecture change [98–100]. In these models, ncRNA-activating (ncRNA-a) or enhancer-like RNAs interact with components of mediator complex, including MED1 and MED12, which bridge enhancer–promoter interactions and activate neighboring gene expression. Although no direct evidence supports the role of R-loop formation in this process, mediator components have been shown to constitute a major pathway to suppress R-loop formation. Yeast strains mutated in mediator components (e.g. med13, med12, med1, med5, and med16) all show significant amounts of genome instability caused by aberrant R-loop formation [19]. Thus, it is attractive to test the hypothesis that R-loop formation is involved in the maintenance of high-order chromatin structure. A recent study showed that in differentiated B cells and pluripotent ESCs, an enhancer RNA (eRNA) transcribed from a distal divergent eRNA-expressing element (lncRNA–CSR) is engaged in long-range DNA interactions that regulate *IgH* 3′ (immunoglobulin heavy chain) super enhancer function [58]. In fact, the eRNAs transcribed from active enhancers form R-loop structures when the RNA exosome machinery (including core and nuclear RNase components) is attenuated, suggesting that the RNA exosome regulates the levels of divergently transcribed eRNA by preventing R-loop formation [58]. However, R-loops formation at eRNA loci, caused by loss of RNA exosome function, is accompanied by loss of the chromatin silencing mark H3K9me2 and HP1 recruitment, which is in contrast to the R-loop-induced gain of H3K9me2 mark at transcription termination sites [49] and at trinucleotide-repeat sequences [56].

More direct evidence supporting R-loop function in higher-order chromatin architecture comes from a yeast study using chromatin conformation capture (3 C) to analyze the role of a lncRNA in facilitating promoter–terminator looping [59]. The yeast *GAL10* gene is known to form gene loops upon transcription activation, but the mechanism is largely unknown [101]. GAL lncRNAs, transcribed from the 3′ end of *GAL10*, reduce Cyc8 repressor binding at *GAL* promoters in a time-dependent manner and form lncRNA–DNA hybrids, a critical step for gene looping and activation during a nutritional switch [59]. Overexpression of RNase H1 disrupts lncRNA–DNA hybrids and abolishes the gene activation, supporting the role of R-loop formation in this regulation.

Conclusions and Perspectives

The study of R-loops is experiencing a renaissance since the recent development of the methods using DNA–RNA hybrid-specific antibody for gene specific and genome-wide R-loop detection. Researchers start to depict the regulatory network of this previously considered rare transcription by-product. However, there are likely many other aspects of R-loop biology that have not been explored.

Crosstalk between R-loops and epigenetic modifications

Emerging evidence suggests that the crosstalk exists between R-loops and epigenetic modifications of chromatin. On the one hand, certain chromatin modification features are preventive for R-loop formation [27,28,80], while on the other hand, their formation can promote chromatin dynamics by altering epigenetic modifications [46,47,55,60,61,65,73,75]. However, it is still unclear why in some cases R-loops promote chromatin condensation but in other cases, R-loops are associated with open chromatin. It is likely that specific epigenetic changes induced by R-loops are context-dependent and are determined by complex regulatory mechanisms, including where it forms (e.g. promoters, enhancers, or termination regions), what leads to its formation (e.g. mRNP biogenesis, DNA topoisomerase, trinucleotide-repeat, or antisense transcription), what are their features (e.g. length, additional structures of single-stranded DNA within R-loop and so on), among others. These factors need to be

considered in experimental design and interpretation of the data. Furthermore, as discussed before, current R-loop mapping in mammalian cells relies on restriction enzymes to fragment the genomic DNA before immunoprecipitation, which limits the resolution to define the lengths of R-loops. It is necessary to develop a high-resolution R-loop mapping strategy to facilitate studies in this direction.

Cell cycle regulation of R-loop dynamics

Unscheduled formation of RNA–DNA hybrids in the genome creates harmful intermediates that can have deleterious consequences on genome stability [102]. Although multiple surveillance protein factors have been identified to control the R-loop formation in general or in a gene-specific manner, it is still unclear how the formation of R-loops is dynamically regulated through cell cycle. To avoid transcription and replication conflicts, cells coordinate transcription and DNA replication such that early firing replication origins tend to be located near transcription start sites and have an open chromatin configuration [103]. Transcriptionally active regions are more likely to cause replication fork stalling, so replicating them at the very beginning of S phase might minimize the possibility of incomplete replication by the end of S phase or delay cell cycle progression. R-loops are mostly found at the promoters of highly transcribed genes, which are also, mainly, the regions of replication firing. The formation of R-loops at these regions has been found to cause replication–transcription conflicts and genome instability [1]. Unfortunately, little is known about specific mechanism that control or limit R-loop formation during cell cycle progression.

Novel factors involved in R-loop biogenesis

Genome-wide screens in budding yeast and human cells have revealed that levels of RNA–DNA hybrids increase when mRNP biogenesis is disturbed at sites of transcription initiation or repression, elongation, splicing, degradation, and export. Additionally, accumulating evidence suggests that DDR pathway proteins are also involved in R-loop regulation. Recently, the yeast homologous recombination protein, Rad51, was found to be required for R-loop formation in *trans* [7]. These findings not only present a paradox in which DNA repair protein can also promote R-loop-dependent DNA damage, but also identifies for the first time that R-loop can form in *trans*. The transient nature of DNA–RNA hybrids complicates their detection and biochemical analysis in a chromatin context, and it is likely that only a fraction of the participating proteins and mechanisms have been identified and characterized.

Acknowledgements

We apologize to the colleagues whose work we are unable to cite due to space constraints.

Funding

This work was supported by the City of Hope Start-up Funding to Dr Yanzhong Yang. Dr Qais Al-Hadid is a postdoctoral fellow supported by the DNA Damage, Response, and Oncogenic Signaling (DDROS) training program at Beckman Research Institute of City of Hope Cancer Center.

References

1. Santos-Pereira JM, Aguilera A. R loops: new modulators of genome dynamics and function. *Nat Rev Genet* 2015, 16: 583–597.
2. Thomas M, White RL, Davis RW. Hybridization of RNA to double-stranded DNA: formation of R-loops. *Proc Natl Acad Sci USA* 1976, 73: 2294–2298.
3. Drolet M, Phoenix P, Menzel R, Masse E, Liu LF, Crouch RJ. Overexpression of RNase H partially complements the growth defect of an *Escherichia coli* delta topA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I. *Proc Natl Acad Sci USA* 1995, 92: 3526–3530.
4. Drolet M, Bi X, Liu LF. Hypernegative supercoiling of the DNA template during transcription elongation *in vitro*. *J Biol Chem* 1994, 269: 2068–2074.
5. Westover KD, Bushnell DA, Kornberg RD. Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* 2004, 119: 481–489.
6. Toriumi K, Tsukahara T, Hanai R. R-loop formation in trans at an AGGAG repeat. *J Nucleic Acids* 2013, 2013: 629218.
7. Wahba L, Gore SK, Koshland D. The homologous recombination machinery modulates the formation of RNA-DNA hybrids and associated chromosome instability. *eLife* 2013, 2: e00505.
8. Nadel J, Athanasiadou R, Lemetre C, Wijetunga NA, Ó Broin P, Sato H, Zhang Z, *et al*. RNA:DNA hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships. *Epigenetics Chromatin* 2015, 8: 46.
9. Bacolla A, Wells RD. Non-B DNA conformations, genomic rearrangements, and human disease. *J Biol Chem* 2004, 279: 47411–47414.
10. Aguilera A, Garcia-Muse T. R loops: from transcription byproducts to threats to genome stability. *Mol Cell* 2012, 46: 115–124.
11. Groh M, Gromak N. Out of balance: R-loops in human disease. *PLoS Genet* 2014, 10: e1004630.
12. Skourti-Stathaki K, Proudfoot NJ. A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev* 2014, 28: 1384–1396.
13. Sollier J, Cimprich KA. Breaking bad: R-loops and genome integrity. *Trends Cell Biol* 2015, 25: 514–522.
14. Wongsurawat T, Jenjaroenpun P, Kwoh CK, Kuznetsov V. Quantitative model of R-loop forming structures reveals a novel level of RNA-DNA interactome complexity. *Nucleic Acids Res* 2012, 40: e16.
15. Roy D, Lieber MR. G clustering is important for the initiation of transcription-induced R-loops *in vitro*, whereas high G density without clustering is sufficient thereafter. *Mol Cell Biol* 2009, 29: 3124–3133.
16. Shrestha P, Xiao S, Dhakal S, Tan Z, Mao H. Nascent RNA transcripts facilitate the formation of G-quadruplexes. *Nucleic Acids Res* 2014, 42: 7236–7246.
17. Sollier J, Stork CT, Garcia-Rubio ML, Paulsen RD, Aguilera A, Cimprich KA. Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol Cell* 2014, 56: 777–785.
18. Chan YA, Aristizabal MJ, Lu PY, Luo Z, Hamza A, Kobor MS, Stirling PC, *et al*. Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-chip. *PLoS Genet* 2014, 10: e1004288.
19. Wahba L, Amon JD, Koshland D, Vuica-Ross M. RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Mol Cell* 2011, 44: 978–988.
20. Luna R, Jimeno S, Marin M, Huertas P, Garcia-Rubio M, Aguilera A. Interdependence between transcription and mRNP processing and export, and its impact on genetic stability. *Mol Cell* 2005, 18: 711–722.
21. Stirling PC, Chan YA, Minaker SW, Aristizabal MJ, Barrett I, Sipahimalani P, Kobor MS, *et al*. R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. *Genes Dev* 2012, 26: 163–175.
22. Huertas P, Aguilera A. Cotranscriptionally formed DNA: RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol Cell* 2003, 12: 711–721.

23. Sikdar N, Banerjee S, Zhang H, Smith S, Myung K. Spt2p defines a new transcription-dependent gross chromosomal rearrangement pathway. *PLoS Genet* 2008, 4: e1000290.
24. Santos-Pereira JM, Herrero AB, Garcia-Rubio ML, Marin A, Moreno S, Aguilera A. The Npl3 hnRNP prevents R-loop-mediated transcription-replication conflicts and genome instability. *Genes Dev* 2013, 27: 2445–2458.
25. El Hage A, French SL, Beyer AL, Tollervey D. Loss of topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev* 2010, 24: 1546–1558.
26. Li M, Pokharel S, Wang JT, Xu X, Liu Y. RECQ5-dependent SUMOylation of DNA topoisomerase I prevents transcription-associated genome instability. *Nat Commun* 2015, 6: 6720.
27. Yang Y, McBride KM, Hensley S, Lu Y, Chedin F, Bedford MT. Arginine methylation facilitates the recruitment of TOP3B to chromatin to prevent R loop accumulation. *Mol Cell* 2014, 53: 484–497.
28. Zhao DY, Gish G, Braunschweig U, Li Y, Ni Z, Schmitges FW, Zhong G, *et al*. SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. *Nature* 2016, 529: 48–53.
29. Cerritelli SM, Crouch RJ. Ribonuclease H: the enzymes in eukaryotes. *FEBS J* 2009, 276: 1494–1505.
30. Hraiky C, Raymond MA, Drolet M. RNase H overproduction corrects a defect at the level of transcription elongation during rRNA synthesis in the absence of DNA topoisomerase I in *Escherichia coli*. *J Biol Chem* 2000, 275: 11257–11263.
31. Majorek KA, Dunin-Horkawicz S, Steczkiewicz K, Muszewska A, Nowotny M, Ginalski K, Bujnicki JM. The RNase H-like superfamily: new members, comparative structural analysis and evolutionary classification. *Nucleic Acids Res* 2014, 42: 4160–4179.
32. Linder P, Jankowsky E. From unwinding to clamping - the DEAD box RNA helicase family. *Nat Rev Mol Cell Biol* 2011, 12: 505–516.
33. Brosh RM Jr. DNA helicases involved in DNA repair and their roles in cancer. *Nat Rev Cancer* 2013, 13: 542–558.
34. Boule JB, Zakian VA. The yeast Pif1p DNA helicase preferentially unwinds RNA DNA substrates. *Nucleic Acids Res* 2007, 35: 5809–5818.
35. Chakraborty P, Grosse F. Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA Repair (Amst)* 2011, 10: 654–665.
36. Skourti-Stathaki K, Proudfoot NJ, Gromak N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol Cell* 2011, 42: 794–805.
37. Becherel OJ, Yeo AJ, Stellati A, Heng EY, Luff J, Suraweera AM, Woods R, *et al*. Senataxin plays an essential role with DNA damage response proteins in meiotic recombination and gene silencing. *PLoS Genet* 2013, 9: e1003435.
38. Britton S, Dernoncourt E, Delteil C, Froment C, Schiltz O, Salles B, Frit P, *et al*. DNA damage triggers SAF-A and RNA biogenesis factors exclusion from chromatin coupled to R-loops removal. *Nucleic Acids Res* 2014, 42: 9047–9062.
39. Bhatia V, Barroso SI, Garcia-Rubio ML, Tumini E, Herrera-Moyano E, Aguilera A. BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature* 2014, 511: 362–365.
40. Gan W, Guan Z, Liu J, Gui T, Shen K, Manley JL, Li X. R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev* 2011, 25: 2041–2056.
41. Schwab RA, Nieminuszczy J, Shah F, Langton J, Lopez Martinez D, Liang CC, Cohn MA, *et al*. The Fanconi anemia pathway maintains genome stability by coordinating replication and transcription. *Mol Cell* 2015, 60: 351–361.
42. Hatchi E, Skourti-Stathaki K, Ventz S, Pinello L, Yen A, Kamieniarz-Gdula K, Dimitrov S, *et al*. BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. *Mol Cell* 2015, 57: 636–647.
43. Kasahara M, Clikeman JA, Bates DB, Kogoma T. RecA protein-dependent R-loop formation *in vitro*. *Genes Dev* 2000, 14: 360–365.
44. Zaitsev EN, Kowalczykowski SC. A novel pairing process promoted by *Escherichia coli* RecA protein: inverse DNA and RNA strand exchange. *Genes Dev* 2000, 14: 740–749.
45. Boguslawski SJ, Smith DE, Michalak MA, Mickelson KE, Yehle CO, Patterson WL, Carrico RJ. Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. *J Immunol Methods* 1986, 89: 123–130.
46. Ginno PA, Lott PL, Christensen HC, Korf I, Chedin F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell* 2012, 45: 814–825.
47. Ginno PA, Lim YW, Lott PL, Korf I, Chedin F. GC skew at the 5′ and 3′ ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome Res* 2013, 23: 1590–1600.
48. Mischo HE, Gomez-Gonzalez B, Grzechnik P, Rondon AG, Wei W, Steinmetz L, Aguilera A, *et al*. Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol Cell* 2011, 41: 21–32.
49. Skourti-Stathaki K, Kamieniarz-Gdula K, Proudfoot NJ. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature* 2014, 516: 436–439.
50. Hage AE, Webb S, Kerr A, Tollervey D. Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria. *PLoS Genet* 2014, 10: e1004716.
51. Alexander RD, Innocente SA, Barrass JD, Beggs JD. Splicing-dependent RNA polymerase pausing in yeast. *Mol Cell* 2010, 40: 582–593.
52. Duttke SH, Lacadie SA, Ibrahim MM, Glass CK, Corcoran DL, Benner C, Heinz S, *et al*. Perspectives on unidirectional versus divergent transcription. *Mol Cell* 2015, 60: 348–349.
53. Lepoivre C, Belhocine M, Bergon A, Griffon A, Yammine M, Vanhille L, Zacarias-Cabeza J, *et al*. Divergent transcription is associated with promoters of transcriptional regulators. *BMC Genomics* 2013, 14: 914.
54. Wu X, Sharp PA. Divergent transcription: a driving force for new gene origination? *Cell* 2013, 155: 990–996.
55. Lim YW, Sanz LA, Xu X, Hartono SR, Chedin F. Genome-wide DNA hypomethylation and RNA:DNA hybrid accumulation in Aicardi-Goutieres syndrome. *eLife* 2015, 4: doi:10.7554/eLife.08007.
56. Colak D, Zaninovic N, Cohen MS, Rosenwaks Z, Yang WY, Gerhardt J, Disney MD, *et al*. Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. *Science* 2014, 343: 1002–1005.
57. Groh M, Lufino MM, Wade-Martins R, Gromak N. R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. *PLoS Genet* 2014, 10: e1004318.
58. Pefanis E, Wang J, Rothschild G, Lim J, Kazadi D, Sun J, Federation A, *et al*. RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. *Cell* 2015, 161: 774–789.
59. Cloutier SC, Wang S, Ma WK, Al Husini N, Dhoondia Z, Ansari A, Pascuzzi PE, *et al*. Regulated formation of lncRNA-DNA hybrids enables faster transcriptional induction and environmental adaptation. *Mol Cell* 2016, 61: 393–404.
60. Nakama M, Kawakami K, Kajitani T, Urano T, Murakami Y. DNA-RNA hybrid formation mediates RNAi-directed heterochromatin formation. *Genes Cells* 2012, 17: 218–233.
61. Castellano-Pozo M, Santos-Pereira JM, Rondon AG, Barroso S, Andujar E, Perez-Alegre M, Garcia-Muse T, *et al*. R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. *Mol Cell* 2013, 52: 583–590.
62. Pfeiffer V, Crittin J, Grolimund L, Lingner J. The THO complex component Thp2 counteracts telomeric R-loops and telomere shortening. *EMBO J* 2013, 32: 2861–2871.
63. Arora R, Lee Y, Wischnewski H, Brun CM, Schwarz T, Azzalin CM. RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. *Nat Commun* 2014, 5: 5220.
64. Deng Z, Norseen J, Wiedmer A, Riethman H, Lieberman PM. TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. *Mol Cell* 2009, 35: 403–413.

65. Porro A, Feuerhahn S, Delafontaine J, Riethman H, Rougemont J, Lingner J. Functional characterization of the TERRA transcriptome at damaged telomeres. *Nat Commun* 2014, 5: 5379.
66. Schmitz KM, Mayer C, Postepska A, Grummt I. Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev* 2010, 24: 2264–2269.
67. Mayer C, Schmitz KM, Li J, Grummt I, Santoro R. Intergenic transcripts regulate the epigenetic state of rRNA genes. *Mol Cell* 2006, 22: 351–361.
68. Bierhoff H, Schmitz K, Maass F, Ye J, Grummt I. Noncoding transcripts in sense and antisense orientation regulate the epigenetic state of ribosomal RNA genes. *Cold Spring Harb Symp Quant Biol* 2010, 75: 357–364.
69. Hartono SR, Korf IF, Chedin F. GC skew is a conserved property of unmethylated CpG island promoters across vertebrates. *Nucleic Acids Res* 2015, 43: 9729–9741.
70. Mutskov V, Felsenfeld G. Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. *EMBO J* 2004, 23: 138–149.
71. Takeshima H, Yamashita S, Shimazu T, Niwa T, Ushijima T. The presence of RNA polymerase II, active or stalled, predicts epigenetic fate of promoter CpG islands. *Genome Res* 2009, 19: 1974–1982.
72. Krajewski WA, Nakamura T, Mazo A, Canaani E. A motif within SET-domain proteins binds single-stranded nucleic acids and transcribed and supercoiled DNAs and can interfere with assembly of nucleosomes. *Mol Cell Biol* 2005, 25: 1891–1899.
73. Chen PB, Chen HV, Acharya D, Rando OJ, Fazzio TG. R loops regulate promoter-proximal chromatin architecture and cellular differentiation. *Nat Struct Mol Biol* 2015, 22: 999–1007.
74. Coffee B, Zhang F, Ceman S, Warren ST, Reines D. Histone modifications depict an aberrantly heterochromatinized FMR1 gene in fragile x syndrome. *Am J Hum Genet* 2002, 71: 923–932.
75. Loomis EW, Sanz LA, Chedin F, Hagerman PJ. Transcription-associated R-loop formation across the human FMR1 CGG-repeat region. *PLoS Genet* 2014, 10: e1004294.
76. Tous C, Aguilera A. Impairment of transcription elongation by R-loops *in vitro*. *Biochem Biophys Res Commun* 2007, 360: 428–432.
77. Belotserkovskii BP, Hanawalt PC. PNA binding to the non-template DNA strand interferes with transcription, suggesting a blockage mechanism mediated by R-loop formation. *Mol Carcinog* 2015, 54: 1508–1512.
78. Campuzano V, Montermini L, Molto MD, Pianese L, Cossee M, Cavalcanti F, Monros E, *et al*. Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* 1996, 271: 1423–1427.
79. Yang Y, Lu Y, Espejo A, Wu J, Xu W, Liang S, Bedford MT. TDRD3 is an effector molecule for arginine-methylated histone marks. *Mol Cell* 2010, 40: 1016–1023.
80. Sims RJ III, Rojas LA, Beck D, Bonasio R, Schuller R, Drury WJ III, Eick D, *et al*. The C-terminal domain of RNA polymerase II is modified by site-specific methylation. *Science* 2011, 332: 99–103.
81. Enerly E, Sheng Z, Li KB. Natural antisense as potential regulator of alternative initiation, splicing and termination. *In Silico Biol* 2005, 5: 367–377.
82. Morrissy AS, Griffith M, Marra MA. Extensive relationship between antisense transcription and alternative splicing in the human genome. *Genome Res* 2011, 21: 1203–1212.
83. Proudfoot NJ. Ending the message: poly(A) signals then and now. *Genes Dev* 2011, 25: 1770–1782.
84. Shinkai Y, Tachibana M. H3K9 methyltransferase G9a and the related molecule GLP. *Genes Dev* 2011, 25: 781–788.
85. Yadon AN, Van de Mark D, Basom R, Delrow J, Whitehouse I, Tsukiyama T. Chromatin remodeling around nucleosome-free regions leads to repression of noncoding RNA transcription. *Mol Cell Biol* 2010, 30: 5110–5122.
86. Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR, Nislow C. A high-resolution atlas of nucleosome occupancy in yeast. *Nat Genet* 2007, 39: 1235–1244.
87. Mavrich TN, Ioshikhes IP, Venters BJ, Jiang C, Tomsho LP, Qi J, Schuster SC, *et al*. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res* 2008, 18: 1073–1083.
88. Powell WT, Coulson RL, Gonzales ML, Crary FK, Wong SS, Adams S, Ach RA, *et al*. R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation. *Proc Natl Acad Sci USA* 2013, 110: 13938–13943.
89. Heo JB, Sung S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* 2011, 331: 76–79.
90. Sun Q, Csorba T, Skourti-Stathaki K, Proudfoot NJ, Dean C. R-loop stabilization represses antisense transcription at the *Arabidopsis* FLC locus. *Science* 2013, 340: 619–621.
91. Chamberlain SJ, Lalande M. Angelman syndrome, a genomic imprinting disorder of the brain. *J Neurosci* 2010, 30: 9958–9963.
92. Huang HS, Allen JA, Mabb AM, King IF, Miriyala J, Taylor-Blake B, Sciaky N, *et al*. Topoisomerase inhibitors unsilence the dormant allele of Ube3a in neurons. *Nature* 2012, 481: 185–189.
93. Boque-Sastre R, Soler M, Oliveira-Mateos C, Portela A, Moutinho C, Sayols S, Villanueva A, *et al*. Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. *Proc Natl Acad Sci USA* 2015, 112: 5785–5790.
94. Doksani Y, de Lange T. The role of double-strand break repair pathways at functional and dysfunctional telomeres. *Cold Spring Harb Perspect Biol* 2014, 6: a016576.
95. Cusanelli E, Chartrand P. Telomeric repeat-containing RNA TERRA: a noncoding RNA connecting telomere biology to genome integrity. *Front Genet* 2015, 6: 143.
96. Hampsey M, Singh BN, Ansari A, Laine JP, Krishnamurthy S. Control of eukaryotic gene expression: gene loops and transcriptional memory. *Adv Enzyme Regul* 2011, 51: 118–125.
97. Allen BL, Taatjes DJ. The mediator complex: a central integrator of transcription. *Nat Rev Mol Cell Biol* 2015, 16: 155–166.
98. Hsieh CL, Fei T, Chen Y, Li T, Gao Y, Wang X, Sun T, *et al*. Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. *Proc Natl Acad Sci USA* 2014, 111: 7319–7324.
99. Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, Shiekhattar R. Activating RNAs associate with mediator to enhance chromatin architecture and transcription. *Nature* 2013, 494: 497–501.
100. Orom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, *et al*. Long noncoding RNAs with enhancer-like function in human cells. *Cell* 2010, 143: 46–58.
101. Laine JP, Singh BN, Krishnamurthy S, Hampsey M. A physiological role for gene loops in yeast. *Genes Dev* 2009, 23: 2604–2609.
102. Stuckey R, Garcia-Rodriguez N, Aguilera A, Wellinger RE. Role for RNA:DNA hybrids in origin-independent replication priming in a eukaryotic system. *Proc Natl Acad Sci USA* 2015, 112: 5779–5784.
103. Helmrich A, Ballarino M, Nudler E, Tora L. Transcription-replication encounters, consequences and genomic instability. *Nat Struct Mol Biol* 2013, 20: 412–418.