R loops: new modulators of genome dynamics and function

José M. Santos-Pereira and Andrés Aguilera

Abstract | R loops are nucleic acid structures composed of an RNA–DNA hybrid and a displaced single-stranded DNA. Recently, evidence has emerged that R loops occur more often in the genome and have greater physiological relevance, including roles in transcription and chromatin structure, than was previously predicted. Importantly, however, R loops are also a major threat to genome stability. For this reason, several DNA and RNA metabolism factors prevent R-loop formation in cells. Dysfunction of these factors causes R-loop accumulation, which leads to replication stress, genome instability, chromatin alterations or gene silencing, phenomena that are frequently associated with cancer and a number of genetic diseases. We review the current knowledge of the mechanisms controlling R loops and their putative relationship with disease.

G-quartets
Planar structures formed by four intra-strand Gs paired with each other.

Genome instability
A cellular pathological condition by which cells undergo mutations, recombination, chromosome rearrangements or chromosome loss at frequencies substantially higher than wild-type levels.

Centro Andaluz de Biología Molecular y Medicina Regenerativa CABIMER, Universidad de Sevilla, Av. Américo Vespucio s/n, Seville 41092, Spain. Correspondence to A.A.
e-mail: aguilo@us.es
doi:10.1038/nrg3961
Published online 15 September 2015

Short RNA–DNA hybrids form during natural processes such as replication, in which they prime DNA synthesis, or transcription at the active site of the RNA polymerase during ribonucleotide triphosphate incorporation. They are more stable than double-stranded DNA (dsDNA) and adopt a 'heteromerous' conformation that is intermediate between the B form of dsDNA and the A form of double-stranded RNA (dsRNA)[1]. A longer form of RNA–DNA hybrid can be generated by hybridization of the nascent RNA molecule with the template DNA strand. Such RNA–DNA hybrids and the resultant displaced single-stranded DNA (ssDNA) are collectively known as R loops (FIG. 1a), which are likely to occur as the nascent RNA exits from the RNA polymerase (known as the 'thread-back' model). Formation and/or stabilization of R loops *in vivo* is favoured by different DNA structural features, such as negative DNA supercoiling, G-content, the presence of DNA nicks or the formation of G-quartets in the displaced ssDNA[2–4].

In vivo R loops are found as natural intermediates during the initiation of DNA replication in mitochondrial DNA, bacterial plasmids, the bacteriophages ColE1 and T4, and in immunoglobulin (Ig) class-switch recombination (CSR)[5] (BOX 1). However, they can also be a source of genome instability[6], which is a hallmark of cancer cells. This was first shown in cells depleted of particular RNA biogenesis and processing factors such as the THO complex in yeast[7] and serine/arginine-rich splicing factor 1 (SRSF1; previously known as ASF and SF2) in vertebrates[8]. During the past decade, evidence has accumulated indicating that R loops occur more often in the genome and have greater physiological relevance than was predicted, with previously unknown roles in transcription[9] and chromatin structure[10].

Recent genome-wide approaches have permitted the detection of R loops in many loci. *In silico* analysis of the human genome identified a high number of putative R-loop-forming sequences (250,000) in 59% of human genes[11]. Consistent with these predictions, DNA–RNA immunoprecipitation followed by sequencing (DRIP–seq) identified thousands of R-loop peaks distributed along the human genome[12,13]. Furthermore, DRIP followed by microarray (DRIP–chip) and chromatin immunoprecipitation followed by sequencing (ChIP–seq) revealed numerous R-loop hotspots along the yeast genome, including the ribosomal DNA (rDNA) region, telomeres, Ty transposons, RNA Pol III-transcribed genes (that is, tRNA genes) and highly expressed RNA Pol II-transcribed genes[14,15] (BOX 2).

Other reviews have focused on different aspects of the structural and functional features of R loops[5,16,17]. Here, we review and update the current understanding of the factors and mechanisms that cells use to control R-loop accumulation and discuss the different physiological roles and consequences of R loops and their putative relation with disease. Our goal is to provide an integrated view of the relevance of R loops in transcription, chromatin structure and genome dynamics.

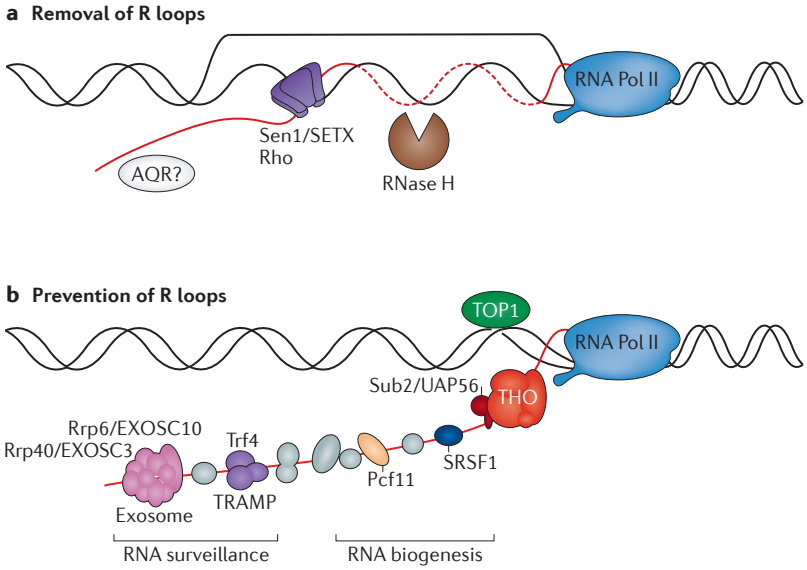


Figure 1 | **Cellular activities limiting R-loop accumulation.** **a** | Removal of R loops occurs by distinct but complementary mechanisms. The RNA moiety can be degraded by RNase H enzymes. Additionally, R loops can be unwound by helicases such as Rho in bacteria, Sen1 in yeast and senataxin (SETX) in humans, all of which have a role in transcription termination. Other putative helicases, such as aquarius (AQR), may also play a part. **b** | R-loop formation is prevented by topoisomerase 1 (TOP1), which resolves the local negative supercoiling behind the elongating RNA polymerase II (RNA Pol II)-transcribed mRNA and by specific RNA-binding proteins that are involved in RNA biogenesis (including the THO complex, serine/arginine-rich splicing factor 1 (SRSF1) and Pcf11) or in RNA surveillance (including the exoribonucleases exosome component 3 (EXOSC3) and EXOSC10 (Rrp40 and Rrp6 in yeast, respectively)) (TABLE 1). TRAMP, Trf4–Air2–Mtr4p polyadenylation complex; UAP56, 56 kDa U2AF65-associated protein (also known as DDX39B; known as Sub2 in yeast).

Holliday junctions
Double-stranded, cruciform structures formed as intermediates of recombinational repair in which both recombining DNA molecules are covalently linked.

Recombinational repair
A double-strand break repair that is active during the S–G_2 phases of the cell cycle and that uses information from a homologous sequence, normally the sister chromatid, to copy DNA.

Okazaki fragment
A discrete fragment created by synthesis of the DNA lagging strand during replication. Okazaki fragments are primed by a short RNA in the form of an RNA–DNA hybrid.

Negative supercoiling
Under-winding of a DNA strand, typically occurring behind an elongating RNA polymerase.

Activities controlling R loops *in vivo*

Cells have evolved various factors and processes to limit or prevent R-loop formation and accumulation (FIG. 1; TABLE 1); these are briefly summarized below.

Removal of R loops by ribonucleases and helicases. R loops can be removed by degradation of the RNA strand. This is achieved by RNase H enzymes, which specifically degrade the RNA moiety of RNA–DNA hybrids[18] (FIG. 1a). There are two types of RNase H enzymes, which differ in structure and substrate specificity. RNase H1 is monomeric, whereas RNase H2 is monomeric in bacteria but composed of three subunits in eukaryotes: RNH2A (the catalytic subunit), RNH2B and RNH2C. Both types of RNase H enzyme have the ability to remove RNA–DNA hybrids in addition to having different specialized roles[14,15,19–24].

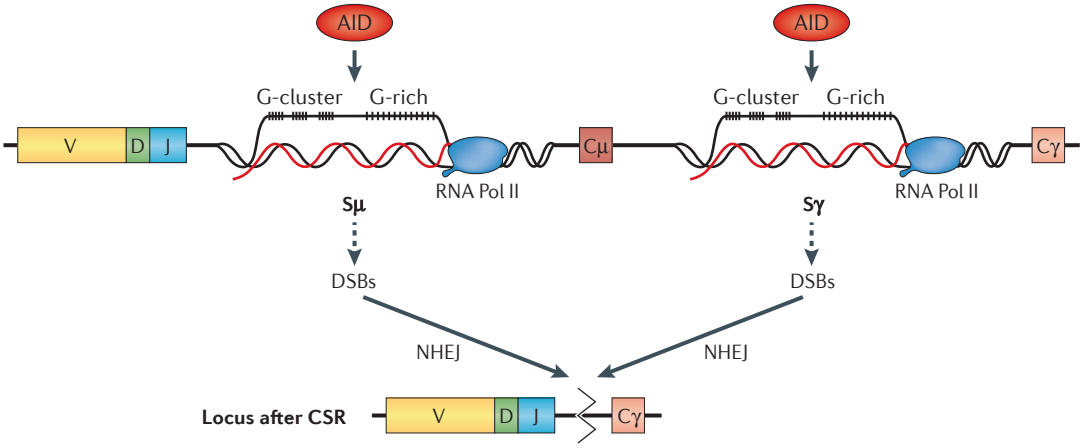
Alternatively, R loops can be removed by RNA–DNA helicases, which unwind RNA–DNA hybrids or limit their formation (FIG. 1a). In *Escherichia coli*, the RecG DNA helicase — which is involved in the migration of Holliday junctions during recombinational repair — and the Rho transcription termination factor unwind R loops *in vitro*[25,26]. The same is the case for the yeast DNA helicase Pif1 — which is involved in mitochondrial DNA maintenance, telomeric DNA synthesis, rDNA replication and Okazaki fragment maturation[27] — and the human DEAH box protein 9 (DHX9) RNA helicase[28].

The yeast protein Sen1 and its human homologue senataxin (SETX) have also been implicated in R-loop homeostasis. *Schizosaccharomyces pombe* Sen1 was initially identified as a DNA and RNA helicase with 5′–3′ RNA–DNA unwinding activity *in vitro*[29]. More recently, it was shown *in vivo* that *Saccharomyces cerevisiae* Sen1 helicase-inactive *sen1-1* mutants accumulated R loops at highly expressed loci[30] and that inactivation of human SETX leads to increased occurrences of R loops at transcription termination pause sites[9]. Furthermore, *Setx*$^{-/-}$ mice also accumulate R loops[31]. The results suggest a role for Sen1/SETX in the dissolution of R loops, in particular during transcription termination. The human aquarius (AQR) protein, which is an RNA helicase of the same family as Sen1/SETX[32], may also participate in R-loop removal *in vivo*, but we currently know little about this protein[33,34].

Negative DNA supercoiling facilitates R-loop formation. Co-transcriptionally generated negative supercoiling behind the RNA polymerase may lead to a transient local unwinding of the DNA strands, which might facilitate the pairing of the nascent RNA with the DNA. This has been suggested by several studies in bacteria and eukaryotes, such as those examining the effects that supercoiling regulation by topoisomerase (TOP) enzymes have on R-loop formation. First, genetic studies in *E. coli* showed that overexpression of RNase H1 suppressed the growth defect of *topA* mutants lacking Top1, and that *topA rnhA* double mutants are inviable[35]. In yeast, the loss of both Top1 and Top2 increases R-loop accumulation at the rDNA locus (FIG. 1b), causing RNA Pol I stalling and defects in precursor ribosomal RNA (pre-rRNA) synthesis[36]. Higher levels of R loops are seen in cells additionally lacking RNase H enzymes. Indeed, apart from mitochondrial DNA, the rDNA locus, located in the nucleolus, is the genomic region where most R loops accumulate in eukaryotic cells, as can be seen directly by immunofluorescence with the S9.6 RNA–DNA-specific hybrid antibody in yeast and human cells[33,37].

R-loop prevention by DNA TOP enzymes is not exclusive to the rDNA locus. Human TOP1-deficient cells show DNA breaks at active genes and replication defects that are suppressed by RNase H1 overexpression[38], suggesting that TOP1 prevents R-loop accumulation at any highly transcribed region. In addition, TOP3B, a subfamily 1A DNA TOP that unwinds negative supercoiled DNA, also counteracts R-loop formation in mammalian cells[39], and *Drosophila melanogaster* Top3B resolves R loops and D loops *in vitro* by cleaving the unpaired strand[40]. In this sense, it has been shown that sumoylation of human DNA TOP by the protein inhibitor of activated STAT protein 1 (PIAS1)–SRSF1 E3 ligase complex in the chromatin fraction containing active RNA Pol II inactivates TOP activity and promotes R-loop accumulation[41], suggesting a concerted regulation of TOP activity during transcription as a way to prevent R-loop formation. These results may also imply a double role for SRSF1 in R-loop homeostasis (see below).

Box 1 | **Immunoglobulin class-switch recombination**



Class-switch recombination (CSR) is a process that occurs at the immunoglobulin (Ig) heavy chain locus in activated mammalian B cells and allows switching to different Ig isotypes without changing the antigen specificity. It consists of a recombination event between two switch (S) regions, which are highly repetitive 1–12 kb sequences with G-rich non-transcribed strands (NTSs) that precede the constant (C) regions of Ig genes[61] (see the figure). R loops act as key intermediates during CSR, in which they form during transcription through S regions owing to the G-rich nature of their NTS[3]. These R loops have been detected both *in vitro* and *in vivo*[117,118].

CSR requires the action of activation-induced cytidine deaminase (AID), as demonstrated by AID transfection in murine lymphoma cells and by the generation of AID-deficient mice (*Aid*$^{-/-}$), the CSR levels of which have been analysed both *in vivo* and *in vitro*[119,120]. AID acts on single-stranded DNA (ssDNA) and can deaminate dC to dU *in vitro* in G-rich substrates, an action that is enhanced by transcription and that occurs preferentially but not exclusively over the NTS[121]. Recently, it has been shown that AID is targeted to S regions by binding to G-quartets formed at intronic switch RNAs[122], providing a role for the switch non-coding RNAs (ncRNAs) generated at S regions. In addition to S regions, AID is recruited to exosome-targeted R-loop-forming ncRNAs in exosome-deficient B cells[123]. The efficiency of CSR *in vivo* using artificial S regions depends on the abundance of G-clusters and AID target sites[124]. It is thought that CSR occurs by non-homologous end-joining (NHEJ) of double-strand breaks (DSBs), but the mechanism leading to these breaks is not completely understood. One possibility is that the dU introduced by AID may be processed either by base-excision repair or by mismatch repair machineries to lead to DSBs or to nicks that would be converted into DSBs during replication. In addition, the formation of G-quartet structures in the NTS could stabilize R loops formed in S regions.
RNA Pol II, RNA polymerase II.

Topoisomerase
(TOP). An enzyme family that can remove (or create) supercoiling in duplex DNA by making transitory breaks in one strand (type 1 TOPs) or both strands (type 2 TOPs) of the DNA backbone.

D loops
(Displacement loops). DNA structures consisting of a main double-stranded DNA that has been separated by a third DNA strand complementary to one of the main strands and that by pairing with it displaces the other main strand.

Bisulfite mutagenesis
A method to mutagenize DNA with sodium bisulfite, which acts exclusively over single-stranded DNA and, therefore, serves to determine whether a particular DNA sequence is present *in vivo* in a single-stranded form (as is the case for the strand displaced by the RNA in the R loop) or in a double-stranded form.

DNA damage response
(DDR). A general cellular response consisting of DNA damage sensing, activation of different checkpoints, the action of the appropriate DNA repair pathway and the arrest of the cell cycle.

Multicopy suppressors
Genes that are able to suppress the phenotype conferred by specific mutations when they are present in high-copy-number plasmids.

mRNP biogenesis protects against R-loop formation. The nascent mRNA is key for the formation of R loops at genes, and deficiencies in messenger ribonucleoprotein particle (mRNP) assembly can strongly stimulate R-loop formation. This may be due to the direct or indirect protection of the nascent mRNA by specific RNA-binding proteins (RBPs) that would prevent its hybridization with the transcribed DNA strand (FIG. 1b). R loops were detected physically in yeast mutants of the THO complex (which is involved in mRNP biogenesis) by purification of RNase H1-sensitive nucleic acids and genetically by suppression of their hyper-recombination phenotype by RNase H1 overexpression. These results provided a first link between R loops, mRNP biogenesis and genomic instability[7] that is conserved in *Caenorhabditis elegans* and human cells[42,43]. Furthermore, another mRNP factor, SRSF1, prevents R-loop formation in vertebrate cells. Chicken DT40 and human HeLa cells depleted of SRSF1 show increased levels of rearrangements suppressed by RNase H1 overexpression, as well as a mutation profile caused by bisulfite mutagenesis, which specifically acts on the displaced ssDNA of R loops, indicating that SRSF1-depleted cells accumulate these structures[8].

Genome-wide screening of genes controlling the DNA damage response (DDR) in yeast and human cells[20,21,34] as well as more focused analyses of specific RNA biogenesis and processing factors have provided further and definitive evidence that a number of nuclear factors involved in RNA metabolism contribute to prevent R-loop accumulation and genome instability (TABLE 1). This conclusion is reinforced by the fact that multicopy suppressors of some of these deficiencies include genes encoding RBPs that are involved in mRNP assembly, such as Tho1 and Sub2 in yeast or RBP with serine-rich domain 1 (RNPS1) in human cells[44,45]. Also, the mitotic regulators zinc finger protein 207 (ZNF207; also known as BuGZ) and BUB3 cause splicing defects and R-loop accumulation in human adenocarcinoma cells[46], consistent with a link between splicing defects and R-loop accumulation. Recently, it has been shown that the yeast ataxin 2 RBP, PAB1-binding protein 1 (Pbp1), interacts with non-coding RNAs (ncRNAs) generated by RNA Pol II in the intergenic spacers of the rDNA locus, which is located between the RNA Pol I-driven rRNA genes, to prevent R-loop formation[47].

The mRNA surveillance system also prevents R-loop accumulation, as shown in yeast cells depleted of the

Trf4 (also known as Pab2) polyadenylation (poly(A)) polymerase of the TRAMP (Trf4–Air2–Mtr4p polyadenylation) complex[48], and in mouse embryonic stem cells depleted of the exoribonucleases exosome component 3 (EXOSC3) and EXOSC10 (Rrp40 and Rrp6 in yeast, respectively) of the RNA exosome. These cells show increased levels of R loops in regions divergently expressing exosome-substrate enhancer RNAs (eRNAs)[49] (FIG. 1b).

It is likely that each of these RNA-binding factors acts in a specific manner to prevent R-loop formation because they act on different steps from transcription to mRNA export and degradation. Interestingly, not all of those factors have the same impact on genome integrity and expression, which suggests that the R loops formed in each case might be different, as discussed below. However, the results indicate that the structure and fate of the nascent mRNP is crucial for R-loop prevention.

R-loop control of RNA Pol II-driven transcription

Evidence shows that R loops can have a role in transcription activation and termination in mammals, in most cases involving antisense RNAs or ncRNAs in association with chromatin modifications; these findings suggest new functions for R loops, which are discussed below.

R loops in transcription activation. R loops may have a natural role in transcription: this can be inferred from the identification through genome-wide mapping of

Box 2 | **Genome-wide mapping of RNA–DNA hybrids**

In vivo detection of R loops had been a challenge before the advent of the S9.6 monoclonal antibody for detection of RNA–DNA hybrids[125], which constituted a considerable advance. This antibody can be used to visualize R loops by immunofluorescence or by immunoprecipitation. However, studies relying solely on this antibody should be complemented with additional experimental controls, including the use of additional detection systems or the suppression of R-loop-dependent structures or phenotypes by RNase H to be sure that other RNA structures are not being detected instead[126].

In the past few years, these methods have been applied to the whole genome by combining immunoprecipitation with next-generation sequencing or microarray (chip) hybridization. For example, DNA–RNA immunoprecipitation followed by sequencing (DRIP–seq) has been used to map R loops in the human genome[12,13]. In this method, genomic DNA is digested with a cocktail of restriction enzymes, which creates small fragments that are then immunoprecipitated using the S9.6 antibody, purified and sequenced. An alternative but less-efficient method, called DNA–RNA *in vitro* enrichment followed by sequencing (DRIVE–seq), relies on a catalytically deficient but binding competent human RNase H1 that is used in affinity pull-down assays to recover R loops, which are then sequenced[13]. Both methods have contributed to the mapping of R loops at promoters and terminator regions of human genes (see the red peaks in the figure; note that the R-loop peaks are artistic representations throughout the figure and not based on specific data). The number of R-loop peaks varies between experiments depending on the method used (from 1,200 to 20,000 peaks). To complement the experimental detection of R loops, a Quantitative Model of R-loop Forming Sequence (Qm-RLFS-finder) has been recently developed to computationally predict the potential of a given sequence to form an R loop[127], with a reported accuracy of 80–90%.

In yeast, chromatin has been crosslinked using formaldehyde and fragmented by sonication before S9.6 immunoprecipitation. Chromatin immunoprecipitation followed by sequencing (ChIP–seq)[15] or DRIP followed by hybridization with tiling microarrays (DRIP–chip)[14] permitted the identification of R-loop genomic hotspots at RNA polymerase II (RNA Pol II)- and RNA Pol III-transcribed genes, ribosomal DNA (rDNA), Ty transposable elements and telomeres (see the blue peaks in the figure).

Finally, a non-denaturing bisulfite modification method can be applied to the whole genome followed by DNA sequencing, as was recently done in bacteria, to infer R loops that have accumulated at high levels in transcription-termination mutants[128]. This method relies on the ability of sodium bisulfite to mutate C to T in the single-stranded DNA (ssDNA) displaced by R loops, followed by PCR amplification and sequencing.

In addition to genome-wide applications, the methods discussed above can be adapted for the detection of R loops at specific genomic regions, such as the R loops that have been identified in pericentromeric regions using DRIP in yeast THO mutants[10] (see the grey peak in the figure). CEN, centromere; LTR, long terminal repeat; ORF, open reading frame; STR, subtelomeric repeat element; TG1–3, telomeric repeats consisting of T and 1–3 Gs; TSS, transcription start site; TTS, transcription termination site.

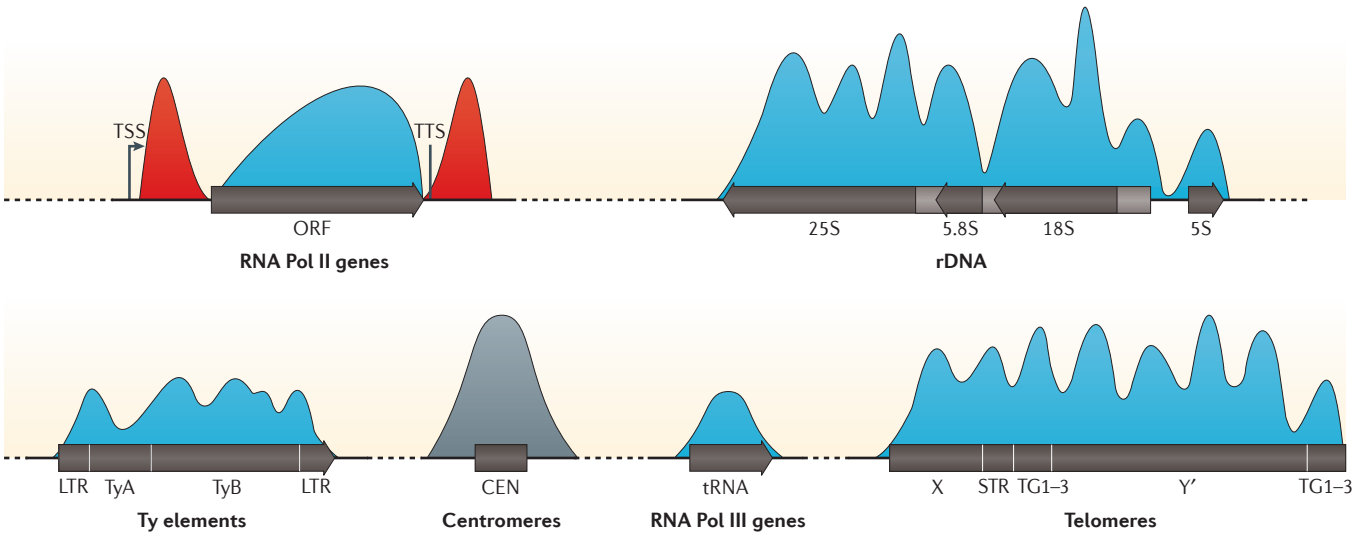


Table 1 | **Representative genes that can cause increased levels of R loops if dysfunctional**

Biological process	Gene	Organism	Function	R-loop-dependent genome instability	Refs
Transcription regulation and chromatin modification	*LEO1*	*Saccharomyces cerevisiae*	PAF complex subunit	+	20,129
	SIN3	*S. cerevisiae*	RPD3 histone deacetylase complex subunit	+	20
	SDS3	*S. cerevisiae*	Rpd3 histone deacetylase complex subunit	+	14,20
	MED12	*S. cerevisiae*	Mediator complex subunit	+	20,129
	TDRD3	*H. sapiens*	Scaffold protein interacting with methylated Arg	ND	39
	Tdrd3	*Mus musculus*	Scaffold protein interacting with methylated Arg	ND	39
Precursor mRNA splicing	*AQR*	*Homo sapiens*	Putative RNA–DNA helicase	+	33,34
	SRSF1	*Gallus gallus*	Splicing factor	+	8
Mitosis regulation — splicing	*ZNF207*	*H. sapiens*	Zinc finger protein	+	46
	BUB3	*H. sapiens*	Spindle assembly checkpoint	+	46
Transcription termination and/or mRNA 3′-end processing	*CLP1*	*S. cerevisiae*	CF1A subunit	+	21
	PCF11	*S. cerevisiae*	CF1A subunit	+	21
	RNA15	*S. cerevisiae*	CF1A subunit	+	21
	CFT2	*S. cerevisiae*	CPF subunit	+	21
	FIP1	*S. cerevisiae*	CPF subunit	+	21
	PBP1	*S. cerevisiae*	Poly(A)-binding protein	+	47
	SEN1	*S. cerevisiae*	Putative RNA–DNA helicase	+	30
	SETX	*H. sapiens*	Putative RNA–DNA helicase	+	9,95
	Setx	*M. musculus*	Putative RNA–DNA helicase	ND	31
	RTT103	*S. cerevisiae*	Termination factor	ND	21
	RPRD1B	*H. sapiens*	Termination factor	+	56
Transcription and RNA export	*HPR1*	*S. cerevisiae*	THO complex subunit	+	7,10
	MFT1	*S. cerevisiae*	THO complex subunit	+	21,64
	THP2	*S. cerevisiae*	THO complex subunit	+	21,85
	THOC1	*H. sapiens*	THO complex subunit	+	43
	DSS1 (also known as *SHFM1*)	*H. sapiens*	TREX2 complex subunit	+	80
	SRM1	*S. cerevisiae*	Ran guanyl-nucleotide exchange factor	+	21
RNA degradation	*KEM1* (also known as *XRN1*)	*S. cerevisiae*	5′–3′ exoribonuclease	+	20,129
	RRP6	*S. cerevisiae*	3′–5′ exoribonuclease	+	20,129
	Exosc3	*M. musculus*	3′–5′ exoribonuclease	ND	49,123
	Exosc10	*M. musculus*	3′–5′ exoribonuclease	ND	49
DNA topology	*TOP1*	*S. cerevisiae*	DNA topoisomerase type 1	ND	36
	TOP2	*S. cerevisiae*	DNA topoisomerase type 2	ND	36
RNA–DNA hybrid processing	*RNH1*	*S. cerevisiae*	Ribonuclease H type 1	+	14,15,20,21
	RNH201	*S. cerevisiae*	Ribonuclease H type 2 subunit	+	14,15,20,21
Nucleosome reorganization	*Spt16*	*S. cerevisiae*	FACT complex subunit	+	94
	SPT16	*H. sapiens*	FACT complex subunit	+	94
	Pob3	*S. cerevisiae*	FACT complex subunit	+	94
	SSRP1	*H. sapiens*	FACT complex subunit	+	94
DNA damage response	*BRCA1*	*H. sapiens*	Breast cancer susceptibility factor	+	80,95,108
	BRCA2	*H. sapiens*	Breast cancer susceptibility factor	+	80

AQR, aquarius; *CF1A*, cleavage factor 1A; *CLP1*, cleavage and polyadenylation factor I subunit 1; *CPF*, cleavage and polyadenylation factor; Exosc, exosome component; ND, not determined; *RNH1*, ribonuclease/angiogenin inhibitor 1; *RPRD1B*, regulation of nuclear pre-mRNA domain containing 1B; *SETX*, senataxin; *SRSF1*, serine/arginine-rich splicing factor 1; *SSRP1*, structure specific recognition protein 1; *TDRD3*, Tudor domain-containing 3; TOP, topoisomerase; TREX2, transcription and export 2; *ZNF207*, zinc finger protein 207.

CpG islands
CpG-rich regions that are usually unmethylated and localized at the 5′ end of genes, where they function as promoter elements.

specific regions where R loops accumulate (BOX 2). Two genome-wide analyses have shown by DRIP–seq in human Ntera2 cells that R loops localize at promoter and terminator regions of numerous human genes, in particular those enriched in CpG islands showing a strong GC skew[12,13]. R loops are formed immediately after the transcription start site[13] (FIG. 2a), where they protect from the action of DNA methyltransferase 3B1 (DNMT3B1), which silences genes; thus, R loops could promote transcription activation. In addition, CpG

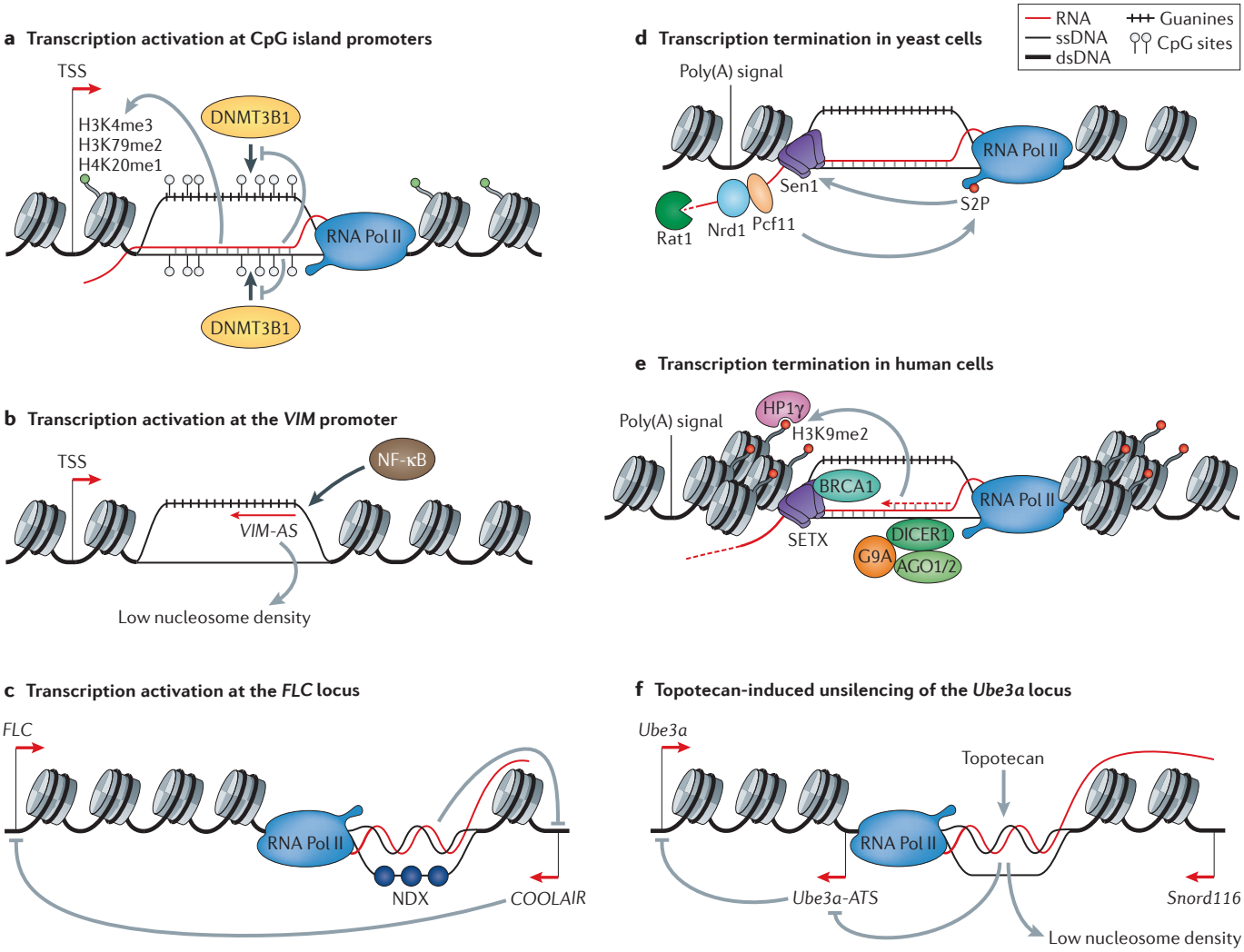


Figure 2 | **R loops as regulators of transcription.** **a** | R loops formed in human CpG island promoters, characterized by an overrepresentation of Gs in the non-transcribed DNA strand (GC skew), protect against *de novo* DNA methylation by DNA methyltransferase 3B1 (DNMT3B1). In addition, CpG island promoters show epigenetic signatures such as histone H3 lysine 4 trimethylation (H3K4me3), H3K79me2 and H4K20me1, all of which are histone marks related to transcription activation. **b** | R loops formed by antisense (AS) transcripts at promoters of specific loci, such as the human vimentin (*VIM*) gene, reduce nucleosome density and promote transcriptional activation of the sense transcript through nuclear factor-κB (NF-κB) binding. However, specific histone modifications mediating this activation have not been described. **c** | Transcription of the *FLOWERING LOCUS C* (*FLC*) locus in *Arabidopsis thaliana* is negatively regulated by the antisense transcript *COOLAIR* during prolonged cold periods. Stabilization of R loops in the promoter region of *COOLAIR* by binding of NODULIN HOMEOBOX (NDX) to the single-stranded DNA (ssDNA) of the R loop represses *COOLAIR* transcription, thus allowing *FLC* expression. **d** | A major mechanism of transcription termination in yeast cells involves cleavage and polyadenylation (poly(A)) of the mRNA followed by its degradation by the exoribonuclease ribonucleic acid-trafficking protein 1 (Rat1) and unwinding of R loops potentially formed at 3′-end regions by the helicase Sen1. In short non-coding RNAs (ncRNAs) transcribed by RNA polymerase II (RNA Pol II), Sen1 is recruited through the action of Pcf11, which binds to the termination factor Nrd1 and promotes RNA Pol II C-terminal domain (CTD) Ser2 phosphorylation (S2P). **e** | In human cells, R loops form in the 3′ region of mRNAs to promote transcription termination via senataxin (SETX) and BRCA1. In this case, the current model indicates that antisense transcription (red dashed arrow) leads to double-stranded RNA (dsRNA) that recruits the RNA interference (RNAi) machinery (for example, DICER1, Argonaute 1 (AGO1) and AGO2) and the histone methyltransferase G9A, which triggers accumulation of the repressive histone mark H3K9me2 and heterochromatin protein 1γ (HP1γ), which reinforce termination. **f** | The topoisomerase inhibitor topotecan induces unsilencing of the paternal copy of ubiquitin protein ligase E3A (*Ube3a*) in neurons by promoting R-loop formation in the terminator region of the small nucleolar RNA, C/D box 116 cluster (*Snord116*) locus. This impedes transcriptional read-through, which leads to reduced transcription of the *Ube3a-ATS* transcript, which in turn leads to *Ube3a* unsilencing. dsDNA, double-stranded DNA; TSS, transcription start site.

island promoters with high GC skew show high levels of histone H3 lysine 4 trimethylation (H3K4me3) and H4K20me1 and H3K79me2 deposition, which are chromatin marks that are linked to transcription initiation and/or elongation[12,13] (FIG. 2a).

Interestingly, R-loop-dependent regulation of mRNA transcription activation seems to be mediated by ncRNAs, at least in some particular cases. In this regard, it has been shown in human colon adenocarcinoma cell lines that antisense RNA forms an R loop at the CpG island-containing promoter of the human vimentin (*VIM*) gene[23]. This R loop activates sense transcription by enhancing chromatin opening, as inferred from a decreased nucleosome occupancy (FIG. 2b), and binding of the transcription factor nuclear factor-κB (NF-κB) to the promoter. This phenomenon is of interest because antisense transcription is commonly associated with transcriptional repression. It might be possible that opening of the DNA due to antisense transcription favours the loading of transcription factors that activate sense transcription. R loops are also detected in mouse embryonic stem cells deficient in exosome components at divergently expressed eRNAs, correlating with increased transcription of these eRNAs and decreased H3K9me2 and heterochromatin protein 1γ (HP1γ) recruitment, both of which are marks of heterochromatin and transcription repression[49].

This effect is opposite to that of the *FLOWERING LOCUS C* (*FLC*) repressor of flower development in *Arabidopsis thaliana,* the expression of which is negatively regulated at its 3′ end by the antisense transcript *COOLAIR*. Prolonged cold periods induce *COOLAIR* transcription, leading to epigenetic silencing of the *FLC* locus. However, stabilization of an R loop at the promoter region of *COOLAIR* by the homeodomain-containing protein NODULIN HOMEOBOX (NDX), which binds to the displaced ssDNA strand, represses *COOLAIR* transcription, thus allowing *FLC* expression[50] (FIG. 2c). How R loops at promoter regions activate or repress transcription is an intriguing question. Probably the epigenetic context or the different length of R loops formed in each case would help to recruit particular chromatin remodelling factors. It is likely that there are numerous unidentified ncRNAs through which R loops contribute to transcription control by modifying chromatin.

R loops in transcription termination. Transcription termination of protein-coding genes occurs when RNA Pol II reaches the poly(A) signal at the 3′ end of genes. Recently, several studies have proposed a function for R loops in this mechanism. As mentioned above, in addition to promoter regions, a subset of human genes show positive GC skew at their 3′ ends that correlates with R-loop accumulation and low DNA methylation[12], suggesting a similar mechanism of protection from DNMTs to that of the promoters. These genes are localized in regions of high gene density, opening the possibility that efficient R-loop-dependent termination avoids transcriptional read-through.

Functional evidence that R loops may play a part during transcription termination has been provided for RNA Pol II-driven genes. In eukaryotes, transcription termination of RNA Pol II-driven genes occurs by two different mechanisms, one of which relies on G-rich sequences (termination pause sites) located after the poly(A) site and involves cleavage of the nascent mRNA attached to the RNA Pol II by the exoribonucleases ribonucleic acid-trafficking protein 1 (Rat1; in yeast) and 5′–3′ exoribonuclease 2 (XRN2; in humans)[51]. In yeast, Rat1 and the Sen1 helicase cooperate to promote transcription termination by degrading the nascent transcript associated with RNA Pol II and RNA Pol I[52], and *sen1-1* mutants accumulate R loops[30]. Human SETX is also necessary for XRN2-dependent termination and prevents R-loop formation at G-rich pause sites[9]. Consistent with this view, termination of ncRNAs generated by RNA Pol II and mediated by the NRD termination complex (consisting of Nrd1, Nab3 and Sen1 in yeast) requires the Pcf11 component of cleavage factor 1A (CF1A). Pcf11 interacts with Nrd1 at the 3′ end of genes, and its absence reduces Sen1 recruitment and Ser2-phosphorylation of RNA Pol II, causing an increase in R loops and RNA Pol II pausing[53] (FIG. 2d).

According to these observations, R loops at termination sites would be required for RNA Pol II pausing, and R-loop removal by Sen1/SETX would help to release the RNA molecule, thus promoting efficient transcription termination[9]. However, Sen1 is not able to unwind a short RNA–DNA duplex generated using a DNA oligonucleotide complementary to the nascent transcript *in vitro*[54]. We probably need to establish more physiological systems to assay whether Sen1/SETX removes R loops in a physiological supercoiled and chromatin context to clarify the mechanism of action of Sen1 *in vivo*, but in principle, we cannot formally discard other roles of Sen1/SETX. Indeed, human SETX has also been implicated in the DDR[55], and the observation that the human termination factor regulation of nuclear pre-mRNA domain-containing protein 1B (RPRD1B) has distinct roles in DNA repair and transcription termination, in which it prevents R-loop accumulation[56], adds further support for a dual role for particular transcription termination factors in R-loop prevention and the DDR.

Transcription termination may additionally be regulated by ncRNAs together with chromatin modifications. In this regard, R loops formed at the G-rich termination pause site of the human β-actin (*ACTB*) gene have been proposed to induce antisense transcription and cause the formation of dsRNA that is able to recruit the RNA interference (RNAi) silencing machinery. As a consequence, repressive heterochromatic sites are formed by H3K9me2 deposition and HP1γ recruitment, which would enhance RNA Pol II pausing before termination[22] (FIG. 2e). A number of interesting questions arise from this study, including how this heterochromatin affects transcription termination or whether factors such as SETX play a part in dsRNA formation after R-loop dissolution. An involvement of ncRNAs in R loops has also been described for mouse ubiquitin protein ligase

GC skew
Asymmetry in the distribution of Gs and Cs between DNA strands, with an overrepresentation of Gs in the non-transcribed DNA strand.

Heterochromatin
A chromosomal region with highly compacted chromatin that is more refractory to the action of enzymes, has a general repressive action on gene transcription and replicates late in the cell cycle.

Homeodomain
A DNA-binding domain that is characteristic of homeobox proteins involved in transcription regulation. It consists of a 60-amino-acid helix–turn–helix structure with three α-helices connected by loop regions.

RNA interference
(RNAi). A mechanism of gene silencing that relies on short non-coding RNAs that have the ability to repress chromatin with the help of additional ancillary factors.

E3A (*Ube3a*), which encodes a brain-specific ubiquitin E3 ligase. The paternal copy of *Ube3a* is epigenetically silenced in neurons by the antisense RNA *Ube3a-ATS*, which is produced by read-through of the small nucleolar RNA, C/D box 116 cluster (*Snord116*) locus located downstream of *Ube3a*. Interestingly, the TOP1 inhibitor topotecan is able to unsilence *Ube3a* by reducing the production of *Ube3a-ATS*[57]; the mechanism involves increased R-loop formation at the termination region of the *Snord116* locus that promotes efficient transcription termination and less *Ube3a-ATS* transcription[58] (FIG. 2f). In this case, R-loop-induced transcription termination is accompanied by nucleosome depletion, similarly to the R loop at the *VIM* promoter that activates its transcription[23].

In summary, mRNA transcription termination can also be regulated by R loops, which could promote termination by slowing down the advance of the RNA polymerase and allowing the recruitment of termination factors such as Sen1/SETX.

R loops in genome dynamics

Even though R loops may work as key intermediates in natural processes, they have a strong capacity to modulate genome structure. They are a source of DNA damage and replication stress and thus constitute an important source of genome instability and genetic variation.

Hypermutation. Despite the number of factors identified as being able to preserve low levels of transcription- and R-loop-dependent genome instability (TABLE 1), their possible effect on mutation rates has been analysed in only a few cases. ssDNA is more mutagenic than dsDNA because it is more accessible to the action of DNA-damaging agents[59], so an increase in R loops is expected to lead to a hypermutation phenotype. Indeed, a computational analysis of nucleotide substitution rates along human genes showed higher rates of cytosine deamination in the non-template strand at the start of the genes, suggesting an ssDNA conformation that could be due to R-loop formation[60].

ssDNA in R loops may also be a target for the action of specific mutagenic agents or enzymes, such as human activation-induced cytidine deaminase (AID), which is involved in Ig CSR and somatic hypermutation in mammalian activated B cells[61] (BOX 1). AID action has also been reported in non-Ig genes, known as AID off-target genes, where AID initiates recurrent mutations or DNA double-strand breaks (DSBs). AID off-target sites correlate with highly transcribed super-enhancers and regulatory clusters where convergent transcription occurs as a result of antisense transcription[62,63]. Indeed, when AID is heterologously expressed in R-loop-accumulating yeast THO mutants, it increases mutations in transcribed genes in a pattern that is analogous to that of the Ig locus of activated human B cells[64]. Therefore, R-loop formation can induce mutagenesis in the exposed ssDNA either spontaneously or through mutagenic agents or enzymes such as AID.

In addition, R loops could lead to hypermutation by promoting mutagenic DNA replication. They can be used as primers to initiate non-canonical replication in *E. coli*[65] and in the rDNA region of yeast cells lacking Top1 and RNase H enzymes in an origin-independent manner[37]. However, little is known about the fidelity of replication primed by R loops. In starved *E. coli* cells, repair of DSBs uses error-prone DNA polymerases, leading to stress-induced mutations (SIMs)[66]. It has been recently proposed that R loops underlie this mechanism, priming replication that, when encountering a nick, generates the DSB that initiates a SIM[67]. Therefore, it is also possible that origin-independent R-loop-dependent DNA replication, whether or not mediated by a DNA break (see below), might be mutagenic. Indeed, break-induced replication (BIR), which is one possible mechanism of repair of breaks putatively induced by R-loops, is mutagenic in yeast[68].

DNA-break-mediated genome instability. R-loop-mediated genome instability is manifested mainly as a form of transcription-associated recombination[6] that arises from DNA breaks. However, we still know little about how R loops lead to these DNA breaks. Understanding the steps by which R loops may evolve into ssDNA gaps or DSBs is crucial to understand how they compromise genome integrity.

R-loop-mediated DNA breaks, recombination and chromosome rearrangements and losses have been demonstrated in yeast and human cells that have increased R-loop occurrences owing to deficiency in different RNA and DNA metabolism factors (TABLE 1). However, replication stress can be caused by DNA damage and by replication-fork stalling due to various types of obstacles and DNA-damage checkpoint failures, and is a major source of spontaneous DNA breaks, recombination and genome rearrangements[69]. Therefore, the capacity of R loops to stall replication-fork progression may be a major cause of DNA breaks (FIG. 3). The interference of R loops with replication progression has been shown from bacteria to humans. In *E. coli*, an rDNA operon transcribed head-on into the direction of replication-fork progression[70] and a transcribed R-loop-forming Ig S region cause replication fork stalling and chromosomal rearrangements[71]. In yeast, interference between replication and transcription has been shown at a GC-rich *lacZ* transgene in THO mutants[72] or by genome-wide analysis of replication in cells lacking THO, Sen1 and the RBP Npl3 (REFS 73–75). In human cells, TOP1 depletion causes transcription to interfere with replication in an R-loop-dependent manner, as detected by DNA combing[38]. Finally, a link between replication-fork stalling and the occurrence of DNA breaks has been shown in *C. elegans thoc-2* mutants, in which pre-meiotic replication is impaired and partly recovered by RNase H microinjection[42]. Therefore, R loops may be a key element that is responsible for transcription–replication collisions as a major source of DNA-break-mediated genome instability. This is supported by the fact that the common fragile site FRA3B, which contains the 1.5 Mb-long fragile histidine triad (*FHIT*) gene, is associated with transcription–replication collisions and R-loop accumulation[76]. However, although R loops may constitute blocks to replication forks, additional structural features may

Replication stress
Any condition in which replication progression slows down and/or stalls, commonly leading to genome instability.

Break-induced replication
(BIR). A mechanism of recombinational repair in which a one-ended double-strand break invades a homologous DNA sequence that is used as template for DNA synthesis to complete repair.

DNA combing
A technique used to produce stretched DNA fibres for multiple applications, including the study of DNA replication by immune detection of modified nucleotides.

Common fragile site
A specific chromosome region that has gaps or constrictions that are visible under the microscope and that tend to break on exposure to replication stress.

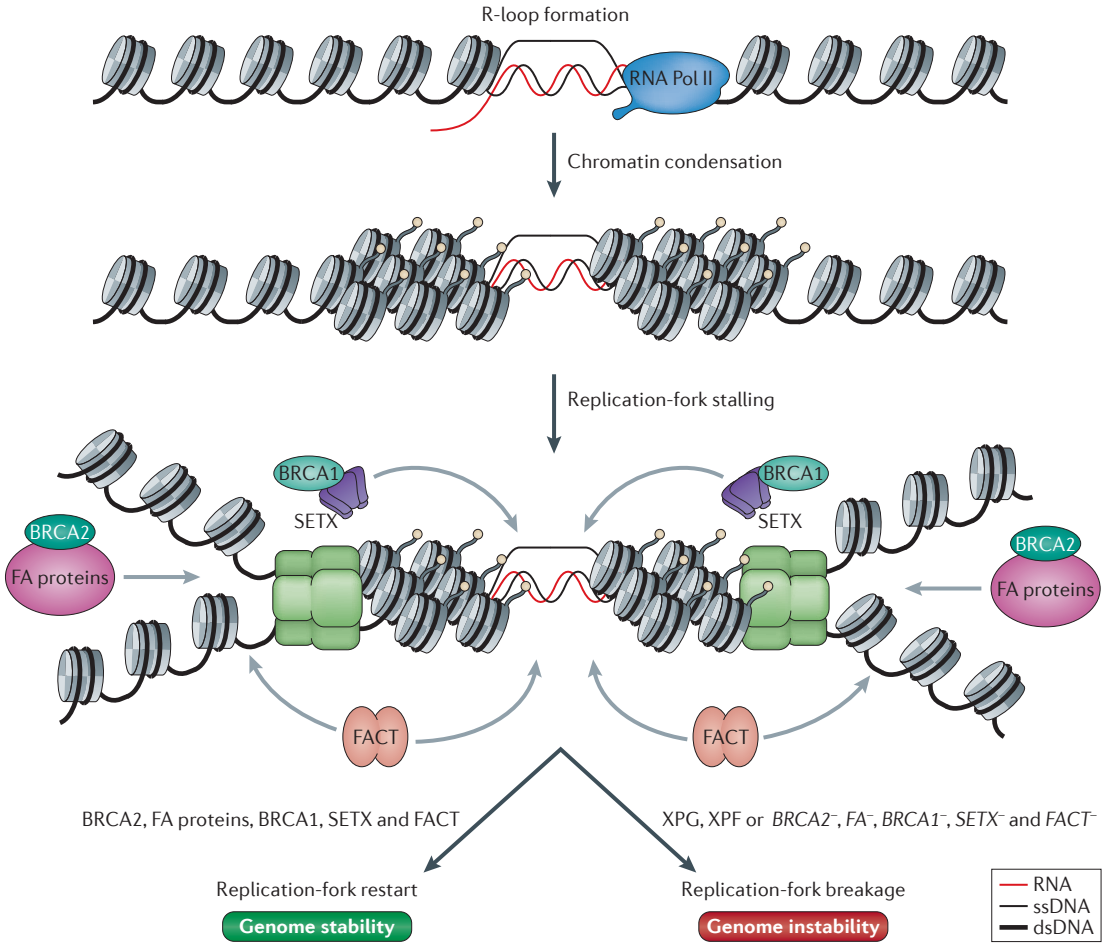


Figure 3 | **A model to explain replication-mediated R-loop-induced genome instability.** R loops are an obstacle for replication progression. They trigger chromatin condensation and heterochromatin formation, which creates an important barrier to the progression of the replication fork. In this scenario, functions that protect stalled replication forks from collapse or that are required for the progression of replication forks through an obstacle would be necessary for replication-fork restart to guarantee genome stability. This includes BRCA2 and potentially other Fanconi anaemia (FA) proteins, chromatin reorganization factors that are required for replication progression through transcribed chromatin (such as the FACT complex), and probably senataxin (SETX), which would allow replication progression through R-loop-accumulating regions in cooperation with BRCA1. As a consequence, failures in these functions will cause the stalled fork to be retained together with the unresolved R loop, leading indirectly to an accumulation of R loops in cells that are deficient in these factors and, furthermore, to replication-fork breakage and genome instability that might be mediated by nucleases such as the nucleotide excision repair endonucleases xeroderma pigmentosum group G (XPG) and XPF. dsDNA, double-stranded DNA; RNA Pol II, RNA polymerase II; ssDNA, single-stranded DNA.

also contribute to such stalling. Thus, because cellular deficiencies that cause an excess of co-transcriptional R loops impair transcription[35,77], we cannot exclude the possibility that an arrested RNA Pol II remains at the site of the R loop, where it contributes to the block of an advancing replisome.

The mechanism by which an R-loop-induced transcription–replication collision leads to a DNA break is in any case unclear. In *E. coli*, it has been proposed that when the replisome co-directionally encounters a backtracked RNA polymerase, the R loop would be used to prime DNA synthesis of the leading strand, while the 5′ end of the RNA moiety would form a nick, which would be converted to a DSB in the next round of replication[78]. But it is also possible that breaks are induced by specific endonucleases regardless of replication. Indeed, the nucleotide excision repair (NER) nucleases xeroderma pigmentosum group G (XPG; also known as ERCC5) and XPF (also known as ERCC4) are required for the processing into DSBs of R loops accumulated in the absence of AQR or SETX, or in the presence of the TOP1 inhibitor camptothecin[33]. However, R-loop processing by the NER machinery could also cause ssDNA breaks that are converted into DSBs after replication. Interestingly, it has been recently reported that transcription-blocking lesions cause spliceosome displacement and subsequent formation of R loops at damaged sites[79]. This causes a specific, non-canonical ataxia telangiectasia mutated (ATM) checkpoint activation as a key event in the DDR of non-replicating cells[79].

Backtracked RNA polymerase
An intermediate state in which an arrested RNA polymerase moves back to allow cleavage of the last ribonucleotide incorporated into the nascent RNA, thus allowing transcription resumption.

Nucleotide excision repair
(NER). A conserved DNA repair pathway that recognizes adducts and repairs them by excision of a short oligonucleotide containing the damage.

The recent finding that the BRCA2 and BRCA1 tumour suppressor and DNA repair genes are involved in R-loop homeostasis opens new perspectives into our understanding of the way R loops can cause replication stress and DNA breaks[80]. Although BRCA2 is a recombinational DSB repair factor, it is also a member of the Fanconi anaemia pathway of repair of interstrand crosslinks and is required for the protection of stalled replication forks to prevent their collapse[81]. It is possible that BRCA2 prevents R-loop accumulation by allowing the restart of replication forks stalled at R loops (FIG. 3), but little is known about how DNA repair mechanisms respond to R-loop-mediated DNA damage. An anti-R-loop mechanism operating at damaged transcribed sites has been recently proposed using two human cell lines[82]. It would involve a transient recruitment of RBPs — including scaffold attachment factor A (SAFA; also known as hnRNPU), FUS (also known as TLS) and 68 kDa TATA-binding protein-associated factor (TAFII68; also known as TAF15 and TAF2N) — to damaged sites containing R loops. Then, after phosphorylation by the DNA damage-induced kinases ATM, ataxia telangiectasia and Rad3-related protein (ATR) and DNA-dependent protein kinase (DNA-PK), these RBPs would be excluded. However, the specific role of these RBPs and their possible connection with DNA repair proteins that might be recruited to R-loop-containing sites needs to be investigated further.

R loops in telomere dynamics. Eukaryotic telomeres are transcribed into non-coding telomeric-repeat-containing RNAs (TERRAs). These RNAs form RNA–DNA hybrids in wild-type yeast but accumulate further in *rat1-1* mutants[83] or in strains depleted of RNase H or THO, the latter of which is recruited to yeast telomeres[84,85]. In telomerase-negative yeast cells, R loops are relevant in telomere dynamics. Loss of telomerase leads to progressive shortening of telomeres and premature cellular senescence, whereas R-loop accumulation in cells additionally lacking RNase H or THO leads to telomere lengthening and delayed senescence. The mechanism is dependent on recombination because additional inactivation of the recombination protein Rad52 causes telomere shortening and accelerated senescence[84]. Thus, R loops could provide an alternative pathway for telomere lengthening in the absence of telomerase owing to their ability to trigger recombination events[86]. Indeed, in human cancer cells with telomerase-independent recombination-mediated alternative lengthening of telomeres, the levels of RNA–DNA-hybrid-forming TERRAs are increased relative to the levels in telomerase-dependent cells, and RNase H1 overexpression specifically reduces telomere recombination[87]. Telomeres thus represent specific loci in which cells take advantage of the capacity of R loops to promote recombination.

R-loop-driven chromatin changes

Recent reports have indicated an association between R loops and specific chromatin modifications, which has changed our view of the way R loops modulate genome function and dynamics. Mounting evidence indicates that chromatin contributes to R-loop-mediated genome instability and that R loops are linked to heterochromatin and chromatin condensation marks in yeast, *C. elegans* and human cells. Related to this, it has been shown that the formation of centromeric heterochromatin in *S. pombe* is mediated by RNA–DNA hybrids formed by association of ncRNAs with chromatin, a phenomenon that requires the RNAi machinery[88]. Notably, yeast and human cells depleted of the THO complex or SETX show an increase of H3S10P, a mark associated with condensed chromosomes during mitosis and also with transcription activation[89–91]. H3S10P is accumulated genome-wide in yeast THO mutants, as detected by ChIP–chip, in an R-loop-dependent manner[10]. This, together with the fact that pericentromeric chromatin is extended in these cells, that the H3K9me2 heterochromatin mark is also accumulated in mitotic and meiotic germ line cells of *C. elegans thoc-2* mutants, and that chromatin is more condensed, as shown by DAPI-staining in these worms[10], suggests that R loops accumulated in the absence of THO or SETX lead to a more closed chromatin.

Interestingly, it is known that replication cannot proceed easily through condensed chromatin, which is the latest to be replicated (FIG. 3). A number of human fragile sites have been shown to undergo premature chromatin condensation, causing DNA breakage and fragility under replicative stress, probably due to failure in completing replication[92]. Importantly, Friedreich ataxia (FRDA) and fragile X syndrome (FXS) occur as a result of repeat expansions in the frataxin (*FXN*) and fragile X mental retardation 1 (*FMR1*) genes, respectively; this leads to gene silencing through H3K9me2 deposition on the expanded regions, which thus become fragile. Such expansions accumulate R loops, providing a new link among heterochromatin, R loops and replication-dependent fragility[93]. The repressive H3K9me2 mark is also associated with R loops at G-rich transcription pause sites of the highly transcribed human β-actin gene[22]. Finally, the FACT (facilitates chromatin transcription) chromatin reorganizer complex of yeast and human HeLa and MRC-5 cells is required for replication of DNA regions that are being transcribed, and FACT-depleted cells accumulate R loops[94], indicating that failures in chromatin reorganization may strongly contribute to replication stress and chromosome instability triggered by R loops.

Despite the evidence showing a connection of R loops with chromatin condensation or heterochromatin and consequently with a closed chromatin state, this may not be a general rule — in other specific cases, R loops correlate with increased chromatin decondensation and accessibility[13,23,58]. These results are consistent with transcription being commonly associated with open chromatin, and with the observation that *in vitro*-generated RNA–DNA hybrids remain nucleosome-free. It is unclear why in some cases R loops are detected with condensed chromatin but with open chromatin in others. One possibility is that the core of the R loop could exclude the nucleosome while the immediate neighbouring region undergoes chromatin compaction. Alternatively, the difference in chromatin impact may imply that there are different types of R loops, as discussed below. Finally, we

Fanconi anaemia pathway
A DNA repair pathway that works on replication forks stalled at interstrand crosslinks and other lesions that block replication.

Table 2 | **Genes related to R-loop metabolism that can cause human diseases if dysfunctional**

Gene	Disease	Cause	Refs
SETX	Ataxia-ocular apraxia type 2 (AOA2) and amyotrophic lateral sclerosis type 4 (ALS4)	Mutations in the RNA–DNA helicase *SETX*	102,103
FXN	Friedreich ataxia (FRDA)	Expansion of GAA repeats in *FXN* gene promotes R-loop formation, H3K9me2 and decreased *FXN* expression	93,98
FMR1	Fragile X syndrome (FXS) and fragile X-associated tremor/ataxia syndrome (FXTAS)	Expansion of CGG repeats in *FMR1* gene promotes R-loop formation, H3K9me2 and decreased *FMR1* expression	93,99, 100
C9orf72	Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD)	Expansion of GGGGCC repeats causes R-loop formation and accumulation of aborted transcripts	101
BRCA1	Cancer	Genome instability caused by R-loop accumulation in BRCA1-deficient cells	80,95, 108
BRCA2	Cancer and Fanconi anaemia (FA)	Genome instability caused by R-loop accumulation in BRCA2-deficient cells	80
FIP1L1	Cancer	Genome instability caused by R-loop accumulation in FIP1L1-deficient cells inferred by the yeast mutant *fip1Δ*	21
BRE1	Cancer	Genome instability caused by R-loop accumulation in BRE1-deficient cells	111
SRSF1	Cancer	Deregulation of cancer-associated genes due to *SRSF1* overexpression	112
ORF57	Kaposi sarcoma-associated herpesvirus (KSHV)	Sequestration of human TREX complex by ORF57 causes R-loop formation and DNA damage	113

C9orf72, chromosome 9 open reading frame 72; FIP1L1, factor interacting with PAPOLA and CPSF1; *FMR1*, fragile X mental retardation 1; *FXN*, frataxin; H3K9me2, histone H3 lysine 9 dimethylation; *SETX*, senataxin; *SRSF1*, serine/arginine-rich splicing factor 1.

cannot discard the possibility that chromatin reorganization might result from the effect of R loops on different DNA metabolic processes, including transcription elongation impairment, transcription termination, replication stalling or R-loop-mediated DNA damage or breakage. However, the fact that R loops detected together with H3S10P and H3K9me2 correlate with DNA breaks and chromosome fragility[10,93], and the fact that the BRCA1 DSB repair factor is recruited together with SETX to the termination pause site of the human β-actin gene[95] favour a model in which R-loop-induced closed chromatin is responsible for replication-fork breakage and genome instability (FIG. 3).

R loops in human disease

R loops have been associated with a number of human neurological disorders and may be responsible for replication stress and genome instability in cancer cells (TABLE 2).

Neurodegenerative disorders. A direct relationship between R loops and disease is becoming established for multiple neurological disorders caused by expansion of trinucleotide repeats. Thus, transcription of CTG repeats leads to R-loop formation both *in vitro* and *in vivo*, which stimulates repeat instability in bacteria and human cells[96], and R loops form during *in vitro* transcription of several disease-associated trinucleotide repeats[97]. FRDA is the most common inherited ataxia and is due to the expansion of an unstable GAA repeat in the first intron of the *FXN* gene. Studies *in vitro* and in bacteria have shown that extensive RNA–DNA hybrids form on GAA templates, leading to RNA polymerase arrest, which is responsible for the decreased FXN expression[98]. FXS and fragile X-associated tremor/ataxia syndrome (FXTAS) are neurodevelopmental diseases caused by the extension of the CGG repeat in the 5′ untranslated region (5′ UTR) of the *FMR1* gene, which leads to reduced *FMR1* expression. Two recent studies showed that *FMR1* gene silencing is mediated by the *FMR1* mRNA, which forms R loops at expanded CGG repeats of the *FMR1* gene during transcription, as determined by DRIP and bisulfite mutagenesis in human dermal fibroblasts[99] and by chromatin isolation by RNA purification (ChIRP) in FXS cell lines[100]. Furthermore, adding a small molecule that inhibits the formation of RNA hairpins at GG repeats suppresses R-loop formation and silencing. Interestingly, *FMR1* R loops induce a decrease in H3K4me2 levels and an increase in H3K9me2 levels[100]. These data, together with the previously discussed observations that R loops form at expanded GAA and CGG repeats in the *FXN* and *FMR1* genes, respectively, and that they colocalize with the H3K9me2 repressive mark, which is responsible for gene silencing[93], suggest that the R loop is a strong contributor to the diseases. In this sense, R loops are also associated with the hexanucleotide repeat expansion GGGGCC in the chromosome 9 open reading frame 72 (*C9orf72*) gene, which is associated with amyotrophic lateral sclerosis (ALS) and frontotemporal dementia; this repeat expansion leads to the accumulation of aborted transcripts in the repeat region and nucleolar stress[101].

The loss of function of R-loop-removing enzymatic activities may lead to different neurodegenerative disorders. This is the case of mutations in the human helicase SETX, which are associated with ataxia-ocular apraxia type 2 (AOA2)[102] and with ALS type 4 (ALS4)[103]. Studies in neuronal cells demonstrated that SETX promotes neuronal differentiation through modulation of fibroblast growth factor 8 (FGF8) expression levels[104], providing an explanation for the clinical manifestation of AOA2 — but not for ALS4. However, it is currently not clear whether AOA2 is due to R-loop accumulation

itself because *Setx*$^{-/-}$ mice, as well as mice depleted of *Atm* (causing ataxia telangiectasia), *Tdp1* (causing spinocerebellar ataxia with axonal neuropathy 1) and *Aptx* (causing AOA1), show increased levels of R loops in germ cells but not in neuronal cells[105].

Cancer. R loops lead to genome instability and replication stress, which are hallmarks of pre-tumoural and tumoural cells[106,107], so R loops are potential drivers of cancer. Indeed, a connection between R loops and cancer has been provided for the DSB repair factors BRCA1 and BRCA2, which contribute to the prevention of R-loop accumulation. Both *BRCA1*$^{-/-}$ cells and *BRCA2*$^{-/-}$ cells show increased R-loop accumulation and DSBs that are partially reduced by RNase H1 overexpression[80,108]. A good example of a tumorigenic process that can be initiated by R loops is provided by Burkitt lymphoma, which is generated by translocations between the *MYC* proto-oncogene and the Ig S regions. AID action is essential for the generation of DSBs at both the GC-rich S region and the transcribed *Myc* region of mice[109], which can potentially form R loops. AID also induces translocations involving heterologous human S and *MYC* transcribed sequences in R-loop-accumulating yeast THO mutants[110]. In addition, TOP3B — which is recruited to highly transcribed genes by Tudor-domain-containing protein 3 (TDRD3) — reduces R-loop accumulation in the human *MYC* promoter by relaxing negative supercoiling, and *Myc–Igh* translocations are increased in *Tdrd3*-null mice[39].

Examples of genes that are mutated in cancer cells and that have a role in preventing R-loop accumulation include factor interacting with PAPOLA and CPSF1 (*FIP1L1*), which encodes a cleavage and poly(A) factor and is affected by translocations between its amino-terminal domain and platelet-derived growth factor receptor-α (PDGFRα) in 10–20% of eosinophilic leukaemia cases[21], and the tumour suppressor *BRE1* (also known as *RNF20*), which is a histone H2B ubiquitin ligase[111]. Additionally, the splicing factor SRSF1 is overexpressed in many types of cancer[44] and induces senescence through p53 in human and mouse fibroblasts[112]; the complexity of phenotypes resulting from altered SRSF1 expression suggests that its connection to cancer is not just due to its role in R-loop prevention. In any case, it has not been tested whether the tumorigenic effects of alterations to these factors are specifically related to R-loop accumulation. Finally, cells infected with the cancer-causing Kaposi's-sarcoma-associated herpesvirus (KSHV) show increased DNA damage and R loops owing to the sequestration of the human transcription and export (TREX) complex by the ORF57 protein of KSHV[113]. Even though we cannot establish a cause–effect relationship in any of these examples, they are in consistent with the idea that R loops are a major source of DNA damage in cancer cells.

Conclusions and perspectives

Research over the past two decades has provided evidence that co-transcriptional R loops form in the genome more often than previously foreseen and may have positive regulatory roles in cell homeostasis — but at the same time they are a source of genome instability and a potential contributor to cancer and disease. This dual effect raises the question of whether there are 'good' and 'bad' R loops, and a notable part of the literature so far suggests the existence of these two different R-loop types. If this is the case, a challenge will be to decipher what determines the positive versus negative effects of R loops; potential explanations could be their genomic context, the stage of the cell cycle at which they form, their subnuclear localization, their size, a difference between newly formed or short-lived R loops and persistent R loops, or the way they can modulate chromatin structure. It is possible that a minimum R-loop size is required to make them sufficiently stable and persistent enough to compromise genome stability and function. R loops with a role in transcription regulation could have shorter RNA–DNA hybrids and form at low levels that are still sufficient to fulfil a relevant role in transcription with a lower impact on genome integrity. However, they could still make those DNA regions where they form hotspots of transcription-associated mutation and transcription-associated recombination[6]. Support for the idea that any R loop has the potential to compromise genome integrity is provided by the observation that the DDR factor and tumour suppressor BRCA1 is recruited to transcription termination sites together with SETX[95]. This suggests that those R loops with a natural function in transcription termination can indeed also compromise genome stability. It would be interesting to know whether this is related to the observation that yeast Sen1 is found together with replication forks at transcribed DNA regions[75]. The observations that BRCA1 and BRCA2 help to remove R loops support the idea that R loops are indeed formed often as a source of genome instability and a potential contributor to cancer[80], and that the DDR contributes to their removal.

In the past several years, the identification of new factors with a role in R-loop prevention and processing has increased considerably, but we have advanced little in our knowledge of the structure of R loops. Most recent analyses have relied on a combination of R-loop detection by the monoclonal S9.6 antibody and on the suppression of R loops or their derived phenotypes by RNase H. We now need to investigate various aspects of R loops, such as their size *in vivo* (a few studies have reported sizes ranging from less than 100 bp to 2 kb at different sites[8,13,114–117]); how long they last (newly formed or short-lived R loops may have a different structure from long-lasting and persistent R loops); the phase of the cell cycle in which they accumulate; and how they alter chromatin structure. We know little about how R loops lead to the replication-fork stalling and DNA breaks that are responsible for genome instability and cancer; it is also unclear whether particular types of DNA damage may enhance R-loop formation. We are only just starting to identify the factors that are responsible for R-loop processing and for the repair of the DNA damage that they induce. Our ability to provide answers to all of these questions will certainly improve our understanding of the physiological role of R loops as key players in genome dynamics and function — and of how they become pathological.

1. Roberts, R. W. & Crothers, D. M. Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. *Science* **258**, 1463–1466 (1992).
2. Roy, D., Zhang, Z., Lu, Z., Hsieh, C. L. & Lieber, M. R. 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**, 146–159 (2010).
3. Roy, D. & Lieber, M. R. 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**, 3124–3133 (2009).
4. Duquette, M. L., Handa, P., Vincent, J. A., Taylor, A. F. & Maizels, N. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev.* **18**, 1618–1629 (2004).
5. Aguilera, A. & Garcia-Muse, T. R loops: from transcription byproducts to threats to genome stability. *Mol. Cell* **46**, 115–124 (2012).
6. Aguilera, A. The connection between transcription and genomic instability. *EMBO J.* **21**, 195–201 (2002).
7. Huertas, P. & Aguilera, A. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol. Cell* **12**, 711–721 (2003).
This paper provides the first demonstration that R loops cause genome instability and that mRNP biogenesis factors prevent R-loop formation. It shows that R loops accumulate in yeast mutants lacking the Hpr1 subunit of the THO complex and that hyper-recombination in these mutants is partially dependent on the nascent mRNA and R-loop accumulation.
8. Li, X. & Manley, J. L. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* **122**, 365–378 (2005).
This is the first evidence in vertebrate cells that depletion of a splicing factor causes genome instability in the form of chromosomal rearrangements and mutagenesis mediated by R loops, as shown for chicken DT40 and human HeLa cells depleted of the SRSF1 protein.
9. Skourti-Stathaki, K., Proudfoot, N. J. & Gromak, N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell* **42**, 794–805 (2011).
This paper connects RNA:DNA helicase SETX deficiency with R-loop accumulation in human cells. The authors propose that R loops form at G-rich termination pausing sites and are resolved by SETX, which would promote 3′ mRNA degradation by XRN2 and transcription termination.
10. Castellano-Pozo, M. *et al.* R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. *Mol. Cell* **52**, 583–590 (2013).
This work shows a functional link between R loops and the H3S10P chromatin condensation mark. Yeast, worm and human cells depleted of THO subunits show increased levels of H3S10P, which is suppressed by RNase H1 overexpression. Enrichment of the H3K9me2 heterochromatic mark is also shown in *C. elegans*. It is proposed that chromatin condensation linked to R loops is a strong barrier to replication progression as a major source of replication stress and genome instability.
11. Wongsurawat, T., Jenjaroenpun, P., Kwoh, C. K. & Kuznetsov, V. Quantitative model of R-loop forming structures reveals a novel level of RNA-DNA interactome complexity. *Nucleic Acids Res.* **40**, e16 (2012).
12. Ginno, P. A., Lim, Y. W., Lott, P. L., 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.* **23**, 1590–1600 (2013).
13. Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I. & Chedin, F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell* **45**, 814–825 (2012).
This is the first genome-wide analysis of R-loop locations. It shows that unmethylated human CpG island promoters are characterized by a positive GC skew and by the formation of R loops, which protect from *de novo* DNA methylation.
14. Chan, Y. A. *et al.* Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-chip. *PLoS Genet.* **10**, e1004288 (2014).
15. El Hage, A., 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.* **10**, e1004716 (2014).
16. Skourti-Stathaki, K. & Proudfoot, N. J. A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev.* **28**, 1384–1396 (2014).
17. Sollier, J. & Cimprich, K. A. Breaking bad: R-loops and genome integrity. *Trends Cell Biol.* http://dx.doi.org/10.1016/j.tcb.2015.05.003 (2015).
18. Cerritelli, S. M. & Crouch, R. J. Ribonuclease H: the enzymes in eukaryotes. *FEBS J.* **276**, 1494–1505 (2009).
19. Cerritelli, S. M. *et al.* Failure to produce mitochondrial DNA results in embryonic lethality in Rnaseh1 null mice. *Mol. Cell* **11**, 807–815 (2003).
20. Wahba, L., Amon, J. D., Koshland, D. & Vuica-Ross, M. RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Mol. Cell* **44**, 978–988 (2011).
21. Stirling, P. C. *et al.* R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. *Genes Dev.* **26**, 163–175 (2012).
22. Skourti-Stathaki, K., Kamieniarz-Gdula, K. & Proudfoot, N. J. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature* **516**, 436–439 (2014).
This paper shows a connection between R loops and chromatin repressive marks in the terminator region of human genes. R loops form at these regions and prime antisense transcription, generating dsRNA that seems to be processed by the RNAi machinery. This event triggers H3K9me2 deposition and heterochromatin formation that facilitates RNA Pol II pausing prior to transcription termination.
23. Boque-Sastre, R. *et al.* Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. *Proc. Natl Acad. Sci. USA* **112**, 5785–5790 (2015).
24. Reijns, M. A. *et al.* Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. *Cell* **149**, 1008–1022 (2012).
25. Hong, X., Cadwell, G. W. & Kogoma, T. *Escherichia coli* RecG and RecA proteins in R-loop formation. *EMBO J.* **14**, 2385–2392 (1995).
26. Harinarayanan, R. & Gowrishankar, J. Host factor titration by chromosomal R-loops as a mechanism for runaway plasmid replication in transcription termination-defective mutants of *Escherichia coli*. *J. Mol. Biol.* **332**, 31–46 (2003).
27. Boule, J. B. & Zakian, V. A. The yeast Pif1p DNA helicase preferentially unwinds RNA DNA substrates. *Nucleic Acids Res.* **35**, 5809–5818 (2007).
28. Chakraborty, P. & Grosse, F. Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA Repair (Amst.)* **10**, 654–665 (2011).
29. Kim, H. D., Choe, J. & Seo, Y. S. The sen1^{+} gene of *Schizosaccharomyces pombe*, a homologue of budding yeast SEN1, encodes an RNA and DNA helicase. *Biochemistry* **38**, 14697–14710 (1999).
30. Mischo, H. E. *et al.* Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol. Cell* **41**, 21–32 (2011).
31. Becherel, O. J. *et al.* Senataxin plays an essential role with DNA damage response proteins in meiotic recombination and gene silencing. *PLoS Genet.* **9**, e1003435 (2013).
32. De, I. *et al.* The RNA helicase Aquarius exhibits structural adaptations mediating its recruitment to spliceosomes. *Nat. Struct. Mol. Biol.* **22**, 138–144 (2015).
33. Sollier, J. *et al.* Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol. Cell* **56**, 777–785 (2014).
This paper provides a mechanism by which R loops accumulated after depletion of human RNA-processing factors such as AQR may be processed into DSBs by the NER nucleases XPG or XPF. Interestingly, this phenomenon seems to be specific to the transcription-coupled NER sub-pathway.
34. Paulsen, R. D. *et al.* A genome-wide siRNA screen reveals diverse cellular processes and pathways that mediate genome stability. *Mol. Cell* **35**, 228–239 (2009).
35. Drolet, M. *et al.* 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* **92**, 3526–3530 (1995).
36. El Hage, A., French, S. L., Beyer, A. L. & Tollervey, D. Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev.* **24**, 1546–1558 (2010).
This work shows that R loops accumulate in the rDNA of yeast Top1 and Top2 mutants and lead to RNA Pol I transcription impairment, truncated rRNA transcripts and reduced rRNA synthesis. In the absence of TOPs, RNase H activity aids in restoring RNA Pol I transcription efficiency.
37. Stuckey, R., Garcia-Rodriguez, N., Aguilera, A. & Wellinger, R. E. Role for RNA:DNA hybrids in origin-independent replication priming in a eukaryotic system. *Proc. Natl Acad. Sci. USA* **112**, 5779–5784 (2015).
38. Tuduri, S. *et al.* Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nat. Cell Biol.* **11**, 1315–1324 (2009).
This paper shows that TOP1-deficient cells accumulate DNA breaks at transcribed genes all over the genome and have slower replication-fork progression owing to the formation of R loops, suggesting a role for TOP1 in avoiding conflicts between replication and transcription.
39. Yang, Y. *et al.* Arginine methylation facilitates the recruitment of TOP3B to chromatin to prevent R loop accumulation. *Mol. Cell* **53**, 484–497 (2014).
This work shows that human TOP3B reduces both negative supercoiling and R-loop formation by binding to highly transcribed genes through TDRD3, which recognizes histone and RNA Pol II carboxy-terminal domain methylation. Interestingly, TDRD3-deficient cells accumulate R loops at the *MYC* locus, and *Tdrd3*-null mice show increased *Myc–Igh* translocations, which are common in Burkitt lymphoma.
40. Wilson-Sali, T. & Hsieh, T. S. Preferential cleavage of plasmid-based R-loops and D-loops by *Drosophila* topoisomerase IIIβ. *Proc. Natl Acad. Sci. USA* **99**, 7974–7979 (2002).
41. Li, M., Pokharel, S., Wang, J. T., Xu, X. & Liu, Y. RECQ5-dependent SUMOylation of DNA topoisomerase I prevents transcription-associated genome instability. *Nat. Commun.* **6**, 6720 (2015).
42. Castellano-Pozo, M., Garcia-Muse, T. & Aguilera, A. R-loops cause replication impairment and genome instability during meiosis. *EMBO Rep.* **13**, 923–929 (2012).
43. Dominguez-Sanchez, M. S., Barroso, S., Gomez-Gonzalez, B., Luna, R. & Aguilera, A. Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. *PLoS Genet.* **7**, e1002386 (2011).
44. Li, X., Niu, T. & Manley, J. L. The RNA binding protein RNPS1 alleviates ASF/SF2 depletion-induced genomic instability. *RNA* **13**, 2108–2115 (2007).
45. Jimeno, S., Luna, R., Garcia-Rubio, M. & Aguilera, A. Tho1, a novel hnRNP, and Sub2 provide alternative pathways for mRNP biogenesis in yeast THO mutants. *Mol. Cell. Biol.* **26**, 4387–4398 (2006).
46. Wan, Y. *et al.* Splicing function of mitotic regulators links R-loop-mediated DNA damage to tumor cell killing. *J. Cell Biol.* **209**, 235–246 (2015).
47. Salvi, J. S. *et al.* Roles for Pbp1 and caloric restriction in genome and lifespan maintenance via suppression of RNA-DNA hybrids. *Dev. Cell* **30**, 177–191 (2014).
48. Gavalda, S., Gallardo, M., Luna, R. & Aguilera, A. R-loop mediated transcription-associated recombination in trf4Δ mutants reveals new links between RNA surveillance and genome integrity. *PLoS ONE* **8**, e65541 (2013).
49. Pefanis, E. *et al.* RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. *Cell* **161**, 774–789 (2015).
50. Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N. J. & Dean, C. R-loop stabilization represses antisense transcription at the *Arabidopsis* FLC locus. *Science* **340**, 619–621 (2013).
51. Proudfoot, N. J. Ending the message: poly(A) signals then and now. *Genes Dev.* **25**, 1770–1782 (2011).
52. Kawauchi, J., Mischo, H., Braglia, P., Rondon, A. & Proudfoot, N. J. Budding yeast RNA polymerases I and II employ parallel mechanisms of transcriptional termination. *Genes Dev.* **22**, 1082–1092 (2008).

53. Grzechnik, P., Gdula, M. R. & Proudfoot, N. J. Pcf11 orchestrates transcription termination pathways in yeast. *Genes Dev.* **29**, 849–861 (2015).
54. Porrua, O. & Libri, D. A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. *Nat. Struct. Mol. Biol.* **20**, 884–891 (2013).
55. Yuce, O. & West, S. C. Senataxin, defective in the neurodegenerative disorder ataxia with oculomotor apraxia 2, lies at the interface of transcription and the DNA damage response. *Mol. Cell. Biol.* **33**, 406–417 (2013).
56. Morales, J. C. *et al.* Kub5-Hera, the human Rtt103 homolog, plays dual functional roles in transcription termination and DNA repair. *Nucleic Acids Res.* **42**, 4996–5006 (2014).
57. Huang, H. S. *et al.* Topoisomerase inhibitors unsilence the dormant allele of Ube3a in neurons. *Nature* **481**, 185–189 (2012).
58. Powell, W. T. *et al.* R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation. *Proc. Natl Acad. Sci. USA* **110**, 13938–13943 (2013).
59. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715 (1993).
60. Polak, P. & Arndt, P. F. Transcription induces strand-specific mutations at the 5′ end of human genes. *Genome Res.* **18**, 1216–1223 (2008).
61. Chaudhuri, J. & Alt, F. W. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nat. Rev. Immunol.* **4**, 541–552 (2004).
62. Meng, F. L. *et al.* Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell* **159**, 1538–1548 (2014).
63. Qian, J. *et al.* B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. *Cell* **159**, 1524–1537 (2014).
64. Gómez-González, B. & Aguilera, A. Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. *Proc. Natl Acad. Sci. USA* **104**, 8409–8414 (2007).
65. Kogoma, T. Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiol. Mol. Biol. Rev.* **61**, 212–238 (1997).
66. Ponder, R. G., Fonville, N. C. & Rosenberg, S. M. A switch from high-fidelity to error-prone DNA double-strand break repair underlies stress-induced mutation. *Mol. Cell* **19**, 791–804 (2005).
67. Wimberly, H. *et al.* R-loops and nicks initiate DNA breakage and genome instability in non-growing *Escherichia coli*. *Nat. Commun.* **4**, 2115 (2013).
68. Deem, A. *et al.* Break-induced replication is highly inaccurate. *PLoS Biol.* **9**, e1000594 (2011).
69. Aguilera, A. & García-Muse, T. Causes of genome instability. *Annu. Rev. Genet.* **47**, 1–32 (2013).
70. Boubakri, H., de Septenville, A. L., Viguera, E. & Michel, B. The helicases DinG, Rep and UvrD cooperate to promote replication across transcription units *in vivo*. *EMBO J.* **29**, 145–157 (2010).
71. Gan, W. *et al.* R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev.* **25**, 2041–2056 (2011).
72. Wellinger, R. E., Prado, F. & Aguilera, A. Replication fork progression is impaired by transcription in hyperrecombinant yeast cells lacking a functional THO complex. *Mol. Cell. Biol.* **26**, 3327–3334 (2006).
73. Gómez-González, B. *et al.* Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. *EMBO J.* **30**, 3106–3119 (2011).
74. Santos-Pereira, J. M. *et al.* The Npl3 hnRNP prevents R-loop-mediated transcription-replication conflicts and genome instability. *Genes Dev.* **27**, 2445–2458 (2013).
75. Alzu, A. *et al.* Senataxin associates with replication forks to protect fork integrity across RNA-polymerase-II-transcribed genes. *Cell* **151**, 835–846 (2012).
This article provides genome-wide evidence that Sen1 (the yeast orthologue of the mammalian RNA–DNA helicase senataxin) accumulates with replication forks at transcribed genes, suggesting that Sen1 protects forks from the formation of recombinogenic damage that can activate the DNA damage checkpoint.
76. Helmrich, A., Ballarino, M. & Tora, L. Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol. Cell* **44**, 966–977 (2011).
In this work, the authors show that the longest human genes take more than a cell cycle to be transcribed and that collision between transcription and replication machineries is inevitable. This leads to hotspots of DNA breaks called common fragile sites, where R loops form and are responsible for the transcription–replication conflicts that generate genome instability.
77. Rõndon, A. G., Jimeno, S., García-Rubio, M. & Aguilera, A. Molecular evidence that the eukaryotic THO/TREX complex is required for efficient transcription elongation. *J. Biol. Chem.* **278**, 39037–39043 (2003).
78. Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M. E. & Nudler, E. Linking RNA polymerase backtracking to genome instability in *E. coli*. *Cell* **146**, 533–543 (2011).
79. Tresini, M. *et al.* The core spliceosome as target and effector of non-canonical ATM signalling. *Nature* **523**, 53–58 (2015).
80. Bhatia, V. *et al.* BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature* **511**, 362–365 (2014).
This paper shows that depletion of BRCA2 leads to genome instability and accumulation of R loops, as detected by the S9.6 antibody, which specifically recognizes these structures, and by an RNase H1 hybrid-binding domain fused to GFP. The manuscript proposes that R loops are a major source of spontaneous replication stress and that BRCA2 and Fanconi anaemia proteins contribute to the elimination of R loops that block replication-fork progression.
81. Schlacher, K., Wu, H. & Jasin, M. A distinct replication fork protection pathway connects Fanconi anemia tumor suppressors to RAD51-BRCA1/2. *Cancer Cell* **22**, 106–116 (2012).
82. Britton, S. *et al.* DNA damage triggers SAF-A and RNA biogenesis factors exclusion from chromatin coupled to R-loops removal. *Nucleic Acids Res.* **42**, 9047–9062 (2014).
83. Luke, B. *et al.* The Rat1p 5′ to 3′ exonuclease degrades telomeric repeat-containing RNA and promotes telomere elongation in *Saccharomyces cerevisiae*. *Mol. Cell* **32**, 465–477 (2008).
84. Balk, B. *et al.* Telomeric RNA-DNA hybrids affect telomere-length dynamics and senescence. *Nat. Struct. Mol. Biol.* **20**, 1199–1205 (2013).
85. Pfeiffer, V., Crittin, J., Grolimund, L. & Lingner, J. The THO complex component Thp2 counteracts telomeric R-loops and telomere shortening. *EMBO J.* **32**, 2861–2871 (2013).
86. Yu, T. Y., Kao, Y. W. & Lin, J. J. Telomeric transcripts stimulate telomere recombination to suppress senescence in cells lacking telomerase. *Proc. Natl Acad. Sci. USA* **111**, 3377–3382 (2014).
87. Arora, R. *et al.* RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. *Nat. Commun.* **5**, 5220 (2014).
88. Nakama, M., Kawakami, K., Kajitani, T., Urano, T. & Murakami, Y. DNA–RNA hybrid formation mediates RNAi-directed heterochromatin formation. *Genes Cells* **17**, 218–233 (2012).
89. Hsu, J. Y. *et al.* Mitotic phosphorylation of histone H3 is governed by Ipl1/aurora kinase and Glc7/PP1 phosphatase in budding yeast and nematodes. *Cell* **102**, 279–291 (2000).
90. Ivaldi, M. S., Karam, C. S. & Corces, V. G. Phosphorylation of histone H3 at Ser10 facilitates RNA polymerase II release from promoter-proximal pausing in *Drosophila*. *Genes Dev.* **21**, 2818–2831 (2007).
91. Zippo, A., De Robertis, A., Serafini, R. & Oliviero, S. PIM1-dependent phosphorylation of histone H3 at serine 10 is required for MYC-dependent transcriptional activation and oncogenic transformation. *Nat. Cell Biol.* **9**, 932–944 (2007).
92. El Achkar, E., Gerbault-Seureau, M., Muleris, M., Dutrillaux, B. & Debatisse, M. Premature condensation induces breaks at the interface of early and late replicating chromosome bands bearing common fragile sites. *Proc. Natl Acad. Sci. USA* **102**, 18069–18074 (2005).
93. Groh, M., Lufino, M. M., Wade-Martins, R. & Gromak, N. R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. *PLoS Genet.* **10**, e1004318 (2014).
These authors used cells from people with FRDA or FXS, diseases that are characterized by the expansion of trinucleotides in the *FXN* and *FMR1* genes, respectively. These expansions constitute rare fragile sites, and the article shows that R loops form at these expanded repeats, leading to accumulation of the repressive mark H3K9me2 and consequent gene silencing, which causes the disease.
94. Herrera-Moyano, E., Mergui, X., Garcia-Rubio, M. L., Barroso, S. & Aguilera, A. The yeast and human FACT chromatin-reorganizing complexes solve R-loop-mediated transcription-replication conflicts. *Genes Dev.* **28**, 735–748 (2014).
95. Hatchi, E. *et al.* BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. *Mol. Cell* **57**, 636–647 (2015).
This work suggests a cooperation between the human DNA repair factor BRCA1 and the RNA–DNA helicase SETX at transcription-termination regions to prevent R-loop formation, as depletion of either factor leads to increased levels of R loops. Interestingly, BRCA1 and SETX physically interact, and SETX binding to termination regions is BRCA1-dependent. BRCA1 binds genome-wide to termination regions of R-loop-accumulating genes, where BRCA1-deficient tumours show increased insertions and deletions.
96. Lin, Y., Dent, S. Y., Wilson, J. H., Wells, R. D. & Napierala, M. R loops stimulate genetic instability of CTG. CAG repeats. *Proc. Natl Acad. Sci. USA* **107**, 692–697 (2010).
97. Reddy, K. *et al.* Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. *Nucleic Acids Res.* **39**, 1749–1762 (2011).
98. Grabczyk, E., Mancuso, M. & Sammarco, M. C. A persistent RNA•DNA hybrid formed by transcription of the Friedreich ataxia triplet repeat in live bacteria, and by T7 RNAP *in vitro*. *Nucleic Acids Res.* **35**, 5351–5359 (2007).
99. Loomis, E. W., Sanz, L. A., Chedin, F. & Hagerman, P. J. Transcription-associated R-loop formation across the human FMR1 CGG-repeat region. *PLoS Genet.* **10**, e1004294 (2014).
100. Colak, D. *et al.* Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. *Science* **343**, 1002–1005 (2014).
101. Haeusler, A. R. *et al.* C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* **507**, 195–200 (2014).
102. Moreira, M. C. *et al.* Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. *Nat. Genet.* **36**, 225–227 (2004).
103. Chen, Y. Z. *et al.* DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). *Am. J. Hum. Genet.* **74**, 1128–1135 (2004).
104. Vantaggiato, C. *et al.* Senataxin modulates neurite growth through fibroblast growth factor 8 signalling. *Brain* **134**, 1808–1828 (2011).
105. Yeo, A. J. *et al.* R-loops in proliferating cells but not in the brain: implications for AOA2 and other autosomal recessive ataxias. *PLoS ONE* **9**, e90219 (2014).
106. Negrini, S., Gorgoulis, V. G. & Halazonetis, T. D. Genomic instability — an evolving hallmark of cancer. *Nat. Rev. Mol. Cell Biol.* **11**, 220–228 (2010).
107. Gaillard, H., Garcia-Muse, T. & Aguilera, A. Replication stress and cancer. *Nat. Rev. Cancer* **15**, 276–289 (2015).
108. Hill, S. J. *et al.* Systematic screening reveals a role for BRCA1 in the response to transcription-associated DNA damage. *Genes Dev.* **28**, 1957–1975 (2014).
109. Ramiro, A. R. *et al.* AID is required for c-myc/IgH chromosome translocations *in vivo*. *Cell* **118**, 431–438 (2004).
110. Ruiz, J. F., Gómez-González, B. & Aguilera, A. AID induces double-strand breaks at immunoglobulin switch regions and c-MYC causing chromosomal translocations in yeast THO mutants. *PLoS Genet.* **7**, e1002009 (2011).

111. Chernikova, S. B. *et al.* Deficiency in mammalian histone H2B ubiquitin ligase Bre1 (Rnf20/Rnf40) leads to replication stress and chromosomal instability. *Cancer Res.* **72**, 2111–2119 (2012).
112. Fregoso, O. I., Das, S., Akerman, M. & Krainer, A. R. Splicing-factor oncoprotein SRSF1 stabilizes p53 via RPL5 and induces cellular senescence. *Mol. Cell* **50**, 56–66 (2013).
113. Jackson, B. R., Noerenberg, M. & Whitehouse, A. A novel mechanism inducing genome instability in Kaposi's sarcoma-associated herpesvirus infected cells. *PLoS Pathog.* **10**, e1004098 (2014).
114. Huang, F. T. *et al.* Sequence dependence of chromosomal R-loops at the immunoglobulin heavy-chain Smu class switch region. *Mol. Cell. Biol.* **27**, 5921–5932 (2007).
115. Huang, F. T., Yu, K., Hsieh, C. L. & Lieber, M. R. Downstream boundary of chromosomal R-loops at murine switch regions: implications for the mechanism of class switch recombination. *Proc. Natl Acad. Sci. USA* **103**, 5030–5035 (2006).
116. Kao, Y. P. *et al.* Detection and characterization of R-loops at the murine immunoglobulin Sα region. *Mol. Immunol.* **54**, 208–216 (2013).
117. Yu, K., Chedin, F., Hsieh, C. L., Wilson, T. E. & Lieber, M. R. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat. Immunol.* **4**, 442–451 (2003).
118. Reaban, M. E. & Griffin, J. A. Induction of RNA-stabilized DNA conformers by transcription of an immunoglobulin switch region. *Nature* **348**, 342–344 (1990).
119. Muramatsu, M. *et al.* Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**, 553–563 (2000).
120. Revy, P. *et al.* Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell* **102**, 565–575 (2000).
121. Chaudhuri, J. *et al.* Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature* **422**, 726–730 (2003).
122. Zheng, S. *et al.* Non-coding RNA generated following lariat debranching mediates targeting of AID to DNA. *Cell* **161**, 762–773 (2015).
123. Pefanis, E. *et al.* Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. *Nature* **514**, 389–393 (2014).
124. Zhang, Z. Z. *et al.* 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**, 557–569 (2014).
125. Boguslawski, S. J. *et al.* Characterization of monoclonal antibody to DNA•RNA and its application to immunodetection of hybrids. *J. Immunol. Methods* **89**, 123–130 (1986).
126. Zhang, Z. Z., Pannunzio, N. R., Hsieh, C. L., Yu, K. & Lieber, M. R. Complexities due to single-stranded RNA during antibody detection of genomic rna:dna hybrids. *BMC Res. Notes* **8**, 127 (2015).
127. Jenjaroenpun, P., Wongsurawat, T., Yenamandra, S. P. & Kuznetsov, V. A. QmRLFS-finder: a model, web server and stand-alone tool for prediction and analysis of R-loop forming sequences. *Nucleic Acids Res.* **43**, W527–W534 (2015).
128. Leela, J. K., Syeda, A. H., Anupama, K. & Gowrishankar, J. Rho-dependent transcription termination is essential to prevent excessive genome-wide R-loops in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **110**, 258–263 (2013).
129. Wahba, L., Gore, S. K. & Koshland, D. The homologous recombination machinery modulates the formation of RNA•DNA hybrids and associated chromosome instability. *eLife* **2**, e00505 (2013).

Acknowledgements
The authors thank T. García-Muse for critical reading of the manuscript and D. Haun for style supervision. Research in A.A.'s laboratory is funded by grants from the Spanish Ministry of Economy and Competitiveness, Junta de Andalucía, European Union (FEDER), Worldwide Cancer Research and PharmaMar. The authors apologize to those whose work could not be cited owing to space limitations.

Competing interests statement
The authors declare no competing interests.