Transcription--replication conflicts: how they occur and how they are resolved

Tatiana García-Muse & Andrés Aguilera

Centro Andaluz de Biología Molecular y Medicina Regenerativa-CABIMER, Universidad de Sevilla, Seville, Spain

corresponding author: aguilo@us.es

Abstract

The frequent occurrence in cells of transcription and DNA replication results in many encounters and thus conflicts between the transcription and replication machineries. These conflicts constitute a major intrinsic source of genome instability, which is a hallmark of cancer cells. How the replication machinery progresses through a DNA occupied by an RNA polymerase is an old question. Here we review recent data on the biological relevance of transcription-replication conflicts and the factors and mechanisms involved in either preventing or resolving them in eukaryotes. With this we aim to provide our current view of how transcription can generate obstacles to replication, including torsional stress and non-B DNA structures, and of the different cellular processes that have evolved to solve them.

Introduction

Genomes are templates for multiple biological processes, including transcription, epigenetic modifications, DNA replication, DNA repair and chromosome segregation. In a number of cases, crosstalk between different processes occurring at the DNA may have a positive effect, as in the case of transcription-coupled repair **[G]** [1]. However, in

other cases, the co-temporal activity of two cellular machineries at the same genomic region may cause a conflict with negative consequences. This is the case of DNA replication and transcription. Research in the past two decades has provided evidence that transcription and replication conflicts constitute a considerable natural ~~intrinsic~~ source of genome instability, which is a hallmark of cancer cells [2]. Given that transcription and replication are two essential processes for cell viability and proliferation~~function~~ and that they occur frequently, a high incidence of encounters between the transcription and replication machineries is to be expected. Although transcription can have a positive ~~role~~ effect on replication initiation through transcription-mediated chromatin changes that may facilitate firing of ~~origins of~~ DNA replication origins [3], collisions are a potential threat to genome integrity and cell viability.

How the replication machinery progresses through a double-stranded DNA occupied by an RNA polymerase is an old question. Alberts and colleagues elegantly addressed this question using the T4 bacteriophage system *in vitro* [4]; but our actual knowledge of the RNA polymerase structure and mechanisms of transcription elongation suggests that the factors and mechanisms used by cells to solve such conflicts are more complex than previously foreseen. The relevance of transcription as a source of genome instability, as measured by the rate of point mutations or of recombination and chromosome rearrangements, and the putative mechanisms by which such instability is mediated have been reviewed recently [5-8]. Here we review recent data on the factors and mechanisms involved in either preventing or resolving transcription--replication collisions, and on their potential consequences. In particular, we discuss how transcription may hinder the progression of the replication forks itself or how transcription activity generates obstacles to replication, including torsional stress **[G]** and non-B DNA **[G]** structures, and the different solutions the cells have evolved to avoid, minimize or resolve these collisions or their consequences.

[H1] How do collisions occur?

A basic difference between the transcription and replication machineries is that the elongating RNA polymerase holoenzyme, which comprises one polymerase subunit, embraces the double-stranded DNA. The nascent RNA chain is synthesized in the active pocket of the RNA polymerase, where it forms a dynamic 9-11-nt RNA-DNA hybrid (Figure 1A). The elongating DNA polymerase holoenzyme, on the other hand, consists of two polymerase subunits (of DNA pol III in bacteria; DNA pol epsilon and

delta in eukaryotes), each working on a single-stranded DNA (ssDNA) template (Figure 1B). Furthermore, whereas several active RNA polymerases can simultaneously transcribe the same gene, replisomes **[G]** move alone and are not followed by a second replication fork. No matter whether collisions are co-directional or in head-on orientation (Figure 2), the replication fork cannot go through an elongating RNA polymerase and so their encounters will cause conflicts [9, 10]. Although replication fork progression may be affected by collisions in both orientations, data suggest that the consequences of collisions are more dramatic in the head-on orientation [11, 12]. When encounters were promoted in yeast artificial systems in a head-on orientation, replication pause sites were detected by two-dimensional (2D)-gel electrophoresis and hyper-recombination was observed [12]. By contrast, co-directional orientation did not lead to replication pauses or high levels of hyper-recombination. This difference can be explained if co-directional encounters may in part be resolved once the RNA polymerase terminates transcription.

Cells have developed different strategies to reduce or prevent collisions. In bacteria, there is a genome-wide bias towards co-orientation of replication and transcription [13] and inverting transcriptional units to provoke head-on collisions causes replication impairment, proliferation defects and genome instability [11]. In eukaryotes a bias towards co-directional replication and transcription is not obvious, but cells seem to have evolved other strategies to reduce head-on collisions. For example, in the *Saccharomyces cerevisiae* highly transcribed ribosomal DNA (rDNA) genes, replication fork blocking (RFB) sites exist that block fork progression and prevent harmful encounters with RNA polymerases [14]. In the mammalian rDNA loci replication and the transcription seem to be ~~efficiently~~ spatially separated in the nucleoli as a way to avoid collisions [15]. In other regions of the genome, transcription and replication seem to be separated temporaly [16]. Analysis of nascent mRNAs in genes encoding replication factors revealed that active genes transcribed during early replication are replicated late in S-phase and *vice versa* [17].

Importantly, however, it is not clear whether the RNA and DNA polymerases ever actually make contact. It is plausible that before the physical connection occurs, transcription- and replication-mediated changes in chromatin and DNA structures attenuate the progression of the polymerases. To fully understand conflicts we therefore need to identify the elements and conditions that affect their occurrence.

[H1] *Cis*-elements affecting collisions

The transcription machinery may constitute a natural obstacle to replication fork progression, but this interference can be direct or indirect, since the transcription process may also generate structural features that have the capacity to hinder replication fork progression, like changes in DNA supercoiling or secondary DNA structures such as hairpins**[G]**, triplex DNA (H-DNA) **[G],** G-quadruplexes**[G]** or RNA-DNA hybrids.

[H3] *DNA supercoiling*

Transcription and replication require the unwinding of the DNA molecule. This unwinding leads to positive and negative supercoiling ahead and behind the RNA polymerase, respectively (Figure 3A). The resulting torsional stress **[G]** is relieved by DNA topoisomerases, which are of Type I or Type II depending on whether they catalyse breakage of one or both DNA strands, respectively. In budding yeast topoisomerase mutants accumulate supercoiling and this torsional stress prevents both transcription and replication of the highly transcribed rDNA [18], suggesting that supercoiling can cause transcription and replication block. Later studies in yeast and human cells have shown that both topoisomerase 1 (human TOP1, yeast Top1) and topoisomerase 2 (human TOP2, yeast Top2) are crucial to prevent transcription-replication collisions [19, 20], indicating that unresolved torsional stress can attenuate the progression of both DNA and RNA polymerases and promote transcription-replication conflicts. Genome-wide analysis of Top1 and Top2 distribution in replicating budding yeast cells revealed an association of these enzymes with moving replication forks [19, 21]. Moreover, *top1 top2* double mutant cells also accumulate DNA damage[19]. Consistently, analyses of replication by DNA combing **[G]** have revealed that replication forks are slower in yeast and human Top1-deficient cells [20]. Interestingly, in *TOP1*-depleted human cells there was an increase in fork stalling that correlated with the accumulation of γ-H2AX foci **[G]** in S-phase and that was suppressed by inhibition of transcription elongation with cordycepin. Therefore, TOP1 activity can prevent transcription-replication conflicts and their harmful consequences [20].

Based on these results we suggest that DNA supercoiling is transiently accumulating between the advancing transcription and replication machineries and may be important in the control of their collisions and their adverse effects (Figure 3A). In theory, this phenomenon should be exacerbated in the case of the head-on orientation. However, convergent transcription, which would create the same topological constraint as a head-on collisions, does not pose a major threat to genome integrity and transcription in budding yeast [22]. Alternatively, inefficient resolution of the negative supercoiling accumulating behind the elongating RNA polymerase may also

facilitate local melting of the DNA duplex and, consequently, formation of non-B DNA structures that can block replication fork progression (see below). Indeed, in yeast, divergent transcription was found to enhance chromosome rearrangements [23].

[H3] *non-B DNA structures and RNA-DNA hybrids*

Some DNA sequences, especially repetitive sequences, can assume non-B DNA structures such as hairpins, triplex DNA (H-DNA) or G-quadruplexes (also termed G-quartets). Such non-B DNA structures have the capacity to stall replication forks and have been correlated with hotspots of DNA double-strand breaks (DSBs) and chromosomal deletions, translocations and other rearrangements [24]. These secondary DNA structures are believed to form preferentially at the ssDNA that is exposed during DNA replication, but they can also be formed during transcription, favoured by the negative supercoiling that is transiently accumulated behind the elongating RNA polymerase (Figure 3). A good example of the putative relevance of non-B structures is provided by G-quadruplexes, which consist of four repeats of at least three guanines that can form four strand-interactions (Figure 3B). G-quadruplexes can form during lagging-strand replication as shown at telomeres [25], as well as during transcription: human cells treated with the G-quadruplex ligand pyridostatin show a tight correlation between pyridostatin binding and γH2AX foci **[G]** formation, which is reduced by treatment with the transcription inhibitor 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) [26]. The idea that G-quadruplex formation can be potentiated behind an elongating RNA Polymerase II (Pol II) has been indirectly inferred in yeast expressing the murine G-quadruplex-prone Sμ Ig switch region, which stimulates recombination in combination with conditions of high transcription levels[27]. The activity of Top1 in these conditions suppresses G-quadruplex-associated recombination, consistent with negative supercoiling enhancing G-quadruplex accumulation [28]. It appears that the genomic instability is higher when the orientation of the G-rich strand of the Sμ sequence, with respect to transcription, leaves the G-rich strand in the non-transcribed strand, suggesting that the ssDNA that allows quadruplex formation originates from transcription. Additional support to the idea that non-B DNA structure may contribute to transcription-mediated replication fork stalling comes from studies in mutants of the budding yeast DNA helicase Pif1 (petite integration frequency) and the fission yeast Pfh1 (Pif1-homolog), which unwind G-quadruplexes *in vitro*. Absence of Pif1 or Pfh1 attenuates or halts replication in regions of high G-quadruplex density and in RNA Pol II- and Pol III-highly transcribed genes [29, 30].

Another type of transcription-mediated structures able to cause fork stalling that can strongly contribute to transcription-replication conflicts are co-transcriptional RNA-DNA hybrids (also known as R loops when formed outside of the transcription bubble; Figure 3C). Although they are natural intermediates in class switching recombination and in initiation of mitochondrial DNA replication, RNA-DNA hybrids are formed in conditions that prevent the proper formation of the ribonucleoprotein particle, as shown in yeast and human cells [31, 32]. Evidence from yeast to mammalian cells suggest that RNA-DNA hybrids can form naturally and may constitute an important transcription intermediate that can provoke replication fork stalling at telomeres, the rDNA regions, CpG islands **[G]** and other sites at specific Pol II-transcribed genes, including 3'-end regions. RNA-DNA hybrids have been thoroughly and extensively reviewed recently [33-36] and will not be discussed further here. It is important to note, however, that an enrichment of sequences with high probability of forming non-B DNA structures or RNA-DNA hybrids and to undergo transcription-replication conflicts are observed at some fragile sites (Box 1).

[H1] Mechanisms for preventing conflicts

Owing to the impact of transcription-replication collisions on genomic stability and thereby potentially on disease, cells have evolved mechanisms to prevent such encounters. The factors that minimize collisions include the transcription machinery itself, as well as factors that help or facilitate replication progression through transcribed DNA.

[H3] *The RNA polymerase transcription apparatus*

Some clues to understand how the RNA polymerase directly contribute to transcription-replication conflicts are starting to emerge from the analysis of several RNA polymerase mutants in bacteria and yeast, but we are still far from having a complete view. A critical step in transcription that seems to be relevant to conflicts is RNA polymerase 'backtracking', which refers to the process by which the RNA polymerase reverses its progression to enzymatically remove the last incorporated ribonucleotide. This allows restarting transcription elongation following a pause provoked by hindrances during transcription elongation or as part of a regulatory process to coordinate the different steps of transcription and RNA processing [37]. A backtracked RNA polymerase is able to block replication progression in *Escherichia coli*. Using specific promoter sequences that allow modulation of the polymerase activity it was shown that a permanently-arrested elongating polymerase causes DSBs [38, 39]. Such

breaks were inferred to be replication-dependent since treatment with the replication inhibitor hydroxyurea prior to transcription activation avoided their appearance. The clash between the replisome and the backtracked RNA polymerase was interpreted to be responsible for the formation of DSBs [39]. Consistent with this view, the *E. coli* transcription elongation factors GreA and GreB , which promote the release of backtracked and stalled RNA polymerases, seem to reduce the consequences of conflicts (Figure 4A). In the absence of GreA and GreB and under substantial transcription activity induced by starvation, replication progression was completely blocked [40]. A similar role was proposed for the yeast transcription elongation factor TFIIS [41], but it remains to be seen if this activity has any effect on putative collisions.

Direct involvement of the transcription apparatus in modulating transcription-replication conflicts was demonstrated in recent studies utilizing RNA polymerase mutants that compromise the stability of transcription complexes. These RNA polymerase mutants were shown to suppress growth defects of *E. coli* cells lacking factors that help resolve collisions such as the DNA helicases Rep (Replicase), UvrD (uv resistant protein D) and DinG (damage-inducible protein G) [42]. These results suggest that less stable transcription complexes may not compromise replication progression since they do not seem to form strong replication obstacles [42]. Also, several yeast RNA Pol II mutants with transcription elongation defects exhibited replication impairment, inferred by 2D-gels, bromodeoxyuridine **[G]** incorporation by DNA polymerases or by altered distribution of Rrm3 (rDNA recombination mutation protein 3), which is a replicative helicase required for replication progression through DNA obstacles [43]. It is likely that following a collision the RNA Pol II is released from the DNA to allow passage of the replisome, as is the case in bacteria [38]. Interestingly, one of these yeast RNA Pol II mutants, the ~~yeast~~ *rpb1-1* mutant (of the largest RNA Pol II subunit), has tighter attachment to chromatin than wild-type RNA Pol II, as determined by chromatin immunoprecipitation, supporting the idea that RNA Pol II mutants with increased attachment to chromatin could aggravate the consequences of a transcription-replication encounter [43]. These results suggest that the transcription machinery, and RNA Pol II itself, may participate in managing transcription-replication conflicts through the feasibility of their eviction from DNA following a collision. The recent observation that PAF1C (RNA polymerase II-associated factor 1 complex) triggers RNA Pol II degradation at sites of collisions [44] supports this view.

[H3] *Replication fork barriers*

Replication forks have to deal along their path with non-nucleosomal protein–DNA complexes that assemble at genes and regulatory elements. In bacteria the barrier formed by the transcription complex is able to pause replication forks, and resumption of replication requires specific DNA helicases [45, 46]. In yeast, different regions that impede replication fork progression *in vivo* have been identified, the most representative being the fork barrier found in the 35S rRNA gene in the rDNA [47]. The rDNA region provides the best model to study the impact of replication stress generated by transcription owing to its high transcription rate and high density of replication origins. The replication barrier consists of DNA replication fork-blocking protein Fob1 bound to the specific RFB sequence, which prevents head-on collisions between RNA and DNA polymerases (Figure 4B). Interestingly, replication fork progression through the RFB–Fob1 complex requires the helicase Rrm3. Deletion of the *rrm3* gene (*rrm3*Δ) increases replication pauses at rDNA, resulting in breakage and accumulation of excised rDNA circles [14, 48]. Although Rrm3 could be seen therefore as a complementary factor acting *in trans* to promote replication fork passage through protein barriers, fork pausing in *rrm3*Δ mutants is also increased in other pause sites such as in tRNA genes or telomeres [49, 50]. Notably, though, other pause sites are found at highly transcribed genes, but are not exacerbated in *rrm3*Δ mutants, suggesting that other factors may have roles in the prevention or resolution of collisions [9].

[H3] *The RecQL5 DNA helicase*

Human RecQ-like ATP-dependent DNA helicase Q5 (RECQL5), a member of the RecQ family **[G]** of DNA helicases, is to date the protein with the best-characterised active role in preventing transcription-replication collisions. RECQL5 forms a stable complex with RNA Pol II and several *in vivo* and *in vitro* studies indicated it has a negative regulatory role in transcription elongation [51]. ChIP-seq analysis with an RNA Pol II antibody revealed that transcription up-regulation in cells lacking *RECQL5* increases transcription pausing, arrest and backtracking, suggesting that uncontrolled and high transcription rates lead to transcriptional stress [52]. RECQL5 associates with the replicative DNA sliding clamp PCNA (proliferating cell nuclear antigen), which suggests that RECQL5 is also involved in replication. Accordingly, *RECQL5*-deficient cells fail to incorporate bromodeoxyuridine in conditions of replication stress and rapidly accumulate DNA damage, effects which can be alleviated by fully arresting replication with the replication inhibitor aphidicolin [53]. Importantly, spontaneous DSBs accumulate in *RECQL5*-depleted cells during replication, but only in association with RNA Pol II transcription, as the spontaneous breaks are located in transcribed genes

and transcription inhibition eliminated their appearance [54]. Furthermore, RECQL5 has been shown to have a role in suppressing genome rearrangements associated preferentially with common fragile sites (Box 1) and transcribed genes [52]. These data have led to the proposal that RECQL5 prevents transcription-replication collisions. Recently, RECQL5 was also ascribed a role in preventing the formation of RNA-DNA hybrids. Apparently, RECQ5 promotes TOP1 SUMOylation by facilitating the interaction with the PIAS1-SRSF1 E3 ligase complex. This modification is necessary for the binding of TOP1 to RNA Pol II and for the efficient recruitment of mRNA processing factors to transcriptionally active sites, thereby reducing the formation of RNA-DNA hybrids, as inferred by the increased levels of such hybrids in cells defective in RECQ5-dependent TOP1 SUMOylation [55]. Therefore, RecQL5 may maintain genome integrity by actively limiting the occurrence of transcription-replication conflicts and/or by reducing the accumulation of non-B DNA structures generated during transcription that could enhance replication blockage (Figure 3).

[H3] *Chromatin remodelling*

In eukaryotes, transcription and replication occur in the context of highly structured chromatin. Following replication the chromatin state is maintained by coupling the deposition of recycled parental histones with newly-synthesized histones on the duplicated DNA, which is carried out by histone chaperones and chromatin remodelling complexes. Even though not much work has been done on the role of chromatin remodelling in diminishing transcription-replication conflicts, evidence exists about the importance of such factors on collisions. This has been clearly shown for the histone chaperone FACT (facilitates chromatin transcription) complex, which was initially found to be required for chromatin remodelling during transcription but is also involved in DNA replication [56-58]. Yeast and human cells lacking FACT complex activity have high levels of transcription–replication collisions, exhibiting fork progression impairment that correlates with increased genomic instability. However, when transcription was inhibited with cordycepin in FACT-depleted human cells, the rate of fork progression was restored to normal [59]. Therefore, chromatin-reorganizing factors, such as FACT, can prevent collisions by promoting the replication fork progression through transcribed regions (Figure 4C). It would certainly be interesting to see whether this view would also apply to the INO80 remodelling complex since it was shown in ~~E. coli~~budding yeast that Ino80 (inositol requiring mutant 80) cooperates with the transcription factor PAF1C to trigger RNA Pol II degradation at sites of collisions [44].

It is worth noting that RNA-DNA hybrids also accumulate in FACT-depleted cells, suggesting that either replication forks stall often at regions containing RNA-DNA hybrids or that RNA-DNA hybrids are formed as a consequence of transcription-replication collisions. The negative DNA supercoiling that locally accumulates behind a stalled RNA polymerase, a putative suboptimal mRNP assembly or an inefficient chromatin remodelling associated with a transcription-replication collision could create the conditions to favour co-transcriptional R-loop formation. The observation that different states of heterochromatin or chromatin condensation, as identified by the phosphorylation of histone H3 Ser10 or the dimethylation of histone H3 Lys9 (H3K9me2) in *S. cerevisiae, Caenorhabditis elegans* and human cells correlate with the formation of co-transcriptional RNA-DNA hybrids [60], suggests that RNA-DNA hybrids can trigger chromatin compaction, which might also contribute to replication fork stalling at transcribed regions. Indeed, the expansion of triplet repeats in the *FXN* and *FMR1* genes (Box 1) induced their silencing through the deposition of H3K9me3 and heterochromatinization, and led to the accumulation of RNA-DNA hybrids [61]. Therefore, chromatin compaction following transcription seems to be a contributor to transcription-replication conflicts that needs further exploring (Figure 4C).

Finally, gene silencing at pericentromeric regions in *Schizosacharomyces pombe* through the establishment of heterochromatin provides additional support to the role of chromatin organization in the occurrence of transcription-replication conflicts. Such silencing requires the RNAi machinery to facilitate recruitment of chromatin modifiers by small interfering RNAs (siRNA), and it has been proposed that the RNAi machinery is involved in the release of RNA polymerase at pericentromeric regions to prevent collisions [62] (Figure 4D). ChIP-seq analysis of the RNAi machinery mutant *dcr1*Δ (dicer protein 1) showed that RNA polymerase accumulates during S-phase at specific pericentromeric repeats, which correlate with sites of siRNA accumulation in normal cells, therefore defining sites of RNA polymerase release to allow replication completion. Importantly, replication fork stalling at such pericentromeric repeats was reduced in mutants of either transcription or replication initiation. Finally, failure to release RNA polymerase due to the absence of RNAi machinery results in DNA damage [62]. These results open the possibility that the capacity of the RNAi machinery to regulate gene expression and modulate chromatin structure may have an important role in transcription-replication conflicts, yet to be deciphered. Consistent with this view, genome-wide analysis has shown that RNAi ~~-dependent silencing transcription termination~~ **[Au: OK? if not please rephrase the original for greater clarity.]** activity in releasing the RNA polymerase is not restricted to heterochromatin regions but also

occurs at sites of replication stress such as at highly transcribed genes and rRNA and tRNA genes [63].

[H1] Conflicts and the DdNA damage response

The major consequence of transcription-replication conflicts is genome instability, mediated in most cases by chromosome breakage occurring because of replication fork blocking and collapse, which can generate DSBs; therefore transcription-replication conflicts are expected to be sensed and resolved by the DNA damage response (DDR) **[G]** and different repair pathways.

[H3] *Preventing transcription-replication conflicts by the DDR*

There is no evidence that the DDR senses transcription-replication collisions directly, but fork blockage could be sufficient to ~~activate~~trigger DNA damage checkpoints since it can result in the uncoupling of leading- and lagging-strand synthesis, generating a long stretch of ssDNA [64, 65] (Figure 5). It is likely that unless a DSB is produced, which activates the ataxia telangiectasia mutated (ATM; Tel1 in budding yeast)-dependent checkpoint, a transcription–replication collision will activate the ataxia telangiectasia and Rad3 related (ATR; Mec1)-dependent replication checkpoint, which senses stretches of ssDNA and protects the integrity of replication forks. Evidence that the ATR/Mec1 checkpoint ~~can prevent~~is involved in transcription-replication collisions has been provided for tRNA transcription in budding yeast. The tRNA transcription cycle involves assembly of a pre-initiation complex (PIC) comprised of RNA Pol III and two DNA-binding factors, TFIIIB and TFIIIC. At tRNA genes, transcription can act as a RFB during normal cell proliferation because the PIC interferes with fork progression [49, 66], but the ~~ATR~~Mec1 replication checkpoint can prevent replication fork stalling by dismantling the PIC in budding yeast [67]. This may be a conserved mechanism to reduce transcription-replication collisions at tRNA genes, as in *D. melanogaster*, removal of the Rpp30 factor (RNase P subunit p30), which is required for correct pre-tRNA processing, leads to increased replication stress and checkpoint activation [68].

A similar involvement of the DDR in reducing transcription-replication conflicts is observed in human cells when using hydroxyurea and doxorubicin, which are genotoxic agents that interfere with DNA replication by depleting dNTP pools and by inhibiting topoisomerase II, respectively, and induce fork stalling preferentially in regions of actively expressed genes. In these conditions, the ATR checkpoint ameliorates transcription-replication collisions by promoting ATR-mediated degradation of the histone chaperone ASF1a (anti-silencing function protein 1 homolog A). This

leads to histone eviction, RNA Pol II release and transcription repression [69]. Related to this, it has been shown in budding yeast that RNA Pol II is removed from transcribed genes located near firing origins of replication after exposure to hydroxyurea to avoid further impairment of replication fork progression. This removal depends on Mec1 in cooperation with the INO80 chromatin remodelling complex and the PAF1C transcription factor [44]. Therefore, a checkpoint-mediated transcription repression, putatively involving RNA Pol II eviction may protect genome integrity by reducing transcription-replication collisions (Figure 5).

Upon its activation, ATR phosphorylates many downstream targets, including the tumour suppressor p53 (TP53, tumour protein p53), to coordinate the DDR. Recently p53 was shown to be involved in preventing transcription-replication conflicts by reducing topological stress, as p53 deficiency increases sensitivity to topoisomerase inhibitors and culminates in replication-dependent DNA damage accumulation, both reversed by transcription inhibition [70]. However, whether this is owing to a specific role of p53 as part of the replication and transcription machineries or by spatially and temporally regulating transcription during replication is still unclear.

Dealing with transcription-replication conflicts in eukaryotes is further complicated by the fact that transcription is coupled with RNA splicing, maturation and nuclear export. A subset of transcribed genes is localized proximally to nuclear pores, presumably to facilitate the nuclear export of their transcripts. This phenomenon is known as “gene gating” and is mediated by RNA binding proteins such as the ~~three prime repair exonuclease 2~~transcription export 2 complex (TREX-2) ~~mRNA biogenesis and export factor~~**[Au: OK?]** ~~complex~~ and nucleoporins [71, 72]. Interestingly, mutations in some of these gene-gating factors partially suppress fork instability in checkpoint mutants, suggesting that in the absence of a functional checkpoint, if replication forks are stalled in loci located at the nuclear pore, chromatin cannot be released from the pores to allow replication restart [73]. It is possible that the persistence of transcribed chromatin at the nuclear periphery and the accumulation of torsional stress due to the fixation of DNA to the pore that restricts supercoil release are impediments to replication fork progression. If this is indeed the case, the Mec1 and Rad53 checkpoint kinases could trigger the release of the fork blocked from the nuclear pore, allowing replication completion [73]. It is interesting to note that ATR or Mec1 activation results in phosphorylation ~~and activation~~**[Au: OK?]** of the Rrm3 and Pif1 DNA helicases, which as mentioned above, assist ~~transcription~~ replication at obstacles in budding yeast. Both helicases have been associated with stalled replication forks and combined removal of both suppresses the increase in fork reversal and cell lethality in the absence of a

functional Rad53 in conditions of replication stress, suggesting that both replication fork reversal and cell lethality depend on the unregulated activity of Rrm3 and Pif1 [74]. Indeed, in conditions of replication stress the expression of phosphomimetic **[G]** *rrm3* mutants lessen the *rad53* mutation defects. Although the role of Rrm3 is not specific for replication forks that are stalled at sites of collisions, and includes other types of protein obstacles that require restart of replication forks, it seems plausible that the ATR checkpoint surveillance mechanism helps reduce transcription-replication collisions and their consequences ~~by~~its ~~action on activating~~via its action on Rrm3.

[H3] *Conflict resolution by the BRCA and Fanconi anemia proteins*

An emerging question is whether specific DNA repair pathways are active at transcription-replication collision sites following activation of the DDR activation by the DNA damage checkpoints. Although thorough analysis is required on this issue, insight into the role of specific DNA repair pathways in preventing transcription-associated genomic instability, including instability caused by RNA-DNA hybrids, is emerging. For example, it has recently been shown that the tumour suppressors BRCA1 (breast cancer type 1 susceptibility protein) and BRCA2, which are involved in DSB repair, help prevent the formation or remove RNA-DNA hybrids[75, 76]. Although this function could be related to the DSB repair function of BRCA proteins, the fact that they are components of the Fanconi anemia pathway, which repair interstrand crosslinks and was recently shown to prevent the collapse of stalled replication forks [77], suggests that these proteins may have a key role in resolving replication fork stalling derived from transcription-replication conflicts (Figure 5). Accordingly, it has recently been demonstrated that the Fanconi anemia repair pathway contributes to preventing transcription-replication conflicts from ~~formi~~leading to DNA lesions, in particular those linked to the accumulation of RNA-DNA hybrids. Reducing the number of RNA-DNA hybrids by inhibiting transcription with cordycepin or by directly removing them with RNaseH alleviated the DNA damage observed in cells lacking Fanconi anemia complex subunits [78, 79]. The Fanconi anemia pathway may function at sites of collisions, in particular those involved with RNA-DNA hybrid accumulation, by repairing the DNA breaks resulting from replication fork arrest caused by RNA-DNA hybrids to facilitate resumption of replication [79]. Specifically, FANCM (Fanconi anemia complementation group M) was shown to have RNA-DNA branch migration activity *in vitro*, which has led to the suggestion that the Fanconi anemia ~~complex~~factors could help remove RNA-DNA hybrids *in vivo* [78], but this has not been demonstrated.

A genome-wide analysis revealed that BRCA1 is enriched in actively transcribed genes at regions of transcription termination, and this enrichment is mediated by the direct interaction with senataxin[80], an RNA-DNA helicase involved in RNA Pol II transcription termination that was shown to supress RNA Pol II collisions at transcription termination sites [81, 82]. Indeed, senataxin and BRCA1 were shown to interact with each other to suppress RNA-DNA hybrids at transcription termination sites [80]. This finding, together with the parallel observation that BRCA2 interacts with TREX-2 complex [75], favours the hypothesis that DNA repair proteins and the transcription apparatus cooperate to respond to transcription-replication conflicts (Figure 5). It would be interesting to explore whether this has anything to do with the purification of the RNA Pol II complex together with DSB repair factors in yeast [83]. It is interesting to note that the yeast mitogen-activated protein kinase Hog1 (high osmolarity glycerol response protein 1), which is also a transcription factor that responds to osmotic stress, was shown to delay replication timing by affecting early-origin firing and replication fork progression through ~~its effector~~direct action on Mrc1 (mediator of the replication checkpoint protein 1), which is a replisome component [84] and a substrate of the replication checkpoint protein Mec1 [85]. These data suggest that in conditions of transcriptional burst, for example in response to environmental stress, the coordinated action on both the transcription and replication machineries could be critical to manage transcription-replication conflicts.

Conclusions and perspectives

We do not yet have a molecular understanding of how replication forks traverse DNA regions undergoing transcription. Evidence accumulated in the past two decades indicates that an important natural source of genome instability stems from transcription-replication conflicts. The biomedical relevance of transcription-replication conflicts is emphasized by a number of cancer prone conditions or human diseases, such as Fanconi anemia, Ataxia-ocular apraxia type 2 or amyotrophic lateral sclerosis type 4, which are caused by mutations in genes that are involved in preventing or solving such conflicts. We have recently started to identify factors and putative mechanisms that may contribute to either diminishing the frequency of collisions or resolving them in a way that limits their negative consequences, such as replication stress and DNA breaks. However, to decipher the mechanisms by which replication forks replicate DNA undergoing transcription without compromising genome integrity, we still need to identify DNA sequences or secondary structures or specific chromatin features at collision hotspots, as well as the role of torsional stress and chromatin

remodelling in either promoting or preventing transcription-replication collisions, or the mechanisms by which the DDR senses such collisions or resolves them. This will necessitate a better understanding of the dynamics of replication and transcription machineries in response to different types of obstacles, from DNA lesions to protein barriers, and will not only help us to understand how cells execute proper replication of their entire genome, bypassing the putative barriers generated by transcription, but to clarify the importance of transcription-replication collisions as a source of DNA damage, in particular of oncogene-induced replication stress and DNA damage (Box 2), and the possibility of using transcription-replication collisions ~~[Au: OK?]~~ as selective targets in cancer therapy.

References

1. Hanawalt, P.C. & Spivak, G. Transcription-coupled DNA repair: two decades of progress and surprises. *Nature Rev. Mol. Cell Biol.* **9**, 958-970 (2008).
2. Gaillard, H., Garcia-Muse, T. & Aguilera, A. Replication stress and cancer. *Nature Rev. Cancer* **15**, 276-289 (2015).
3. Barlow, J.H. & Nussenzweig, A. Replication initiation and genome instability: a crossroads for DNA and RNA synthesis. *Cell. Mol. Life Sci.* **71**, 4545-4559 (2014).
4. Bedinger, P., Hochstrasser, M., Jongeneel, C.V. & Alberts, B.M. Properties of the T4 bacteriophage DNA replication apparatus: the T4 dda DNA helicase is required to pass a bound RNA polymerase molecule. *Cell* **34**, 115-123 (1983).
5. Bermejo, R., Lai, M.S. & Foiani, M. Preventing replication stress to maintain genome stability: resolving conflicts between replication and transcription. *Mol. Cell* **45**, 710-718 (2012).
6. Helmrich, A., Ballarino, M., Nudler, E. & Tora, L. Transcription-replication encounters, consequences and genomic instability. *Mol. Cell* **20**, 412-418 (2013).
7. Aguilera, A. & Garcia-Muse, T. Causes of genome instability. *Annu. Rev. Genet.* **47**, 1-32 (2013).
8. Gaillard, H. & Aguilera, A. Transcription as a threat to genome integrity. *Annu. Rev. Biochem.* **85** (2016).
9. Azvolinsky, A., Giresi, P.G., Lieb, J.D. & Zakian, V.A. Highly transcribed RNA polymerase II genes are impediments to replication fork progression in *Saccharomyces cerevisiae*. *Mol. Cell* **34**, 722-734 (2009).
   In this study, ChIP-chip analysis of the budding yeast genome identified ~~transcribed genes as~~ sites of DNA polymerase occupancy~~,~~. Among these sites transcribed genes were enriched, suggesting that they are prone to transcription-replication conflicts.
10. Merrikh, H., Machón, C., Grainger, W.H., Grossman, A.D. & Soultanas, P. Co-directional replication-transcription conflicts lead to replication restart. *Nature* **470**, 554-557 (2011).

This works shows that co-directional conflicts at highly transcribed rRNA genes can stall replication *in vivo* in *B. subtilis.*

11. Srivatsan, A., Tehranchi, A., MacAlpine, D.M. & Wang, J.D. Co-orientation of replication and transcription preserves genome integrity. *PLoS Genet.* **6**, e1000810 (2010).

12. Prado, F. & Aguilera, A. Impairment of replication fork progression mediates RNA pol II transcription-associated recombination. *EMBO J.* **24**, 1267-1276 (2005).
This study demonstrates that head-on transcription and replication encounters cause replication fork impairment that leads to genome instability.

13. Merrikh, H., Zhang, Y., Grossman, A.D. & Wang, J.D. Replication-transcription conflicts in bacteria. *Nature Rev. Microbiol.* **10**, 449-458 (2012).
14. Ivessa, A.S., Zhou, J.Q. & Zakian, V.A. The *Saccharomyces* Pif1p DNA helicase and the highly related Rrm3p have opposite effects on replication fork progression in ribosomal DNA. *Cell* **100**, 479-489 (2000).
15. Smirnov, E. *et al*. Separation of replication and transcription domains in nucleoli. *J. Struct. Biol.* **188**, 259-266 (2014).
16. Gilbert, D.M. Replication timing and transcriptional control: beyond cause and effect. *Curr. Opin. Cell Biol.* **14**, 377-383 (2002).
17. Meryet-Figuiere, M. *et al*. Temporal separation of replication and transcription during S-phase progression. *Cell Cycle* **13**, 3241-3248 (2014).
18. Brill, S.J., DiNardo, S., Voelkel-Meiman, K. & Sternglanz, R. Need for DNA topoisomerase activity as a swivel for DNA replication for transcription of ribosomal RNA. *Nature* **326**, 414-416 (1987).
19. Bermejo, R. *et al*. Top1- and Top2-mediated topological transitions at replication forks ensure fork progression and stability and prevent DNA damage checkpoint activation. *Genes Dev.* **21**, 1921-1936 (2007).

20. Tuduri, S. *et al*. Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nature Cell Biol.* **11**, 1315-1324 (2009).
This genome-wide analysis of TOP1-deficient cells shows correlation between fork stalling and DNA breaks at transcribed genes, suggesting that TOP1 reduces R-loops-dependent transcription–replication conflicts.

21. Bermejo, R. *et al*. Genome-organizing factors Top2 and Hmo1 prevent chromosome fragility at sites of S phase transcription. *Cell* **138**, 870-884 (2009).
22. García-Rubio, M.L. & Aguilera, A. Topological constraints impair RNA polymerase II transcription and causes instability of plasmid-borne convergent genes. *Nucleic Acids Res.* **40**, 1050-1064 (2012).
23. Pannunzio, N.R. & Lieber, M.R. Dissecting the roles of divergent and convergent transcription in chromosome instability. *Cell Rep.* **14**, 1025-1031 (2016).
24. Zhao, J., Bacolla, A., Wang, G. & Vasquez, K.M. Non-B DNA structure-induced genetic instability and evolution. *Cell Mol. Life Sci.* **67**, 43-62 (2010).
25. Audry, J. *et al*. RPA prevents G-rich structure formation at lagging-strand telomeres to allow maintenance of chromosome ends. *EMBO J.* **34**, 1942-1958 (2015).
26. Rodriguez, R. *et al*. Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nature Chem. Biol.* **8**, 301-310 (2012).

27. Kim, N. & Jinks-Robertson, S. Guanine repeat-containing sequences confer transcription-dependent instability in an orientation-specific manner in yeast. *DNA Repair (Amst)* **10**, 953-960 (2011).
28. Yadav, P. *et al*. Topoisomerase I plays a critical role in suppressing genome instability at a highly transcribed G-quadruplex-forming sequence. *PLoS Genet.* **10**, e1004839 (2014).
29. Sabouri, N., McDonald, K.R., Webb, C.J., Cristea, I.M. & Zakian, V.A. DNA replication through hard-to-replicate sites, including both highly transcribed RNA Pol II and Pol III genes, requires the *S. pombe* Pfh1 helicase. *Genes Dev.* **26**, 581-593 (2012).
30. Paeschke, K., Capra, J.A. & Zakian, V.A. DNA replication through G-quadruplex motifs is promoted by the *Saccharomyces cerevisiae* Pif1 DNA helicase. *Cell* **145**, 678-691 (2011).
31. Huertas, P. & Aguilera, A. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol. Cell* **12**, 711-721 (2003).
32. Li, X. & Manley, J.L. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* **122**, 365-378 (2005).
33. Santos-Pereira, J.M. & Aguilera, A. R loops: new modulators of genome dynamics and function. *Nature Rev. Genet.* **16**, 583-597 (2015).
34. 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).
35. Sollier, J. & Cimprich, K.A. Breaking bad: R-loops and genome integrity. *Trends Cell Biol.* **25**, 514-522 (2015).
36. Costantino, L. & Koshland, D. The Yin and Yang of R-loop biology. *Curr. Opin. Cell Biol.* **34**, 39-45 (2015).
37. Nudler, E. RNA polymerase backtracking in gene regulation and genome instability. *Cell* **149**, 1438-1445 (2012).
38. Pomerantz, R.T. & O'Donnell, M. Direct restart of a replication fork stalled by a head-on RNA polymerase. *Science* **327**, 590-592 (2010).
39. 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).
40. Tehranchi, A.K. *et al*. The transcription factor DksA prevents conflicts between DNA replication and transcription machinery. *Cell* **141**, 595-605 (2010).
This work illustrates the function of RNA polymerases, transcription factors and DNA repair proteins in preventing and/or resolving transcription–replication conflicts.
41. Dutta, A. *et al*. Ccr4-Not and TFIIS function cooperatively to rescue arrested RNA polymerase II. *Mol. Cell Biol.* **35**, 1915-1925 (2015).
42. Baharoglu, Z., Lestini, R., Duigou, S. & Michel, B. RNA polymerase mutations that facilitate replication progression in the *rep uvrD recF* mutant lacking two accessory replicative helicases. *Mol. Microbiol.* **77**, 324-336 (2010).
43. Felipe-Abrio, I., Lafuente-Barquero, J., García-Rubio, M.L. & Aguilera, A. RNA polymerase II contributes to preventing transcription-mediated replication fork stalls. *EMBO J.* **34**, 236-250 (2015).
44. Poli, J. *et al*. Mec1, INO80, and the PAF1 complex cooperate to limit transcription replication conflicts through RNAPII removal during replication stress. *Genes Dev.* **30**, 337-354 (2016).
45. Merrikh, C.N., Brewer, B.J. & Merrikh, H. The *B. subtilis* accessory helicase PcrA facilitates DNA replication through transcription units. *PLoS Genet.* **11**, e1005289 (2015).

46. Gupta, M.K. *et al*. Protein-DNA complexes are the primary sources of replication fork pausing in *Escherichia coli*. *Proc. Natl. Acad. Sci.* **110**, 7252-7257 (2013).
47. Brewer, B.J. & Fangman, W.L. A replication fork barrier at the 3' end of yeast ribosomal RNA genes. *Cell* **55**, 637-43 (1988).
48. Torres, J.Z., Bessler, J.B. & Zakian, V.A. Local chromatin structure at the ribosomal DNA causes replication fork pausing and genome instability in the absence of the *S. cerevisiae* DNA helicase Rrm3p. *Genes Dev.* **18**, 498-503 (2004).
49. Ivessa, A.S. *et al*. The *Saccharomyces cerevisiae* helicase Rrm3p facilitates replication past nonhistone protein-DNA complexes. *Mol. Cell* **12**, 1525-1536 (2003).
50. Ivessa, A.S., Zhou, J.-Q., Schulz, V.P., Monson, E.K. & Zakian, V.A. *Saccharomyces* Rrm3p, a 5' to 3' DNA helicase that promotes replication fork progression through telomeric and subtelomeric DNA. *Genes Dev.* **16**, 1383-1396 (2002).
51. Popuri, V., Tadokoro, T., Croteau, D.L. & Bohr, V.A. Human RECQL5: guarding the crossroads of DNA replication and transcription and providing backup capability. *Crit. Rev. Biochem. Mol. Biol.* **48**, 289-299 (2013).

52. Saponaro, M. *et al*. RECQL5 controls transcript elongation and suppresses genome instability associated with transcription stress. *Cell* **157**, 1037-1049 (2014).
This study shows that RECQL5 attenuates transcription elongation and suppresses genome rearrangements at common fragile sites, indicating that RECQL5 prevents transcription-replication conflicts.

53. Hu, Y., Lu, X., Zhou, G., Barnes, E.L. & Luo, G. Recql5 plays an important role in DNA replication and cell survival after camptothecin treatment. *Mol. Biol. Cell* **20**, 114-123 (2009).
54. Li, M., Xu, X. & Liu, Y. The SET2-RPB1 interaction domain of human RECQ5 is important for transcription-associated genome stability. *Mol. Cell. Biol.* **31**, 2090-2099 (2011).
55. Li, M., Pokharel, S., Wang, J.-T., Xu, X. & Liu, Y. RECQ5-dependent SUMOylation of DNA topoisomerase I prevents transcription-associated genome instability. *Nature Commun.* **6**, 6720 (2015).
56. Orphanides, G., Wu, W.H., Lane, W.S., Hampsey, M. & Reinberg, D. The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins. *Nature* **400**, 284-288 (1999).
57. Foltman, M. *et al*. Eukaryotic replisome components cooperate to process histones during chromosome replication. *Cell Rep.* **3**, 892-904 (2013).
58. Abe, T. *et al*. The histone chaperone facilitates chromatin transcription (FACT) protein maintains normal replication fork rates. *J. Biol. Chem.* **286**, 30504-30512 (2011).

59. 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).
This work provides compelling evidence of the role of chromatin-reorganizing complex FACT in the resolution of transcription-replication conflicts.

60. Castellano-Pozo, M. *et al*. R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. *Mol. Cell* **52**, 583-590 (2013).

61. 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).

62. Zaratiegui, M. *et al*. RNAi promotes heterochromatic silencing through replication-coupled release of RNA Pol II. *Nature* **479**, 135-138 (2011).
The authors show that at pericentromeric regions the co-transcriptional RNA interference machinery releases the RNA polymerase to avoid conflicts with the replication machinery.

63. Castel, S.E. *et al*. Dicer promotes transcription termination at sites of replication stress to maintain genome stability. *Cell* **159**, 572-583 (2014).
64. Lopes, M., Foiani, M. & Sogo, J.M. Multiple mechanisms control chromosome integrity after replication fork uncoupling and restart at irreparable UV lesions. *Mol. Cell* **21**, 15-27 (2006).
65. Heller, R.C. & Marians, K.J. Replication fork reactivation downstream of a blocked nascent leading strand. *Nature* **439**, 557-562 (2006).
66. Deshpande, A.M. & Newlon, C.S. DNA replication fork pause sites dependent on transcription. *Science* **272**, 1030-1033 (1996).
67. Nguyen, V.C. *et al*. Replication stress checkpoint signaling controls tRNA gene transcription. *Nat. Struct. Mol. Biol.* **17**, 976-981 (2010).
68. Molla-Herman, A., Vallés, A.M., Ganem-Elbaz, C., Antoniewski, C. & Huynh, J.-R. tRNA processing defects induce replication stress and Chk2-dependent disruption of piRNA transcription. *EMBO J.* (2015).
69. Im, J.S. *et al*. ATR checkpoint kinase and CRL1betaTRCP collaborate to degrade ASF1a and thus repress genes overlapping with clusters of stalled replication forks. *Genes Dev.* **28**, 875-887 (2014).
70. Yeo, C.Q. *et al*. p53 Maintains Genomic stability by preventing interference between transcription and replication. *Cell Rep.* **15**, 132-146 (2016).
71. Cabal, G.G. *et al*. SAGA interacting factors confine sub-diffusion of transcribed genes to the nuclear envelope. *Nature* **441**, 770-773 (2006).
72. Blobel, G. Gene gating: a hypothesis. *Proc. Natl. Acad. Sci. USA* **82**, 8527-8529 (1985).
73. Bermejo, R. *et al*. The replication checkpoint protects fork stability by releasing transcribed genes from nuclear pores. *Cell* **146**, 233-246 (2011).
74. Rossi, S.E., Ajazi, A., Carotenuto, W., Foiani, M. & Giannattasio, M. Rad53-mediated regulation of Rrm3 and Pif1 DNA helicases contributes to prevention of aberrant fork transitions under replication stress. *Cell Rep.* **13**, 80-92 (2015).
75. Bhatia, V. *et al*. BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature* 511, 362-365 (2014).
76. 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).
77. Schlacher, K. *et al*. Double-strand break repair-independent role for BRCA2 in blocking stalled replication fork degradation by MRE11. *Cell* **145**, 529-542 (2011).

78. Schwab, R.A. *et al*. The Fanconi anemia pathway maintains genome stability by coordinating replication and transcription. *Mol. Cell* **60**, 351-361 (2015).
79. García-Rubio, M.L. *et al*. The Fanconi anemia pathway protects genome integrity from R-loops. *PLoS Genet.* **11**, e1005674 (2015).
References 78 and 79 provide evidence for a novel role of Fanconi anemia factors in preventing transcription-replication conflicts mediated by R-loops.

80. 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).
81. Yüce, Ö. & 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).
82. Alzu, A. *et al*. Senataxin associates with replication forks to protect fork integrity across RNA-polymerase-II-transcribed genes. *Cell* **151**, 835-846 (2012).
83. Maldonado, E. *et al*. A human RNA polymerase II complex associated with SRB and DNA-repair proteins. *Nature* **381**, 86-89 (1996).
84. Duch, A. *et al*. Coordinated control of replication and transcription by a SAPK protects genomic integrity. *Nature* **493**, 116-119 (2013).
85. Alcasabas, A.A. et al. Mrc1 transduces signals of DNA replication stress to activate Rad53. *Nature Cell Biol.* **3**, 958-965 (2001).
86. Glover, T.W., Berger, C., Coyle, J. & Echo, B. DNA polymerase alpha inhibition by aphidicolin induces gaps and breaks at common fragile sites in human chromosomes. *Hum. Genet.* **67**, 136-142 (1984).
87. Zhang, H. & Freudenreich, C.H. An AT-rich sequence in human common fragile site FRA16D causes fork stalling and chromosome breakage in *S. cerevisiae*. *Mol. Cell* **27**, 367-379 (2007).
88. Gerhardt, J. *et al*. The DNA replication program is altered at the FMR1 locus in fragile X embryonic stem cells. *Mol. Cell* **53**, 19-31 (2014).
89. Letessier, A. *et al*. Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. *Nature* **470**, 120-123 (2011).
90. Ozeri-Galai, E. *et al*. Failure of origin activation in response to fork stalling leads to chromosomal instability at fragile sites. *Mol. Cell* **43**, 122-131 (2011).
91. Le Tallec, B. *et al*. Common fragile site profiling in epithelial and erythroid cells reveals that most recurrent cancer deletions lie in fragile sites hosting large genes. *Cell Rep.* **4**, 420-428 (2013).
92. Wilson, T.E. *et al*. Large transcription units unify copy number variants and common fragile sites arising under replication stress. *Genome Res.* **25**, 189-200 (2015).
93. 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).
This work documents transcription-replication conflicts in very long human genes and shows that the instability of common fragile sites located within those genes is dependent on transcription and R-loop formation.
94. Barlow, J.H. *et al*. Identification of early replicating fragile sites that contribute to genome instability. *Cell* **152**, 620-632 (2013).
95. Hoffman, E.A., McCulley, A., Haarer, B., Arnak, R. & Feng, W. Break-seq reveals hydroxyurea-induced chromosome fragility as a result of unscheduled conflict between DNA replication and transcription. *Genome Res.* **25**, 402-412 (2015).
96. Halazonetis, T.D., Gorgoulis, V.G. & Bartek, J. An oncogene-induced DNA damage model for cancer development. *Science* **319**, 1352-1355 (2008).
97. Miron, K., Golan-Lev, T., Dvir, R., Ben-David, E. & Kerem, B. Oncogenes create a unique landscape of fragile sites. *Nature Commun.* **6**, 7094 (2015).
98. Bartkova, J. *et al*. DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature* **434**, 864-870 (2005).
99. Valovka, T. *et al*. Transcriptional control of DNA replication licensing by Myc. *Sci. Rep.* **3**, 3444 (2013).

100. Srinivasan, S.V., Dominguez-Sola, D., Wang, L.C., Hyrien, O. & Gautier, J. Cdc45 is a critical effector of myc-dependent DNA replication stress. *Cell Rep.* **3**, 1629-1639 (2013).
101. Jones, R.M. *et al*. Increased replication initiation and conflicts with transcription underlie Cyclin E-induced replication stress. *Oncogene* **32**, 3744-53 (2013).

Acknowledgements

The authors would like to thank B. Gómez-González for her comments on 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, the Junta de Andalucía, the European Union (FEDER), Worldwide Cancer Research, and the European Research Council. The authors apologize to those whose work could not be cited owing to space limitations.

Competing interests statement

The authors declare no competing interests.

BOX 1. Fragile sites as hotspots of transcription-replication collisions

Fragile sites are genomic regions exhibiting constrictions or gaps in metaphase chromosomes following replication stress. They are categorized into two classes: rare fragile sites are found in <5% of individuals and arise from trinucleotide repeat expansion, and common fragile sites (CFSs), which are found in all individuals and are not associated with repeat expansion [86]. Fragile sites are frequently enriched in sequences that can stall DNA replication, such as AT-dinucleotide-rich sequences of high DNA flexibility in CFSs, as shown in yeast at FRA16D [87], or in other replication-attenuating sequences, as shown in human cells expressing the rare fragile site FRAXA at the *FMR1* locus [87, 88]. Scarcity in origins of DNA replication or inefficient replication activation in large genomic regions may explain the fragility of some CFSs [89, 90]. Interestingly, mapping of CFSs in several human cell lines revealed they are located mostly within large genes. Also, there is a high correlation between CFSs and recurrent chromosomal rearrangements observed in cancer cells, and a similar correlation emerged from the analysis of copy number variants, whose hotspots matched CFSs when located in large regions of active transcription in both human and mouse cells [91, 92]. These data suggest that concomitant transcription and replication may lead to fragility. Related to this is the observation that RNA-DNA hybrids

accumulate in the long *FHIT*, *WWOX* or *IMMP2L* genes, which harbour the CFSs FRA3B, FRA16D and FRA7K, respectively [93], as well as in the rare fragile sites FXN and FRAXA [61], among others [2]. Recently, chromatin immunoprecipitation analysis with the single-strand DNA binding protein replication protein A (RPA) following replication stress has allowed the identification and mapping of a new class of fragile sites, termed early replication fragile sites (ERFSs), which contrary to CFSs are located near replication origins and within actively transcribed genes, strengthening the possibility that their fragility result from transcription-replication conflicts [94]. Similarly, using the Break-seq **[G]** technique in cells following exposure to the replication-stress agent hydroxyurea, replication-induced double strand breaks were mapped preferentially at genes whose expression is~~[Au: OK?]~~ induced under replication stress conditions [95]. Therefore, genomes contain hotspots for transcription-replication collisions, which can manifest as different forms of fragility.

BOX 2. Oncogenes and transcription-replication collisions.

An oncogene refers to a gene that when mutated contributes to the development or progression of cancer, whereas the term proto-oncogene is reserved to its wild-type allele. Oncogenes generally regulate cell division, cell differentiation and/or cell death. This is the case of the oncogene c-Myc, which regulates transcription of several genes that control cell growth and cell cycle progression[2]. As replication stress and genomic instability are hallmarks of cancer cells [96], it seems plausible that oncogenes may increase the rate of transcription-replication conflicts, which will serve as a source of genomic instability. Altered expression of cyclin E or oncogenic Ras induces chromosomal fragility at sites that co-localize with large genes and only partially overlap with the canonical, replication stress-induced fragile sites [97]. Oncogene expression can negatively affect replication by promoting replication origin activation, as shown for c-Myc or cyclin E [98, 99]. Using DNA combing and cell-free extracts derived from *Xenopus laevis* eggs it has been shown that c-Myc increases activation of early-replicating origins, resulting in elevated fork collapse and subsequent DNA damage accumulation [100]. Although such replication fork collapses may occur independently of transcription, the excess of active replication forks may increase the probability of collisions. Consistent with this view, DNA damage resulting from replication impairment by cyclin E overexpression was partially suppressed by the transcription inhibitor cordycepin, suggesting that collisions can indeed contribute to oncogene-induced replication stress[101]. It would be important to determine the general relevance of this

phenomenon in cancer cells and to explore the possibility of using transcription-replication collisions as a selective target in cancer therapy.

Figure legends

Figure 1. Transcription and replication. A. A small portion of the double DNA helix is unwound by the RNA polymerase (RNAP) to enable transcription (known as "transcription bubble"). DNA unwinding by the RNAP generates positive and negative supercoiling**[G]**, which is alleviated by topoisomerases. In eukaryotes, transcription also involves chromatin modification and remodelling. The progression of RNA polymerase requires the activity of transcription elongation factors. The nascent RNA is co-transcriptionally processed by different factors. **B**. At the replication fork the DNA helicase minichromosome maintenance complex (MCM) opens the double helix and the DNA polymerases Polε and Polδ extend the leading and lagging strand, respectively. Synthesis of each new DNA molecule is initiated by the Polα-Primase complex (Polα-Pri). Lagging strand synthesis leads to the formation of ssDNA, which is coated with replication protein A (RPA). Fork progression requires the activity of several replication cofactors, including the clamp proliferating cell nuclear antigen (PCNA). DNA unwinding by the replication fork generates positive supercoiling, which is alleviated by topoisomerases. Replication also entails reassembly of recycled and de novo-synthesized nucleosomes at the newly synthesized DNA. ~~Dashed a~~Arrows indicate the direction of ~~fork progression and~~RNA and DNA polymerases synthesis.

Figure 2. Head-on and co-directional transcription-replication collisions. A. Progression in opposite directions of an RNA Polymerase (RNAP) and a replication fork leads to head-on collisions, which induce pausing and blockage of the replication fork and may lead to its collapse and the formation of DNA breaks . **B.** Progression of an RNA Polymerase and a replication fork in the same direction leads to co-directional collisions if the fork moves faster than the RNA Polymerase. Co-directional collisions can be resolved by displacement of the RNA Polymerase from the DNA. MCM, minichromosome maintenance complex; Polε, DNA polymerase ε; Polδ, DNA polymerase δ.

Figure 3. Conditions that affect the occurrence of transcription-replication collisions. **A.** Convergence of an RNA Polymerase (RNAP) and a replication fork when oriented head-on can lead to the accumulation of positive DNA supercoiling

between them, which induces pausing of the fork. **B.** The partial unwinding of DNA by the negative supercoiling generated behind the RNA polymerase can enable the formation of non-B DNA structures, such as G-quadruplexes, which may constitute an obstacle for replication fork progression. **C.** Other non-B DNA structures include RNA-DNA hybrids, which also may constitute an obstacle for fork progression. **(B and C)** Once a stable non-B DNA structure capable of blocking fork progression is co-transcriptionally formed, the direction of transcription or the presence of the RNA Polymerase itself would be in principle irrelevant for the formation of the transcription-replication collision. MCM, minichromosome maintenance complex; Polε, DNA polymerase ε; Polδ, DNA polymerase δ.

Figure 4. Mechanisms preventing transcription-replication collisions. A. Pausing of RNA polymerase (RNAP) is normally resolved by backtracking, which disengages the 3'-end of the RNA molecule from the active site and leads to back and forth sliding of the RNA polymerase. The GreA and GreB RNA cleavage factors stimulate the removal of the extruded RNA and the reactivation of transcription. Without GreA and GreB the RNA polymerase might stall and become an obstacle for the replication fork, leading to transcription-replication collisions. **B.** The replication fork barrier (RFB) site is a DNA sequence located near the 3'-end of rRNA genes and prevents transcription–replication conflicts in the budding yeast. DNA replication fork-blocking protein Fob1 is required for RFB activity as without it there is no replication fork arrest at the RFB, resulting in transcription-replication collisions. **C.** Chromatin remodelling by the FACT complex facilitates transcription as well as replication fork progression. Without FACT altered chromatin reorganization results in transcription-replication collisions. **D.** At pericentromeric regions co-transcriptional RNAi releases the RNA polymerase thereby allowing completion of DNA replication. Without the RNAi machinery, failure to release the RNA polymerase during S-phase results in transcription–replication collisions. DnaB, DNA replicase B; Polα, DNA polymerase III α; MCM, minichromosome maintenance complex; Polε, DNA polymerase ε; Polδ, DNA polymerase δ.

Figure 5. Resolving transcription–replication collisions to avoid genome instability~~by the DNA damage response[Au: OK?]~~. A stalled replication fork can activate the ATR-dependent checkpoint to solve collisions and avoid their consequences. Specific factors are recruited to resolve or prevent the obstacle, including DNA helicases (Rrm3), mRNA processing proteins or chromatin remodelling complexes (FACT)~~, DNA helicases (Rrm3) or mRNA processing proteins [Au: THO is not mentioned in the main text nor explained here, so we removed it, OK?]~~. In

addition, the RNA ~~Pol~~polymerase (RNAP) could be released at transcription termination sites, with the help of BRCA1 and helicase SETX, or the INO80 histone remodelling complex and transcription factor PAF1C complex, as a way to avoid the collisions. Similarly, at tRNAs transcription sites de RNA Pol III is directly evicted during S-phase. Stabilization and resumption of stalled forks at transcribed DNA regions can occur via DNA repair factors, such as those of the Fanconi anemia repair pathway including tumour suppressors BRCA1 and BRCA2. Finally different DNA repair pathways can act at collision sites, if these degenerate into DNA lesions. Although in some examples direct involvement of the ATR-dependent checkpoint has been reported, in other cases is yet unknown. MCM, minichromosome maintenance complex; Polε, DNA polymerase ε; Polδ, DNA polymerase δ.

Glossary:

Break-seq. Technique to map chromosome breaks based on DSB labelling and next generation sequencing.

Bromodeoxyuridine. Synthetic analogue of the thymidine nucleoside used to follow DNA synthesis.

CpG islands. Chromosomal regions with high density of non-methylated CpG sequences, often located at gene promoters.

DNA combing. A method for the analysis of single DNA molecules; used for studying DNA replication.

DNA damage response (DDR). Network of DNA damage repair and checkpoint factors that together to deal with DNA lesions.

G-quadruplexes. Four repeats of at least three guanines that can interact to form four-stranded DNA structures.

Hairpins. DNA structures in which a strand folds on itself and forms intrastrand base pairing.

non-B DNA. Any DNA structure that is different from right-handed double helix with 10 nucleotides per turn.

Phosphomimetic. Proteins with amino acid substitutions that simulate their phosphorylated state.

RecQ family. DNA helicase proteins defined by their helicase domain, which is essential for ATP binding and hydrolysis and the RecQ domain, which is required for DNA binding.

Replisome. A protein complex with helicase, primase and DNA polymerase activities that conducts DNA replication.

Supercoiling. Over- or under-winding of the DNA helix.

Torsional stress. Physical stress at the DNA molecule generated by over-rotation of the double helix; manifested as the accumulation of positive or negative supercoils.

Transcription-coupled repair. Subpathway of the nucleotide excision repair pathway that removes lesions from the template DNA strands at actively transcribed genes.

Triplex DNA (H-DNA). A single-stranded DNA region bound to the major groove of the DNA duplex ~~and~~ forming a three-stranded helix, normally at sequences with mirror symmetry.

γH2AX foci. Histone 2 variant that is phosphorylated (γH2AX) and forms nuclear foci, which are generally accepted as markers of DNA double-strand breaks.

Biographies:

Andrés Aguilera is Professor of Genetics in the University of Seville and Director of the CABIMER Research Centre, Seville (Spain). He obtained his Ph.D. in Seville in 1983. After two postdoctoral stays in Darmstadt (Germany) and New York (USA), he started his laboratory in 1991. His main research interests are the mechanisms by which replication stress, transcription and RNA processing and export interfere with genome stability, in particular the mechanisms mediated by R-loops. He is a member of EMBO, of a number of Research Centre International Scientific Evaluation Committees and Advisory Boards, as well as member of the editorial boards of several scientific journals.

Tatiana García-Muse is assistant professor of Genetics in Seville University, Spain. She obtained her Molecular Biology Ph.D. in the University Autónoma of Madrid in 2003. After a postdoctoral stay in Dr. Simon Boulton´s group in Clare Hall Cancer Research, UK, she joined Prof. Andrés Aguilera's group in the CABIMER in 2008. Her current research interests include the mechanisms underlying DNA damage response and DNA repair.

Online summary:

- Transcription and replication occur at high frequency in cells. Since they share the same DNA template a high incidence of encounters is expected between the transcription and replication machineries, which can cause transcription–replication conflicts, DNA damage and genomic instability.
- Cells have developed different strategies to reduce or prevent transcription-replication encounters, from genome organization favouring co-orientation of replication and transcription to specific mechanisms to avoid or resolve such collisions.
- Transcription-replication collisions can occur owing to *cis* structural features such as changes in DNA supercoiling, or secondary DNA structures including hairpins, G-quadruplexes or RNA-DNA hybrids, which have the capacity to hinder replication fork progression.
- The factors that minimize collisions include the transcription machinery itself and mRNA processing proteins, as well as factors that help or facilitate replication progression like DNA helicases and topoisomerases or chromatin remodelling complexes.
- The DNA damage response is able to sense a stalled replication fork caused by the transcription-replication conflicts and to promote different mechanisms that solve the collisions. This includes for example the removal of the RNA polymerase and the action of different repair pathways such as Fanconi anemia pathway.
- A better understanding of the dynamics of replication and transcription machineries will help clarify the importance of transcription-replication collisions as a source of ~~DNA damage and replication stress~~genomic instability and to open the possibility of using them as selective targets in cancer therapy.