Transcription as a Threat to Genome Integrity

Hélène Gaillard and Andrés Aguilera

Centro Andaluz de Biología Molecular y Medicina Regenerativa CABIMER, Universidad de Sevilla, Av. Américo Vespucio s/n, Sevilla 41092, Spain.

E-mails: gaillard@us.es and aguilo@us.es

Corresponding author: Andrés Aguilera, aguilo@us.es

Abstract

Genomes undergo different types of sporadic alterations including DNA damage, point mutations and genome rearrangements that constitute the basis for evolution. However, these changes may occur at high levels as a result of a cell pathology resulting in genome instability, a hallmark of cancer and a number of genetic diseases. In the last two decades evidence has accumulated that transcription constitutes an important natural source of DNA metabolic errors that can compromise the integrity of the genome. Transcription can create the conditions for high levels of mutations and recombination by its ability to open the DNA structure and remodel chromatin to make it more accessible to DNA insulting agents, and to become a barrier to DNA replication. Here we review the molecular basis of such events from a mechanistic perspective with particular emphasis on the role of transcription as a genome instability determinant.

Contents

INTRODUCTION

Genomes are subjected to different types of changes from point mutations to large genome rearrangements and chromosome number variation that constitute the basis for evolution. These changes accompany sexual reproduction via a developmentally controlled process, which reached its evolutionary peak in eukaryotic meiosis. However, all type of cells, including mitotically dividing cells, undergo genome changes that, even though occurring at extremely low level, may have considerable consequences in development and fitness. Such changes respond to a stochastic process triggered by lesions in the DNA generated in most cases by either endogenous metabolites or external genotoxic agents. Cells possess powerful systems to respond to such insults based on the sensing and repair of DNA damages and coupling to the cell cycle to warrant proper cell proliferation or apoptosis, which are grouped under the general term of DNA Damage Response (DDR) (1). However, when such systems fail DNA damage accumulates at high levels and the genome becomes highly unstable, which constitutes a hallmark of cancer and a number of genetic diseases (2).

DNA is not only the substrate of Replication, Repair and Recombination (3Rs), it is also the substrate of Transcription. Any DNA lesion or secondary structure able to stall a DNA polymerase is potentially able to block an RNA polymerase (RNAP) (3). Cells use this in their favor through the mechanism of Transcription-Coupled Repair (**TC-NER**, see the sidebar titled Transcription-coupled DNA Repair) (4). Nevertheless, transcription also creates the conditions for high levels of mutations (Transcription-Associated Mutation, TAM) and recombination (Transcription-Associated Recombination, TAR) as well as DNA breaks, chromosome rearrangements and chromosome loss by its ability to open the DNA structure and promote chromatin remodeling making DNA more accessible to genotoxic agents, and to become a barrier to DNA replication (5). Here we review the molecular basis of all such types

of events under the general term of Transcription-Associated Genome Instability (TAGIN). Recent reviews have been devoted specifically to TAM, TAR or to TAGIN as a general phenomenon (6-8) emphasizing the resulting mutation and rearrangement events. Here we review TAGIN from a mechanistic perspective referred to as the transcriptional activity in normal cells, but also as pathology of cells defective in different DNA and RNA nuclear processes, the potential role of TAGIN in tumorigenesis and genetic diseases being discussed at the end.

TRANSCRIPTION-ASSOCIATED GENOME INSTABILITY EVENTS

Single mutations

In the early 70's, transcription was shown to increase the reversion rates of the ICR191 mutagen in *Escherichia coli* (9), providing a first indication for TAM. Later work established the mutagenic potential of active transcription in bacteria and yeast by measuring the rates of spontaneous reversion on specific alleles under the control of inducible promoters (10, 11). A major limitation of reversion assays is that they detect only a specific subset of mutations, although this is also their strength since it enables the analysis of specific events. The reversion system originally used in yeast, which is based on a *LYS2* allele and detects one-base pair (bp) frameshift reversion, demonstrated that TAM frequency augments with increasing levels of transcription (11). The system used in *E. coli* (10) detects C-to-T transitions on the non-transcribed strand (NTS) of a mutant allele of the kanamycin-resistance gene (*KanR*) under the control of a regulated promoter in either orientation, allowing the monitoring of the reversion frequencies when the *KanR* allele was on the transcribed strand (TS) or the NTS. The results obtained with this system demonstrated that the frequency of C-to-T transitions rises with transcription and occurs preferentially on the NTS. Cytosine, which is the most-unstable base of nucleic acids, is frequently converted to uracil by spontaneous

deamination in living cells. Uracil persisting in the DNA during replication can mispair with adenine and cause C-to-T transitions (**Figure 1a**). Further studies using *LYS2*-based frameshift reversion assays in yeast and reversion assays that detect other kinds of mutations than C-to-T transitions in bacteria led to similar results, demonstrating that transcription induces mutations, particularly on the NTS (6, 8).

Consistent with these conclusions, transcriptional induction of regulated genes in response to starvation in bacteria correlates with increased mutagenesis within these genes (12). The empirical evidence of TAM has been reinforced by subsequent comparative genomic studies pointing into similar conclusions. Such studies revealed a bias in the distribution of specific base substitutions toward the coding strand in mammalian genes as well (13, 14), indicating again that base substitutions occur more frequently in the NTS during evolution. Specific TS repair mechanisms such as the one exemplified by **TC-NER** (see the sidebar titled Transcription-coupled DNA Repair) could contribute to reduce the frequency of TAM at the TS.

Forward mutation assays have also been used to gain insights into the spectrum of transcription-associated mutations. In these assays the majority of mutations were base substitutions under low transcription and insertion-deletions of few nucleotides within short tandem repeat sequences under high transcription, as shown in yeast (15) (**Figure 1a**). Altogether, these data plus posterior reports reviewed recently (6, 8) indicate that transcription may be mutagenic.

DNA breaks, recombination and chromosomal rearrangements

Evidence that recombination events are promoted by transcription in bacteria was obtained in *E. coli* by analyzing lambda-phage site-specific recombination (16). However the first example that transcription enhances homologous recombination (HR) was reported in

Saccharomyces cerevisiae, in which a genetic screen for genomic fragments stimulating mitotic recombination in a direct-repeat system identified a segment of the rDNA that was named *HOT1* (17). Later work has shown that the stimulation of recombination induced by the insertion of *HOT1* at ectopic locations in the yeast genome was caused by its ability to promote transcription by RNAPI (18). T1 and T4-phage mediated transduction is enhanced by transcription in bacteria (19), and RNAPII-driven transcription also induces TAR as first shown in yeast with a direct-repeat recombination system under the control of the inducible *pGAL* promoter (20) and in mammalian cells using two alleles transcribed at high and low levels (21). Analogous results were also obtained with an RNAPIII-transcribed system in *S. cerevisiae* (22). In the late 90's, *HOT1* and different recombination systems transcribed by RNAPII have been used as tools to provide evidence of TAR as a general phenomenon and to gain insights into its mechanistic basis (5). The vast majority of TAR events have been detected genetically in direct-repeat assays as deletions occurring majorly via single strand annealing (SSA) (**Figure 1***b*). As such events results from double-strand breaks (DSB), their increase is interpreted as the result of an increase of spontaneous DNA breaks. Specific genetic studies performed in yeast support this conclusion (23). Many other reports confirm that transcription enhances spontaneous recombination and DNA breaks at different regions in bacteria, yeast and human cells, as collected in a recent review (8). Collectively, the results support a model in which TAR would be caused by DNA breaks within the actively transcribed region.

The induction of DNA breaks, recombination and genome rearrangements by transcription has been shown in different situations, pointing to complex and varied sources of TAGIN that can be modulated by different factors. Thus, exogenous genotoxic agents increases TAR in direct-repeat recombination assays in yeast (24), convergent transcription significantly enhances deletions in chimeric *E. coli* plasmids containing the M13 phage (25),

and TAR is strongly stimulated when transcription collides with replication (26). In addition, specific yeast hyper-recombination mutants such as *hpr1*, altered in transcription elongation and messenger ribonucleoprotein particle (mRNP) biogenesis, specifically stimulate TAR (27), indicating that TAR may not be just the result of high transcription, as discussed next.

CHANGES IN DNA STRUCTURE POTENTIALLY RELATED TO TRANSCRIPTION-ASSOCIATED GENOME INSTABILITY

Topological constraints and DNA topoisomerases

Transcription induces positive supercoiling (over-winding) ahead and negative supercoiling (under-winding) behind the advancing RNAP (28) (**Figure 2a**). These local topological changes, which can be further constrained by DNA-binding proteins (e.g. transcription factors at promoters) and chromatin loops, may accumulate in highly transcribed regions and at sites of transcription start or termination, in particular at divergent or convergent genes, respectively. Such topological constrains may explain the increased TAR observed under convergent transcription in *E. coli* mentioned above (25).

Negative supercoils associated with transcription are primarily resolved by Topoisomerase 1 (Top1) (29). Concordantly, regions of under-wound DNA are enriched with Top1 and map to transcriptionally active chromatin in the human genome (30). Top2 activities are required to relax positive supercoils generated ahead of an elongating RNAP and act in addition to Top1 at highly transcribed genes and promoter regions (31). In the absence of Top1 and Top2 activities, torsional stress accumulates at transcribed regions leading to transient underwound DNA that might result in stretches of single-stranded DNA (ssDNA) behind the RNAP (**Figure 2a**). Indeed, co-transcriptional negative supercoiling produced in the absence of Top1 and Top2 is linked to increased recombination, as observed at the highly

repetitive rDNA locus and within a region located between convergent genes in yeast (32, 33). It is likely that this excess in negative supercoiling leads to more DNA breaks, probably due to a potential increase in ssDNA stretches that could increase the susceptibility to genotoxic agents or could facilitate the formation of non-B DNA structures that would evolve into a DSB after replication. Nevertheless, no definitive evidence has yet been provided that transcription is required for the increase in recombination caused by Top1-deficiency (see below), as the topological constrains generated during replication may be in principle sufficient to explain this phenomenon.

Also important in the context of genome instability, the reaction catalyzed by topoisomerases involves DNA breakage and covalent binding of the enzyme to one of the DNA termini, which are both reversed upon completion of the reaction. Failures in finishing the reaction lead to a DNA break with the topoisomerase trapped to one of the ends, a structure called topoisomerase cleavage complex (Top1cc or Top2cc). Such DNA breaks are, whether directly or indirectly, presumably responsible for the increased recombination and chromosome rearrangements observed upon treatment with topoisomerase inhibitors, such as camptothecin, but whether such Topcc structures are formed at higher levels in transcribed DNA sequences as a major source of TAR is not known yet. Nevertheless, the processing of abortive Top1-iniciated reactions has been proposed as a mechanism responsible for the specific TAM signature consisting of 2-bp deletions at short tandem repeats in yeast (34, 35) (**Figures 1*a* and 2*b***). Removal of the Top1cc intermediate would generate a 2-nucleotide gap that would be converted into a ligatable nick upon realignment of the DNA strands, leading to the 2-bp deletion after replication.

Finally, Top1 also functions as a site-specific endoribonuclease (36). In yeast, this activity on ribonucleotides incorporated erroneously in the DNA leads to modified ssDNA nicks (ribonucleoside-2'-3'-caclic phosphate-terminated nicks) that can be repaired in an error-

free fashion or can generate 2-nucleotide deletions when it occurs within a dinucleotide repeat sequence (37, 38). However, evidence for the contribution of this mechanism in TAM or TAR is yet scarce.

Non-B DNA structures

DNA can adopt alternative conformations that differ from the canonical B structure. These non-B DNA structures include hairpins, cruciforms, left-handed Z-DNA, triplex DNA (also called H-DNA), G-quadruplexes, and RNA-DNA hybrids. The formation of non-B DNA conformations requires unwinding of the DNA sequence, as occurs during replication, transcription or recombination, and is strongly favored by negative supercoiling (**Figure 2a**). Once unwound, the ability to adopt such structures depends on the DNA sequence. Sequences containing motives prone to form non-B DNA are widespread in the human genome and are associated with genome instability, behaving as hotspots for DSB that may subsequently trigger chromosomal rearrangements and being at the origin of several human genetic diseases (39, 40). Part of the genetic instability associated with non-B DNA structures relies in the interference with replication, by either blocking the replisome, providing substrates to nucleases, promoting DNA polymerase slippage, or interfering with Mismatch Repair (41).

However, transcription also appears to play a role in the formation of non-B DNA structures. In the first place, transcription increases the probability of non-B structures to be formed as a consequence of DNA supercoiling. Thus, the rates of instability at different trinucleotide repeats sequences – which are prone to form non-B structures such as hairpin, cruciforms or triplex DNA– is increased by negative supercoiling in *E. coli* (42). In the second place, studies of *E. coli* and yeast have shown that transcription of sequences capable to adopt non-B DNA conformation can induce genomic instability (43-45). In human cells, transcription promotes contraction of CAG and expansion of GAA repeat tracts (46, 47).

Noteworthy, these repeat-length changes were observed in the absence of replication. Replication-independent instability induced by large tracts of GAA repeats was also observed in yeast (48) and trinucleotide instability takes place in non-proliferative human tissues (e.g. brain) (40). This instability could rely on the accumulation of stalled RNAPs at repeated sequences, as non-B DNA structures formed within gene bodies can impede transcription elongation (3).

Finally, since RNAP stalling serves as a signal for TC-NER (4) prolonged stalling of the RNAP at a non-B DNA structure, in the absence of DNA lesions could trigger 'gratuitous' repair (**Figure 2***b*). Although NER normally proceeds error-free, its action in the absence of DNA damage puts genome stability at risk because it involves an ssDNA intermediate that can easily suffer damage and gap-filling DNA synthesis that might introduce errors. This is particularly relevant for DNA sequences containing repeats, as the putative formation of non-B DNA structure might impede the completion of the repair reaction and increase the error rates of DNA polymerases. In support of this idea, NER was shown to promote triplex-induced mutagenesis and contraction at CAG repeats (49, 50). Consistently, a reduction of genomic instability induced by cruciform-forming short inverted repeats was observed in yeast and mammalian cells lacking the NER XPF-ERCC1 (Rad1-Rad10 in yeast) nucleases (51). Other enzymes involved in NER, including the XPB and XPD helicases were shown to bind (and unwind in the case of XPD) G-quadruplexes (52). Therefore, DNA repair coupled to transcription could be misled by non-B DNA structures as a potential source of TAGIN (**Figure 2***b*). More work will be required to dissect the role of NER, and possibly other repair pathways, in TAGIN potentially induced by non-B DNA structures.

DNA DAMAGE ACCUMULATED AS A CONSEQUENCE OF TRANSCRIPTION

Non-enzymatically driven base damage during transcription

As indicated, the accumulation of negative supercoils behind the elongating polymerase favors DNA unwinding and may create stretches of ssDNA, to which genotoxic agents may access more efficiently. These regions of ssDNA in the gene body, which may be stabilized by the formation of secondary structures or RNA-DNA hybrids, are likely responsible for the increased vulnerability of transcribed DNA (**Figure 2**). Accordingly, treatment with DNA damaging drugs such as 4-NQO and MMS increased recombination of transcribed genes in yeast (24) and UV-C and bezo[a]pyrenediolepoxide-induced mutagenesis was significantly increased when a reporter gene was transcribed in mouse embryonic stem cells (53). The results are consistent with the idea that transcribed DNA is more easily damaged by external genotoxic agents. In support of a role for transcription-associated DNA damage in TAM, the analysis of transcription-dependent reversion rates in different *S. cerevisiae* mutant background revealed that TAM largely depends on the activity of Rev3, the catalytic subunit of the error-prone translesion DNA polymerase ζ (11) and increases in DNA repair deficient strains (54).

Notably, accumulation of apurinic/apyrimidinic (AP) sites is enhanced in highly-transcribed DNA and largely derived from the removal of uracil from the DNA (55). Uracil can be present in DNA as a result of spontaneous cytidine deamination or of mis-incorporation of deoxyuridine monophosphate (dUMP) during replication. In support of the latter, a reduction in TAM has been observed in conditions in which intra-cellular dUTP concentrations were kept low. Interestingly, increased cytidine deamination occurs within UV-induced photolesions located on the TS of active human genes (56), in agreement with the idea that transcription-dependent conformational changes in the DNA increases its susceptibility to being damaged.

Enzyme-mediated instability

Specialized cytidine deaminases belonging to the AID/APOBEC (activation-induced cytidine deaminase/apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like) superfamily provide the best example of enzyme-mediated DNA damage at active genes in vertebrates. AID expression is restricted to B cells, where it initiates the processes of somatic hypermutation (SHM) and class switch recombination (CSR) at the switch (S) regions of the immunoglobulin genes required for antibody diversification (57). Transcription is required for both processes, enabling AID-catalyzed deamination of cytidine to uridine in the transiently exposed ssDNA that will result in a DSB responsible for CSR. The formation of R loops behind the RNAP that are stabilized by G-quadruplexes presumably generated at the G-rich displaced ssDNA (58), may contribute to make the DNA strands more accessible to AID at the S regions. Thus, SHM is a landmark for TAM that relies on transcription-associated increase of DNA accessibility to a mutagenic agent, in this case a nucleotide-modifying enzyme.

Beyond its normal roles on immunoglobulin genes, AID can erroneously act on 'off-target' genes promoting chromosome translocations and generating mutations that may activate oncogenes (59). Off-target AID activity mostly occurs downstream of the transcriptional start site of a subset of transcribed genes, in agreement with the requirement of transcription for its activity. However, transcriptional activity may not be sufficient for AID targeting. Physical association of AID with the Spt5 transcription elongation factor enables AID targeting at stalled RNAPII (60), which may be enriched at off-target genes. In support of this hypothesis, off-target AID activity was recently found to occur within regions of overlapping sense and antisense transcription (61) as well as at transcribed enhancers and promoters (62). Indeed, a positive correlation between mutation rates and transcription was

demonstrated in mammalian cells using a reversion assay, in hyper-mutating cell lines expressing AID (63), consistent with the idea that the accumulation of negative supercoils behind the elongating polymerase favors DNA unwinding, creating the stretches of ssDNA on which AID would act (64). This has been recapitulated with the ectopic expression of AID in *E. coli* and yeast, supporting the importance of co-transcriptionally formed secondary structures and ssDNA stretches in AID-mediated genomic instability as determined by single mutations, deletions and translocations (65-67).

APOBEC family members can be induced in several cell types as part of the innate immune response and are important to restrain viral infections and retrotransposon mobility. Under some yet poorly defined circumstances, APOBEC enzymes - which act specifically on cytidines within ssDNA - may target nuclear DNA (68). Due to their sequence preferences, APOBEC deaminases generate characteristic mutation signatures that in some cases correlate with APOBEC expression levels in cancerous cells (69, 70) or are enriched in the NTS of highly expressed genes (71). This together with the recent finding that expression of APOBEC proteins in yeast leads to clustered mutations in active genes across the genome (72) suggests that the impact on genome instability of APOBEC family members might also occur in association with transcription.

AID and APOBEC are good examples of enzymes that better access transcribed DNA and are potentially responsible for TAGIN but these enzymes are specific for vertebrates. Studies on whether transcription may directly affect the accessibility to general nucleases and DNA modifying enzymes are scarce. However, evidence shows that transcription may also generate structures that are target of specific nucleases, as in the case of cruciform DNA (51) or R loops (73). Importantly, the potential of transcription to block the progression replication forks (RF) that in turn could demand the action of nucleases or other enzymes to facilitate

replication resumption may also be an important source of enzyme-mediated instability associated with transcription, as discussed next.

REPLICATION AS A MAJOR PLAYER IN TRANSCRIPTION-ASSOCIATED RECOMBINATION AND REARRANGEMENTS

Replication fork stall and restart at transcribed DNA regions

Obstacles on the template DNA, such as DNA lesions, secondary DNA structures or DNA-bound non-histone proteins can impede RF progression. Transcription complexes represent a prominent example of DNA-bound proteins hindering RF progression, as first shown in bacteria and yeast (74, 75). Replication impairment is a major issue in dividing cells and needs to be solved to avoid persistent RF stalling, a manifestation of replication stress that represents a major source of genomic instability (**Figure 3a**). Replication checkpoint factors act on stalled RF and promote their stabilization ensuring that the replisome remains associated with the template and is proficient in resuming replication, eventually implicating translesion DNA synthesis, transitory template switching to the sister chromatid, or transient replication fork reversal to bypass the obstacle (41). However, RF may break thus activating the intra-S checkpoint, and/or DNA polymerases can dissociate from the DNA leading to RF collapse (**Figure 3a**). RF breakage and increased HR can indeed be detected using systems in which ssDNA breaks are induced either enzymatically or by specific mutations in the Rad3/XPD DNA helicase, as shown in *S. cerevisiae* (76, 77). However, it is also likely that a blocked RF does not necessarily break, as has been shown in *Schizosaccharomyces pombe* using a strong protein barrier (78) in which case replication restart may proceed via template switching. In *S. pombe* DNA synthesis at HR-restarted forks is error-prone at an early stage, RF slippage and template switching occurring frequently, and thus contributing to the generation of GCR (79-81).

The repair of a broken RF via HR initiated by invasion of the resected 3'-end into the homologous sister chromatid would lead to RF restart (**Figure 3a**). However, HR may occur at ectopic sites sharing homologies, leading to a genetically detectable hyper-recombination phenotype as well as chromosomal rearrangements depending on the location of the recipient molecule (**Figure 1b**). Transcription may either enhance damage capable of blocking RF progression or may by itself constitute an impediment to normal RF progression as the source of transcription-associated DNA breaks and recombination **(Figure 3b)**. The importance of transcription in reducing RF progression as a source of instability has been shown in human cells depleted of Top1 or the chromatin reorganizing factor FACT in which either RF progression as determined by DNA combing or DNA breaks as determined by single-cell electrophoresis were suppressed by inhibiting transcription (82, 83) **(Figure 3b).**

Finally, repair of one-ended DSB as the one putatively generated at a broken RF may also occur by break-induced replication (BIR), a form of non-reciprocal HR that frequently leads to loss of heterozygosity and chromosomal rearrangements (84-86). Although it is unlikely that BIR represents a *bona fide* form of RF restart, because it is initiated by DSB not necessarily associated with replication, BIR may also occur at microhomologies between a resected single-stranded 3'-end and an ssDNA gap, structures that are also found at RFs and stalled transcription complexes. Interestingly, DNA synthesis by BIR has been shown to be mutagenic (87, 88), which reveals that this process may also be a potential source of instability in eukaryotes potentially associated with transcription.

Head-on versus co-directional collisions between replication and transcription

Depending on the orientation of a particular gene relative to the origin of replication, collisions between the transcriptional and replication machineries can occur in a head-on orientation (genes encoded on the lagging strand) or a co-directional orientation (genes

16

encoded on the leading strand). The detrimental potential of head-on conflicts has long been recognized, and they constitute a stronger hindrance for replisome progression than co-directional ones in vitro (89). A stronger impact of head-on versus co-directional collisions both on recombination and RF progression has been shown in direct-repeat constructs in *S. cerevisiae* in which RF stalling visualized by two-dimensional gel electrophoresis correlated with an increase in TAR (26). TAR could be a consequence of RF restart after a transcription-replication encounter and indeed, such a restart has been shown to require the RecA HR factor in *Bacillus subtilis* (90). Co-directional conflicts are frequent in prokaryotes as the replication machinery proceeds more rapidly than transcription complexes, and RF progression does not seem to be profoundly altered in normal conditions, as recently reviewed (91). The organization of bacterial genomes, in which most genes and virtually all rDNA loci are found in co-directional orientation with replication, contributes to keep head-on collisions low. This may be relevant for evolution as genes located in head-on orientation show faster adaptive variations due to increased TAM (92) and the doubling time of *E. coli* strains in which the highly transcribed *rrn* operon is replicated in head-on orientation, opposite to normal, is twice that of normal cells (93).

In vitro studies have shown that a replisome blocked as a consequence of a frontal encounter with transcription remains stable and can resume replication in the presence of accessory factors such as the T4 Dda helicase or the Mfd translocase (89, 94). In *E. coli*, the Rep, UvrD and DinG accessory helicases may displace proteins ahead of advancing RF, and are required for replication through the rDNA locus placed in head-on orientation (95) whereas the *B. subtilis* accessory helicase PcrA assists RF progression through protein-coding, rRNA and tRNA genes (96). As in bacteria, accessory DNA helicases act at RF and enable their progression by removing obstacles in eukaryotes. The best-characterized example is the budding yeast Rrm3 helicase, which is part of the replisome and promotes replication of

protein-bound DNA, including tRNA genes (97, 98). RF pausing occurs frequently in yeast, most pausing sites coinciding with the coding region of highly transcribed RNAPII genes (99). The relevance of transcription as an important barrier to RF progression and recombination is even more evident in transcription-elongation defective THO mutants (100) or in the absence of the Rrm3 helicase (26, 99). Therefore, it seems that collisions with transcription, in particular in head-on orientation, demand the action of specific factors to allow RF progression and maintenance of genome integrity. Indeed, the fact that replication termination takes place at fork pausing elements, of which the majority coincides with transcription clusters (101), supports the idea that transcription may suppose an important barrier to RF progression that cells can also use to their benefit.

The deleterious effects of potential collisions, as manifested in different forms of TAGIN may be enhanced by the accumulation of torsional stress, which can also compromise the integrity of the RF (26, 82, 83, 102). This may be particularly relevant in cases in which transcribed genes are gated to the nuclear envelope, as shown in yeast (103), since it considerably restrains the rotation freedom of the template DNA. A mechanism has been described to minimize the torsional stress resulting from concomitant replication and transcription at the nuclear pores, in which the activation of the replication checkpoint leads to the release of transcribed genes from the nuclear envelope (104). Thus, the DNA would gain more flexibility in resolving the transcription-blocked RF. The possibility that a number of transcription-associated DNA breaks may occur in the proximity of the nuclear periphery is yet an open question that needs to be properly addressed.

Avoiding collisions that threaten genome stability

In eukaryotes, initiation of DNA replication takes place at multiple origins on each chromosome and is a tightly controlled event restricted to the S phase of the cell cycle, thus

allowing a potential temporal separation of transcription and replication. Some genes, however, need to be transcribed during S phase, including histone, rRNA and tRNA genes, which creates a challenge to the cell. In bacteria, the challenge would be even greater since transcription and replication occur concomitantly. Cells use different ways to avoid potentially deleterious transcription-replication collisions.

A strategy to avoid such collisions is to restrain the amount of arrested RNAPs, which may furthermore lead to backed-up arrays of stalled transcription complexes. In support of this idea, bacterial transcription elongation factors (e.g. DskA, GreA/B, Mfd) are required to resolve replication-transcription conflicts in vivo (105, 106). Similarly, mutants deficient in Rho-dependent transcription termination suffer DSB that depend on replication, suggesting that termination factors function in the release of obstructing RNAP during replication (107). Replisome pausing, which is primarily caused by DNA-bound RNAP, is a frequent event during bacterial replication (108), and only remains innocuous if replication can resume normally owing to the activities of accessory helicases and transcription co-factors. In situations in which transcription elongation is compromised, stalling and subsequent backtracking of the RNAP can lead to an arrested RNAP. Collision of the replication machinery with such arrested RNAP was shown to generate DSB and thus represents a threat to genomic stability (109). Indeed, proteins of the replication restart machinery accumulate at the highly transcribed rDNA locus in *B. subtilis*, indicating that co-directional collisions are not always harmless (110).

In eukaryotes, one way to avoid collisions between transcription and replication is to restrain replication at transcribed regions. This can be achieved by the presence of an RF barrier (RFB) downstream of highly transcribed genes, the best-characterized example being found at the budding yeast rDNA locus, the region at which transcription raises the major challenge to replication. The budding yeast rDNA locus is composed of about 150 rDNA

repeats, each one comprising the RNAPI-transcribed 35S and the convergent RNAPIII-transcribed 5S pre-rRNA genes. About half of the repeats are transcribed at a given time and both the active and the repressed fractions are important for the stability of the locus (111, 112). The rDNA intergenic spacer contains an origin of replication (rARS) and a DNA element acting as RFB downstream of the 35S gene (113). The RFB possess a polar RF blocking activity that is mediated by binding of a specialized protein called Fob1 and allows avoiding frontal collisions between RNAPI and the replication machinery at the 35S gene (114). The sophisticated organization of the rDNA region evidences the importance of avoiding transcription-replication collisions, and how frequent can such collisions potentially be in the rDNA region. Interestingly, only a subset of the rARS are actually fired to initiate replication in a given S phase, active rARS localizing downstream of transcriptionally active genes (115).

Recent work has shown that the Sir2 and the Rpd3 histone deacetylases control the activation of rARS, Sir2 having a repressive effect that can be counteracted by Rpd3 (116). Consequently, Sir2-deficient yeast cells have a higher density of active RF at the rDNA regions, a feature particularly relevant given the fact that *sir2* mutants show a strong increase in rDNA recombination (117). Even though in the past such rDNA hyper-recombination in *sir2* cells was believed to be due to a better accessibility of the HR machinery into a more open chromatin at the rDNA locus, rDNA hyper-recombination in *sir2* cells may be a consequence of a significant increase in the frequency of collisions. Nevertheless, Sir2 is also known to silence a bidirectional RNAPII promoter called E-pro, which locates close to the RFB in the intergenic rDNA spacer. A low rDNA copy number has been shown to induce transcription from E-pro, which in turn leads to cohesin dislodgment (118, 119). Transcription-induced dissociation of cohesin consequently promotes unequal recombination

events between rDNA repeats allowing their amplification, providing an additional mechanism by which *sir2* cells enhance rDNA copy number change.

Another strategy to avoid collisions is to act on the replication machinery in conditions requiring extensive transcription, as recently described in conditions of osmostress, in which a large number of stress-responsive genes are induced. The stress-activated protein kinase Hog1 was shown to achieve a delay in origin firing and a slowing down of RF progression by directly targeting the Mrc1 component of the replication complex (120).

However, opposite strategies in which transcription is repressed prior to the passage of the replication machinery have also been described. Thus, replication and intra-S checkpoint proteins act on the Maf1 transcriptional repressor and consequently lead to a global repression of tRNA gene transcription in budding yeast (121). Particularly interesting is the case of the RecQL5 DNA helicase involved in genome integrity maintenance. In addition to its functions at RF and/or recombination intermediates, the mammalian RecQL5 helicase has been proposed to down-regulate transcription during S phase by means of direct binding to RNAPII (122). Importantly, RecQL5 and the transcription elongation factor TFIIS compete for the same binding site on RNAPII and RecQL5 promotes elongation, avoiding RNAPII stalling and backtracking (123, 124). This indicates that RecQL5 could contribute to the avoidance of collisions by a mechanism that is not necessarily coupled to replication. Along the same line, a slowdown of RF progression was observed at a transcribed locus in RNAPII mutants in which transcription is ineffective and correlates with an increase in recombination (125). Therefore transcription of a DNA region may block RF progression as a major cause of TAGIN.

R LOOP-MEDIATED TRANSCRIPTION-ASSOCIATED GENOME INSTABILITY

Co-transcriptional R loops as a source of genome instability

One important source of TAGIN is co-transcriptional R loops, which are formed by reinvasion of the nascent RNA molecule into the DNA template (**Figure 4a**). Although R loops are not necessarily responsible for TAGIN, evidence suggests that their relevance in genome instability is higher than previously foreseen. They were first shown in *S. cerevisiae* as responsible of the high increase in transcription-dependent recombination of *hpr1* mutants of the THO complex involved in mRNP biogenesis and export (126). Transcription-elongation impairment is also a feature of these mutants, suggesting that such impairment may also contribute to TAR. In chicken DT40 and human HeLa cells, depletion of the SRSF1 RNA processing factor was also shown to cause genome rearrangements and accumulation of R loops (127).

One possible explanation for the role of these RNA-binding proteins in R loop prevention is its putative role in protecting the RNA and assembling the mRNP, so that this would reduce its capacity to react with the DNA template. This is supported by the observation that overexpression of other RNA-binding proteins such as yeast Tho1 and Sub2 and human RNPS1 suppresses the instability associated with yeast THO mutants and human SRSF1-depleted cells (128, 129). Further research in yeast and human cells have generated an important body of results showing that a number of RNA metabolism and processing factors cause both R loop accumulation and genome instability as measured by DSB accumulation detected by γH2AX, genome rearrangements or minichromosome loss as well as RNA-DNA hybrid detection via the specific anti-RNA-DNA monoclonal S9.6 antibody (130-134). Therefore, even though R loops also have a physiological role in transcription activation and termination as well as in replication initiation as reviewed recently (135), they are a major contributor to TAGIN.

From co-transcriptional R loops to genome instability via replication

R loops can enhance the occurrence of single mutations, however most experimentally reported cases indicate that R loops are related to DNA break-induced instability, such as recombination, genome rearrangements and chromosome instability. The ssDNA present in an R loop would be in principle highly susceptible to DNA genotoxic agents, but the way by which an R loop can lead to a recombinogenic DNA break may be diverse. The most accepted mechanism is the one relying on the potential capacity of an R loop to stall or block RF progression (**Figure 4***a*). RF stalling and/or blockage has been shown to occur in yeast THO mutants at the *lacZ* 3'-end region (136), where R loops are preferentially accumulated (126). In addition, impaired replication in *Caenorhabditis elegans thoc2* mutants is partially suppressed by microinjection of RNase H1, which degrades the RNA moiety of RNA:DNA hybrids (137). Similarly, T7 promoter-driven transcription of S regions of the immunoglobulin genes in bacterial systems have been shown to pause or stall replication (138). In bacteria, defective replication through the *rrn* locus is suppressed by RNase H1 overexpression, suggesting that transcription-replication collisions are promoted by R loops (95). Other studies indeed show that R loops mediate genome instability as a result of transcription-replication conflicts both in human and yeast cells (82, 83, 100, 137, 139-141), or that R loops accumulate at triplet repeats often present at rare **fragile sites** (see the sidebar titled Fragile Sites) when they are transcribed under the T7 promoter in bacteria (142, 143), provided that fragile sites are mediated by stressed replication. Importantly, the activity of Top1 was also shown to restrain the deleterious effect of transcription-replication collision by preventing the formation of R loops (82, 144). Interestingly, the bacterial replisome can use the nascent RNA transcript as a primer to restart replication after colliding with transcription (145) and persistent R loops have been recently shown to initiate unscheduled replication in

the rDNA locus (146), revealing a novel kind of transcription-dependent threat to genome stability. It would be important to know whether or not this type of event can contribute to TAGIN in RNAPII-transcribed genes.

An unsolved question is whether an R loop is sufficient to provoke RF stalling or a more complex chromatin structure is involved. The observations that R loops in *S. cerevisiae*, *C. elegans* and human cells are linked to chromatin marks of condensation, such as histone H3S10P that correlates with R loops and can be diminished by RNAse H1 overexpression suggests that it may not be just the R loop but an R loop-mediated chromatin reorganization causing a more compacted or closed chromatin that is responsible for RF stalling and genome instability (147) (**Figure 4a**). Indeed, the FACT reorganizing complex is required for the RF to move through a transcribed region and in its absence R loops accumulated, favoring the idea that R loop-mediated TAR and DNA breaks are in part caused by an irregular chromatin structure (83). Importantly, rare fragile sites such as FRAXA in the *FMR1* gene and frataxin (*FXN*), which are associated with fragile X syndrome and Friedreich ataxia, respectively, accumulate R loops and undergo histone H3 K9me2 responsible for the silencing that causes the disease (148, 149). A similar, but non-pathological, R-loop-dependent silencing mechanism occurs over mammalian gene terminators (150).

If an R loop-containing region is encountered by an RF, the cell needs to actively promote the passage of the fork over that region. The observation that the DSB repair factor and tumor suppressors BRCA1 and BRCA2 partially prevent the accumulation of R loops in human cells suggests that the DDR may participate in this process (151, 152). Since BRCA2 plays a role stabilizing RF and is the FANCD1 component of the Fanconi Anemia (FA) pathway of repair, it is possible that the FA pathway contributes to removing R loops by promoting resumption of the RF stalled at the R-loop-containing region (M. García-Rubio, C. Pérez-Calero, S. Barroso, E. Tumini, E. Herrera-Moyano, I. Rosado and A. Aguilera,

unpublished data). The possibility exists therefore that replication restart and HR can indirectly contribute to the removal of a fraction of R loops.

Co-transcriptional R loop-mediated instability in the absence of replication

TAGIN mediated by R loops seems to also be possible in the absence of replication. This could be supported by the fact that NER factors, including TC-NER functions and the XPG/Rad2 nuclease are required for the formation of DNA breaks at R loops (73) (**Figure 4a**). It is likely that such factors could act at the R loop generating ssDNA breaks or gaps in the absence of replication, which can eventually lead to DSB upon replication. Consistent with this idea, co-transcriptional R loops can initiate DNA breaks in non-growing bacteria (153). Also, the requirement for BRCA1 to recruit the SETX RNA-DNA helicase at transcription termination sites to potentially remove R loops accumulated at the 3' end of human genes, supports indeed that R loops with a potential physiological role can compromise genome integrity without a link to replication (154). In favor of this idea, it has recently been found that the ATM checkpoint protein, which sense DSB, is activated in an R loop-dependent manner by transcription blocking lesions, such as those generated by UV light, as a key event of the DDR in non-replicating cells (155). Therefore, it is likely that an R loop can lead to DNA breaks by different means. This could depend on the cell cycle stage at which they are accumulated, their local effect on the chromatin structure, their location along the genes (whether at 5' or 3'-ends), or their size (135). These features would need to be explored further not only to understand how R loops contribute to TAGIN but also to determine what is the percentage of TAGIN events that are mediated by R loops in normal cells and the key cellular functions responsible for their prevention or processing.

TRANSCRIPTION-ASSOCIATED INSTABILITY IN HUMAN DISEASE AND CANCER

Since replication stress has been reported in pre-cancerous and cancerous cells and evidence supports the idea that it can drive tumorigenic processes (156, 157), it is expected that transcription conditions leading to replication stress and TAGIN might also be related to cancer. This could be for instance the case of tumorigenesis caused by expression of the cyclin E oncogene (158). In cells overexpressing the cyclin E oncogene, concomitant enhancement of transcriptional activity and replication could generate a large amount of torsional stress, impeding RF progression and possibly leading to abundant reversed forks (159). Transcription-replication encounters might also be relevant in the case of MYC oncogene expression, which leads to both global increases in transcriptional activity and to the induction of premature origin firing (160). However, whether transcription-replication interference contributes to the replication stress and TAGIN associated with oncogene expression remains to be addressed.

In addition, mutations in a number of genes important for RF restart and repair, such as *BRCA1* and *BRCA2,* lead to strong cancer predisposition. Appealing new data converge in a role for both BRCA proteins in the prevention or repair of damage generated by co-transcriptional R loops formation (151, 152, 154), suggesting that a major function of BRCA proteins may be to prevent accumulation of R loops that would otherwise lead to substantial genomic instability. Therefore, co-transcriptional R loops may be a source of TAGIN as well as an important tumorigenic event, and some tumor suppressors may act by preventing their occurrence.

At a different level, breakpoint analysis of common translocations associated with cancer (161), has revealed a strong correlation with **fragile sites** (see the sidebar titled Fragile

Sites). Provided the evidence discussed above, the finding that active transcription contributes to expression of both early and common fragile sites (141, 162) emphasizes the importance of TAGIN in cancer. Compelling evidence of the connection between TAGIN and cancer is provided by the human RECQL5 helicase, which acts on both the RNA polymerase and the RF to avoid deleterious collisions (124), since *Recql5* knockout mice are cancer prone and show increased chromosomal instability (163).

TAGIN may also be associated with a number of neurodegenerative diseases. A large number of human neurodegenerative and neuromuscular diseases including Huntington's disease, myotonic dystropohy 1, Fragile X syndrome and Fridreich's ataxia, are caused by expansions of unstable trinucleotide repeats at specific genome locations (40). Noteworthy, the concerned repeats all locate in transcribed regions, albeit not necessarily within coding sequences. Using engineered repeat-containing constructs it was shown that transcription promotes both trinucleotide contractions and expansions (46, 47) and that CAG repeat expansions correlate with R loops (143, 164) and with transcription in different mouse tissues (165). As with cancer, the data suggest that TAGIN is an important element promoting the appearance of the disease, even though it is unlikely to be responsible for the clinical manifestations.

SUMMARY POINTS

1. DNA regions undergoing transcription have a higher probability of mutating or suffering DNA breaks resulting in recombination and genome rearrangements.

2. Transcription-associated genome instability is of great relevance in cell proliferation and differentiation, provided the link of genome instability with cancer and a number of cancer-prone genetic diseases; it supposes that RNA-coding DNA sequences may accumulate more variability and changes than the rest of the genome.

3. TAGIN may be due to a major accessibility of genotoxic agents to transcribed DNA, caused by the probable opening of DNA strands promoted by local co-transcriptional negative supercoiling or chromatin remodeling.

4. Transcription and derived DNA and chromatin structural changes may have the ability to block the progression of the RF and to cause genome instability; transcription may be a natural source of replication stress and a potential activator of the DDR, in particular under pathological conditions.

5. TAGIN may also be partially due to the possible aberrant action of different DNA repair pathways on co-transcriptional structures mistaken as false DNA damage.

FUTURE ISSUES

1. We need to define the most prominent co-transcriptional structures triggering RF breakage, whether a stacked RNAP, a closed chromatin structure, an RNA-DNA hybrid, a co-transcriptionally formed ssDNA break or DNA damage, or a transcription-mediated defect in the DDR. It is likely that such a structure is different depending on the cell pathology involved.

2. We need to establish which cellular functions, from transcription factors to RNA processing an export factors and from DNA replication proteins to DDR factors, trigger TAGIN when they are defective.

3. We need to understand how transcription may cause different types of events from deletions to large chromosomal rearrangements, whether there is a preferential cell cycle phase favoring one type of event versus another or whether this has to do with either the specific structural features of the transcribed DNA or with the type of RNA generated (whether tRNA, rRNA, mRNA, lncRNA, etc.).

4. Deciphering the mechanisms of TAGIN is key to gain a more complete understanding of the causes and consequences of genome instability and its putative role in cancer as well as of the biochemical steps and pathways that may initiate a tumorigenic process.

DISCLOSURE STATEMENT

ACKNOWLEDGMENTS

LITERATURE CITED

1. Jackson SP, Bartek J. 2009. The DNA-damage response in human biology and disease. *Nature* 461: 1071-8

2. Gaillard H, Garcia-Muse T, Aguilera A. 2015. Replication stress and cancer. *Nat Rev Cancer* 15: 276-89

3. Belotserkovskii BP, Mirkin SM, Hanawalt PC. 2013. DNA sequences that interfere with transcription: implications for genome function and stability. *Chem Rev* 113: 8620-37

4. Hanawalt PC, Spivak G. 2008. Transcription-coupled DNA repair: two decades of progress and surprises. *Nat Rev Mol Cell Biol* 9: 958-70

5. Aguilera A. 2002. The connection between transcription and genomic instability. *EMBO J* 21: 195-201

6. Jinks-Robertson S, Bhagwat AS. 2014. Transcription-associated mutagenesis. *Annu Rev Genet* 48: 341-59

7. Kim N, Jinks-Robertson S. 2012. Transcription as a source of genome instability. *Nat Rev Genet* 13: 204-14

8. Gaillard H, Herrera-Moyano E, Aguilera A. 2013. Transcription-associated genome instability. *Chem Rev* 113: 8638-61

9. Herman RK, Dworkin NB. 1971. Effect of gene induction on the rate of mutagenesis by ICR-191 in Escherichia coli. *J Bacteriol* 106: 543-50

10. Beletskii A, Bhagwat AS. 1996. Transcription-induced mutations: increase in C to T mutations in the nontranscribed strand during transcription in Escherichia coli. *Proc Natl Acad Sci U S A* 93: 13919-24

11.   Datta A, Jinks-Robertson S. 1995. Association of increased spontaneous mutation rates with high levels of transcription in yeast. *Science* 268: 1616-9

12.   Wright BE, Longacre A, Reimers JM. 1999. Hypermutation in derepressed operons of Escherichia coli K12. *Proc Natl Acad Sci U S A* 96: 5089-94

13.   Green P, Ewing B, Miller W, Thomas PJ, Green ED. 2003. Transcription-associated mutational asymmetry in mammalian evolution. *Nat Genet* 33: 514-7

14.   Polak P, Arndt PF. 2008. Transcription induces strand-specific mutations at the 5' end of human genes. *Genome Res* 18: 1216-23

15.   Lippert MJ, Freedman JA, Barber MA, Jinks-Robertson S. 2004. Identification of a distinctive mutation spectrum associated with high levels of transcription in yeast. *Mol Cell Biol* 24: 4801-9

16.   Ikeda H, Matsumoto T. 1979. Transcription promotes recA-independent recombination mediated by DNA-dependent RNA polymerase in Escherichia coli. *Proc Natl Acad Sci U S A* 76: 4571-5

17.   Keil RL, Roeder GS. 1984. Cis-acting, recombination-stimulating activity in a fragment of the ribosomal DNA of S. cerevisiae. *Cell* 39: 377-86

18.   Voelkel-Meiman K, Keil RL, Roeder GS. 1987. Recombination-stimulating sequences in yeast ribosomal DNA correspond to sequences regulating transcription by RNA polymerase I. *Cell* 48: 1071-9

19.   Dul JL, Drexler H. 1988. Transcription stimulates recombination. II. Generalized transduction of Escherichia coli by phages T1 and T4. *Virology* 162: 471-7

20.   Thomas BJ, Rothstein R. 1989. Elevated recombination rates in transcriptionally active DNA. *Cell* 56: 619-30

21.   Nickoloff JA. 1992. Transcription enhances intrachromosomal homologous recombination in mammalian cells. *Mol Cell Biol* 12: 5311-8

22. de la Loza MC, Wellinger RE, Aguilera A. 2009. Stimulation of direct-repeat recombination by RNA polymerase III transcription. *DNA Repair (Amst)* 8: 620-6

23. Gonzalez-Barrera S, Garcia-Rubio M, Aguilera A. 2002. Transcription and double-strand breaks induce similar mitotic recombination events in Saccharomyces cerevisiae. *Genetics* 162: 603-14

24. Garcia-Rubio M, Huertas P, Gonzalez-Barrera S, Aguilera A. 2003. Recombinogenic effects of DNA-damaging agents are synergistically increased by transcription in Saccharomyces cerevisiae. New insights into transcription-associated recombination. *Genetics* 165: 457-66

25. Vilette D, Uzest M, Ehrlich SD, Michel B. 1992. DNA transcription and repressor binding affect deletion formation in Escherichia coli plasmids. *EMBO J* 11: 3629-34

26. Prado F, Aguilera A. 2005. Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *EMBO J* 24: 1267-76

27. Prado F, Piruat JI, Aguilera A. 1997. Recombination between DNA repeats in yeast hpr1delta cells is linked to transcription elongation. *EMBO J* 16: 2826-35

28. Liu LF, Wang JC. 1987. Supercoiling of the DNA template during transcription. *Proc Natl Acad Sci U S A* 84: 7024-7

29. Brill SJ, Sternglanz R. 1988. Transcription-dependent DNA supercoiling in yeast DNA topoisomerase mutants. *Cell* 54: 403-11

30. Naughton C, Avlonitis N, Corless S, Prendergast JG, Mati IK, et al. 2013. Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nat Struct Mol Biol* 20: 387-95

31. Sperling AS, Jeong KS, Kitada T, Grunstein M. 2011. Topoisomerase II binds nucleosome-free DNA and acts redundantly with topoisomerase I to enhance recruitment of RNA Pol II in budding yeast. *Proc Natl Acad Sci U S A* 108: 12693-8

32. Christman MF, Dietrich FS, Fink GR. 1988. Mitotic recombination in the rDNA of S. cerevisiae is suppressed by the combined action of DNA topoisomerases I and II. *Cell* 55: 413-25

33. Garcia-Rubio ML, Aguilera A. 2012. Topological constraints impair RNA polymerase II transcription and causes instability of plasmid-borne convergent genes. *Nucleic Acids Res* 40: 1050-64

34. Lippert MJ, Kim N, Cho JE, Larson RP, Schoenly NE, et al. 2011. Role for topoisomerase 1 in transcription-associated mutagenesis in yeast. *Proc Natl Acad Sci U S A* 108: 698-703

35. Takahashi T, Burguiere-Slezak G, Van der Kemp PA, Boiteux S. 2011. Topoisomerase 1 provokes the formation of short deletions in repeated sequences upon high transcription in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* 108: 692-7

36. Sekiguchi J, Shuman S. 1997. Site-specific ribonuclease activity of eukaryotic DNA topoisomerase I. *Mol Cell* 1: 89-97

37. Kim N, Huang SN, Williams JS, Li YC, Clark AB, et al. 2011. Mutagenic processing of ribonucleotides in DNA by yeast topoisomerase I. *Science* 332: 1561-4

38. Sparks JL, Burgers PM. 2015. Error-free and mutagenic processing of topoisomerase 1-provoked damage at genomic ribonucleotides. *EMBO J* 34: 1259-69

39. Wang G, Vasquez KM. 2014. Impact of alternative DNA structures on DNA damage, DNA repair, and genetic instability. *DNA Repair (Amst)* 19: 143-51

40. Lopez Castel A, Cleary JD, Pearson CE. 2010. Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat Rev Mol Cell Biol* 11: 165-70

41. Aguilera A, Garcia-Muse T. 2013. Causes of genome instability. *Annu Rev Genet* 47: 1-32

42.   Napierala M, Bacolla A, Wells RD. 2005. Increased negative superhelical density in vivo enhances the genetic instability of triplet repeat sequences. *J Biol Chem* 280: 37366-76

43.   Kohwi Y, Panchenko Y. 1993. Transcription-dependent recombination induced by triple-helix formation. *Genes Dev* 7: 1766-78

44.   Mochmann LH, Wells RD. 2004. Transcription influences the types of deletion and expansion products in an orientation-dependent manner from GAC*GTC repeats. *Nucleic Acids Res* 32: 4469-79

45.   Wierdl M, Greene CN, Datta A, Jinks-Robertson S, Petes TD. 1996. Destabilization of simple repetitive DNA sequences by transcription in yeast. *Genetics* 143: 713-21

46.   Lin Y, Dion V, Wilson JH. 2006. Transcription promotes contraction of CAG repeat tracts in human cells. *Nat Struct Mol Biol* 13: 179-80

47.   Ditch S, Sammarco MC, Banerjee A, Grabczyk E. 2009. Progressive GAA.TTC repeat expansion in human cell lines. *PLoS Genet* 5: e1000704

48.   Tang W, Dominska M, Greenwell PW, Harvanek Z, Lobachev KS, et al. 2011. Friedreich's ataxia (GAA)n*(TTC)n repeats strongly stimulate mitotic crossovers in Saccharomyces cerevisae. *PLoS Genet* 7: e1001270

49.   Lin Y, Wilson JH. 2007. Transcription-induced CAG repeat contraction in human cells is mediated in part by transcription-coupled nucleotide excision repair. *Mol Cell Biol* 27: 6209-17

50.   Wang G, Seidman MM, Glazer PM. 1996. Mutagenesis in mammalian cells induced by triple helix formation and transcription-coupled repair. *Science* 271: 802-5

51.   Lu S, Wang G, Bacolla A, Zhao J, Spitser S, Vasquez KM. 2015. Short Inverted Repeats Are Hotspots for Genetic Instability: Relevance to Cancer Genomes. *Cell Rep* 10: 1674-80

52. Gray LT, Vallur AC, Eddy J, Maizels N. 2014. G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD. *Nat Chem Biol* 10: 313-8

53. Hendriks G, Calleja F, Vrieling H, Mullenders LH, Jansen JG, de Wind N. 2008. Gene transcription increases DNA damage-induced mutagenesis in mammalian stem cells. *DNA Repair (Amst)* 7: 1330-9

54. Morey NJ, Greene CN, Jinks-Robertson S. 2000. Genetic analysis of transcription-associated mutation in Saccharomyces cerevisiae. *Genetics* 154: 109-20

55. Kim N, Jinks-Robertson S. 2009. dUTP incorporation into genomic DNA is linked to transcription in yeast. *Nature* 459: 1150-3

56. Hendriks G, Calleja F, Besaratinia A, Vrieling H, Pfeifer GP, et al. 2010. Transcription-dependent cytosine deamination is a novel mechanism in ultraviolet light-induced mutagenesis. *Curr Biol* 20: 170-5

57. Chaudhuri J, Alt FW. 2004. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nat Rev Immunol* 4: 541-52

58. Yu K, Chedin F, Hsieh CL, Wilson TE, Lieber MR. 2003. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat Immunol* 4: 442-51

59. Robbiani DF, Nussenzweig MC. 2013. Chromosome translocation, B cell lymphoma, and activation-induced cytidine deaminase. *Annu Rev Pathol* 8: 79-103

60. Pavri R, Gazumyan A, Jankovic M, Di Virgilio M, Klein I, et al. 2010. Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell* 143: 122-33

61. Meng FL, Du Z, Federation A, Hu J, Wang Q, et al. 2014. Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell* 159: 1538-48

62. Qian J, Wang Q, Dose M, Pruett N, Kieffer-Kwon KR, et al. 2014. B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. *Cell* 159: 1524-37

63. Bachl J, Carlson C, Gray-Schopfer V, Dessing M, Olsson C. 2001. Increased transcription levels induce higher mutation rates in a hypermutating cell line. *J Immunol* 166: 5051-7

64. Parsa JY, Ramachandran S, Zaheen A, Nepal RM, Kapelnikov A, et al. 2012. Negative supercoiling creates single-stranded patches of DNA that are substrates for AID-mediated mutagenesis. *PLoS Genet* 8: e1002518

65. Sohail A, Klapacz J, Samaranayake M, Ullah A, Bhagwat AS. 2003. Human activation-induced cytidine deaminase causes transcription-dependent, strand-biased C to U deaminations. *Nucleic Acids Res* 31: 2990-4

66. Gomez-Gonzalez B, Aguilera A. 2007. Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. *Proc Natl Acad Sci U S A* 104: 8409-14

67. Ruiz JF, Gomez-Gonzalez B, Aguilera A. 2011. AID induces double-strand breaks at immunoglobulin switch regions and c-MYC causing chromosomal translocations in yeast THO mutants. *PLoS Genet* 7: e1002009

68. Suspene R, Aynaud MM, Guetard D, Henry M, Eckhoff G, et al. 2011. Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism. *Proc Natl Acad Sci U S A* 108: 4858-63

69. Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, et al. 2013. APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* 494: 366-70

70. Roberts SA, Lawrence MS, Klimczak LJ, Grimm SA, Fargo D, et al. 2013. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat Genet* 45: 970-6

71.   Nordentoft I, Lamy P, Birkenkamp-Demtroder K, Shumansky K, Vang S, et al. 2014. Mutational context and diverse clonal development in early and late bladder cancer. *Cell Rep* 7: 1649-63

72.   Lada AG, Kliver SF, Dhar A, Polev DE, Masharsky AE, et al. 2015. Disruption of Transcriptional Coactivator Sub1 Leads to Genome-Wide Re-distribution of Clustered Mutations Induced by APOBEC in Active Yeast Genes. *PLoS Genet* 11: e1005217

73.   Sollier J, Stork CT, Garcia-Rubio ML, Paulsen RD, Aguilera A, Cimprich KA. 2014. Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol Cell* 56: 777-85

74.   French S. 1992. Consequences of replication fork movement through transcription units in vivo. *Science* 258: 1362-5

75.   Deshpande AM, Newlon CS. 1996. DNA replication fork pause sites dependent on transcription. *Science* 272: 1030-3

76.   Cortes-Ledesma F, Aguilera A. 2006. Double-strand breaks arising by replication through a nick are repaired by cohesin-dependent sister-chromatid exchange. *EMBO Rep* 7: 919-26

77.   Moriel-Carretero M, Aguilera A. 2010. A postincision-deficient TFIIH causes replication fork breakage and uncovers alternative Rad51- or Pol32-mediated restart mechanisms. *Mol Cell* 37: 690-701

78.   Lambert S, Mizuno K, Blaisonneau J, Martineau S, Chanet R, et al. 2010. Homologous recombination restarts blocked replication forks at the expense of genome rearrangements by template exchange. *Mol Cell* 39: 346-59

79.   Iraqui I, Chekkal Y, Jmari N, Pietrobon V, Freon K, et al. 2012. Recovery of arrested replication forks by homologous recombination is error-prone. *PLoS Genet* 8: e1002976

80.   Lambert S, Watson A, Sheedy DM, Martin B, Carr AM. 2005. Gross chromosomal rearrangements and elevated recombination at an inducible site-specific replication fork barrier. *Cell* 121: 689-702

81.   Mizuno K, Miyabe I, Schalbetter SA, Carr AM, Murray JM. 2013. Recombination-restarted replication makes inverted chromosome fusions at inverted repeats. *Nature* 493: 246-9

82.   Tuduri S, Crabbe L, Conti C, Tourriere H, Holtgreve-Grez H, et al. 2009. Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nat Cell Biol* 11: 1315-24

83.   Herrera-Moyano E, Mergui X, Garcia-Rubio ML, Barroso S, Aguilera A. 2014. The yeast and human FACT chromatin-reorganizing complexes solve R-loop-mediated transcription-replication conflicts. *Genes Dev* 28: 735-48

84.   Pardo B, Aguilera A. 2012. Complex chromosomal rearrangements mediated by break-induced replication involve structure-selective endonucleases. *PLoS Genet* 8: e1002979

85.   Smith CE, Llorente B, Symington LS. 2007. Template switching during break-induced replication. *Nature* 447: 102-5

86.   Costantino L, Sotiriou SK, Rantala JK, Magin S, Mladenov E, et al. 2014. Break-induced replication repair of damaged forks induces genomic duplications in human cells. *Science* 343: 88-91

87.   Deem A, Keszthelyi A, Blackgrove T, Vayl A, Coffey B, et al. 2011. Break-induced replication is highly inaccurate. *PLoS Biol* 9: e1000594

88.   Sakofsky CJ, Roberts SA, Malc E, Mieczkowski PA, Resnick MA, et al. 2014. Break-induced replication is a source of mutation clusters underlying kataegis. *Cell Rep* 7: 1640-8

89.   Liu B, Alberts BM. 1995. Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. *Science* 267: 1131-7

90.   Million-Weaver S, Samadpour AN, Merrikh H. 2015. Replication Restart after Replication-Transcription Conflicts Requires RecA in Bacillus subtilis. *J Bacteriol* 197: 2374-82

91.   Merrikh H, Zhang Y, Grossman AD, Wang JD. 2012. Replication-transcription conflicts in bacteria. *Nat Rev Microbiol* 10: 449-58

92.   Paul S, Million-Weaver S, Chattopadhyay S, Sokurenko E, Merrikh H. 2013. Accelerated gene evolution through replication-transcription conflicts. *Nature* 495: 512-5

93.   Ivanova D, Taylor T, Smith SL, Dimude JU, Upton AL, et al. 2015. Shaping the landscape of the Escherichia coli chromosome: replication-transcription encounters in cells with an ectopic replication origin. *Nucleic Acids Res*

94.   Pomerantz RT, O'Donnell M. 2010. Direct restart of a replication fork stalled by a head-on RNA polymerase. *Science* 327: 590-2

95.   Boubakri H, de Septenville AL, Viguera E, Michel B. 2010. The helicases DinG, Rep and UvrD cooperate to promote replication across transcription units in vivo. *EMBO J* 29: 145-57

96.   Merrikh CN, Brewer BJ, Merrikh H. 2015. The B. subtilis Accessory Helicase PcrA Facilitates DNA Replication through Transcription Units. *PLoS Genet* 11: e1005289

97.   Azvolinsky A, Dunaway S, Torres JZ, Bessler JB, Zakian VA. 2006. The S. cerevisiae Rrm3p DNA helicase moves with the replication fork and affects replication of all yeast chromosomes. *Genes Dev* 20: 3104-16

98.   Ivessa AS, Lenzmeier BA, Bessler JB, Goudsouzian LK, Schnakenberg SL, Zakian VA. 2003. The Saccharomyces cerevisiae helicase Rrm3p facilitates replication past nonhistone protein-DNA complexes. *Mol Cell* 12: 1525-36

99.   Azvolinsky A, Giresi PG, Lieb JD, Zakian VA. 2009. Highly transcribed RNA polymerase II genes are impediments to replication fork progression in Saccharomyces cerevisiae. *Mol Cell* 34: 722-34

100.   Gomez-Gonzalez B, Garcia-Rubio M, Bermejo R, Gaillard H, Shirahige K, et al. 2011. Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. *EMBO J* 30: 3106-19

101.   Fachinetti D, Bermejo R, Cocito A, Minardi S, Katou Y, et al. 2010. Replication termination at eukaryotic chromosomes is mediated by Top2 and occurs at genomic loci containing pausing elements. *Mol Cell* 39: 595-605

102.   Gottipati P, Cassel TN, Savolainen L, Helleday T. 2008. Transcription-associated recombination is dependent on replication in Mammalian cells. *Mol Cell Biol* 28: 154-64

103.   Cabal GG, Genovesio A, Rodriguez-Navarro S, Zimmer C, Gadal O, et al. 2006. SAGA interacting factors confine sub-diffusion of transcribed genes to the nuclear envelope. *Nature* 441: 770-3

104.   Bermejo R, Capra T, Jossen R, Colosio A, Frattini C, et al. 2011. The replication checkpoint protects fork stability by releasing transcribed genes from nuclear pores. *Cell* 146: 233-46

105.   Trautinger BW, Jaktaji RP, Rusakova E, Lloyd RG. 2005. RNA polymerase modulators and DNA repair activities resolve conflicts between DNA replication and transcription. *Mol Cell* 19: 247-58

106.   Tehranchi AK, Blankschien MD, Zhang Y, Halliday JA, Srivatsan A, et al. 2010. The transcription factor DksA prevents conflicts between DNA replication and transcription machinery. *Cell* 141: 595-605

107.   Washburn RS, Gottesman ME. 2011. Transcription termination maintains chromosome integrity. *Proc Natl Acad Sci U S A* 108: 792-7

108.   Gupta MK, Guy CP, Yeeles JT, Atkinson J, Bell H, et al. 2013. Protein-DNA complexes are the primary sources of replication fork pausing in Escherichia coli. *Proc Natl Acad Sci U S A* 110: 7252-7

109.   Dutta D, Shatalin K, Epshtein V, Gottesman ME, Nudler E. 2011. Linking RNA polymerase backtracking to genome instability in E. coli. *Cell* 146: 533-43

110.   Merrikh H, Machon C, Grainger WH, Grossman AD, Soultanas P. 2011. Co-directional replication-transcription conflicts lead to replication restart. *Nature* 470: 554-7

111.   Ide S, Miyazaki T, Maki H, Kobayashi T. 2010. Abundance of ribosomal RNA gene copies maintains genome integrity. *Science* 327: 693-6

112.   Kobayashi T, Heck DJ, Nomura M, Horiuchi T. 1998. Expansion and contraction of ribosomal DNA repeats in Saccharomyces cerevisiae: requirement of replication fork blocking (Fob1) protein and the role of RNA polymerase I. *Genes Dev* 12: 3821-30

113.   Brewer BJ, Fangman WL. 1988. A replication fork barrier at the 3' end of yeast ribosomal RNA genes. *Cell* 55: 637-43

114.   Takeuchi Y, Horiuchi T, Kobayashi T. 2003. Transcription-dependent recombination and the role of fork collision in yeast rDNA. *Genes Dev* 17: 1497-506

115.   Muller M, Lucchini R, Sogo JM. 2000. Replication of yeast rDNA initiates downstream of transcriptionally active genes. *Mol Cell* 5: 767-77

116. Yoshida K, Bacal J, Desmarais D, Padioleau I, Tsaponina O, et al. 2014. The histone deacetylases sir2 and rpd3 act on ribosomal DNA to control the replication program in budding yeast. *Mol Cell* 54: 691-7

117. Gottlieb S, Esposito RE. 1989. A new role for a yeast transcriptional silencer gene, SIR2, in regulation of recombination in ribosomal DNA. *Cell* 56: 771-6

118. Kobayashi T, Horiuchi T, Tongaonkar P, Vu L, Nomura M. 2004. SIR2 regulates recombination between different rDNA repeats, but not recombination within individual rRNA genes in yeast. *Cell* 117: 441-53

119. Kobayashi T, Ganley AR. 2005. Recombination regulation by transcription-induced cohesin dissociation in rDNA repeats. *Science* 309: 1581-4

120. Duch A, Felipe-Abrio I, Barroso S, Yaakov G, Garcia-Rubio M, et al. 2013. Coordinated control of replication and transcription by a SAPK protects genomic integrity. *Nature* 493: 116-9

121. Nguyen VC, Clelland BW, Hockman DJ, Kujat-Choy SL, Mewhort HE, Schultz MC. 2010. Replication stress checkpoint signaling controls tRNA gene transcription. *Nat Struct Mol Biol* 17: 976-81

122. Islam MN, Fox D, 3rd, Guo R, Enomoto T, Wang W. 2010. RecQL5 promotes genome stabilization through two parallel mechanisms--interacting with RNA polymerase II and acting as a helicase. *Mol Cell Biol* 30: 2460-72

123. Kassube SA, Jinek M, Fang J, Tsutakawa S, Nogales E. 2013. Structural mimicry in transcription regulation of human RNA polymerase II by the DNA helicase RECQL5. *Nat Struct Mol Biol* 20: 892-9

124. Saponaro M, Kantidakis T, Mitter R, Kelly GP, Heron M, et al. 2014. RECQL5 controls transcript elongation and suppresses genome instability associated with transcription stress. *Cell* 157: 1037-49

125. Felipe-Abrio I, Lafuente-Barquero J, Garcia-Rubio ML, Aguilera A. 2015. RNA polymerase II contributes to preventing transcription-mediated replication fork stalls. *EMBO J* 34: 236-50

126. Huertas P, Aguilera A. 2003. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol Cell* 12: 711-21

127. Li X, Manley JL. 2005. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* 122: 365-78

128. Jimeno S, Luna R, Garcia-Rubio M, Aguilera A. 2006. Tho1, a novel hnRNP, and Sub2 provide alternative pathways for mRNP biogenesis in yeast THO mutants. *Mol Cell Biol* 26: 4387-98

129. Li X, Niu T, Manley JL. 2007. The RNA binding protein RNPS1 alleviates ASF/SF2 depletion-induced genomic instability. *RNA* 13: 2108-15

130. Dominguez-Sanchez MS, Barroso S, Gomez-Gonzalez B, Luna R, Aguilera A. 2011. Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. *PLoS Genet* 7: e1002386

131. Paulsen RD, Soni DV, Wollman R, Hahn AT, Yee MC, et al. 2009. A genome-wide siRNA screen reveals diverse cellular processes and pathways that mediate genome stability. *Mol Cell* 35: 228-39

132. Wahba L, Amon JD, Koshland D, Vuica-Ross M. 2011. RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Mol Cell* 44: 978-88

133. Stirling PC, Chan YA, Minaker SW, Aristizabal MJ, Barrett I, et al. 2012. R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. *Genes Dev* 26: 163-75

134. Mischo HE, Gomez-Gonzalez B, Grzechnik P, Rondon AG, Wei W, et al. 2011. Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol Cell* 41: 21-32

135. Santos-Pereira JM, Aguilera A. 2015. R loops: new modulators of genome dynamics and function. *Nat Rev Genet*

136. Wellinger RE, Prado F, Aguilera A. 2006. Replication fork progression is impaired by transcription in hyperrecombinant yeast cells lacking a functional THO complex. *Mol Cell Biol* 26: 3327-34

137. Castellano-Pozo M, Garcia-Muse T, Aguilera A. 2012. R-loops cause replication impairment and genome instability during meiosis. *EMBO Rep* 13: 923-9

138. Gan W, Guan Z, Liu J, Gui T, Shen K, et al. 2011. R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev* 25: 2041-56

139. Santos-Pereira JM, Herrero AB, Garcia-Rubio ML, Marin A, Moreno S, Aguilera A. 2013. The Npl3 hnRNP prevents R-loop-mediated transcription-replication conflicts and genome instability. *Genes Dev* 27: 2445-58

140. Alzu A, Bermejo R, Begnis M, Lucca C, Piccini D, et al. 2012. Senataxin Associates with Replication Forks to Protect Fork Integrity across RNA-Polymerase-II-Transcribed Genes. *Cell* 151: 835-46

141. Helmrich A, Ballarino M, Tora L. 2011. Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol Cell* 44: 966-77

142. Grabczyk E, Mancuso M, Sammarco MC. 2007. 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: 5351-9

143. Lin Y, Dent SY, Wilson JH, Wells RD, Napierala M. 2010. R loops stimulate genetic instability of CTG.CAG repeats. *Proc Natl Acad Sci U S A* 107: 692-7

144. El Hage A, French SL, Beyer AL, Tollervey D. 2010. Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev* 24: 1546-58

145. Pomerantz RT, O'Donnell M. 2008. The replisome uses mRNA as a primer after colliding with RNA polymerase. *Nature* 456: 762-6

146. Stuckey R, Garcia-Rodriguez N, Aguilera A, Wellinger RE. 2015. Role for RNA:DNA hybrids in origin-independent replication priming in a eukaryotic system. *Proc Natl Acad Sci U S A* 112: 5779-84

147. Castellano-Pozo M, Santos-Pereira JM, Rondon AG, Barroso S, Andujar E, et al. 2013. R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. *Mol Cell* 52: 583-90

148. Groh M, Lufino MM, Wade-Martins R, Gromak N. 2014. R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. *PLoS Genet* 10: e1004318

149. Loomis EW, Sanz LA, Chedin F, Hagerman PJ. 2014. Transcription-associated R-loop formation across the human FMR1 CGG-repeat region. *PLoS Genet* 10: e1004294

150. Skourti-Stathaki K, Kamieniarz-Gdula K, Proudfoot NJ. 2014. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature* 516: 436-9

151. Bhatia V, Barroso SI, Garcia-Rubio ML, Tumini E, Herrera-Moyano E, Aguilera A. 2014. BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature* 511: 362-5

152.	Hill SJ, Rolland T, Adelmant G, Xia X, Owen MS, et al. 2014. Systematic screening reveals a role for BRCA1 in the response to transcription-associated DNA damage. *Genes Dev* 28: 1957-75

153.	Wimberly H, Shee C, Thornton PC, Sivaramakrishnan P, Rosenberg SM, Hastings PJ. 2013. R-loops and nicks initiate DNA breakage and genome instability in non-growing Escherichia coli. *Nat Commun* 4: 2115

154.	Hatchi E, Skourti-Stathaki K, Ventz S, Pinello L, Yen A, et al. 2015. BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. *Mol Cell* 57: 636-47

155.	Tresini M, Warmerdam DO, Kolovos P, Snijder L, Vrouwe MG, et al. 2015. The core spliceosome as target and effector of non-canonical ATM signalling. *Nature* 523: 53-8

156.	Gorgoulis VG, Vassiliou LV, Karakaidos P, Zacharatos P, Kotsinas A, et al. 2005. Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature* 434: 907-13

157.	Bartkova J, Horejsi Z, Koed K, Kramer A, Tort F, et al. 2005. DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature* 434: 864-70

158.	Jones RM, Mortusewicz O, Afzal I, Lorvellec M, Garcia P, et al. 2013. Increased replication initiation and conflicts with transcription underlie Cyclin E-induced replication stress. *Oncogene* 32: 3744-53

159.	Neelsen KJ, Zanini IM, Herrador R, Lopes M. 2013. Oncogenes induce genotoxic stress by mitotic processing of unusual replication intermediates. *J Cell Biol* 200: 699-708

160.	Dominguez-Sola D, Gautier J. 2014. MYC and the control of DNA replication. *Cold Spring Harb Perspect Med* 4

161. Bignell GR, Greenman CD, Davies H, Butler AP, Edkins S, et al. 2010. Signatures of mutation and selection in the cancer genome. *Nature* 463: 893-8

162. Barlow JH, Faryabi RB, Callen E, Wong N, Malhowski A, et al. 2013. Identification of early replicating fragile sites that contribute to genome instability. *Cell* 152: 620-32

163. Hu Y, Raynard S, Sehorn MG, Lu X, Bussen W, et al. 2007. RECQL5/Recql5 helicase regulates homologous recombination and suppresses tumor formation via disruption of Rad51 presynaptic filaments. *Genes Dev* 21: 3073-84

164. Reddy K, Tam M, Bowater RP, Barber M, Tomlinson M, et al. 2011. Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. *Nucleic Acids Res* 39: 1749-62

165. Goula AV, Stys A, Chan JP, Trottier Y, Festenstein R, Merienne K. 2012. Transcription elongation and tissue-specific somatic CAG instability. *PLoS Genet* 8: e1003051

166. Epshtein V, Kamarthapu V, McGary K, Svetlov V, Ueberheide B, et al. 2014. UvrD facilitates DNA repair by pulling RNA polymerase backwards. *Nature* 505: 372-7

167. Venema J, Mullenders LH, Natarajan AT, van Zeeland AA, Mayne LV. 1990. The genetic defect in Cockayne syndrome is associated with a defect in repair of UV-induced DNA damage in transcriptionally active DNA. *Proc Natl Acad Sci U S A* 87: 4707-11

168. Spivak G, Ganesan AK. 2014. The complex choreography of transcription-coupled repair. *DNA Repair (Amst)* 19: 64-70

169. Durkin SG, Glover TW. 2007. Chromosome fragile sites. *Annu Rev Genet* 41: 169-92

170. Letessier A, Millot GA, Koundrioukoff S, Lachages AM, Vogt N, et al. 2011. Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. *Nature* 470: 120-3

171.   Cha RS, Kleckner N. 2002. ATR homolog Mec1 promotes fork progression, thus averting breaks in replication slow zones. *Science* 297: 602-6

Glossary

DNA damage response (DDR): Set of mechanisms that allows the detection of DNA lesions, their signaling and subsequent repair.

Chromatin remodeling: Dynamic changes in the chromatin structure at a DNA region that involve post-translational histone modification or alteration of nucleosome architecture.

Homologous recombination (HR): DSB repair mechanism that uses identical or homologous stretches of DNA as template to restore the broken DNA sequence.

Single strand annealing (SSA): Homology-directed DSB repair pathway involving DNA-end resection and annealing of the resulting ssDNA to promote recombination between tandem DNA repeats.

Messenger ribonucleoprotein particle (mRNP): Complex of mRNA and proteins forming during eukaryotic transcription that is involved in processing, export and translation of the transcript.

G-quadruplexes: Four-stranded nucleic acid structures in which tetrads of hydrogen-bonded guanine bases are stacked on top of each other.

R loop: non-B DNA structure formed by an RNA-DNA hybrid and the displaced ssDNA of the original DNA duplex.

Replication fork (RF): DNA area that is unwound by the replicative helicase to allow DNA synthesis of the complementary strands.

Replication stress: Situation in which replication progression is impaired, leading to RF slow-down or stalling.

Replication checkpoint: Checkpoint activated in case of impaired RF progression that leads to origin firing inhibition and stalled RF stabilization.

Template switching: Process by which a replicating DNA polymerase continues DNA synthesis using a different DNA molecule as template.

Replication fork (RF) reversal: Regression of the RF that causes the leading strand to hybridize with the complementary lagging strand forming a four-branched structure.

Intra-S checkpoint: Checkpoint activated by DSB during S phase that leads to DNA synthesis down-regulation and promotes DSB repair.

FACT: Chromatin remodeler complex that disrupts the interaction between the H2A/H2B dimers and H3/H4 tetramer of nucleosomes during transcription elongation.

Break-induced replication (BIR): One-ended DSB repair pathway that is initiated by strand invasion into a homologous DNA duplex followed by replication.

Loss of heterozygosity: The loss of the normal allele at a originally heterozygous locus caused by a deletion or mutational event.

Cohesin: Protein complex that ensures sister chromatid cohesion during replication and at sites of DSB, thus favoring equal recombination events.

Fanconi Anemia (FA) pathway: A DNA repair pathway that removes interstrand crosslinks and other types of DNA damages that block replication fork progression.

Sidebars

TRANSCRIPTION-COUPLED DNA REPAIR

Transcription-Coupled Repair refers to the processes that enable efficient repair of the transcribed strand of active genes as compared to elsewhere in the genome. Its hallmark is TC-NER, a sub-pathway of NER. TC-NER is conserved from bacteria to humans and initiates when an RNAP encounters a DNA lesion that impairs its progression (4). The stalled RNAP subsequently triggers, with the assistance of specialized factors, the recruitment of the NER machinery. The Mfd protein couples transcription and repair in *E. coli*. Alternatively, the reaction can be mediated by the coordinated action of the UvrD helicase and the NusA transcription elongation factor (166). In eukaryotes, a number of factors cooperate to enable TC-NER, the best-characterized being the human Cockayne's syndrome protein B (CSB) (167). Other factors affecting TC-NER include CSA, the UV-stimulated scaffold protein A (UVSSA), ubiquitin-specific protease 7 (USP7), XPA-binding protein 2 (XAB2) and FACT in humans and proteins affecting transcription elongation in yeast (168), but the specific function of these factors, whether direct or indirect, is unclear. The actions taking place at a stalled RNAP during TC-NER have not yet been resolved and several, not necessarily exclusive, models involving RNAP degradation, remodeling, backtracking or forward translocation have been proposed.

FRAGILE SITES

Fragile sites refer to specific genomic locations that are prone to suffer gaps, breaks and constrictions in metaphase chromosomes upon mild inhibition of replication in higher eukaryotes (169). Rare fragile sites are found in a small subset of the population and are often due to the expansion of nucleotide repeats. These are in some cases associated with human

genetic disorders (e.g. Fragile X syndrome). Common fragile sites (CFS) represent a normal component of chromosomes present in all human genomes. CFS are typically late replicating regions characterized by their paucity in replication origins that forces a given RF to travel over long distances (170), these features making them highly vulnerable to replication stress. CFS often overlap with large tumor suppressor genes (e.g. FRA3B and FRA16D) and have been associated with recurrent cancer-specific breakpoints (161). Early replicating fragile sites are a new class that are not observable cytologically and coincide with clusters of highly expressed genes, many of which are related to cancer (162). Regions highly prone to break under replication stress exist in all organisms. They are called differently but may be functionally equivalent to CFS, such as the Slow Replicating Zones in yeast (171) or other types of fragile sites.

Figure 1

Genome instability events induced by transcription. (*a*) Examples of transcription-associated mutation. Spontaneous cytidine deamination can lead to C-to-T transition if the resulting uracil is not repaired prior to replication (*top*). Two-base pairs (2-bp) deletion can occur following topoisomerase 1 (Top1) cleavage at sequences containing short tandem repeats (*bottom*). The proposed model involves a short single-stranded DNA (ssDNA) gap as intermediate, whose ends are ligated after realignment of the sequence (enabled by the presence of tandem repeats). (*b*) Examples of transcription-associated recombination and rearrangements. Transcription of a sequence containing direct repeats can lead to the deletion of a region containing one of the repeats and the intervening sequence if single-strand annealing (SSA) is used to repair a double strand break (DSB) (*top*). This type of event can be easily followed genetically if the intervening sequence contains a marker. Transcription can lead to DSB or to ssDNA gaps, which may be converted to DSB by replication or nuclease activities (*bottom*). Subsequent DSB repair may lead to chromosome rearrangements if error-prone DSB repair mechanisms are involved, such as Non-homologous End-Joining (NHEJ) or Break-Induced Replication (BIR), or to different rearrangements (e.g. deletion, inversion, translocation) if homologous recombination (HR) takes place with an ectopic homologous DNA sequence as template.

Figure 2

Changes in primary DNA structure mediated by transcription. (*a*) Transcription requires a transient unwinding of the DNA that generates positive supercoiling (+) ahead of the RNA polymerase (RNAP) and negative supercoiling (-) behind it. The negative supercoiling favors

DNA unwinding and leads to a major susceptibility of the DNA strands to spontaneous base modifications (e.g. cytidine deamination), or to the actions of either genotoxic agents leading to base damage (*red star*) or enzymes that could cause single-stranded DNA (ssDNA) breaks or base damages. DNA unwinding also favors the formation of non-B DNA structures (e.g. hairpins, G-quadruplexes (G4) DNA, or DNA:RNA hybrids). (*b*) DNA nicks or ssDNA gaps can arise as a consequence of enzyme-driven processes at transcribed DNA. They could result from unfinished topoisomerisation reactions and subsequent removal of the trapped Topoisomerase (Top1cc) (*left*), from stalling of RNAP at a non-B DNA structure that may trigger gratuitous transcription-coupled repair (TC-NER) (*middle*), or from non-B DNA structures formed during transcription that could promote nuclease-mediated DNA cleavage (*right*).

Figure 3

Transcription as a threat to replication fork integrity. (*a*) Collisions between transcription and replication machineries may impede replication fork (RF) progression leading to RF stalling (*top left*). Activation of the replication checkpoint acts on the arrested fork, preventing RF collapse and triggering fork restart (*left*). The presence of single-stranded DNA (ssDNA) gaps in transcribed DNA (*top right*), or a direct collision between a transcribing RNAP and the replisome (*top left*) can also lead to RF collapse, and therewith to a double strand break (DSB) whether or not nuclease-mediated. The occurrence of DSB in replicating cells triggers the intra-S checkpoint (*right*). Upon repair of the DSB, replication may be restarted. (*b*) Examples of RF progression impairment, seen by DNA combing analysis as short DNA synthesis tracts (*red*: first labeling pulse of DNA synthesis; *green*: second labeling pulse of DNA synthesis; *blue*: DNA), and DNA break accumulation, seen by single-cell electrophoresis as comet tails, in transcriptionally active HeLa cells depleted of the FACT complex. Both phenotypes are

suppressed in the presence of the transcription inhibitor cordycepin (83), as visualized by shorter DNA synthesis tracts and non-detectable comet tails (images provided by E. Herrera-Moyano).

Abbreviations: RPA, Replication Protein A; ATR, Ataxia Telangiectasia and Rad3 Related; CHK1, Checkpoint Kinase 1; ATM, Ataxia Telangiectasia Mutated; CHK2, Checkpoint kinase 2.

Figure 4

R loops as a source of replication stress and genome instability. (*a*) The single-stranded nature of the displaced DNA strand within the R loop makes it more vulnerable to genotoxic agents and enzymatic activities such as Activation-Induced cytidine Deaminase (AID) and nucleases (*top*). Base damage (*red star*) can lead to mutagenesis, whereas ssDNA nicks or gaps lead to replication fork (RF) collapse during replication, as depicted in Figure 3. Co-transcriptionally formed R loops, with or without the transcribing RNAP (*middle*), whether or not accompanied by heterochromatinization (*bottom*), can impede RF progression, leading to fork stalling, potential collapse and breakage (see Figure 3).
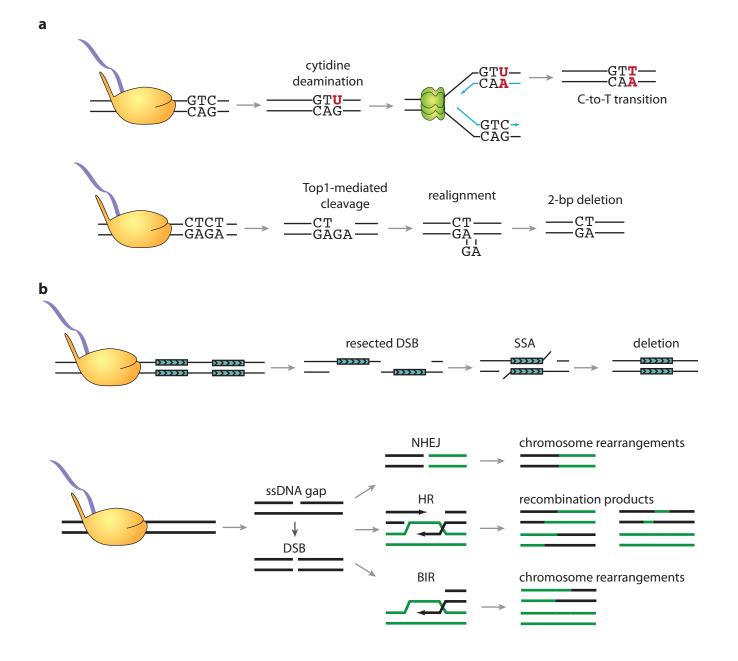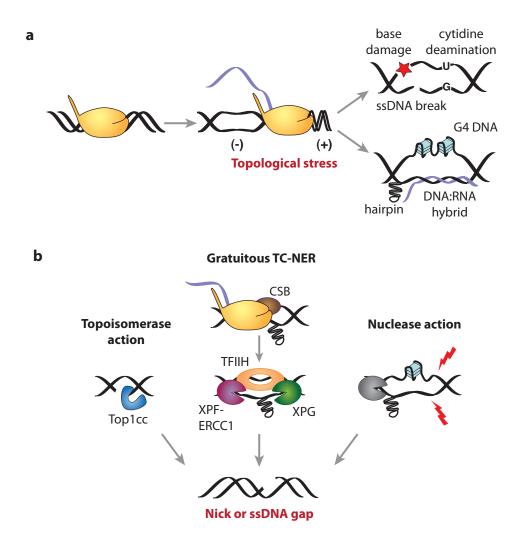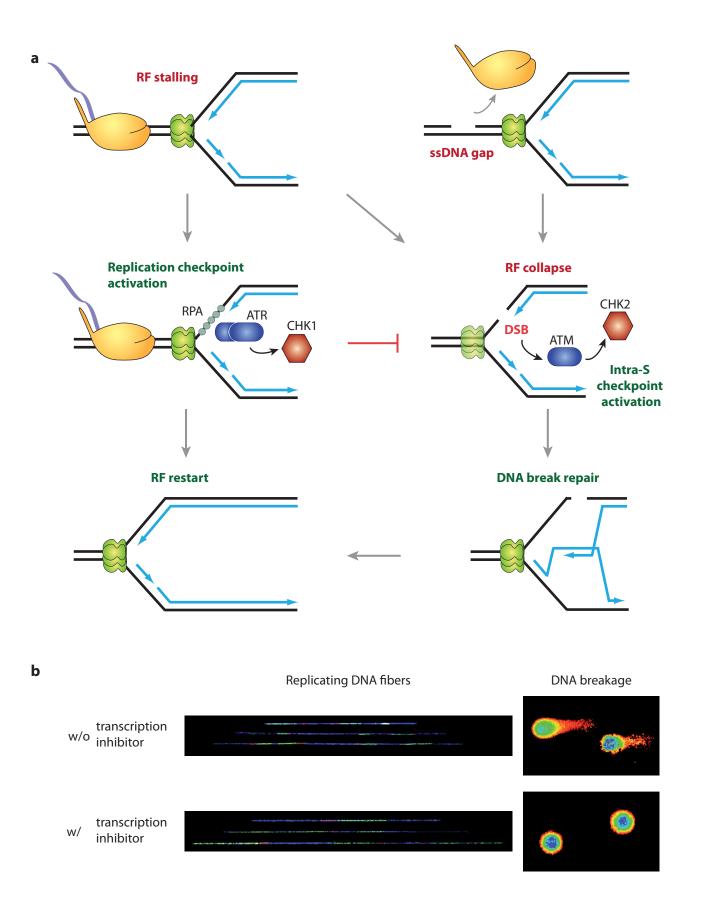
Figure 1

a

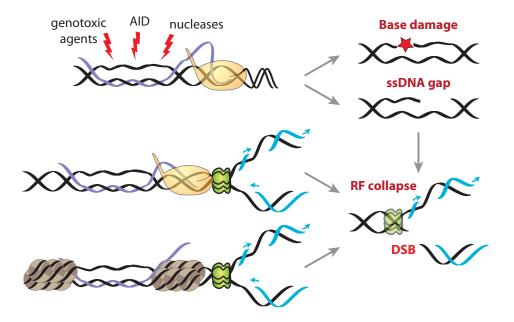base damage   cytidine deamination

ssDNA break

G4 DNA

(-)   (+)
Topological stress

hairpin   DNA:RNA hybrid

b

Gratuitous TC-NER

CSB

Topoisomerase action

TFIIH

Nuclease action

Top1cc

XPF-ERCC1   XPG

Nick or ssDNA gap

Figure 2

a

RF stalling

ssDNA gap

Replication checkpoint activation

RPA   ATR   CHK1

RF collapse

DSB   ATM   CHK2

Intra-S checkpoint activation

RF restart

DNA break repair

b

Replicating DNA fibers

DNA breakage

w/o transcription inhibitor

w/ transcription inhibitor

Figure 3

Figure 4