Published in final edited form as:

J Mol Biol. 2017 October 27; 429(21): 3156–3167. doi:10.1016/j.jmb.2016.08.005.

Ribonucleotides and transcription-associated mutagenesis in yeast

Jang-Eun Cho and Sue Jinks-Robertson

Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC

Abstract

High levels of transcription stimulate mutation rates in microorganisms and this occurs primarily through an enhanced accumulation of DNA damage. The major source of transcription-associated damage in yeast is topoisomerase I (Top1), an enzyme that removes torsional stress that accumulates when DNA strands are separated. Top1 relieves torsional stress by nicking and re-sealing one DNA strand, and some Top1-dependent mutations are due to trapping and processing of the covalent cleavage intermediate. Most, however, reflect enzyme incision at ribonucleotides, which are the most abundant non-canonical component of DNA. In either case, Top1 generates a distinctive mutation signature comprised of short deletions in tandem repeats; in the specific case of ribonucleotide-initiated events, mutations reflect sequential cleavage by the enzyme. Top1-dependent mutations do not require highly activated transcription, but their levels are greatly increased by transcription, which partially reflects an interaction of Top1 with RNA polymerase. Recent studies have demonstrated that Top1-dependent mutations exhibit a strand bias, with the nature of the bias differing depending on the transcriptional status of the underlying DNA. Under low-transcription conditions, most Top1-dependent mutations arise in the context of replication and reflect incision at ribonucleotides incorporated during leading-strand synthesis. Under high-transcription conditions, most Top1-dependent events arise when the enzyme cleaves the non-transcribed strand of DNA. In addition to increasing genetic instability in growing cells, Top1 activity in transcriptionally active regions may be a source of mutations in quiescent cells.

Graphical abstract

Corresponding Author: Sue Jinks-Robertson, Department of Molecular Genetics and Microbiology, Duke University Medical Center, 384 CARL Bldg. DUMC 3020, 213 Research Dr., Durham, NC 27710, Telephone: +1-919-681-7273, Fax: +1-919-681-7273, sue.robertson@duke.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Topoisomerase 1; transcription; ribonucleotides; mutagenesis

Mutations are permanent changes to DNA that arise spontaneously or are induced by exogenous DNA damaging agents. Basic mechanisms of mutagenesis have been intensively studied because of their relevance to evolutionary processes and to tumorigenesis. With regard to cancer, mutation patterns can reveal information about the driver(s) of genetic instability, both in terms of primary DNA damage and underlying abnormalities in repair processes [1]. This information, in turn, can inform overall prognosis and tumor management. Mutations can be confined to one or a few base pairs, or can affect overall chromosome structure. Small-scale changes include base substitutions, small insertions/deletions and more complex mutations comprised of multiple, closely linked changes. Large-scale alterations such as translocations, inversions and other types of gross chromosomal rearrangements primarily reflection replication fork breakage/repair and will not be considered here.

Most spontaneous mutations in actively dividing cells arise as errors during DNA replication. Replicative DNA polymerases (DNAPs) are highly accurate, but nevertheless insert incorrect nucleotides at a low rate or slip in repetitive regions such as mononucleotide repeats. The resulting base-base mismatches or insertion/deletion loops are efficiently proofread by DNAP exonuclease activity or are removed by the post-replicative mismatch repair (MMR) machinery (reviewed in [2]). If an error escapes repair it becomes fixed as a permanent mutation at the next round of replication. In addition to simple mistakes made when replicating “normal” DNA, damaged DNA can force errors by replicative DNAPs or engage specialized, but error-prone, translesion synthesis (TLS) DNAPs (reviewed in [3]). Endogenous sources of DNA damage that contribute to spontaneous mutagenesis include water, reactive oxygen species, reactive nitrogen species and alkylating agents [4]. For decades, abasic sites and oxidative damage were considered to be the predominant abnormalities in DNA. Recent work has demonstrated, however, that ribonucleotides are the most abundant noncanonical component of DNA (reviewed in [5]). The effects of

ribonucleotides on mutagenesis in *Saccharomyces cerevisiae*, especially in the context of transcription, will be the focus of this review.

Ribonucleotide incorporation into DNA

DNAPs discriminate efficiently between rNTPs and dNTPs, but the high concentration of rNTPs relative to dNTPs in the nucleotide pool leads to frequent ribonucleotide insertion during DNA replication [6-8]. In yeast, it is estimated that DNAPs introduce 10,000-15,000 ribonucleotides during each duplication of the 12 Mb genome [8]. A similar density of genomic ribonucleotides has been inferred in cultured mouse cells [9,10]. In budding yeast, mutant DNAPs that insert either fewer or more ribonucleotides than wild type have been particularly useful for studying physiological consequences of ribonucleotides embedded in DNA [7,11]. In addition to ribonucleotide insertion by replicative DNAPs during genome duplication, primase-generated ribonucleotide chains are required to prime the Okazaki fragments that characterize lagging-strand synthesis. If not completely removed during Okazaki fragment maturation, ligase can inefficiently seal a nick flanked by a single 5' ribonucleotide. Such ligation is frequently aborted to generate a highly toxic 5'-adenylated ribonucleotide that is reversed by aprataxin [12]. Ribonucleotides that persist as nicks are converted into genome destabilizing double-strand breaks at the next round of DNA replication. In addition to their insertion during genome replication, ribonucleotides are likely introduced even more frequently during repair that occurs outside of S phase, when rNTP:dNTP ratios are elevated. Finally, RNA transcripts can be used to prime DNA synthesis [13] or as a template to repair double-strand breaks [14], providing additional potential sources of ribonucleotides in DNA.

Error-free removal of ribonucleotides from DNA

One or a few ribonucleotides embedded in duplex DNA are efficiently removed by the ribonucleotide excision repair (RER) pathway, which has been reconstituted *in vitro* (Figure 1;[15]). RER initiates with incision at the phosphodiester bond 5' of the ribonucleotide by the trimeric RNase H2 complex, which in yeast is comprised of the catalytic Rnh201 and accessory Rnh202 and Rnh203 proteins [16]. *In vitro*, RNase H2 incision provides a 3' OH to prime DNA synthesis, resulting in Pol δ or Pol ϵ displacement of a ribonucleotide-containing flap that is subsequently removed by the flap endonuclease Rad27/Fen1 or by the Exo1 5'>3' exonuclease. Ligation of the remaining nick completes the repair process. Given their abundance in genomic DNA, it has been speculated that ribonucleotides have important physiological functions. In fission yeast, for example, a site-specific di-ribonucleotide is used to initiate mating type switching [17]. In budding yeast, ribonucleotide-associated nicks provide a very weak strand-discrimination signal for MMR [18,19], and may be relevant to relieving torsional stress in the wake of the replication fork [20].

Ribonucleotides may be important in some circumstances, but seminal studies in yeast showed that in the absence of RNase H2, their persistence is highly mutagenic and is associated with a distinctive mutation signature. This signature is comprised of 2-5 bp deletions that reflect loss of a repeat unit from a low copy-number tandem repeat [7]. These events were originally proposed to reflect stalling of DNAP at template ribonucleotides, but

the lack of intermediate removal by MMR suggested a replication-independent deletion mechanism [6]. As will be elaborated below, ribonucleotide-dependent deletions require the activity of Topoisomerase I (Top1), an enzyme that removes both replication- and transcription-associated supercoils [21].

Transcription as a source of genetic instability

Using mutation accumulation as an evolutionary time clock makes the simplifying assumption that genomic DNA mutates at a constant rate. Early studies in bacteria and budding yeast, however, demonstrated that induction of gene expression affects induced mutation rates in the corresponding gene [22,23]. Subsequent studies using reporters fused to highly inducible promoters established a firm connection between transcription and elevated mutagenesis of the underlying DNA template [24-26]. This phenomenon is referred to as transcription-associated mutagenesis or TAM (reviewed in [27]). In *Escherichia coli* and yeast, transcription increases multiple types of base substitutions [28,29]. Using a yeast-based frameshift-reversion assay, it was demonstrated that TAM is reduced in the absence of the primary error-prone, translesion synthesis DNAP and elevated in the absence of error-free repair/bypass pathways [25,30]. These genetic studies suggested that TAM is primarily a consequence of damage to the DNA template. In addition, it was shown in yeast that dUMP is incorporated into transcriptionally active DNA at an elevated level, leading to the formation of mutagenic abasic sites [31]. The physiological reason for transcription-associated dUMP incorporation is unclear, but could reflect increased repair synthesis outside of S phase, when dUTP levels are elevated.

How might the process of transcription sensitize DNA to damage? Transcription requires the transient separation of DNA strands, and single-stranded DNA (ssDNA) is chemically more reactive and more vulnerable to DNA damage than is double-stranded DNA (dsDNA) [4,32]. There also is an inherent asymmetry between the two DNA strands in terms of their single-stranded character. Specifically, the strand that serves as the template for RNA polymerase (RNAP) – the transcribed strand or TS – is relatively protected due to base pairing with the nascent RNA. Within a migrating transcription bubble, there are ~10 nucleotides of ssDNA [33]. The exposure of the NTS is especially acute when the RNA threads back after exiting RNAP and stably base pairs with the template strand, forming a structure known as an R-loop. Consistent with a transcription-associated strand asymmetry during TAM, there is more spontaneous cytosine deamination on the non-transcribed strand (NTS) than on the TS in bacterial cells [24]. In eukaryotes, R-loop formation renders the NTS a better target than the TS for cytosine deaminases that prefer ssDNA [34].

In addition to the inherent TS-NTS asymmetry during transcription, the requisite unwinding of duplex DNA generates positive supercoils ahead of and negative supercoils behind an advancing RNAP [35]. Negative supercoils reflect an underwound state of DNA strands that gives both the TS and NTS single-stranded character and promotes the formation of R-loops [36]. Both positive and negative supercoils are removed by Top1 [21], leading to an early prediction that Top1 would likely contribute to the stability of transcriptionally active DNA. This is true in the case of transcription-associated recombination [37-39], where R-loops are a major driver in replication-transcription conflicts that result in fork breakage (reviewed in

[40]). Genome-wide analysis in yeast has confirmed an enrichment of R-loops at highly transcribed regions [41-43]. Consistent with active transcription exposing the NTS as ssDNA, alterations in primary sequence that increase nascent RNA folding, which is predicted to reduce R-loop formation, mitigate TAM [44]. It should be noted, however, that direct evidence of an NTS-bias in the accumulation of endogenous DNA damage has not been reported in yeast.

In addition to results in defined reporter assays, analysis of mutation accumulation lines in *S. cerevisiae* and comparative genome analyses between different yeasts suggest a global correlation between mutation rates and highly elevated transcription [45]. Given the conserved structure of DNA and basic DNA metabolic processes, it seems likely that TAM is a universal phenomenon. TAM has not been documented in higher eukaryotes, however, except in the case of somatic hypermutation in the vertebrate immune system [46]. This failure could reflect a lack of appropriate reporters. Comparative analysis of mammalian genomes has nevertheless revealed strand asymmetries in some mutation types (e.g., C>T occurs more frequently than G>A), and these have been positively correlated with germline gene expression levels [47]. The favored hypothesis is that such biases reflect more efficient repair of DNA damage on the TS via transcription-coupled nucleotide excision repair, but they could also reflect more damage on the NTS.

Top1 activity is the major source of TAM in yeast

As mentioned previously, studies of TAM using frameshift-reversion assays implicated DNA damage, but subsequent experiments were unable to pinpoint the precise nature of the damage. Frameshift as well as base-substitution reversion assays, however, only identify a very narrow range of mutation types. A breakthrough came with the use of forward mutation assays, which detected a novel mutation signature associated with high levels of transcription. In a *LYS2* forward mutation assay, 2-bp deletions were enriched under high-transcription conditions [48]. Although their numbers were small, the sequence context of these deletions matched the weak consensus sequence for Top1 (5'-A/T-G/C-T/A-T-3'; [49]). Subsequent studies using the *CAN1* forward mutation assay confirmed that 2-5 bp deletions are a unique mutation signature of high transcription, comprising ~50% of mutations [29,50]. These deletions accumulate specifically at low-copy tandem repeats, with the deletion size matching the repeat unit size. There also are clear hotspots for these events, with only a small subset of tandem repeats accumulating deletions. Most importantly, the studies at *CAN1* demonstrated that the transcription-associated deletions require the catalytic activity of Top1. It was suggested that Top1-dependent deletion hotspots reflect an efficient Top1 cleavage site within or adjacent to the relevant tandem repeat, with subsequent trapping of the covalent Top1 cleavage complex (Top1cc). Processing of the Top1cc would then give rise to a small gap that, if within a tandem repeat, facilitates realignment with the complementary strand to convert the gap to a readily ligated nick.

Top1 resolves torsional stress that accumulates during transcription, replication and chromatin remodeling by creating a transient single-strand break in one DNA strand [21,51]. The active site tyrosine of the enzyme forms a phosphotyrosine bond with the 3'-end of the nick, leaving a 5' hydroxyl (5'-OH) on the other side. Controlled rotation of duplex DNA

downstream of the Top1-bound nick resolves the torsional stress, and Top1-assisted attack of the phosphotyrosyl bond by the 5'-OH releases Top1 and restores DNA integrity. Because perfect alignment of the 5'-OH is critical for the religation step, 5'-OH displacement stabilizes the Top1cc. *In vitro*, nearby base-base mismatches or DNA damage efficiently trap the Top1cc [52] as does collision with RNAP on the TS [53]. As noted above, it was proposed that error-prone repair of a trapped Top1cc is likely responsible for the signature, short deletions. This model was later supported genetically through expression of a mutant form of Top1 (Top1-T722A) that shifts the cleavage-ligation equilibrium of the enzyme and stabilizes the Top1cc *in vitro* [54]. As predicted, replacing wild-type Top1 with Top1-T722A increased 2-bp deletions when an appropriate hotspot was monitored in isolation [55].

It should be noted that Top1-dependent deletions are not limited to highly transcribed DNA. When individual deletion hotspots identified in *CAN1* are transplanted into a more sensitive *lys2*-based frameshift-reversion assay that efficiently detects 2-bp deletions, these events are abundant even under low-transcription conditions [29,55]. Their rates are elevated several hundred fold, however, when transcription of the reporter is driven by a highly active tetracycline/doxycycline-regulated promoter instead of the endogenous *LYS2* promoter. The stimulatory effect of high transcription likely reflects increased Top1 recruitment to actively transcribed genes through its interaction with the phosphorylated C-terminal domain of elongating RNAP II [56,57].

Top1 is required for ribonucleotide-dependent deletions

As noted previously, the ribonucleotide persistence associated with the elimination of RNase H2 is accompanied by the accumulation of short deletions in low-copy tandem repeats [7]. Ribonucleotide-dependent deletions were initially observed in the *URA3* forward mutation reporter, which has a low level of transcription [6,7]. Because of the striking similarity between these events and the Top1-dependent deletions observed during TAM, an obvious question was whether there is any relationship between the two. Subsequent experiments demonstrated that the ribonucleotide-dependent events require the activity of Top1 [20,58].

Ribonucleotide-dependent deletion hotspots are operationally defined as those that are highly elevated in the absence of RNase H2 and are most often studied in *rmh201* backgrounds devoid of the catalytic subunit. Whether these hotspots reflect a trapped Top1cc or low-abundance ribonucleotides that persist when RNase H2 is present has not been examined. RNase H2 cleaves at single rNMPs embedded in duplex DNA but also degrades the RNA component of R-loops; the other RNase H enzyme in yeast, RNase H1, requires at least four contiguous ribonucleotides and is assumed to primarily target R-loops (reviewed in [59]). That ribonucleotide-dependent deletions are the result of single ribonucleotides embedded in DNA was demonstrated using a separation-of function mutant of RNase H2 (Rnh201-P45D,Y219A) that retains only the ability to process multiple ribonucleotides. The mutant behaves as a null with respect to limiting short deletions [60].

In the absence of RNase H2, short deletions dramatically increase in the presence of an rNTP-permissive form of DNA Pol ϵ encoded by the *pol2-M644G* allele [7]. Expression of comparable forms of Pol α or Pol δ , however, has little effect on the 2-5 bp deletion rate

[11]. It is generally accepted that Pol ϵ is responsible for most leading-strand synthesis in yeast, while lagging-strand synthesis is catalyzed by Pol α and Pol δ (reviewed in [61]). The genetic results thus suggest that most ribonucleotide-dependent deletions are initiated at ribonucleotides incorporated by Pol ϵ during leading-strand synthesis. It has been speculated that this asymmetry reflects a need for Top1 to remove supercoils that accumulate behind the replication fork during leading-strand synthesis (Figure 2); the nicks that mark Okazaki fragments would presumably serve this role during lagging-strand synthesis [11]. A similar argument has been made with regard to a role for ribonucleotides in MMR, where RER-introduced nicks would aid in strand discrimination during continuous leading-strand synthesis [18,19]. It should be noted that Pol ϵ is naturally more rNTP permissive than Pol δ [8], consistent with physiological significance for ribonucleotides in the removal of supercoils and replication errors.

One anomaly of experiments using the ribonucleotide-permissive forms of Pol δ and Pol ϵ is that the difference in the rate of 2-5 bp deletions is 5-10 fold greater than the difference in ribonucleotide incorporation by these two DNAPs. This suggests there may be active targeting of Top1 to cleave the nascent leading strand rather than its template, which contains ribonucleotides inserted by Pol δ at the prior round of replication. With regard to this asymmetry, Top1 interacts physically with the CMG replicative helicase complex [62], which travels on the leading-strand template.

In addition to the replication-associated effects observed in *rnh201* mutants, RNase H2 loss dramatically elevates TAM at *CAN1* [55]. Upon close examination, however, the locations of ribonucleotide-associated deletion hotspots identified under high-transcription conditions do not precisely match the Top1-dependent deletions identified in the presence of RNase H2. By examining individual hotspots in *pTET-lys2* reporter constructs, it was confirmed that ribonucleotide-dependent deletions require Top1, but that not all Top1-dependent hotspots are enhanced by loss of RNase H2. These results demonstrate the existence of two independent classes of Top1-dependent deletions: those reflecting repair of a trapped Top1cc and those reflecting Top1 incision at a ribonucleotide. It is possible that some hotspots reflect both types of events. As will be discussed further below, ribonucleotide-dependent mutagenesis in the context of transcription occurs primarily when Top1 cleaves the NTS and, therefore, is not governed by which DNAP inserted the ribonucleotide (Figure 2).

Mechanism of Top1-dependent deletions at ribonucleotides

Biochemical studies done almost 20 years ago used vaccinia Top1 (vcTop1) to investigate the consequences of substituting a ribonucleotide for the deoxyribonucleotide at the scissile phosphate [63]. Cleavage occurred efficiently and was followed by attack of the phosphotyrosine bond by the 2'-OH of ribose. This released vcTop1 from the DNA, leaving a single-strand nick flanked by a 2',3' cyclic phosphate and a 5'-OH. It was proposed that a similar nick is the initial DNA damage leading to ribonucleotide-dependent deletions in yeast. Consistent with this idea, it was confirmed that human Top1 has activity similar to vcTop1 on DNA fragments containing Top1-dependent hotspots identified *in vivo* [58]. In contrast to the stimulatory effect of Top1-T722A expression on ribonucleotide-independent deletions, however Top1-T722A expression reduced the rate of ribonucleotide-dependent

deletions *in vivo* [55]. This suggested that proficient ligation by Top1 is important for ribonucleotide-initiated deletions, leading to the proposal of a sequential-cleavage model for their formation (Figure 1; yellow arrows). In this model, Top1 incision at and release from a ribonucleotide is followed by a second Top1 incision event at an upstream (5') deoxyribonucleotide. Spontaneous dissociation from the complementary strand of the short oligonucleotide (oligo) flanked by the Top1 incisions creates a gap and traps the Top1cc. If the gap falls within a tandem repeat, realignment between complementary strands converts the gap to a nick, bringing the 5'-OH close to the Top1cc and allowing efficient Top1-mediated ligation.

The sequential-cleavage model proposes that Top1 is sufficient for deletion formation, and this has been confirmed *in vitro* using purified yeast or human Top1 [64,65]. Subsequent *in vitro* and *in vivo* analyses done in parallel have provided additional insight into the molecular mechanism [66]. As predicted by the model, varying the distance between Top1 cleavage sites *in vitro* produces a corresponding change in deletion size *in vivo*, and up to 7-bp deletions have been detected. Furthermore, there is an inverse relationship between deletion size and rate *in vivo*, which could reflect the length of the leaving oligo, the ease of strand realignment after oligo release and/or sequential cleavage by a single Top1. Finally, the position of the Top1 cleavage sites relative to the repeat units is important, with a repeat linked to the 5'-OH (produced by initial cleavage at the ribonucleotide) realigning more readily with the complementary strand than a repeat linked to the Top1cc (produced by the second cleavage). This difference in end mobility likely reflects the clamping down of Top1 on duplex DNA during incision [67].

Given the highly mutagenic potential of single ribonucleotides in DNA, it is perhaps not surprising that mechanisms exist that limit associated deletions. An early genetic study that monitored a 4-bp deletion hotspot demonstrated synergistic effects of RNase H2 and Rad27/Fen1 loss [68], suggesting that Rad27-mediated 5'-flap removal reduces ribonucleotide-dependent deletions. A more recent analysis demonstrated that the Srs2 3'→5' helicase and the Exo1 5'→3' exonuclease work together to suppress ribonucleotide-dependent deletions *in vivo*. *In vitro*, their combined action removes the 5'-OH flanking the nick created by Top1 incision at a ribonucleotide [69]. Removal of the 5'-OH precludes subsequent Top1-mediated ligation after the second cleavage, thereby preventing rNMP-dependent deletions. Finally, the Sgs1 3'→5' helicase and the Tdp1 tyrosyl-DNA phosphodiesterase were identified in a candidate gene approach as additional activities that likely suppress ribonucleotide-dependent deletions [70]. Sgs1 may act similarly to Srs2, and Tdp1 presumably removes the Top1cc generated by the second cleavage event. Additional activities that limit ribonucleotide-dependent deletions may exist, but the redundancy in end-processing activities in yeast [71] will likely hinder their identification and characterization.

Ribonucleotide-dependent deletions in the context of TAM

The combined effect of highly activated transcription and RNase H2 loss on Top1-dependent deletions in the *CANI* forward mutation assay is multiplicative, suggesting sequential action of RNase H2 and Top1 on ribonucleotides embedded in DNA [55]. Transcriptional status does not affect the amount of ribonucleotides present in the reporter in the absence of RNase

H2 and Top1, consistent with transcription serving to simply recruit more Top1. A greater than multiplicative effect has been observed, however, when hotspots are monitored in isolation [55], suggesting that transcriptional effects may extend beyond simple Top1 recruitment by RNAP II.

A striking aspect of ribonucleotide-dependent deletions is that those occurring in the context of high levels of transcription follow very different rules than those occurring under low-transcription conditions. Under low-transcription conditions (e.g., the endogenous *URA3*, *CAN1* or *LYS2* promoter) only those ribonucleotides incorporated by the Pol ϵ leading-strand DNAP highly stimulate Top1-dependent deletions; those incorporated by the Pol α or Pol δ lagging-strand DNAPs have relatively little effect [11]. When a given hotspot is strongly affected by the direction of DNA replication, the strand cleaved by Top1 can be deduced. At one specific hotspot, this logic was used to identify the (TG)₂-containing, rather than the complementary (CA)₂-containing, strand as the strand targeted by Top1 [72]. Insertion of *pLYS2*-driven (TG)₂ reporter next to a well-defined origin of replication confirmed the strong replication-orientation bias for 2-bp deletions at this particular hotspot. Additional use of rNMP-permissive forms of Pol ϵ and Pol δ demonstrated that the replication-related bias was maintained regardless of whether the (TG)₂ sequence was on the NTS or TS of the low-transcription reporter (Figure 3, boxes with red dashed lines). When the reporter was highly transcribed from *pTET*, however, there was a striking NTS bias that completely eclipsed the replication-orientation bias observed under low-transcription conditions. That is, deletions accumulated primarily when Top1 cleaves the NTS, regardless of whether this strand is generated during leading- or lagging-strand synthesis (Figure 3, boxes with solid black lines). Thus, in contrast to low-transcription conditions where only those ribonucleotides incorporated by Pol ϵ are highly mutagenic [11], Pol δ can be the major source of deletion-causing rNMPs under high-transcription conditions [72].

It has been speculated that the replication-orientation bias for ribonucleotide-dependent deletions reflects a need to remove supercoils that accumulate during leading-strand synthesis [11]. What, then is responsible for the striking NTS bias observed under high-transcription conditions? One possibility is that nicked/gapped intermediates arising on the TS are subject to transcription-coupled nucleotide excision repair (NER), which would preclude their maturation into a deletion event. Elimination of NER, however, does not alter the NTS bias [72]. A second possibility is that Top1 is only able to cleave the NTS, perhaps because the TS strand is shielded as part of an RNA:DNA hybrid. This also is unlikely, however, as overexpression of RNase H1, which should reduce R-loops, does not affect the NTS bias. Top1 could also be specifically targeted to the NTS, but how this might occur is not immediately obvious. A final possibility is that a trapped Top1cc on the NTS is “pushed” by RNAP towards the 5′-OH to promote ligation; when on the TS, RNAP would “pull” the 5′-OH away from the Top1cc and prevent religation (Figure 4).

In addition to a strong NTS bias for the (TG)₂ hotspot under high-transcription conditions, a similar transcription-related bias has been observed at a (CCCTT)₂ hotspot that accumulates ribonucleotide-dependent, 5-bp deletions [66]. (CCCTT)₂ matches the consensus cleavage site for vcTop1 [73], allowing its effect on yeast mutagenesis to be examined in a *top1* background. Expression of vcTop1 generates only ribonucleotide-independent 5-bp

deletions (our unpublished observations), suggesting that a Top1cc is the initiator of the deletion process. Interestingly, a strong NTS bias has also been observed for rNMP-independent vcTop1-dependent events. This suggests there may be a closer mechanistic similarity of ribonucleotide-independent events to ribonucleotide-dependent, sequential Top1-cleavage events than has been previously thought. In the case of Top1cc-initiated events, subsequent Top1 incision would require proteolytic digestion and/or enzymatic removal of the trapped Top1. Varying the concentration of Top1 *in vivo* may allow distinction between a requirement for one versus two Top1 cleavages for the formation of Top1cc-dependent deletions.

Ribonucleotide-dependent TAM events that do not depend on Top1

A novel type of ribonucleotide-dependent TAM that is independent of Top1 activity has been reported in a *lys2*-based frameshift-reversion assay [74]. Under high-transcription conditions in an *rnh201* background, these mutations are comprised of multiple, closely linked sequence alterations. Their recurrent nature suggests that they are likely templated, and indeed, most can be explained by the conversion of an imperfect inverted repeat or quasi-palindrome (QP) into a perfect inverted repeat. In bacteria, such “QP mutations” generally reflect a template switch during DNA synthesis [75,76], and consistent with a similar origin, the yeast QP events were strongly affected by the direction of replication through the reporter. Furthermore, QP mutations accumulated at the same position in each of three *lys2* reversion alleles examined, even though the primary sequence varied slightly, suggesting that a common, position-dependent replication block initiates the template switch. In addition to Top1 independence, QP mutations depend on RNase H1, which requires at least four consecutive ribonucleotides to process an RNA:DNA hybrid; they are limited by components of the NER machinery; and they are promoted by TLS polymerases. Based on these unusual genetic requirements, it was proposed that the RNA:DNA hybrids degraded by RNase H1 might be transcript remnants that re-prime leading-strand synthesis following a co-directional encounter of RNAP with the leading-strand DNAP. Alternatively, the relevant RNA:DNA species might reflect Okazaki fragment remnants that accumulate during lagging-strand synthesis. The involvement of NER could reflect either removal of secondary structure or of ribonucleotides that remain in DNA [77], either of which might present a block during replication. Finally TLS polymerases might be needed to extend an unstable or mismatched 3' end during a template switch.

Summary and concluding remarks

Our current understanding of TAM has come from genetic studies in yeast and bacteria. In microorganisms, transcription rates vary as a function of growth conditions; in higher eukaryotes, transcription patterns vary during development and are expected to exhibit tissue specificity. Studies have shown that transcription elevates damage in the underlying DNA template and is associated with elevated uracil incorporation, leading to the accumulation of diverse mutation types. It should be noted that most TAM studies have relied on reversion assays, each of which detects only a limited range of mutation types; use of other reporters may reveal additional mutation types. In the yeast system, the most abundant source of TAM in a relatively unbiased forward mutation assay reflects the activity of Top1, a type IB

enzyme that is unique to eukaryotes. Whether other topoisomerases that remove transcription-associated supercoils are similarly mutagenic has not been examined.

In yeast, the short-deletion signature of TAM is similar to that associated with persistent ribonucleotides in DNA and indeed, most Top1-dependent TAM reflects initial incision of the enzyme at a ribonucleotide. Ribonucleotide-dependent mutagenesis requires two Top1 cleavage events on the same DNA strand and, in the context of transcription, mutations arise primarily when the enzyme cleaves the non-transcribed strand of the reporter. The basis of this distinctive strand bias is unknown and remains a subject for future investigation.

Although the focus here has been on ribonucleotide-dependent TAM in yeast, at least some events are ribonucleotide independent, with genetic studies suggesting involvement of a stabilized Top1cc. Of particular interest is whether chemotherapeutic drugs that stabilize the Top1cc (e.g., camptothecin and its derivatives) increase Top1-dependent mutagenesis. If so, this should be considered when using these drugs in the clinic.

The traditional view is that most mutations arise during DNA replication, and TAM thus far has only been studied in growing cells. Whether TAM is limited to S phase or occurs throughout the cell cycle is important, as the ratio of rNTPs to dNTPs is elevated outside of S phase. Because transcription damages the underlying DNA template, DNA synthesis associated with damage repair/bypass outside of S phase may be a particularly potent source of ribonucleotides in DNA. Also important are the potential contributions of transcription to genetic instability in non-dividing cells. At least in mammalian cells, there is evidence that transcription increases the instability of trinucleotide repeats in non-growing cells [78]. TAM may be physiologically important for adaptation to stressful conditions and contribute to “targeted” mutagenesis in microorganisms; in humans, TAM could be a driver of genetic instability in oncogenes and tumor suppressors during tumor evolution. Finally, yeast mitochondria contain Top1 [79] but are devoid of RNase H2 [80,81] and this cellular compartment may be particularly vulnerable to Top1-mediated genetic instability. There remain many unanswered questions and likely are additional TAM mechanisms to be discovered, ensuring that this phenomenon will remain an active area of research.

Acknowledgements

Research in the lab of SJR was supported by grants from the National Institutes of Health (GM101690 and GM118077).

References

1. Alexandrov LB, Stratton MR. Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. *Curr. Opin. Genet. Dev.* 2014; 24:52–60. [PubMed: 24657537]
2. Ganai RA, Johansson E. DNA replication-a matter of fidelity. *Mol. Cell.* 2016; 62:745–755. [PubMed: 27259205]
3. Sale JE. Translesion DNA synthesis and mutagenesis in eukaryotes. *Cold Spring Harb. Perspect. Biol.* 2013; 5:a012708. [PubMed: 23457261]
4. Lindahl T. Instability and decay of the primary structure of DNA. *Nature.* 1993; 362:709–715. [PubMed: 8469282]
5. Williams JS, Lujan SA, Kunkel TA. Processing ribonucleotides incorporated during eukaryotic DNA replication. *Nat. Rev. Mol. Cell. Biol.* 2016; 17:350–363. [PubMed: 27093943]

6. Clark AB, Lujan SA, Kissling GE, Kunkel TA. Mismatch repair-independent tandem repeat sequence instability resulting from ribonucleotide incorporation by DNA polymerase ϵ . *DNA Repair*. 2011; 10:476–482. [PubMed: 21414850]
7. Nick McElhinny SA, Kumar D, Clark AB, Watt DL, Watts BE, Lundstrom EB, Johansson E, Chabes A, Kunkel TA. Genome instability due to ribonucleotide incorporation into DNA. *Nat. Chem. Biol.* 2010; 6:774–781. [PubMed: 20729855]
8. Nick McElhinny SA, Watts BE, Kumar D, Watt DL, Lundstrom EB, Burgers PM, Johansson E, Chabes A, Kunkel TA. Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proc. Natl. Acad. Sci. USA*. 2010; 107:4949–4954. [PubMed: 20194773]
9. Hiller B, Achleitner M, Glage S, Naumann R, Behrendt R, Roers A. Mammalian RNase H2 removes ribonucleotides from DNA to maintain genome integrity. *J. Exp. Med.* 2012; 209:1419–1426. [PubMed: 22802351]
10. Reijns MA, Rabe B, Rigby RE, Mill P, Astell KR, Lettice LA, Boyle S, Leitch A, Keighren M, Kilanowski F, et al. Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. *Cell*. 2012; 149:1008–1022. [PubMed: 22579044]
11. Williams JS, Clausen AR, Lujan SA, Marjavaara L, Clark AB, Burgers PM, Chabes A, Kunkel TA. Evidence that processing of ribonucleotides in DNA by topoisomerase 1 is leading-strand specific. *Nat. Struct. Mol. Biol.* 2015; 22:291–297. [PubMed: 25751426]
12. Tumbale P, Williams JS, Schellenberg MJ, Kunkel TA, Williams RS. Aprataxin resolves adenylated RNA-DNA junctions to maintain genome integrity. *Nature*. 2014; 506:111–115. [PubMed: 24362567]
13. Stuckey R, Garcia-Rodriguez N, Aguilera A, Wellinger RE. Role for RNA:DNA hybrids in origin-independent replication priming in a eukaryotic system. *Proc. Natl. Acad. Sci. USA*. 2015; 112:5779–5784. [PubMed: 25902524]
14. Keskin H, Shen Y, Huang F, Patel M, Yang T, Ashley K, Mazin AV, Storici F. Transcript-RNA-templated DNA recombination and repair. *Nature*. 2014; 515:436–439. [PubMed: 25186730]
15. Sparks JL, Chon H, Cerritelli SM, Kunkel TA, Johansson E, Crouch RJ, Burgers PM. RNase H2-Initiated Ribonucleotide Excision Repair. *Mol. Cell*. 2012; 47:980–986. [PubMed: 22864116]
16. Jeong HS, Backlund PS, Chen HC, Karavanov AA, Crouch RJ. RNase H2 of *Saccharomyces cerevisiae* is a complex of three proteins. *Nucleic Acids Res.* 2004; 32:407–414. [PubMed: 14734815]
17. Vengrova S, Dalgaard JZ. The wild-type *Schizosaccharomyces pombe* mat1 imprint consists of two ribonucleotides. *EMBO Rep.* 2006; 7:59–65. [PubMed: 16299470]
18. Ghodgaonkar MM, Lazzaro F, Olivera-Pimentel M, Artola-Boran M, Cejka P, Reijns MA, Jackson AP, Plevani P, Muzi-Falconi M, Jiricny J. Ribonucleotides misincorporated into DNA act as strand-discrimination signals in eukaryotic mismatch repair. *Mol. Cell*. 2013; 50:323–332. [PubMed: 23603115]
19. Lujan SA, Williams JS, Clausen AR, Clark AB, Kunkel TA. Ribonucleotides are signals for mismatch repair of leading-strand replication errors. *Mol. Cell*. 2013; 50:437–443. [PubMed: 23603118]
20. Williams JS, Smith DJ, Marjavaara L, Lujan SA, Chabes A, Kunkel TA. Topoisomerase 1-mediated removal of ribonucleotides from nascent leading-strand DNA. *Mol. Cell*. 2013; 49:1010–1015. [PubMed: 23375499]
21. Wang JC. Cellular roles of DNA topoisomerases: a molecular perspective. *Nat. Rev. Mol. Cell. Biol.* 2002; 3:430–440. [PubMed: 12042765]
22. Herman RK, Dworkin NB. Effect of gene induction on the rate of mutagenesis by ICR-191 in *Escherichia coli*. *J. Bacteriol.* 1971; 106:543–550. [PubMed: 4929867]
23. Savic DJ, Kanazir DT. The effect of a histidine operator-constitutive mutation on UV-induced mutability within the histidine operon of *Salmonella typhimurium*. *Mol. Gen. Genet.* 1972; 118:45–50. [PubMed: 4576007]
24. Beletskii A, Bhagwat AS. Transcription-induced mutations: increase in C to T mutations in the nontranscribed strand during transcription in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*. 1996; 93:13919–13924. [PubMed: 8943036]

25. Datta A, Jinks-Robertson S. Association of increased spontaneous mutation rates with high levels of transcription in yeast. *Science*. 1995; 268:1616–1619. [PubMed: 7777859]
26. Kim N, Abdulovic AL, Gealy R, Lippert MJ, Jinks-Robertson S. Transcription-associated mutagenesis in yeast is directly proportional to the level of gene expression and influenced by the direction of DNA replication. *DNA Repair*. 2007; 6:1285–1296. [PubMed: 17398168]
27. Jinks-Robertson S, Bhagwat AS. Transcription-associated mutagenesis. *Annu. Rev. Genet.* 2014; 48:341–359. [PubMed: 25251854]
28. Klapacz J, Bhagwat AS. Transcription-dependent increase in multiple classes of base substitution mutations in *Escherichia coli*. *J. Bacteriol.* 2002; 184:6866–6872. [PubMed: 12446637]
29. Lippert MJ, Kim N, Cho JE, Larson RP, Schoenly NE, O'Shea SH, Jinks-Robertson S. Role for topoisomerase I in transcription-associated mutagenesis in yeast. *Proc. Natl. Acad. Sci. USA*. 2011; 108:698–703. [PubMed: 21177427]
30. Morey NJ, Greene CN, Jinks-Robertson S. Genetic analysis of transcription-associated mutation in *Saccharomyces cerevisiae*. *Genetics*. 2000; 154:109–120. [PubMed: 10628973]
31. Kim N, Jinks-Robertson S. dUTP incorporation into genomic DNA is linked to transcription in yeast. *Nature*. 2009; 459:1150–1153. [PubMed: 19448611]
32. Frederico LA, Kunkel TA, Shaw BR. A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. *Biochemistry*. 1990; 29:2532–2537. [PubMed: 2185829]
33. Westover KD, Bushnell DA, Kornberg RD. Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science*. 2004; 303:1014–1016. [PubMed: 14963331]
34. Gomez-Gonzalez B, Aguilera A. Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. *Proc. Natl. Acad. Sci. USA*. 2007; 104:8409–8414. [PubMed: 17488823]
35. Liu LF, Wang JC. Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. USA*. 1987; 84:7024–7027. [PubMed: 2823250]
36. Drolet M. Growth inhibition mediated by excess negative supercoiling: the interplay between transcription elongation, R-loop formation and DNA topology. *Mol. Microbiol.* 2006; 59:723–730. [PubMed: 16420346]
37. El Hage A, French SL, Beyer AL, Tollervey D. Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev.* 2010; 24:1546–1558. [PubMed: 20634320]
38. Tuduri S, Crabbe L, Conti C, Tourriere H, Holtgreve-Grez H, Jauch A, Pantesco V, De Vos J, Thomas A, Theillet C, et al. Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nat. Cell Biol.* 2009; 11:1315–1324. [PubMed: 19838172]
39. Kim N, Jinks-Robertson S. Guanine repeat-containing sequences confer transcription-dependent instability in an orientation-specific manner in yeast. *DNA Repair*. 2011; 10:953–960. [PubMed: 21813340]
40. Aguilera A, Garcia-Muse T. R loops: from transcription byproducts to threats to genome stability. *Mol. Cell*. 2012; 46:115–124. [PubMed: 22541554]
41. Chan YA, Aristizabal MJ, Lu PY, Luo Z, Hamza A, Kobor MS, Stirling PC, Hieter P. Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-chip. *PLoS Genet.* 2014; 10:e1004288. [PubMed: 24743342]
42. El Hage A, Webb S, Kerr A, Tollervey D. Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria. *PLoS Genet.* 2014; 10:e1004716. [PubMed: 25357144]
43. Wahba L, Costantino L, Tan FJ, Zimmer A, Koshland D. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes Dev.* 2016; 30:1327–1338. [PubMed: 27298336]
44. Chen X, Yang JR, Zhang J. Nascent RNA folding mitigates transcription-associated mutagenesis. *Genome Res.* 2016; 26:50–59. [PubMed: 26518484]
45. Park C, Qian W, Zhang J. Genomic evidence for elevated mutation rates in highly expressed genes. *EMBO Rep.* 2012; 13:1123–1129. [PubMed: 23146897]

46. Peters A, Storb U. Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. *Immunity*. 1996; 4:57–65. [PubMed: 8574852]
47. McVicker G, Green P. Genomic signatures of germline gene expression. *Genome Res*. 2010; 20:1503–1511. [PubMed: 20686123]
48. Lippert MJ, Freedman JA, Barber MA, Jinks-Robertson S. Identification of a distinctive mutation spectrum associated with high levels of transcription in yeast. *Mol. Cell. Biol*. 2004; 24:4801–4809. [PubMed: 15143174]
49. Been MD, Burgess RR, Champoux JJ. Nucleotide sequence preference at rat liver and wheat germ type 1 DNA topoisomerase breakage sites in duplex SV40 DNA. *Nucleic Acids Res*. 1984; 12:3097–3114. [PubMed: 6326051]
50. Takahashi T, Burguiere-Slezak G, Van der Kemp PA, Boiteux S. Topoisomerase 1 provokes the formation of short deletions in repeated sequences upon high transcription in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA*. 2011; 108:692–697. [PubMed: 21177431]
51. Champoux JJ. DNA topoisomerases: structure, function, and mechanism. *Annu. Rev. Biochem*. 2001; 70:369–413. [PubMed: 11395412]
52. Pourquier P, Ueng LM, Kohlhagen G, Mazumder A, Gupta M, Kohn KW, Pommier Y. Effects of uracil incorporation, DNA mismatches, and abasic sites on cleavage and religation activities of mammalian topoisomerase I. *J. Biol. Chem*. 1997; 272:7792–7796. [PubMed: 9065442]
53. Wu J, Liu LF. Processing of topoisomerase I cleavable complexes into DNA damage by transcription. *Nucleic Acids Res*. 1997; 25:4181–4186. [PubMed: 9336444]
54. Megonigal MD, Fertala J, Bjornsti MA. Alterations in the catalytic activity of yeast DNA topoisomerase I result in cell cycle arrest and cell death. *J. Biol. Chem*. 1997; 272:12801–12808. [PubMed: 9139740]
55. Cho JE, Kim N, Li YC, Jinks-Robertson S. Two distinct mechanisms of Topoisomerase 1-dependent mutagenesis in yeast. *DNA Repair*. 2013; 12:205–211. [PubMed: 23305949]
56. Baranello L, Wojtowicz D, Cui K, Devaiah BN, Chung HJ, Chan-Salis KY, Guha R, Wilson K, Zhang X, Zhang H, et al. RNA Polymerase II regulates Topoisomerase 1 activity to favor efficient transcription. *Cell*. 2016; 165:357–371. [PubMed: 27058666]
57. Wu J, Phatnani HP, Hsieh TS, Greenleaf AL. The phosphoCTD- interacting domain of Topoisomerase I. *Biochem. Biophys. Res. Commun*. 2010; 397:117–119. [PubMed: 20493173]
58. Kim N, Huang SN, Williams JS, Li YC, Clark AB, Cho JE, Kunkel TA, Pommier Y, Jinks-Robertson S. Mutagenic processing of ribonucleotides in DNA by yeast topoisomerase I. *Science*. 2011; 332:1561–1564. [PubMed: 21700875]
59. Cerritelli SM, Crouch RJ. Ribonuclease H: the enzymes in eukaryotes. *FEBS J*. 2009; 276:1494–1505. [PubMed: 19228196]
60. Chon H, Sparks JL, Rychlik M, Nowotny M, Burgers PM, Crouch RJ, Cerritelli SM. RNase H2 roles in genome integrity revealed by unlinking its activities. *Nucleic Acids Res*. 2013; 41:3130–3143. [PubMed: 23355612]
61. Lujan SA, Williams JS, Kunkel TA. DNA Polymerases divide the labor of genome replication. *Trends Cell Biol*. 2016
62. Gambus A, Jones RC, Sanchez-Diaz A, Kanemaki M, van Deursen F, Edmondson RD, Labib K. GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nat. Cell Biol*. 2006; 8:358–366. [PubMed: 16531994]
63. Sekiguchi J, Shuman S. Site-specific ribonuclease activity of eukaryotic DNA topoisomerase I. *Mol. Cell*. 1997; 1:89–97. [PubMed: 9659906]
64. Huang SY, Ghosh S, Pommier Y. Topoisomerase I alone is sufficient to produce short DNA deletions and can also reverse nicks at ribonucleotide sites. *J. Biol. Chem*. 2015; 290:14068–14076. [PubMed: 25887397]
65. Sparks JL, Burgers PM. Error-free and mutagenic processing of topoisomerase 1-provoked damage at genomic ribonucleotides. *EMBO J*. 2015; 34:1259–1269. [PubMed: 25777529]
66. Cho JE, Huang SN, Burgers PM, Shuman S, Pommier Y, Jinks-Robertson S. Parallel analysis of ribonucleotide-dependent deletions produced by yeast Top1 *in vitro* and *in vivo*. *Nucleic Acids Res*. 2016

67. Shuman S, Turner J. Site-specific interaction of vaccinia virus topoisomerase I with base and sugar moieties in duplex DNA. *J. Biol. Chem.* 1993; 268:18943–18950. [PubMed: 7689562]
68. Chen JZ, Qiu J, Shen B, Holmquist GP. Mutational spectrum analysis of RNase H(35) deficient *Saccharomyces cerevisiae* using fluorescence- based directed termination PCR. *Nucleic Acids Res.* 2000; 28:3649–3656. [PubMed: 10982888]
69. Potenski CJ, Niu H, Sung P, Klein HL. Avoidance of ribonucleotide-induced mutations by RNase H2 and Srs2-Exo1 mechanisms. *Nature.* 2014; 511:251–254. [PubMed: 24896181]
70. Niu H, Potenski CJ, Epshtein A, Sung P, Klein HL. Roles of DNA helicases and Exo1 in the avoidance of mutations induced by Top1-mediated cleavage at ribonucleotides in DNA. *Cell Cycle.* 2016; 15:331–336. [PubMed: 26716562]
71. Pommier Y, Barcelo JM, Rao VA, Sordet O, Jobson AG, Thibaut L, Miao ZH, Seiler JA, Zhang H, Marchand C, et al. Repair of topoisomerase I-mediated DNA damage. *Prog. Nucleic Acid Res. Mol. Biol.* 2006; 81:179–229. [PubMed: 16891172]
72. Cho JE, Kim N, Jinks-Robertson S. Topoisomerase 1-dependent deletions initiated by incision at ribonucleotides are biased to the non-transcribed strand of a highly activated reporter. *Nucleic Acids Res.* 2015; 43:9306–9313. [PubMed: 26271994]
73. Shuman S, Prescott J. Specific DNA cleavage and binding by vaccinia virus DNA topoisomerase I. *J. Biol. Chem.* 1990; 265:17826–17836. [PubMed: 2170398]
74. Kim N, Cho JE, Li YC, Jinks-Robertson S. RNA:DNA hybrids initiate quasi-palindrome-associated mutations in highly transcribed yeast DNA. *PLoS Genet.* 2013; 9:e1003924. [PubMed: 24244191]
75. Lovett ST. Encoded errors: mutations and rearrangements mediated by misalignment at repetitive DNA sequences. *Mol. Microbiol.* 2004; 52:1243–1253. [PubMed: 15165229]
76. Ripley LS. Frameshift mutation: determinants of specificity. *Annu. Rev. Genet.* 1990; 24:189–213. [PubMed: 2088167]
77. Vaisman A, McDonald JP, Huston D, Kuban W, Liu L, Van Houten B, Woodgate R. Removal of misincorporated ribonucleotides from prokaryotic genomes: an unexpected role for nucleotide excision repair. *PLoS Genet.* 2013; 9:e1003878. [PubMed: 24244177]
78. Swami M, Hendricks AE, Gillis T, Massood T, Mysore J, Myers RH, Wheeler VC. Somatic expansion of the Huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset. *Hum. Mol. Genet.* 2009; 18:3039–3047. [PubMed: 19465745]
79. Tua A, Wang J, Kulpa V, Wernette CM. Mitochondrial DNA topoisomerase I of *Saccharomyces cerevisiae*. *Biochimie.* 1997; 79:341–350. [PubMed: 9310183]
80. Cerritelli SM, Crouch RJ. The balancing act of ribonucleotides in DNA. *Trends Biochem. Sci.* 2016; 41:434–445. [PubMed: 26996833]
81. Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK. Global analysis of protein localization in budding yeast. *Nature.* 2003; 425:686–691. [PubMed: 14562095]

Research Highlights

Transcription increases damage and elevates mutations in the DNA template

Transcription produces a distinctive deletion signature that requires Top1 activity

Most Top1-dependent mutations initiate at ribonucleotides embedded in DNA

Ribonucleotide-dependent deletions occur via sequential Top1 cleavage of DNA

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.

Error-free and error-prone removal of ribonucleotides from DNA.

In the error-free pathway (gray arrows on the left side), RNase H2 incises 5' of the ribonucleotide (red rectangle) to generate a 3'-OH and 5'-phosphate (P) end. Following strand displacement synthesis by Pol δ or Pol ε (blue pentagon) the 5' flap is removed by Rad27 or Exo1 and the remaining nick is sealed by ligase. Black lines indicate DNA and arrowheads designate 3'-OH ends; dotted lines indicate new DNA synthesis. As an alternative to strand displacement after RNase H2 incision, abortive ligation generates a toxic 5'-AMP attached to the ribose, which can be reversed by aprataxin. The error-prone pathway of ribonucleotide removal (yellow arrows on the right side) generates deletions via sequential Top1 (yellow oval) cleavage. In contrast to RNase H2 incision, Top1 cleaves 3' of

the ribonucleotide to generate a 5'-OH and a covalent linkage of the enzyme to a 3'-phosphate. Nucleophilic attack of the Top1-DNA covalent bond by the 2'-OH of the ribonucleotide releases Top1, leaving behind a nick flanked by a 5'-OH and 2',3'-cyclic phosphate (red triangle attached to the red rectangle). Top1 either facilitates ligation of the ends through the reverse reaction or cleaves DNA a second time. The second cleavage generates a short oligo that is released, resulting in trapping of the Top1cc. Spontaneous realignment of complementary strands brings the 5'-OH and Top1cc into close proximity to facilitate Top1-mediated ligation. Removal of 5'-OH by Srs2 and Exo1 prevents Top1-mediated ligation.

Figure 2.

Ribonucleotide-dependent deletions during replication and transcription in the absence of RNase H2. Black lines are DNA and arrowheads designate 3'-OH ends. Dotted lines are DNA synthesized during replication and solid red lines are RNA primers for DNA synthesis. Red and blue pentagons correspond to Pol ϵ and Pol δ , respectively. "R" corresponds to ribonucleotides; those inserted by Pol ϵ and Pol δ are red and blue, respectively. The red line trailing RNAP is the RNA transcript. Top1 (yellow oval) is tethered to components of the replication (CMG helicase in green) and transcription machineries (purple oval moving in the direction of the purple arrow) to relieve supercoils formed during each. During replication, supercoils form on the leading strand of replication and Top1 incision generates deletions primarily at ribonucleotides incorporated by Pol ϵ . During transcription, deletions reflect incision at ribonucleotides located on the NTS, which can be inserted by either Pol ϵ or Pol δ .

Figure 3.

Strand bias of Top1 cleavage at ribonucleotides during replication and transcription. The experimental set-up used to determine strand specificities of Top1 cleavage in the absence of RNase H2 is illustrated. The white arrow indicates the NTS of a hotspot-containing reporter. The yellow box represents a Top1 cleavage site and the white box corresponds to the non-scissile, complementary strand. The red and black lines correspond to DNA synthesized by Pol ϵ and Pol δ , respectively, and the ribonucleotide (R) incorporated by each is similarly colored. A small fragment containing yellow/white boxes was inverted within the reporter to switch the strand on which the cleavage site resides; this switches the cleavage site between the TS-NTS and between the strands synthesized by Pol ϵ and Pol δ . The direction of replication was reversed by inverting the entire reporter, thereby switching the strands synthesized by Pol ϵ and Pol δ . Under low-transcription conditions, deletions were observed at a high level only in the two configurations marked with a red dotted box. The common feature is that the yellow box contains a red R incorporated by Pol ϵ . Under high-transcription conditions, deletions accumulated only in configurations marked with black box. In both cases, the yellow box is on NTS.

Figure 4.

Encounters with RNAP promote deletions only when Top1 cleaves the NTS. Black lines correspond to DNA and arrowheads indicate 3' ends. A deletion hotspot with three repeat units (boxes) is shown in each panel; inversion of the hotspot moves the cleavage sites from the NTS to the TS in the panels on the left and right, respectively. The Top1cc (yellow oval) is trapped at the end of the red repeat; the repeat within the gap created by sequential Top1 cleavage is in gray. The 5'-OH that borders the gap was generated by the first Top1 incision at a ribonucleotide. The blue arrow depicts the direction of RNAP (blue oval) movement and the trailing red line is the nascent RNA. When the Top1cc is trapped on the NTS, the approaching RNAP pushes it (yellow arrow) and the attached repeat towards the 5'-OH to facilitate ligation. When the Top1cc is trapped on the TS, an advancing RNAP pulls the 5'-OH away from the Top1cc, preventing Top1-mediated ligation.