

Contractions of the C-Terminal Domain of *Saccharomyces cerevisiae* Rpb1p Are Mediated by Rad5p

Taylor Stewart,* Alexandra E. Exner,* Paras Patnaik,* and Stephen M. Fuchs*^{†,‡,1}

*Department of Biology, Tufts University, Medford, MA 02155, [†]Institute for Protein Innovation, Boston, MA 02115, and

[‡]Program in Cellular and Molecular Medicine, Boston Children's Hospital, Boston, MA

ORCID ID: 0000-0001-7261-9214 (S.M.F.)

ABSTRACT The C-terminal domain (CTD) is an essential domain of the largest subunit of RNA polymerase II, Rpb1p, and is composed of 26 tandem repeats of a seven-amino acid sequence, YSPTSPS. Despite being an essential domain within an essential gene, we have previously demonstrated that the CTD coding region is genetically unstable. Furthermore, yeast with a truncated or mutated CTD sequence are capable of promoting spontaneous genetic expansion or contraction of this coding region to improve fitness. We investigated the mechanism by which the CTD contracts using a tet-off reporter system for *RPB1* to monitor genetic instability within the CTD coding region. We report that contractions require the post-replication repair factor Rad5p but, unlike expansions, not the homologous recombination factors Rad51p and Rad52p. Sequence analysis of contraction events reveals that deleted regions are flanked by microhomologies. We also find that G-quadruplex forming sequences predicted by the QGRS Mapper are enriched on the noncoding strand of the CTD compared to the body of *RPB1*. Formation of G-quadruplexes in the CTD coding region could block the replication fork, necessitating post-replication repair. We propose that contractions of the CTD result when microhomologies misalign during Rad5p-dependent template switching via fork reversal.

KEYWORDS

RNA polymerase
tandem repeat
template
switching
repeat
polymorphism

RNA polymerase II is an essential eukaryotic protein complex that is responsible for the transcription of mRNA. The largest subunit of this complex, Rpb1p, has a C-terminal domain (CTD) that serves as an essential binding domain for numerous transcription factors as well as proteins involved in chromatin remodeling and DNA repair (Egloff and Murphy 2008; Fuchs *et al.* 2009). The CTD is composed of tandem repeats of a seven-amino acid sequence, YSPTSPS. The number of repeats varies across organisms and generally increases with increasing organismal complexity: budding yeast typically have 26 repeats, while mammals can have up to 52. Studies have shown, however, that as few as eight repeats can support growth in yeast

(Nonet *et al.* 1987). Despite being an essential region of an essential gene, we have demonstrated that the CTD coding region is genetically unstable. In addition to repeat length variation across organisms, previous work by our lab has shown that the CTD varies in length somewhat across strains of yeast, and yeast with a suboptimal CTD length or mutated CTD sequence are capable of promoting spontaneous expansion or contraction of the coding region in order to improve fitness (Morrill *et al.* 2016). Repeat instability may therefore serve as a mechanism for reducing mutagenesis in an essential sequence by promoting the removal or templated repair of damaged repeats while maintaining overall length.

Tandem repeats are well known to be highly unstable, with mutation rates 10 to 100,000 times higher than the genomic average. Variable tandem repeats are found in a variety of genomic locations, including promoter regions as well as within coding regions, and tandem repeat instability is well known to be associated with disease (Gatchel and Zoghbi 2005; Mirkin 2006; Gemayel *et al.* 2010). For example, trinucleotide repeats are known to be capable of undergoing expansions that can lead to a variety of neurodegenerative diseases. These repeats form stable hairpins that can lead to errors during DNA

Copyright © 2020 Stewart *et al.*

doi: <https://doi.org/10.1534/g3.120.401409>

Manuscript received January 6, 2020; accepted for publication May 26, 2020; published Early Online May 28, 2020.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supplemental material available at figshare: <https://doi.org/10.25387/g3.12343940>.

¹Corresponding author: E-mail: stephen.fuchs@proteininnovation.org

repair and replication (Kim and Mirkin 2013). In addition to disease, there are numerous documented cases in which changes in repeat copy number in coding regions results in variable phenotypes that enable organisms to adapt to different environmental conditions (Fidalgo *et al.* 2006, 2008; Gemayel *et al.* 2015; Press and Queitsch 2017). Tandem repeat instability may therefore be an important driver of evolution (Gemayel *et al.* 2010).

Unlike trinucleotide repeats, little is known about the mechanisms by which complex tandem repeats expand and contract. We are using the CTD, which is comprised of a degenerate 21-base pair repeat, as a model for complex tandem repeats. The CTD, as well as other complex tandem repeats, differs from the simpler trinucleotide repeats because it is not perfectly repetitive and therefore likely does not form canonical hairpin structures. Our lab has developed a tet-off reporter system to monitor expansions and contractions in the CTD, and genetic studies performed by our lab have revealed that homologous recombination factors are required for expansions but not contractions of the CTD (Morrill *et al.* 2016). We therefore sought to determine the mechanism(s) by which the CTD undergoes contractions.

In this manuscript, we measured the frequency of contractions of the CTD in the absence of key DNA repair proteins and determined that contractions require the post-replication repair factor Rad5p but not Rad52p or Rad51p. We also analyzed the sequence of contraction events and found that microhomologies flank the repair junctions. Based on these findings, we propose that template switching via fork reversal mediates contractions of the variable CTD of RNA Polymerase II independently of homologous recombination factors.

MATERIALS AND METHODS

Yeast strains and plasmids

Strains used in this study were derived from GRY3019 (*MATa his3Δ leu2Δ lys2Δ met15Δ trp1Δ::hisG URA::CMV-tTA kanRPtetO7-TATA-RPB1*) (Malagon *et al.* 2006). DNA repair mutants were constructed by heterologous gene replacement and verified by PCR with the primers in Supplemental Table 1. Yeast were grown on synthetic complete (SC) dropout medium or YPD at 30°. Doxycycline (+DOX, 50 µg/ml) was added to plates to control the expression of genomic *RPB1* when appropriate. Plasmids used in this work were described previously in Morrill *et al.* 2016. Plasmids were freshly transformed into yeast and maintained on SC media lacking leucine (SC-Leu).

Spotting assays

For phenotypic growth assays, yeast expressing the appropriate plasmid were grown overnight in SC-Leu. Saturated overnight cultures were used to start fresh cultures in the same medium at an A_{600} of 0.2. The cells were allowed to double at least two times before approximately 1.0×10^7 cells were harvested and resuspended in sterile water in a 96-well plate. Cells were serially diluted fivefold five times and then spotted onto SC-Leu plates with and without doxycycline using a 48-pin replicating tool. Plates were incubated at 30° and imaged after three days.

Suppressor analysis

Individual colonies expressing p4stop were grown overnight in SC-Leu in a 96-well plate. Cultures were serially diluted ten-fold four times in sterile water, and each of the four dilutions were spotted onto SC-Leu+DOX plates. The range of dilutions ensured that single colonies for each culture could be identified for analysis by colony PCR. The plates were allowed to grow for three to four days, or until

colonies were sufficiently large for colony PCR. Primers flanking the CTD were used to amplify the CTD coding region (Table S1). To ensure each suppressor is the result of an independent mutagenic event, only one colony per culture was analyzed. Colony PCR products were visualized by gel electrophoresis using 1% agarose in TBE, and a subset of products were sequenced by traditional Sanger sequencing. Suppressors from at least three independent plasmid transformations were analyzed, and contraction frequencies were calculated from an aggregate of the total number of contractions observed relative to the total number of suppressors analyzed.

Data availability

Strains and plasmids are available upon request. Table S1 lists all primer sequences used in this work. Figure S1 is a multiple sequence alignment of the C-terminal domain coding region of *RPB1* from 93 strains (Strope *et al.* 2015). The number in parenthesis indicates multiple strains have identical sequence. Figures S2-S10 are raw data used to determine the frequencies of CTD contractions. Supplemental material available at figshare: <https://doi.org/10.25387/g3.12343940>.

RESULTS

The CTD coding region of RPB1 is highly polymorphic

Others have studied the evolution of the repetitive CTD of RNA Polymerase II across species (Chapman *et al.* 2008; Liu *et al.* 2010; Yang and Stiller 2014; Simonti *et al.* 2015). Using budding yeast as a model, we set out to more closely examine how this sequence has evolved. Sequence alignments of the CTD from 93 *S. cerevisiae* genomes (Strope *et al.* 2015) revealed that while the total length of the CTD is generally conserved between 24 and 26 repeats, there have been multiple, independent rearrangements within the CTD repeat region (Figure 1 and Figure S1). More specifically, Figure 1A shows spontaneous repeat expansions and contractions in closely related strains of *S. cerevisiae*. It also shows that more distantly related strains have in many cases nearly identical CTD architectures. This suggests that there is a history of both expansions and contractions within the CTD, and as many have shown previously, there is strong selective pressure to conserve a CTD with a particular length. Interestingly, Figure 1B shows the high prevalence of DNA microhomologies within these polymorphic regions suggesting they may serve as templates for these rearrangements.

The repetitive nature of the CTD coding sequence likely enables it to rapidly evolve. While the mechanisms of repeat instability in simple repeat sequences, such as trinucleotide repeats, are well studied (Kim and Mirkin 2013; Usdin *et al.* 2015), little is known about how more complex repeat sequences like the CTD promote variability. Our lab previously determined that expansions of the CTD require Rad52p, suggesting that expansions result from homologous recombination (Morrill *et al.* 2016). In this work, we sought to determine the mechanism(s) by which the CTD contracts.

CTD contraction frequency in the absence of key DNA repair proteins

We previously reported the mutation rate for cells expressing a plasmid-based mutant CTD construct designed to monitor contractions, p4stop (Morrill *et al.* 2016). This construct encodes all 26 yeast CTD repeats, with repeats 8–11 each containing a stop codon (Tyr₁ → stop) and a non-coding Ser₂ → Trp mutation (Figure 2A). Previous studies have determined that a minimum of eight CTD repeats are required for efficient growth (West and Corden 1995), and we have demonstrated that the protein produced from our 4stop mutant

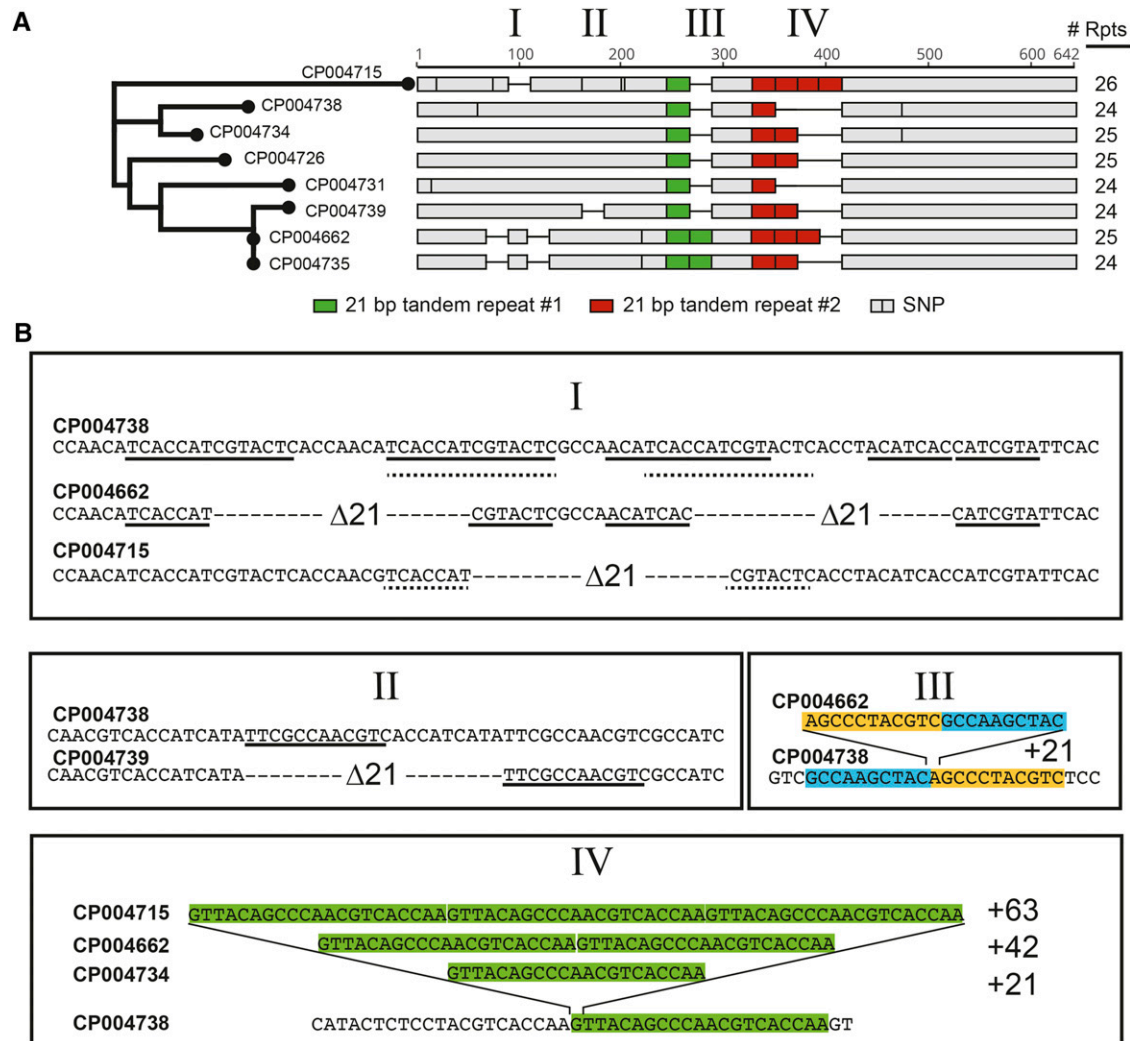


Figure 1 Sequence comparison of *S. cerevisiae* strains reveals multiple, independent instances of expansion and contraction within the repetitive C-terminal domain coding region. A) Sequence alignment of a subset of strains from the analysis by Strobe *et al.* demonstrating four regions of expansion and contraction (I–IV). Red and green blocks indicate two genetically distinct versions of the 21bp motif, and contraction events are represented by horizontal black lines. The tree is based on sequence similarity in the core region of *RPB1*. B) Analysis of the four regions showing that variable-length regions are flanked by regions of homology (underlined) and that all indels are some multiple of 21bp. In region I, two distinct contraction events have occurred in different strains. Microhomologies underlined with a solid line flank contraction event #1 while dashed underlined microhomologies flank contraction event #2. Region II shows a precise 21bp deletion again flanked by microhomologies (underlined). Region III shows an expansion in which the inserted sequence is comprised of the end of one adjacent repeat unit and the beginning of the other, resulting in one additional 21bp motif. Region IV represents three expansions relative to the reference strain in which the highlighted sequence is duplicated. Duplicated sequences are multiples of 21bp, corresponding perfectly to the conserved 7 amino acid repeat sequence of the CTD.

contains only seven CTD repeats (Morrill *et al.* 2016). The expression of this construct is controlled by a tet-off system that we developed based on the work of Strathern and others (Malagon *et al.* 2006). Briefly, the 4stop variant is under the control of the native *RPB1* promoter on a CEN/ARS-containing plasmid that is transformed into yeast. The genomic copy of *RPB1* is under the control of a tetracycline-responsive promoter, and doxycycline is added to the growth medium when appropriate in order to repress transcription of the genomic copy of *RPB1* (Figure 2B). Under these conditions, the cell must rely on the 4stop variant of *Rpb1p* for transcription. Thus, only cells that have acquired a spontaneous mutation that bypasses this selection, often through rearrangement of the *RPB1* coding sequence itself, grow on plates containing doxycycline (Figure 2C).

We sought to determine the mechanism(s) through which the CTD contracts first by measuring the frequency of contractions in the absence of key DNA repair proteins. The p4stop plasmid was transformed into a series of isogenic strains in which different DNA repair genes were deleted, and cells were grown to large colonies to allow for the accumulation of spontaneous mutations. Cultures started from these colonies were then spotted on doxycycline to select for suppressors (Figure 3A). These fast-growing suppressors represent three types of events: 1) contractions within the plasmid that removed the four repeats containing stop codons, 2) homologous recombination events where the mutant plasmid copy of *RPB1* underwent a rearrangement with the doxycycline-regulated copy of *RPB1* in the genome, and 3) mutations elsewhere in the genome (Morrill *et al.* 2016). The overall rate of suppressor formation, which encompasses

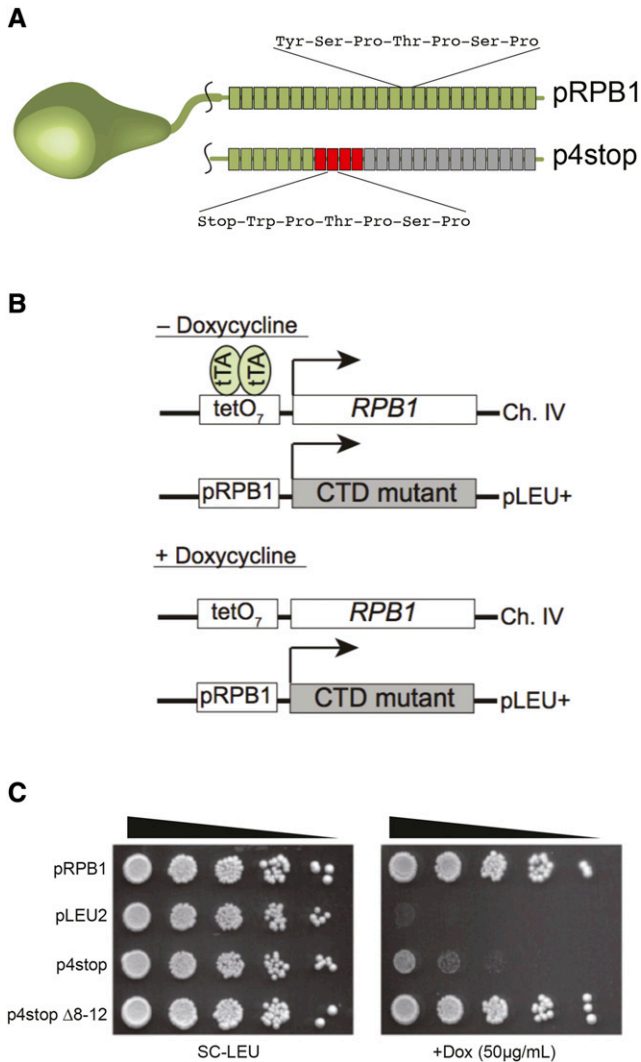


Figure 2 A genetic system for measuring changes in CTD repeat length. A) Schematic of the 4stop mutant CTD constructs. Each block represents one 21-bp repeat encoding the seven-amino acid CTD consensus sequence. pRPB1 encodes all 26 wild-type yeast CTD repeats, whereas p4stop encodes 26 repeats but is interrupted by four stop codons, indicated by red blocks. The protein produced from p4stop has only seven functional CTD repeats. B) Tet-Off system for monitoring contractions within the CTD. In the absence of doxycycline, the Tet transactivator (tTA) binds to tetO₇ sites upstream of the genomic copy of *RPB1*, allowing transcription. In the presence of doxycycline, transcription of the genomic copy is repressed, and cells rely on a plasmid-based copy of *RPB1* under the control of its endogenous promoter. C) Spotting assay demonstrating the effectiveness of our Tet-Off system. pLEU2 is pRS315. p4stop produces a protein product that results in poor yeast viability in the presence of doxycycline. This phenotype is rescued by a contraction event that removes the four repeats containing stop codons. Spotting assays are representative examples of at least three independent trials.

all of these events, was measured by fluctuation analysis (Supplemental Table 2). Although the global mutation rate is variable across different DNA repair-deficient strains, we are interested in how often the mutations that do arise are contractions. Zhao *et al.* reported deletion frequency as a measure of the role of the nuclease ERCC1-XPF in H-DNA processing in human cells. DNA breakage and

deletions occur near H-DNA structure-forming sequences as a result of H-DNA processing, and a reduced deletion frequency in XPF-deficient cells indicates that XPF is involved in H-DNA processing (Zhao *et al.* 2018). Similarly, a decreased contraction frequency in the absence of a given DNA repair pathway indicates that this pathway is important for contractions of the CTD. In the present study, unique suppressors were therefore characterized by colony PCR in order to determine the fraction of suppressors that were contractions.

In the parent strain we find that 56% (84 contractions / 149 suppressors) of fast-growing cells suppress the slow growth phenotype of 4stop by contraction within this region, removing the stop codons (Figure 3B). The frequency of contractions in a *ku70Δ* background is 60% (75/126, Figure S2 and S3), which is not significantly different from wild type based on a two-proportion Z test. However, in both *rad52Δ* (85/87) and *rad51Δ* (30/35, Figure S4) backgrounds, contractions occur significantly more frequently compared to wild type (98 and 86%, respectively). We similarly observed a high frequency of contractions (89%) in the *pol32Δ* background (77/87, Figure S5).

Rad5p, but not other translesion synthesis factors, is required for contractions within the CTD coding region

In addition to the primary DNA repair pathways, we also investigated the role of post-replication repair (PRR) in contractions. PRR consists of two pathways, Translesion Synthesis (TLS) and Template Switching (TS), both of which require *Rad5p*. In a *rad5Δ* mutant background, contractions of the CTD are almost completely abolished (1/105, Figures 4A, S6), defining PRR as a key pathway in spontaneous rearrangements within the CTD coding region. In TLS, *Rad5p* physically interacts with *Rev1p*, which acts as a scaffold for translesion polymerases, such as DNA polymerase zeta, that bypass DNA lesions by inserting a base, potentially incorrectly, across from a lesion (Xu *et al.* 2016). While this pathway is often mutagenic, it generates point mutations and is unlikely to be a source of repeat-length instability. Nonetheless, to confirm that TLS does not play a role in contractions of the CTD, we measured the frequency of contractions in the absence of the scaffolding protein *Rev1p* and the catalytic subunit of polymerase zeta, *Rev3p*. We found that the contraction frequency is not significantly different from wild type in either a *rev1Δ* or a *rev3Δ* background (Figures 4A, S7, S8).

In addition to the DNA lesions that are readily bypassed by TLS, there are many other types of replication barriers, such as transcriptional machinery, DNA-bound proteins, and DNA secondary structures. When the replication fork collides with obstacles that cannot be bypassed by TLS, cells employ the other PRR pathway, Template Switching, to avoid fork collapse (Branzei and Szakal 2016). In addition to its role in TLS, *Rad5p* is important for the initiation of TS through its role in the polyubiquitination of PCNA (Torres-Ramos *et al.* 2002; Gangavarapu *et al.* 2006). Because repetitive sequences of DNA are prone to forming secondary structures such as hairpins, triplex DNA, and G-quadruplexes that impede DNA replication (Usdin *et al.* 2015), we hypothesized that the CTD forms secondary structures that necessitate bypass by *Rad5p*-mediated TS.

While trinucleotide repeats readily form hairpins, the CTD consists of a degenerate 21-bp repeat whose structure-forming ability is not immediately apparent. Capra *et al.* previously reported G₄-DNA near the 3' end of the *RPB1* coding sequence (Capra *et al.* 2010). G₄-DNA is characterized by GG, GGG, or GGGG motifs that are connected by short loops and can form G-quadruplex structures, which consist of stacked G-tetrads (Kikin *et al.* 2006). Our lab recently demonstrated that several segments of the CTD are capable of forming G-quadruplex structures *in vitro*

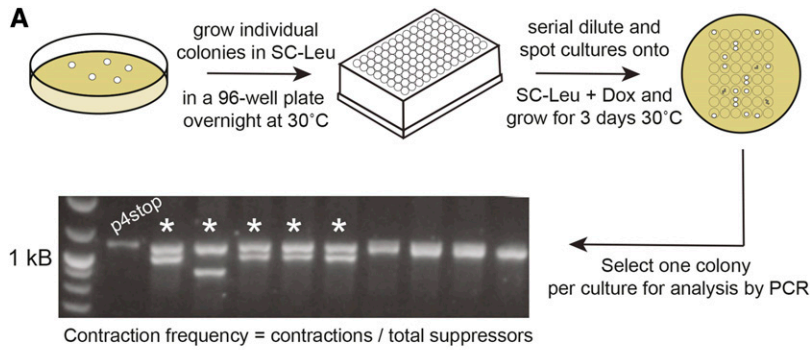
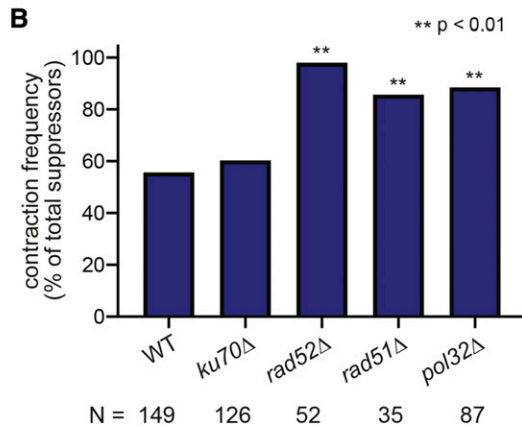


Figure 3 Analysis of CTD contraction frequencies. A) Suppressor generation and characterization assay. p4stop is transformed into yeast and maintained on SC-Leu, and individual colonies are used to start overnight cultures. Saturated overnight cultures are serially diluted and spotted on SC-Leu+DOX. One colony, representing a unique suppressor event, is selected from each spot for analysis by colony PCR. The contraction frequency is calculated by dividing the number of contractions observed by the total number of suppressors analyzed (N). B) Contraction frequencies of the CTD in the absence of key DNA repair proteins. Frequencies are determined by the accumulation of data from three independent plasmid transformations, and statistical significance was determined using a two-proportions z-test.



(Morrill *et al.* 2016). Consistent with our previous results and the results of Capra *et al.* we found that GG motifs are enriched on the noncoding strand of the CTD compared to the body of the *RPB1* gene (Figure 4B). Sequence analysis of the complete *RPB1* gene using the QGRS Mapper developed by Kikin *et al.* also confirms that predicted G-quadruplex forming sequences are enriched in the CTD. Despite the fact that the CTD represents only 10% of the total length of *RPB1*, this region has 33% of the predicted G-quadruplex forming sequences (Figure 4B). In addition to being able to form secondary structures, *RPB1* is a long gene, and increased transcription time may result in collisions between the replicative and transcriptional machineries.

Microhomologies flank deleted sequences in contraction events

We observed microhomologies at sites of rearrangements within the CTD across 93 lab and wild isolates (Figure 1), and in order to verify that our system is a representative model of natural instability, we sequenced p4stop contraction events. Sequence analysis of the repair junctions revealed the presence of microhomologies flanking the deleted regions in all contraction events in our system, which result in the loss of four, five, six, or eight repeats. The most commonly observed contraction event is the in-frame deletion of only the four repeats containing stop codons, repeats 8–11, flanked by a microhomology of 14 bp. Less frequent events involve deletion of up to four additional repeats preceding those containing stop codons, resulting in the loss of up to eight total repeats. Microhomologies flanking contractions ranged in size from 11 bp to 14 bp, but no correlation was observed between microhomology length and the frequency of an observed contraction. Sequences from deletions of four and eight repeats are shown in Figures 5A and 5B, respectively.

The microhomologies bordering the repair junctions that we have observed in the CTD are consistent with repair by microhomology-mediated end-joining (MMEJ). In order to determine if this pathway plays a role in contractions of the CTD, we measured the contraction frequency in *rad1Δ* and *rad10Δ* backgrounds (Figures 5C, S9, S10). Rad1p/Rad10p acts as a complex to cleave the flaps generated during MMEJ (Ma *et al.* 2003; Lee and Sang 2007). In the absence of Rad1p and Rad10p, contractions significantly increased from 56% in the wild type to 79 (22/28) and 83 (35/42) percent, respectively. We therefore conclude that MMEJ does not play a role in contractions of the CTD.

DISCUSSION

Repetitive sequences have generally been dismissed as inconsequential; however, it is becoming increasingly clear that repetitive regions can play important roles in modulating protein function. Repeat copy number in most repetitive coding regions often varies not only at the organismal level, but from generation to generation (Babokhov *et al.* 2018). When found in open reading frames, variable repeats can lead to variable phenotypes, and repetitive sequences throughout the genome may serve as hotspots for evolution. Trinucleotide repeat instability is well known to play a role both in disease and beneficial phenotypic variation, and the mechanisms by which these repeats expand and contract are well studied (Gemayel *et al.* 2010, 2015; Kim and Mirkin 2013; Khristich and Mirkin 2020). While more complex tandem repeats are also known to be variable and to contribute to variable phenotypes (Verstrepen *et al.* 2005; Chapman *et al.* 2008; Fidalgo *et al.* 2008), less is known about the molecular mechanisms that contribute to variation.

The CTD of RNA polymerase II is a fascinating model for exploring repeat copy number variation as the sequence acts as a

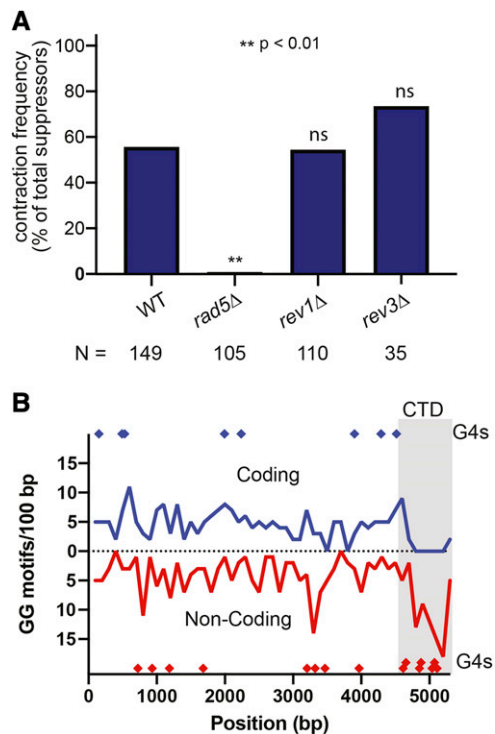


Figure 4 Template switching bypasses G-quadruplexes and mediates contractions of the CTD. A) Contraction frequencies of the CTD in the absence of DNA repair proteins involved in post-replication repair. Frequencies are determined by the accumulation of data from three independent plasmid transformations, and statistical significance was determined using a two-proportions z-test. B) The CTD is enriched in predicted G-quadruplex-forming sequences. GG motifs on the coding (orange) and noncoding (blue) strands of *RPB1*. G-quadruplex-forming sequences predicted by the QGRS Mapper are highlighted. The region between 4624 and 5202 bp is the CTD.

scaffolding domain that is essential for transcription, and its amino acid sequence is likewise highly conserved across eukaryotes. Despite this, the coding sequence is both highly polymorphic and under strong purifying selection to retain function. Using the heptapeptide repeat of the CTD of RNA polymerase II as a model, we set out to identify the mechanisms by which complex tandem repeats promote variability and drive evolution. In addition to variable repeat copy number across organisms, with humans having twice as many repeats as yeast, there is significant evidence of rearrangements of the CTD across strains of yeast (Figure 1). We previously determined that expansions of the CTD require *Rad52p* (Morrill *et al.* 2016); in this work, we sought to determine the mechanism by which the CTD undergoes contractions.

In a *ku70Δ* background, the contraction frequency was not significantly different from wild type, suggesting that contractions do not occur by non-homologous end-joining. The contraction frequency was significantly increased in *pol32Δ*, *rad52Δ*, and *rad51Δ* backgrounds, indicating that contractions are not a result of break-induced replication or canonical homologous recombination. In fact, our data support that homology-directed repair is a competing mechanism for dealing with genetic instability. Ruling out the predominant DNA repair pathways, we hypothesized that PRR mediates contractions. Both PRR pathways (TS and TLS) require the ubiquitin ligase *Rad5p*. In a *rad5Δ* background, the frequency of contractions was reduced from 56 to less than 1%,

indicating that contractions are mediated by one of the PRR pathways. TLS, which requires *REV1* and *REV3*, primarily results in point mutations and was not expected to lead to repeat instability. In both a *rev1Δ* and a *rev3Δ* background, the contraction frequency of the CTD was not significantly different from wild type, confirming that TLS does not play a role in contractions of the CTD. Because contractions require *Rad5p* but are not mediated by TLS, we propose that contractions of the CTD occur by the other PRR pathway, template switching. In order to verify that TS mediates contractions, we attempted to measure the contraction frequency of the CTD in a *rad18Δ* background. We were not able to make this deletion directly in our haploid strain, and a heterozygous diploid could not be sporulated to generate the desired mutant. However, we believe that the combination of the *rad5Δ*, *rev1Δ*, and *rev3Δ* mutants is sufficient to isolate TS.

As demonstrated in Figure 4B, the CTD-coding region of *RPB1* is rich in GG motifs, and in previous work we showed that oligonucleotides derived from this region can form G-quadruplex-like structures (Morrill *et al.* 2016). Template switching allows cells to bypass replication obstacles such as G-quadruplexes and can occur by two mechanisms: 1) a strand invasion mechanism in which the nascent strand invades into the sister chromatid in order to bypass the obstacle and 2) a fork reversal mechanism in which the replication fork regresses, and the nascent strands dissociate from their respective templates and anneal to each other (Branzei and Szakal 2016). Template switching by strand invasion requires the *RAD52* epistasis group in addition to *RAD5* (Ball *et al.* 2009; Vanoli *et al.* 2010); however, we have determined that neither *Rad51p* nor *Rad52p* is required for contractions of the CTD. *Rad5p* can mediate template switching by fork reversal independently of homologous recombination factors (Blastyák *et al.* 2007; Shin *et al.* 2018; Bryant *et al.* 2019). Our working model is therefore that contractions of the CTD occur as a result of template switching by fork reversal, which cells may employ to stabilize the replication fork during collisions with G-quadruplexes (or related structures).

Template switching events are typically thought to be error-free due to the use of a homologous template as the source for bypassing the replication obstacle (Lawrence and Christensen 1979; Zhang and Lawrence 2005; Branzei and Szakal 2016). Repetitive sequences are unusual, however, due to the presence of multiple homologous templates located in close proximity, which can result in repair events that change repeat copy number while maintaining the open reading frame and primary amino acid sequence. We examined the repair junctions of contraction events and discovered microhomologies ranging from 11 to 14 bp in length that could serve as sites of misalignment during template switching via fork reversal (Figure 5A and B). This misalignment would result in a segment of the CTD coding region looping out and being cleaved by an endonuclease, leading to a contraction (Figure 6A). We also considered that the microhomologies present in the CTD could act as templates for repair by MMEJ. Unexpectedly, contractions increased in the absence of *Rad1p* and *Rad10p*, suggesting that contractions of the CTD do not require MMEJ. Instead, we postulate that the microhomologies that we have observed flanking contraction junctions both in our system and in a diverse set of yeast strains are sites of misalignment during fork reversal.

Our previously published data indicates that expansions of the CTD require *Rad52p* (Morrill *et al.* 2016). We therefore initially hypothesized that expansions are mediated by canonical homologous recombination. We also found, however, that the frequency of

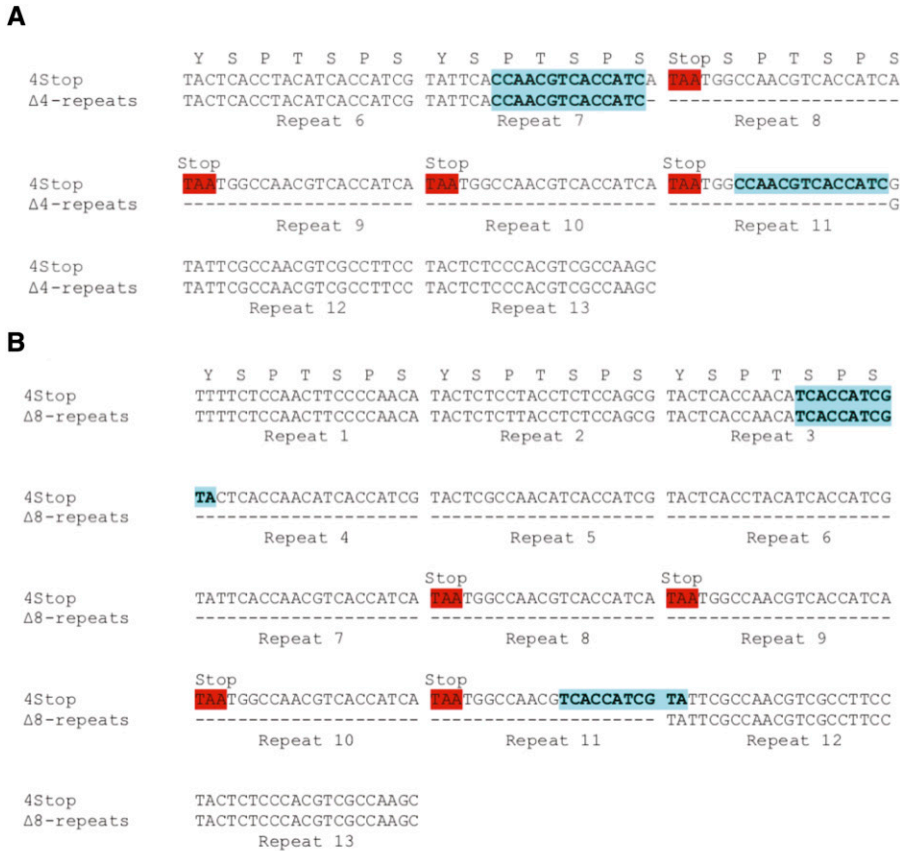
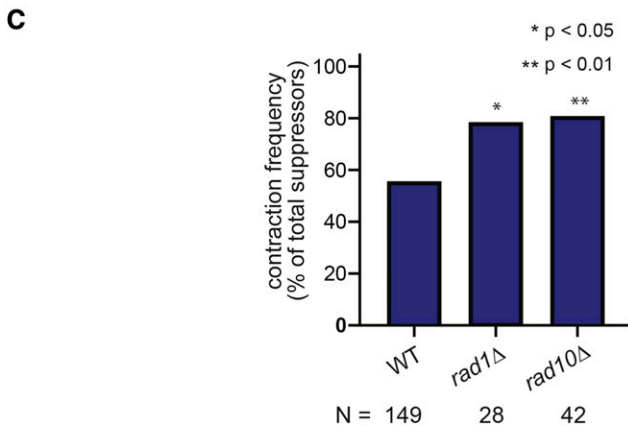


Figure 5 Microhomologies within the CTD serve as templates for misalignment during TS. A) Sequence analysis of a contraction event in which the four repeats containing stop codons were deleted. A 14-bp microhomology is present before and after the deletion. B) Sequence analysis of a contraction event in which eight total repeats, including those containing stop codons, were deleted. An 11-bp microhomology is present before and after the deletion. C) Contraction frequencies of the CTD in the absence of DNA repair proteins involved in microhomology-mediated end-joining. Frequencies are determined by the accumulation of data from three independent plasmid transformations, and statistical significance was determined using a two-proportions z-test.



expansions significantly decreased in a *rad5*Δ background (unpublished data). Based on this result and our new evidence that contractions of the CTD are mediated by template switching, we propose that expansions of the CTD may also be mediated by TS, but through a Rad52p-mediated strand invasion mechanism as opposed to fork reversal (Figure 6B). This hypothesis is supported by the fact that expansions increased in a *pif1-m2* background (Morrill *et al.* 2016). Pif1p is a DNA helicase that plays a role in unwinding G-quadruplex structures (Paeschke *et al.* 2013), and in the absence of functional Pif1p, G-quadruplex structures formed by the CTD sequence may block the replication fork more frequently, promoting template switching. Pif1p was also recently shown to play a direct role in processing regions of ssDNA in preparation for template switching (García-Rodríguez *et al.* 2018), which is consistent with the increased expansion frequency in

the *pif1-m2* background if expansions are indeed mediated by template switching.

The role of the CTD as a signaling and binding domain demands a repetitive sequence, and the enrichment of GG motifs and predicted G-quadruplex forming sequences are isolated to the CTD. The CTD sequence therefore is highly prone to instability compared to the body of the essential *RPB1* gene. This instability may serve as a mechanism to reduce mutagenesis in an essential sequence by allowing for the templated removal of damaged repeats while maintaining overall length. We have observed this phenomenon in our system through the removal of artificially introduced stop codons, and sequence analysis of 93 yeast strains from a wide range of sources demonstrates that this phenomenon occurs in wild populations as well (Figures 1 and S1). The microhomologies that we observed flanking repair junctions provide evidence for such templated rearrangements. We

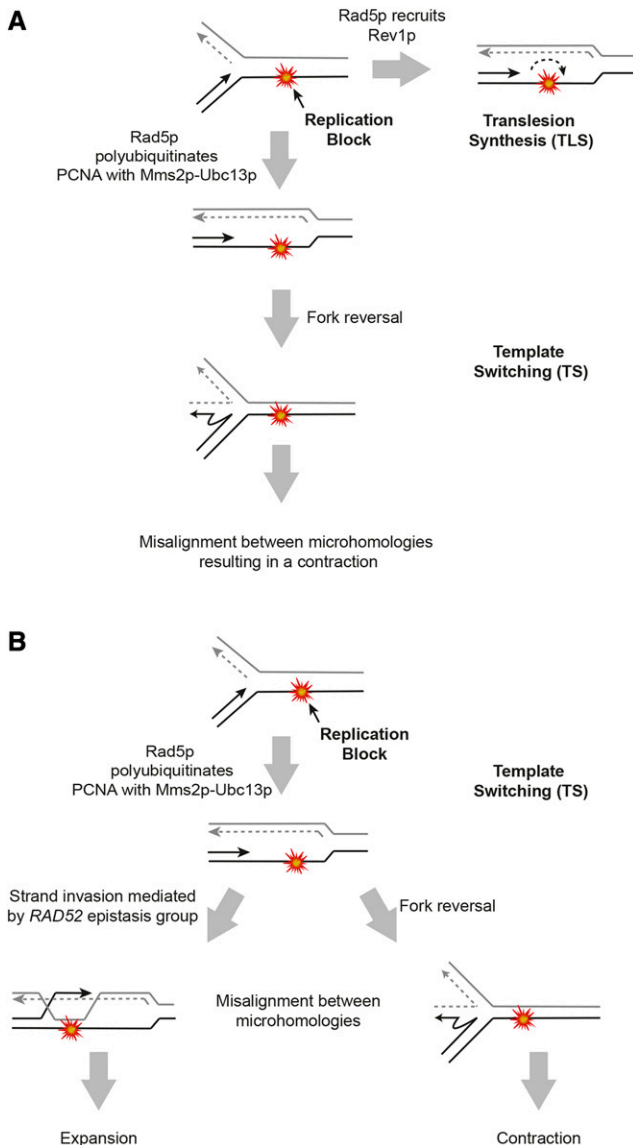


Figure 6 Template switching during post-replication repair mediates repeat instability of the CTD. A) During post-replication repair, the replication fork encounters an obstacle that it cannot bypass on its own. One repair option is TLS, in which translesion polymerases replace the replicative polymerase to bypass the damage. Cells also employ template switching to bypass the obstacle using the sister chromatid as a template. Due to the presence of microhomologies throughout the CTD, we propose that misalignment during fork reversal leads to contractions of the CTD. B) When the replication fork encounters an obstacle, such as a G-quadruplex, template switching can be used to bypass it. TS is initiated by Rad5p-mediated polyubiquitination of PCNA and can occur by two mechanisms—strand invasion and fork reversal. During strand invasion, microhomologies in the CTD result in the re-replication of one or more repeats, leading to expansions. This process requires Rad52p. During fork reversal, microhomologies align, causing a segment of the CTD to loop out, resulting in a contraction.

report that contractions of the CTD require Rad5p but not Rev1p, Rev3p, Rad52p, or Rad51p, indicating that contractions are mediated by a PRR pathway that does not require translesion synthesis or strand invasion. Combined with the presence of secondary structure-forming sequences in the CTD, this data supports a fork reversal

mechanism. Our work demonstrates that the alignment of microhomologies during template switching events can lead to variability in complex tandem repeats, enabling these sequences to both promote evolution and maintain functionality.

ACKNOWLEDGMENTS

The authors would like to acknowledge Summer Morrill, Mohammad Mosaheb, and Bradley Reinfeld for early contributions to this project. We also thank M. Parmenter, K.A. McElroy, and other members of the Fuchs Lab for their support and helpful discussions during the preparation of this manuscript. This work was supported by the Army Research Office [W911NF-16-1-0175 and W911NF-19-1-0299 to S.M.F.].

LITERATURE CITED

- Babokhov, M., B. I. Reinfeld, K. Hackbarth, Y. Bentov, and S. M. Fuchs, 2018 Tandem repeats drive variation of intrinsically disordered regions in budding yeast. *bioRxiv*. (Preprint posted on June 05, 2018)<https://doi.org/10.1101/339663>
- Ball, L. G., K. Zhang, J. A. Cobb, C. Boone, and W. Xiao, 2009 The yeast Shu complex couples error-free post-replication repair to homologous recombination. *Mol. Microbiol.* 73: 89–102. <https://doi.org/10.1111/j.1365-2958.2009.06748.x>
- Blastyák, A., L. Pintér, I. Unk, L. Prakash, S. Prakash *et al.*, 2007 Yeast Rad5 Protein Required for Postreplication Repair Has a DNA Helicase Activity Specific for Replication Fork Regression. *Mol. Cell* 28: 167–175. <https://doi.org/10.1016/j.molcel.2007.07.030>
- Branzei, D., and B. Szakal, 2016 DNA damage tolerance by recombination: Molecular pathways and DNA structures. *DNA Repair (Amst.)* 44: 68–75. <https://doi.org/10.1016/j.dnarep.2016.05.008>
- Bryant, E. E., I. Sunjevaric, L. Berchowitz, R. Rothstein, and R. J. D. Reid, 2019 Rad5 dysregulation drives hyperactive recombination at replication forks resulting in cisplatin sensitivity and genome instability. *Nucleic Acids Res.* 47: 9144–9159. <https://doi.org/10.1093/nar/gkz631>
- Capra, J. A., K. Paeschke, M. Singh, and V. A. Zakian, 2010 G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in *Saccharomyces cerevisiae*. *PLOS Comput. Biol.* 6: e1000861. <https://doi.org/10.1371/journal.pcbi.1000861>
- Chapman, R. D., M. Heidemann, C. Hintermair, and D. Eick, 2008 Molecular evolution of the RNA polymerase II CTD. *Trends Genet.* 24: 289–296. <https://doi.org/10.1016/j.tig.2008.03.010>
- Egloff, S., and S. Murphy, 2008 Cracking the RNA polymerase II CTD code. *Trends Genet.* 24: 280–288. <https://doi.org/10.1016/j.tig.2008.03.008>
- Fidalgo, M., R. R. Barrales, J. I. Ibeas, and J. Jimenez, 2006 Adaptive evolution by mutations in the FLO11 gene. *Proc. Natl. Acad. Sci. USA* 103: 11228–11233. <https://doi.org/10.1073/pnas.0601713103>
- Fidalgo, M., R. R. Barrales, and J. Jimenez, 2008 Coding repeat instability in the FLO11 gene of *Saccharomyces yeasts*. *Yeast* 25: 879–889. <https://doi.org/10.1002/yea.1642>
- Fuchs, S. M., R. N. Larabee, and B. D. Strahl, 2009 Protein modifications in transcription elongation. *Biochim. Biophys. Acta. Gene Regul. Mech.* 1789: 26–36. <https://doi.org/10.1016/j.bbgrm.2008.07.008>
- Gangavarapu, V., L. Haracska, R. E. Johnson, L. Prakash, S. Prakash *et al.*, 2006 Mms2-Ubc13-Dependent and -Independent Roles of Rad5 Ubiquitin Ligase in Postreplication Repair and Translesion DNA Synthesis in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 26: 7783–7790. <https://doi.org/10.1128/MCB.01260-06>
- García-Rodríguez, N., R. P. Wong, and H. D. Ulrich, 2018 The helicase Pif1 functions in the template switching pathway of DNA damage bypass. *Nucleic Acids Res.* 46: 8347–8356. <https://doi.org/10.1093/nar/gky648>
- Gatchel, J. R., and H. Y. Zoghbi, 2005 Diseases of unstable repeat expansion: mechanisms and common principles. *Nat. Rev. Genet.* 6: 743–755. <https://doi.org/10.1038/nrg1691>
- Gemayel, R., S. Chavali, K. Pougach, M. Legendre, B. Zhu *et al.*, 2015 Variable Glutamine-Rich Repeats Modulate Transcription Factor Activity. *Mol. Cell* 59: 615–627. <https://doi.org/10.1016/j.molcel.2015.07.003>

- Gemayel, R., M. D. Vences, M. Legendre, and K. J. Verstrepen, 2010 Variable Tandem Repeats Accelerate Evolution of Coding and Regulatory Sequences. *Annu. Rev. Genet.* 44: 445–477. <https://doi.org/10.1146/annurev-genet-072610-155046>
- Khristical, A. N., and S. M. Mirkin, 2020 On the wrong DNA track: Molecular mechanisms of repeat-mediated genome instability. *J. Biol. Chem.* 295: 4134–4170.
- Kikin, O., L. D'Antonio, and P. S. Bagga, 2006 QGRS Mapper: A web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.* 34: W676–W682. <https://doi.org/10.1093/nar/gkl253>
- Kim, J. C., and S. M. Mirkin, 2013 The balancing act of DNA repeat expansions. *Curr. Opin. Genet. Dev.* 23: 280–288. <https://doi.org/10.1016/j.gde.2013.04.009>
- Lawrence, C. W., and R. B. Christensen, 1979 Metabolic suppressors of trimethoprim and ultraviolet light sensitivities of *Saccharomyces cerevisiae* rad6 mutants. *J. Bacteriol.* 139: 866–876. <https://doi.org/10.1128/JB.139.3.866-876.1979>
- Lee, K., and E. L. Sang, 2007 *Saccharomyces cerevisiae* Sae2- and Tel1-dependent single-strand DNA formation at DNA break promotes microhomology-mediated end joining. *Genetics* 176: 2003–2014. <https://doi.org/10.1534/genetics.107.076539>
- Liu, P., J. M. Kenney, J. W. Stiller, and A. L. Greenleaf, 2010 Genetic organization, length conservation, and evolution of RNA polymerase II carboxyl-terminal domain. *Mol. Biol. Evol.* 27: 2628–2641. <https://doi.org/10.1093/molbev/msq151>
- Ma, J.-L., E. M. Kim, J. E. Haber, and S. E. Lee, 2003 Yeast Mre11 and Rad1 Proteins Define a Ku-Independent Mechanism To Repair Double-Strand Breaks Lacking Overlapping End Sequences. *Mol. Cell. Biol.* 23: 8820–8828. <https://doi.org/10.1128/MCB.23.23.8820-8828.2003>
- Malagon, F., M. L. Kireeva, B. K. Shafer, L. Lubkowska, M. Kashlev *et al.*, 2006 Mutations in the *Saccharomyces cerevisiae* RPB1 gene conferring hypersensitivity to 6-azauracil. *Genetics* 172: 2201–2209. <https://doi.org/10.1534/genetics.105.052415>
- Mirkin, S. M., 2006 DNA structures, repeat expansions and human hereditary disorders. *Curr. Opin. Struct. Biol.* 16: 351–358. <https://doi.org/10.1016/j.sbi.2006.05.004>
- Morrill, S. A., A. E. Exner, M. Babokhov, B. I. Reinfeld, and S. M. Fuchs, 2016 DNA instability maintains the repeat length of the yeast RNA polymerase II C-terminal domain. *J. Biol. Chem.* 291: 11540–11550. <https://doi.org/10.1074/jbc.M115.696252>
- Nonet, M. L., C. Scafe, J. Sexton, and R. Young, 1987 Eucaryotic RNA polymerase conditional mutant that rapidly ceases mRNA synthesis. *Mol. Cell. Biol.* 7: 1602–1611. <https://doi.org/10.1128/MCB.7.5.1602>
- Paeschke, K., M. L. Bochman, P. Daniela Garcia, P. Cejka, K. L. Friedman *et al.*, 2013 Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature* 497: 458–462. <https://doi.org/10.1038/nature12149>
- Press, M. O., and C. Queitsch, 2017 Variability in a Short Tandem Repeat Mediates Complex Epistatic Interactions in *Arabidopsis thaliana*. *Genetics* 205: 455–464. <https://doi.org/10.1534/genetics.116.193359>
- Shin, S., K. Hyun, J. Kim, and S. Hohng, 2018 ATP Binding to Rad5 Initiates Replication Fork Reversal by Inducing the Unwinding of the Leading Arm and the Formation of the Holliday Junction. *Cell Rep.* 23: 1831–1839. <https://doi.org/10.1016/j.celrep.2018.04.029>
- Simonti, C. N., K. S. Pollard, S. Schröder, D. He, B. G. Bruneau *et al.*, 2015 Evolution of lysine acetylation in the RNA polymerase II C-terminal domain. *BMC Evol. Biol.* 15: 35. <https://doi.org/10.1186/s12862-015-0327-z>
- Strope, P. K., D. A. Skelly, S. G. Kozmin, G. Mahadevan, E. A. Stone *et al.*, 2015 The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Res.* 25: 762–774. <https://doi.org/10.1101/gr.185538.114>
- Torres-Ramos, C. A., S. Prakash, and L. Prakash, 2002 Requirement of RAD5 and MMS2 for Postreplication Repair of UV-Damaged DNA in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 22: 2419–2426. <https://doi.org/10.1128/MCB.22.7.2419-2426.2002>
- Usdin, K., N. C. M. House, and C. H. Freudenreich, 2015 Repeat instability during DNA repair: Insights from model systems. *Crit. Rev. Biochem. Mol. Biol.* 50: 142–167. <https://doi.org/10.3109/10409238.2014.999192>
- Vanoli, F., M. Fumasoni, B. Szakal, L. Maloisel, and D. Branzei, 2010 Replication and recombination factors contributing to recombination-dependent bypass of DNA lesions by template switch. *PLoS Genet.* 6: e1001205. <https://doi.org/10.1371/journal.pgen.1001205>
- Verstrepen, K. J., A. Jansen, F. Lewitter, and G. R. Fink, 2005 Intragenic tandem repeats generate functional variability. *Nat. Genet.* 37: 986–990. <https://doi.org/10.1038/ng1618>
- West, M. L., and J. L. Corden, 1995 Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics* 140: 1223–1233.
- Xu, X., A. Lin, C. Zhou, S. R. Blackwell, Y. Zhang *et al.*, 2016 Involvement of budding yeast Rad5 in translesion DNA synthesis through physical interaction with Rev1. *Nucleic Acids Res.* 44: 5231–5245. <https://doi.org/10.1093/nar/gkw183>
- Yang, C., and J. W. Stiller, 2014 Evolutionary diversity and taxon-specific modifications of the RNA polymerase II C-terminal domain. *Proc. Natl. Acad. Sci. USA* 111: 5920–5925. <https://doi.org/10.1073/pnas.1323616111>
- Zhang, H., and C. W. Lawrence, 2005 The error-free component of the RAD6/RAD18 DNA damage tolerance pathway of budding yeast employs sister-strand recombination. *Proc. Natl. Acad. Sci. USA* 102: 15954–15959. <https://doi.org/10.1073/pnas.0504586102>
- Zhao, J., G. Wang, I. M. del Mundo, J. A. McKinney, X. Lu *et al.*, 2018 Distinct Mechanisms of Nuclease-Directed DNA-Structure-Induced Genetic Instability in Cancer Genomes. *Cell Rep.* 22: 1200–1210. <https://doi.org/10.1016/j.celrep.2018.01.014>

Communicating editor: G. Brown