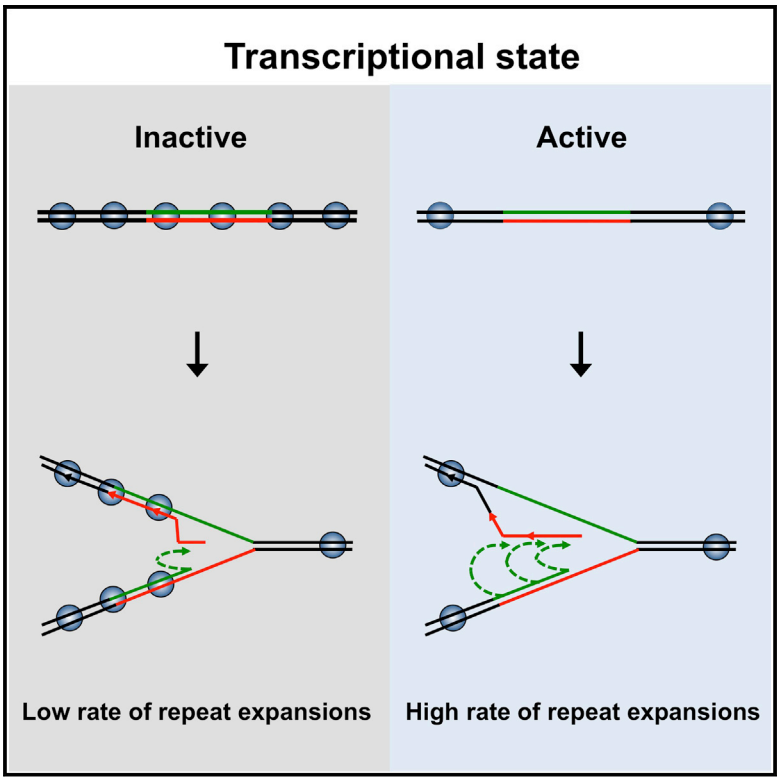
Report

Cell Reports

Coupling Transcriptional State to Large-Scale Repeat Expansions in Yeast

Graphical Abstract



Highlights

Large-scale repeat expansions can occur in the lack of transcription

Transcriptional activation of a locus elevates the rate of repeat expansions

Replication defects affect expansions irrespective of a repeat's transcriptional state

Nucleosome density acts as an inherent modulator of repeat expansions

Authors

Kartik A. Shah, Ryan J. McGinty, Vera I. Egorova, Sergei M. Mirkin

Correspondence

sergei.mirkin@tufts.edu

In Brief

The molecular pathways of replication and transcription are implicated in the expansion of simple DNA repeats, but it has been difficult to distinguish their relative contributions. Shah et al. design and utilize genetic systems to investigate repeat instability in the absence of transcription, showing that transcriptional state rather than transcription through the repeat per se affects this process. The authors present a model linking transcriptional state to the density of nucleosomes in the repeat-containing region, which inherently acts as a modulator of large-scale expansions.

Shah et al., 2014, Cell Reports 9, 1594–1602
December 11, 2014 ©2014 The Authors
http://dx.doi.org/10.1016/j.celrep.2014.10.048

OPEN ACCESS
CellPress

Cell Reports
Report

Coupling Transcriptional State to Large-Scale Repeat Expansions in Yeast

Kartik A. Shah,[1,2] Ryan J. McGinty,[1] Vera I. Egorova,[1] and Sergei M. Mirkin[1,*]
[1]Department of Biology, Tufts University, Medford, MA 02155, USA
[2]Present address: Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
*Correspondence: sergei.mirkin@tufts.edu
http://dx.doi.org/10.1016/j.celrep.2014.10.048

SUMMARY

Expansions of simple DNA repeats cause numerous hereditary disorders in humans. Replication, repair, and transcription are implicated in the expansion process, but their relative contributions are yet to be distinguished. To separate the roles of replication and transcription in the expansion of Friedreich's ataxia $(GAA)_n$ repeats, we designed two yeast genetic systems that utilize a galactose-inducible *GAL1* promoter but contain these repeats in either the transcribed or nontranscribed region of a selectable cassette. We found that large-scale repeat expansions can occur in the lack of transcription. Induction of transcription strongly elevated the rate of expansions in both systems, indicating that active transcriptional state rather than transcription through the repeat per se affects this process. Furthermore, replication defects increased the rate of repeat expansions irrespective of transcriptional state. We present a model in which transcriptional state, linked to the nucleosomal density of a region, acts as a modulator of large-scale repeat expansions.

INTRODUCTION

Expansions of simple DNA repeats cause numerous human hereditary disorders, including Huntington's disease, myotonic dystrophy type 1, Friedreich's ataxia, fragile X syndrome, and many others (reviewed in Pearson et al., 2005). Expandable repeats can be located in either the protein-coding or regulatory regions of their carrier genes (Gatchel and Zoghbi, 2005). Repeats in the protein-coding parts usually expand on a relatively low scale: 8–20 repeats is the norm whereas 30–100 repeats result in diseased state, whereas repeats in the noncoding areas of the gene undergo cataclysmic expansions, resulting in the addition of hundreds or thousands of repeats during intergenerational transmissions (McMurray, 2010).

The mechanisms of small- and large-scale repeat expansions have been intensively studied in many experimental systems. It is becoming increasingly clear that, in dividing cells, these events occur primarily in the course of DNA replication or postreplication repair (Mirkin, 2007). However, expansions also occur in nondividing and/or terminally differentiated cells where the DNA repair machinery is believed to play a role in the expansion process (McMurray, 2010). More recently, it was proposed that transcription across a repeat could stimulate expansions (Lin et al., 2009). This idea was based on the observation that transcription across a $(CAG)_n$ repeat increased its instability in human cells (Lin et al., 2006, 2009), which depended on transcription-coupled nucleotide excision repair and/or mismatch repair (Lin et al., 2009). It was proposed that formation of transient RNA:DNA hybrid structures during transcriptional elongation through the repeat could be responsible for the repeat instability (Lin and Wilson, 2012). Yet another possibility is that changes in the chromatin structure of a repeat-containing region could promote expansions through any of the aforementioned mechanisms (Debacker et al., 2012; House et al., 2014; Yang and Freudenreich, 2010). Despite these insights however, very few studies have quantitatively addressed the effect of transcription on repeat expansions. The main limitation has been the lack of an experimental system that can distinguish the contributions of transcription from DNA replication and DNA repair.

In humans, expansion of $(GAA)_n$ repeats in the first intron of the frataxin (*FXN*) gene (Campuzano et al., 1996) causes Friedreich's ataxia (FRDA). It was suggested that transcription through long $(GAA)_n$ repeats could be compromised due to the formation of a DNA triplex and/or extended R-loop structure (Bidichandani et al., 1998; Grabczyk and Usdin, 2000), likely resulting in the observed repeat instability. Expanded $(GAA)_n$ repeats are known to cause local heterochromatin formation (Al-Mahdawi et al., 2008; Greene et al., 2007) and gene silencing (Kumari and Usdin, 2012), which can be reversed by histone deacetylase inhibitors in FRDA cell lines (Herman et al., 2006; Tomassini et al., 2012). Most recently, it was found that the propensity of $(GAA)_n$ repeats to expand in human cell lines increased when the repeat was transcribed (Ditch et al., 2009), although the exact mechanistic details are far from clear. These attributes prompted us to focus on the role of transcription in the large-scale expansions of $(GAA)_n$ repeats in budding yeast *S. cerevisiae*.

Previously, we have described a system to study the large-scale expansions of $(GAA)_n$ repeats within the intron of an artificially split *URA3* gene (Shah et al., 2012; Shishkin et al., 2009). Here, we modified this system by utilizing a strong inducible promoter, thereby allowing us to compare expansion rates in different transcriptional states. We also developed a system to study large-scale expansions, placing repeats at a location in which there is little to no transcription. We found that large-scale

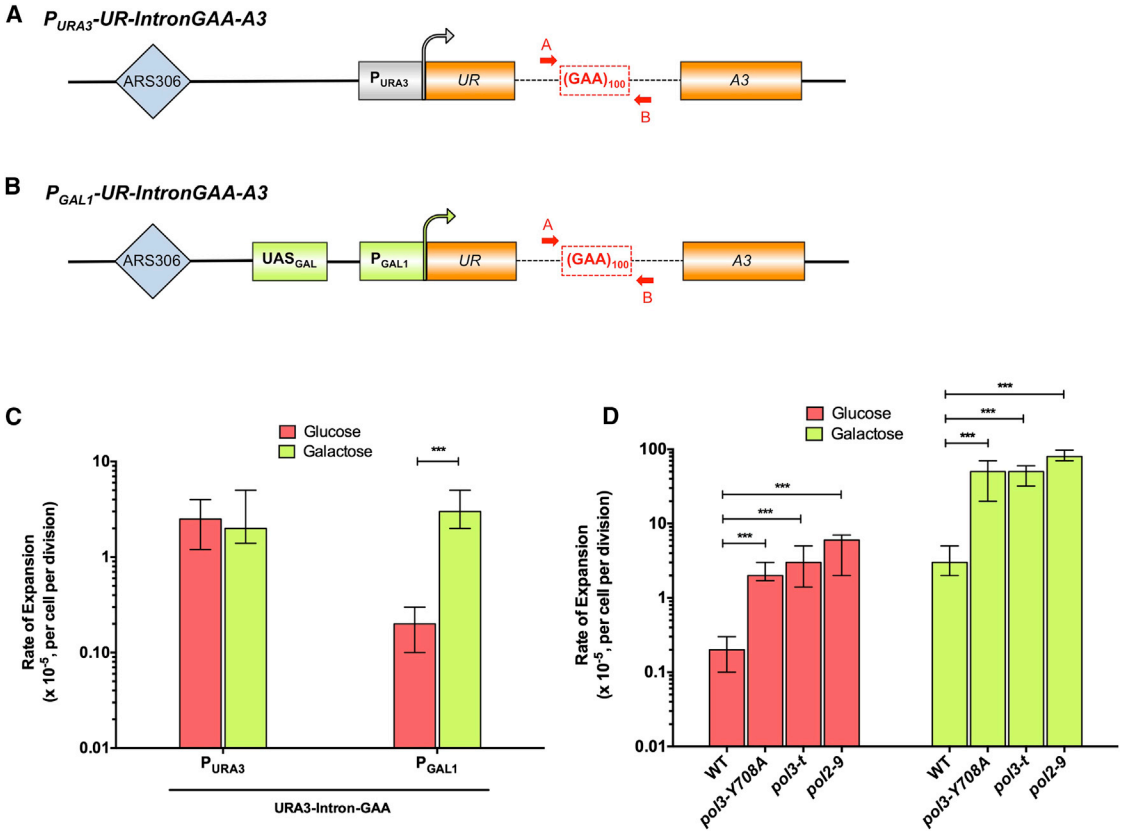


Figure 1. Inducible System to Study $(GAA)_n$ Repeat Expansions under Different Transcriptional States

(A and B) Constructs with a chimeric *URA3* marker containing $(GAA)_{100}$ repeats in the intron were put under the control of either (A) its native promoter (P_{URA3}) or (B) the galactose-inducible *GAL1* promoter (P_{GAL1}). In both these constructs, GAA repeats are transcribed by RNA polymerase II.

(C) Rate of expansion (per cell per division) in constructs driven either by P_{URA3} or P_{GAL1} grown in different transcriptional states (glucose or galactose).

(D) Rate of expansion (per cell per division) in the WT and DNA polymerase mutant strains containing the P_{GAL1}-UR-IntronGAA-A3 construct grown in different transcriptional states (glucose or galactose). Rates are presented as medians, and error bars denote 95% confidence intervals. ***$p \leq 0.001$ based on Mann-Whitney U test.

repeat expansions can occur in both systems, indicating that transcription is not necessary for this process. Surprisingly though, induction of transcription drastically elevated the rate of expansions in both systems. These results suggest that expansions are not generated during transcription elongation through the repeat per se. Rather it is the transcriptional activation of the locus that stimulates expansions. Finally, mutations in replicative DNA polymerases affected expansion rates irrespective of transcriptional state, lending further support to the replication-dependent mechanism of repeat expansions. We propose a model where the transcriptional status, linked to the nucleosomes density of a repeat-containing region, inherently modulates the rate of expansions during DNA replication.

RESULTS

Transcription through $(GAA)_n$ Repeats Elevates Expansion Rates

Previously, our lab developed an experimental system that mimics large-scale expansions of $(GAA)_n$ repeats in budding yeast *S. cerevisiae* (Figure 1A; Shah et al., 2012; Shishkin et al., 2009). In this system, $(GAA)_n$ repeats were cloned into the intron of an artificially split *URA3* gene. Repeat expansions that increased the intron's length beyond ~1 kb precluded *URA3* splicing, which allowed us to detect these events on media containing 5-fluorouracil (5-FOA). Using this system, we observed large-scale expansions, the rates of which were dependent on the initial length of the repetitive tract. In this setting, the *URA3* gene is constitutively expressed, making it impossible to distinguish the relative contribution of transcription and replication to the expansion process. Therefore, we modified our selectable system by replacing the native *URA3* promoter (P_{URA3}) with the galactose-inducible *GAL1* promoter (P_{GAL1}) (Figure 1B). The native *GAL1* promoter consists of an upstream activating sequence (UAS_{GAL}), followed by the upstream repressor sequence (URS_{GAL}) and the basal promoter (Giniger et al., 1985; Reagan and Majors, 1998). The UAS_{GAL} contains four binding sites for the Gal4 activator and is positioned ~150 bp upstream of the TATA box of the basal promoter. The URS_{GAL} contains three binding sites for the glucose repressor Mig1p. For this study, we removed the repressor element and extended the distance between the UAS_{GAL} and basal promoter

to 200 bp (see rationale below). When cells are grown on glucose, the level of transcription from P_{GAL1} is minimal whereas growth on galactose results in its dramatic induction (Giniger et al., 1985; St John and Davis, 1981).

To understand the effect of transcription levels on repeat expansions, yeast strains carrying the P_{GAL1}-UR-IntronGAA-A3 construct (Figure 1B) were first grown in nonselective media that contained either glucose or galactose as the carbon source. Appropriate dilutions of cells from isolated colonies were then replated onto selective media that contained galactose and 5-FOA to score for 5-FOA-resistant cells (that originated during growth on nonselective media). Based on our experience from previous studies, the duration of growth on nonselective media was minimized as much as possible so as to select cells that underwent only primary expansion events. PCR analysis was then used to check for the size of the expanded allele in 5-FOA-resistant colonies. The rate of repeat expansions, determined from these fluctuation assays, is presented in Figure 1C (P_{GAL1}). We found large-scale repeat expansions even under noninduced conditions (glucose), although the rate of expansions was quite low ($\sim 2 \times 10^{-6}$ per cell per division). Remarkably, however, when transcription was induced (galactose), the rate of expansion increased 10-fold ($\sim 3 \times 10^{-5}$ per cell per division), similar to what we observed for the constitutively transcribed P_{URA3} construct ($\sim 2.5 \times 10^{-5}$ per cell per division). These results suggested that the repeat tracts are more stable under conditions of low transcription. The scale of expansions remained large in either condition: 45–65 repeats were added to the initial repeat, consistent with what we have previously reported (Shah et al., 2012). To rule out the possibility that galactose induction might have an indirect effect on the expansion rates, we performed an experiment in which the cells containing the previously developed P_{URA3}-UR-IntronGAA-A3 expansion construct (Figure 1A) were grown on galactose. The results (shown in Figure 1C, P_{URA3}) confirm that galactose had no effect on the rate of repeat expansions in this system. Thus, the elevated rate of expansions described above, obtained using the P_{GAL1}-driven selectable construct on galactose, is specific to the transcriptional activation of the expandable cassette.

Replication Defects Elevate Rate of Repeat Expansions Irrespective of Transcriptional Status

We have previously shown that, in yeast, mutations in *POL3* and *POL2* genes, encoding the catalytic subunits of replicative DNA polymerases delta and epsilon, respectively, drastically elevated the rate of $(GAA)_n$ expansions (Shah et al., 2012). To investigate the effect of these replication defects under different transcriptional states, we transformed the P_{GAL1}-UR-IntronGAA-A3 construct containing $(GAA)_{100}$ repeats into yeast strains carrying one of the following mutations: *pol3-Y708A*, *pol3-t*, or *pol2-9*. Mutations *pol3-Y708A* and *pol3-t* destabilize the lagging strand polymerase whereas mutation *pol2-9* destabilizes the putative leading strand polymerase (Shah et al., 2012 and references therein).

Wild-type (WT) and mutant strains containing the P_{GAL1}-UR-IntronGAA-A3 construct were grown in nonselective media containing either glucose or galactose, followed by plating onto selective media with galactose and 5-FOA. A PCR analysis was carried out on 5-FOA^R colonies to score for the expansion rates, as described above. Figure 1D shows that, irrespective of the transcriptional state of the repetitive tract, DNA polymerase mutants caused a marked increase (10- to 20-fold) in the rate of repeat expansions relative to the WT strain. The elementary step of expansion was still in the range of 45–60 repeats, similar to what was observed for the WT strain. Thus, whereas the transcriptional state affected the rate of repeat expansions, replication defects further elevated this rate similar-fold in both transcriptional states. These observations indicate that whereas transcriptional state modulates the rate of repeat expansions, DNA replication is the causative mechanism behind their occurrence.

A System for Studying Repeat Expansions in a Nontranscribed Location

In yeast, transcriptional enhancers or upstream activating sequences are usually positioned several hundred bp upstream of the TATA box of the core promoter (Dobi and Winston, 2007). In the endogenous *GAL1* regulon, a UAS_{GAL} is situated around 150 bp upstream of the basal promoter P_{GAL1}. It was previously found that increasing the distance between the UAS_{GAL} and TATA box beyond ~800 bp disrupted transcriptional activation (Dobi and Winston, 2007). Because the DNA region between the UAS_{GAL} and basal promoter is occupied by a nucleosome, it inhibits access to RNA polymerase (Lohr and Lopez, 1995; Lohr et al., 1995). We reasoned therefore that a $(GAA)_{100}$ repeat positioned between the UAS_{GAL} and TATA box is unlikely to be transcribed, particularly in cells grown on a glucose-containing media. Furthermore, if the repeat expands, the distance between the UAS_{GAL} and basal promoter would exceed the 800 bp distance threshold, blocking activation of a downstream selectable marker. Based on these considerations, we generated two selectable cassettes presented in Figures 2A and 2B. The repeatless cassette consists of the modified *GAL1* promoter (described above) driving expression of the *CAN1* selectable marker. Strains carrying this cassette were resistant to canavanine when grown on glucose but sensitive to it when grown on galactose. The distance between the UAS_{GAL} and TATA box in this construct is ~200 bp. The expandable cassette contains the $(GAA)_{100}$ repeat placed in between the UAS_{GAL} and TATA box (Figure 2B). Addition of this repeat increased the distance between the two elements to ~650 bp, which still falls below the promoter inactivation threshold of 800 bp. Consequently, strains carrying the expandable cassette were also resistant to canavanine when grown on glucose but sensitive to it when grown on galactose. Note that the above two cassettes were integrated at our standard chromosomal location (~1 kb downstream of ARS306), in a strain with the endogenous *CAN1* gene deleted (*can1Δ*).

To ensure that the repeats positioned between UAS_{GAL} and TATA are not transcribed, we analyzed transcription through this regulatory region by quantitative RT-PCR (qRT-PCR). Recent studies have shown that transcription in yeast is much more pervasive than earlier believed (Djebali et al., 2012; Nagalakshmi et al., 2008), and these cryptic and/or noncoding transcripts are quickly degraded by a ribonuclease encoded by the *RRP6* gene. Hence, we decided to conduct the RT-PCR

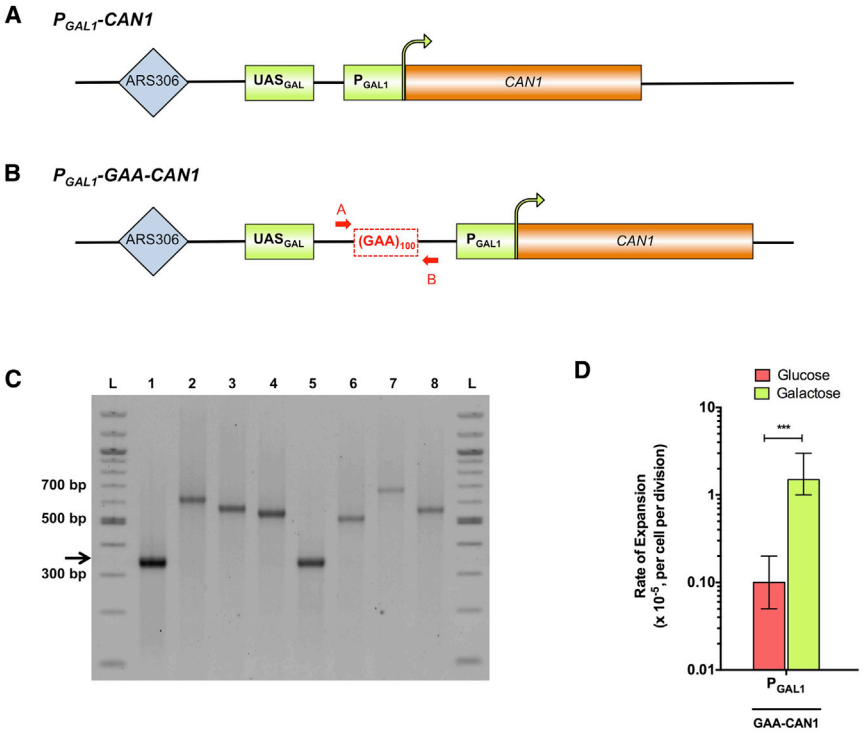


Figure 2. System for Studying $(GAA)_n$ Repeat Expansions in a Nontranscribed Location

The *CAN1* marker was put under the control of galactose-inducible P_{GAL1}.

(A) Repeatless construct in which the distance between UAS_{GAL} and TATA box is ~200 bp.

(B) Repeat-containing construct, where addition of $(GAA)_{100}$ repeats between UAS_{GAL} and TATA box increases the distance to ~650 bp. Expansion of the repeat extends this distance further, disrupting expression of *CAN1* and making cells resistant to canavanine in the media.

(C) Representative agarose gel showing PCR amplification of repeat alleles from CAN^R colonies. Arrow points to the initial length of $(GAA)_{100}$ repeats, which is ~350 bp. A majority of the colonies (~75%) contained expansions, but colonies with mutations in *CAN1* were also observed (lanes 1 and 5).

(D) Rate of expansion (per cell per division) in the P_{GAL1}-GAA-CAN1 construct from different transcriptional states (glucose or galactose). Rates are presented as medians, and error bars denote 95% confidence intervals. ***$p \leq 0.001$ based on Mann-Whitney U test.

analysis in an *rrp6Δ* background. Using primer pair A-B, situated on either side of the repeat, we have previously shown that the repeat in the original P_{URA3}-UR-IntronGAA-A3 cassette is actively transcribed (Shishkin et al., 2009). Using the same primer set here for RT-PCR, we were unable to detect transcription through the repeat situated between the UAS_{GAL} and TATA box in the new cassette (Figure 3F), irrespective of whether the cells were grown on glucose or galactose.

We have previously found in a different system and chromosomal setting that $(GAA)_n$ runs can serve as weak promoters (Zhang et al., 2012). To determine whether there is transcription downstream of the $(GAA)_{100}$ run in our current construct, we conducted qRT-PCR with three more primer pairs (Figures 3E and 3F): primer pair C-D for the region immediately downstream of the repeat, primer pair E-F for the region located further downstream of the repeat but upstream of the TATA box, and primer pair G-H for the 5′ end of the *CAN1* open reading frame (ORF). Traditional RT-PCRs were also carried out separately and run on an agarose gel for comparison (Figures 3C and 3D). Low levels of C-D and E-F transcript were indeed observed in the region downstream of the repeats (Figure 3D), consistent with the idea that they can serve as a weak promoter. Note, however, that the relative levels of these transcripts were at least 100-fold lower than for the G-H transcript. Overall, our results show that, even in an *rrp6Δ* background, the levels of transcription in the vicinity of the repeat are negligible and likely due to noncanonical or gratuitous transcription. We also carried out traditional RT-PCRs in the WT background (presented in Figure S1), where primer pairs C-D and E-F did not detect transcription between the repeat and *CAN1* gene, irrespective of whether the strains were grown on glucose or galactose. Altogether, these results indicate that the $(GAA)_{100}$ repeat in our system is indeed positioned in a nontranscribed location on the chromosome.

Our RT-PCR and qRT-PCR data from Figures 3C–3F show that, when grown on galactose, the P_{GAL1}-CAN1 constructs generated high levels of G-H transcript. Importantly, however, this transcript was visible even when the strains were grown on glucose, albeit at a much lower level (~150-fold). The latter observation can be attributed to leakage from the *GAL1* promoter, most likely due to the lack of glucose repressor elements (URS_{GAL}) in our cassettes.

Large-Scale Expansion of Repeats Positioned in a Nontranscribed Location

When compared to the P_{GAL1}-CAN1 construct, the levels of transcript from primer pair G-H are lower in the P_{GAL1}-GAA-CAN1 construct (~2-fold; Figures 3E and 3F). This decrease in the G-H transcript is likely due to the greater distance between UAS_{GAL} and TATA box (200 bp in P_{GAL1}-CAN1 as opposed to 650 bp in P_{GAL1}-GAA-CAN1). We reasoned therefore that a repeat expansion event would increase the distance between UAS_{GAL} and TATA box beyond the 800 bp threshold, blocking transcriptional activation and resulting in a canavanine-resistant phenotype (Can^R). To investigate, strains carrying the P_{GAL1}-GAA-CAN1 construct were grown on either glucose or galactose before being transferred to galactose and canavanine-containing media. As described above, the duration of growth on nonselective media was minimized as much as possible to account only for primary expansion events. We found Can^R colonies from cells grown on either condition (glucose or galactose), and a majority of them carried expanded GAA tracts (Figure 2C). The fact that large-scale repeat expansions were observed, even

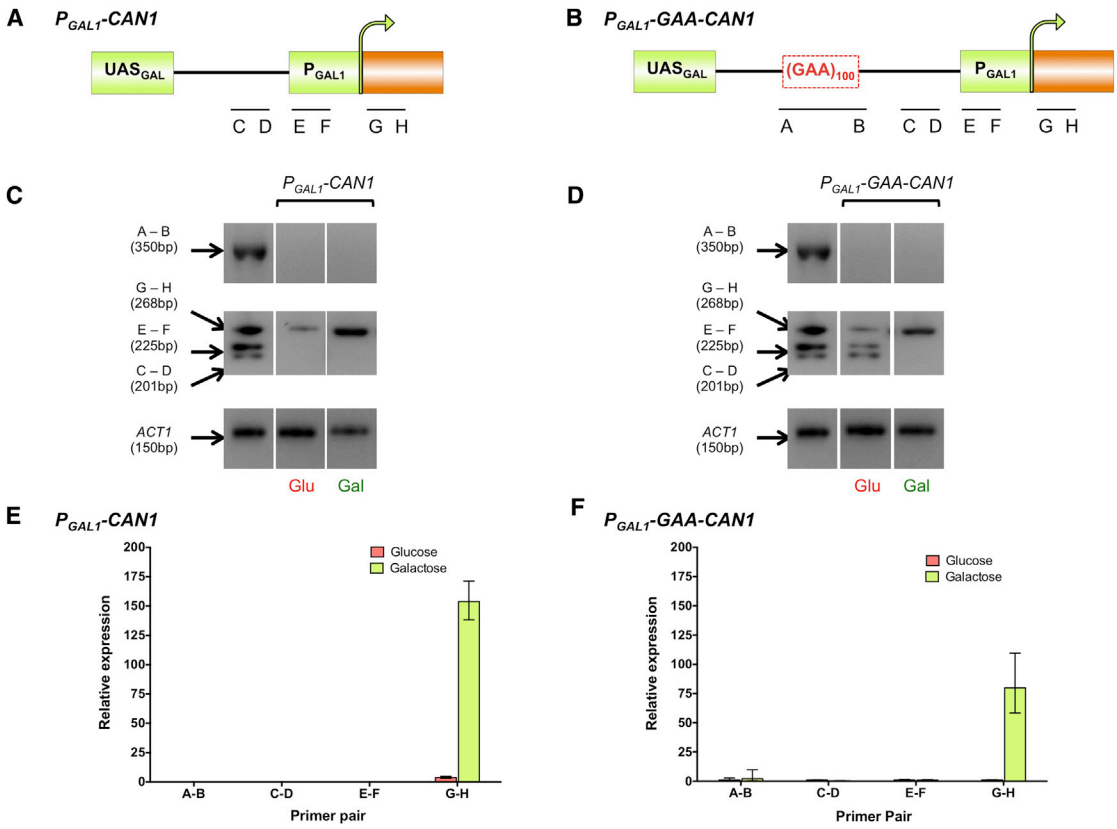


Figure 3. Analysis of Transcription in the Region between UAS_{GAL} and TATA-box

(A–D) Black lines show regions amplified by primer pairs used for qRT-PCR analysis in the (A) P_{GAL1}-CAN1 and (B) P_{GAL1}-GAA-CAN1 constructs. RT-PCRs of primer pairs A-B, C-D, E-F, and G-H from strains carrying the (C) P_{GAL1}-CAN1 and (D) P_{GAL1}-GAA-CAN1 constructs. A primer pair homologous to the *ACT1* gene was used for normalization. Leftmost lanes show expected size of primer pairs (amplified from genomic DNA).

(E and F) qRT-PCRs of the primer pairs described above. Threshold cycle (C_t) values were normalized to *ACT1* expression before plotting. Bars depict mean relative expression from three technical replicates, and error bars represent SD.

All the above RT-PCRs were performed on cDNA extracted from cells with an *rrp6Δ* background.

when cells carrying our P_{GAL1}-GAA-CAN1 construct were grown on glucose, confirms that repeat expansions can occur in the lack of transcription. That said, when transcription was induced by galactose, the rate of expansions increased ~10-fold compared to what was observed in cells grown on glucose: 1.5×10^{-5} versus 1.5×10^{-6} per replication, respectively (Figure 2D). Thus, induction of transcription, even when the repeat was placed in its "nontranscribed" region, increased the rate of expansions to exactly the same extent, as when the repeat was in the transcribed region of the selectable cassette.

Previous studies have shown that the native UAS_{GAL} element is hypersensitive to micrococcal nuclease (MNase) digestion, indicating that it remains free of nucleosomes irrespective of the carbon source (Bryant et al., 2008). In the noninduced state, two nucleosomes are positioned downstream of the UAS_{GAL} and upstream of the ORF. These nucleosomes get disrupted upon galactose induction through a process mediated by Gal4p (Lohr et al., 1995). We decided to investigate if this is indeed the case in our modified constructs through the MNase-qPCR assay (Infante et al., 2012). Strains containing the P_{GAL1}-CAN1 and P_{GAL1}-GAA-CAN1 constructs were grown in either glucose or galactose before isolating chromatin and digesting with MNase to extract mononucleosome-sized DNA fragments. We designed a series of overlapping primers across our constructs (with amplicons ranging from 50 to 100 bp) to perform qPCR analyses on the mononucleosomal DNA. Genomic DNA was also extracted simultaneously and used for normalization. The results of this assay are presented in Figure 4. In the P_{GAL1}-CAN1 construct, the UAS_{GAL} element was hypersensitive to MNase digestion (Figure 4A), consistent with results from previous studies (Bryant et al., 2008). The region immediately downstream of UAS_{GAL} was less sensitive (or relatively enriched), indicating that it was occupied by a nucleosome in the noninduced state (glucose), but not in the induced state (galactose). The region further downstream of UAS_{GAL} and including the transcription start site (TSS) of *CAN1* was highly enriched, suggesting that it was strongly occupied by a nucleosome in either condition, although to a lower level upon induction. The latter results are consistent with the previous genome-wide analyses of phased nucleosomes positioned at TSSs (Jiang and Pugh, 2009). In the P_{GAL1}-GAA-CAN1 construct, we observed similar results—the UAS_{GAL} element was hypersensitive to

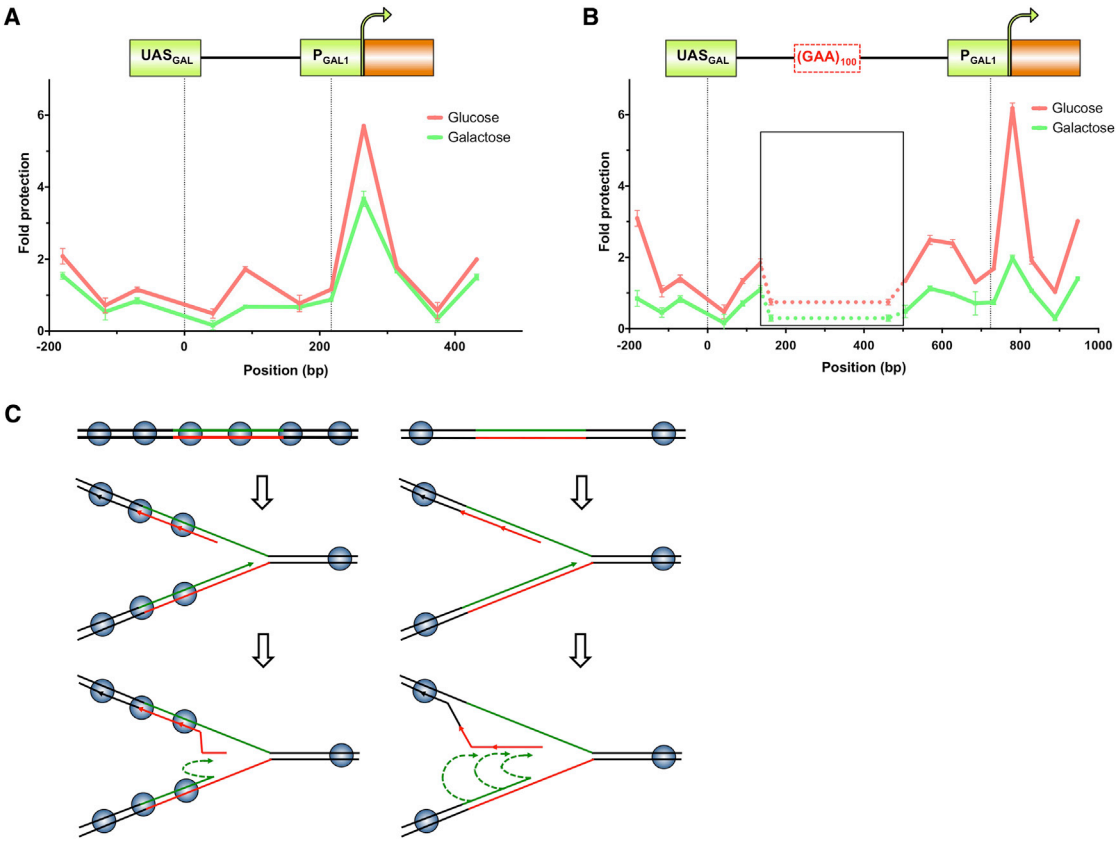


Figure 4. Link between Transcriptional Status, Nucleosome Density, and Repeat Expansions

(A and B) Nucleosome density in the (A) P_{GAL1}-CAN1 and (B) P_{GAL1}-GAA-CAN1 constructs under various transcriptional states. Figures plot mean fold protection values derived from mononucleosome-sized fragments against position of each primer pair, and error bars represent SD from biological replicates. Data shown within the inset in (B) are derived from trinucleosome-sized fragments. Vertical dashed lines represent the position of UAS_{GAL} and TATA-box elements in each construct.

(C) A high density of nucleosomes around a DNA fragment limits the propensity of a repetitive segment to template switch during replication fork progression (left). Whereas large-scale repeat expansions can occur in this context, the rate of repeat expansion is low. Nucleosomal density decreases upon transcriptional activation, promoting template switching of the repetitive segment at a replication fork (right). In this context, the rate of repeat expansion increases irrespective of whether the repeat itself is transcribed or not.

MNase digestion and the TSS was highly enriched (Figure 4B). Note that the distance between UAS_{GAL} and TSS in this construct is longer (~650 bp) than in the repeatless construct. The repetitive nature of the GAA tract precluded us from using overlapping primers across this segment. Hence, we decided to extract trinucleosome-sized fragments from the above MNase digests and use the A-B primer pair to compare relative protection under glucose and galactose conditions. We observed that the $(GAA)_{100}$ repeat-containing region was relatively enriched in the noninduced (glucose) over the induced condition (galactose; Figure 4B, inset). These results confirm that the nucleosomal density in the $(GAA)_{100}$ repeat-containing region changes upon transcriptional activation of UAS_{GAL} being higher in the noninduced versus the induced state.

DISCUSSION

Repeat expansions can occur in both dividing and nondividing cells. Furthermore, expandable repeats always occur in the transcribed regions of their carrier genes. This has led to several hypotheses on the role of transcription in repeat expansions (Lin and Wilson, 2011; Lin et al., 2009). At the same time, a large amount of data from model systems has pointed to the role of DNA replication and/or postreplication repair in the expansion process (Mirkin, 2007). Thus, it seemed pertinent to evaluate the relative contributions of transcription and replication to the expansion process. Here, we designed two yeast genetic systems to study the role of transcription in large-scale repeat expansions. In the first system, we modulated transcription through a $(GAA)_{100}$ repeat placed within the transcribed region (intron) of a selectable marker. We found a direct correlation between level of transcription and rate of repeat expansions. Our data are consistent with previous observations made in human cell lines (Ditch et al., 2009), in which positioning transcriptional terminators upstream of a repetitive $(GAA)_n$ tracts resulted in a sharp decline in their propensity to expand. Our results are also in agreement with data from individuals with FRDA, in which tissues with higher levels of *FXN* expression showed a bias

toward repeat expansions (Campuzano et al., 1996; De Biase et al., 2007).

Because the causative mechanism behind FRDA is a repeat-mediated epigenetic silencing of the *FXN* locus, what would happen if *FXN* is not silenced? Ditch et al. (2009) found frequent and progressive expansions in human embryonic kidney 293 cell lines when the expanded GAA repeats cloned into an expression cassette remained constitutively transcribed. More recently, $(GAA)_n$ repeats were shown to expand upon reprogramming of the FRDA patient fibroblasts into induced pluripotent stem cells with a near 100% probability (Ku et al., 2010; Mirkin, 2010). Several studies have pointed to the therapeutic potential of reversing the epigenetic silencing at the *FXN* locus (Chapdelaine et al., 2013; Herman et al., 2006; Tomassini et al., 2012; Tremblay et al., 2012). Given our results, it might be of relevance to track changes in repeat length at the *FXN* locus in cells that were exposed to these therapeutic strategies.

We found that whereas transcription through the $(GAA)_n$ repeat elevated its expansion rate, mutations in replicative DNA polymerases Pol δ and Pol ε had the same destabilizing effects on the repeat's stability, irrespective of whether the repeat was transcribed or not. These data indicated that DNA replication plays a central role in the expansion process, both in noninduced as well as induced transcriptional states, consistent with our previous findings (Shah et al., 2012; Shishkin et al., 2009; Zhang et al., 2012). We have also previously shown that large-scale expansions of GAA repeats are infrequent in nondividing cells (Zhang et al., 2012).

To investigate the role of transcription further, we developed a selectable system, utilizing an inducible promoter to drive expression of the selectable marker, but positioned $(GAA)_{100}$ repeats in the region between the UAS_{GAL} and TATA box of the promoter. Since the discovery of pervasive transcription in yeast and higher eukaryotes, many cryptic unstable transcripts (CUTs) and stable unannotated transcripts (SUTs) have been discovered (Marquardt et al., 2011; Thompson and Parker, 2007). These noncoding RNAs are quickly degraded in the nuclei by various RNA surveillance pathways. In budding yeast, the *RRP6* gene codes for a 3′ to 5′ ribonuclease involved in the rapid degradation of CUTs and SUTs (Hazelbaker et al., 2013). Hence, we confirmed that the repeat tract located between UAS_{GAL} and TATA box in our construct is not actively transcribed in either the WT or *rrp6Δ* background. We found large-scale $(GAA)_n$ repeat expansions in this system, effectively blocking expression of the downstream *CAN1* marker. Because these events were observed under conditions where little to no transcription was detected by qRT-PCR, our results strongly suggest that transcription is not necessary for repeat expansions. Remarkably though, the rate of expansions increased 10-fold when transcription was induced, quantitatively similarly to what we observed upon induction when the repeat was located in a heavily transcribed region. We conclude that it is the transcriptional state of a repeat-containing region, rather than transcription elongation through the repeat tract per se, that accounts for the elevated rate of expansions.

On one hand, our results suggest that whereas transcription through a repeat is not required for expansions, the transcriptional state of a repeat-containing region affects its expansion rate. On the other hand, increases in the expansion rate due to replication defects were identical irrespective of transcriptional state. The seemingly paradoxical nature of our results prompted us to wonder if the arrangement of nucleosomes at a repeat-containing region was modulating the rate of repeat expansions. Recently, genome-wide analysis of lagging strand DNA synthesis revealed that the size of Okazaki fragments correspond (on average) to a mononucleosome-sized length of ~155 bp in yeast (Smith and Whitehouse, 2012). The authors argue that positioning of nascent nucleosomes systematically terminates synthesis of each Okazaki fragment on the lagging strand. Data from our lab and others indicate that repeat expansions are inextricably linked to Okazaki fragment synthesis and likely involve some form of template switching (Shah et al., 2012; Shishkin et al., 2009) or fork reversal (Follonier et al., 2013; Kerrest et al., 2009) at the replication fork.

If this is the case, then our previously described template-switch model generates a testable prediction: the rate of repeat expansions depends on the density of nucleosomes around the repeat-containing region. To investigate, we used MNase-qPCR to analyze the density of nucleosomes in our system. Previous studies have shown that strong transcriptional induction of the native *GAL1* promoter wipes out nucleosomes from the region downstream of UAS_{GAL} (Angermayr and Bandlow, 2003; Bryant et al., 2008). Consistent with these data, our results show that the nucleosomal density in the repeat-containing region of the system is lower in the induced than the noninduced state. Taken together, in a transcriptionally noninduced state, the repeat-containing region contained a higher density of nucleosomes and the rate of expansion remained low. Upon transcriptional induction, nucleosomes were disrupted and the rate of expansions was elevated by ~10-fold. How exactly could nucleosome density affect template switching? Our model (Figure 4C) proposes that a high density of nucleosomes limits the portion of the leading strand available for switching, or in other words, the switching window of nascent strands is limited to an area the size of just one nucleosome (~155 bp). Decreasing the density of nucleosomes increases the size of this switching window, thereby increasing the probability of a template switch. Because template switching is the first in a series of steps that ultimately leads to an expanded repeat, affecting its probability affects the overall rate of expansions.

This model could explain why expandable repeats are found only in the transcribed regions of the human genome (Mirkin, 2007). It could also explain the destabilizing effects of nuclear reprogramming on repeat stability (Ku et al., 2010; Mirkin, 2010). Whereas it is replication centric, it does not exclude the role of mismatch repair or transcription-coupled repair, both of which were implicated in the repeat expansion process (Lin and Wilson, 2012). Finally, we want to emphasize that our system allows one to study the large-scale expansions or contractions of any DNA microsatellite, including the ones that are prone to deletions when placed in a transcribed location or block transcription elongation, even at short lengths.

EXPERIMENTAL PROCEDURES

Strains and Plasmids

Saccharomyces cerevisiae WT strain CH1585 (*MAT*a, *leu2-Δ1*, *trp1-Δ63*, *ura3–52*, and *his3–200*) was used in this study. Construction of various cassettes and strains is described in Supplemental Experimental Procedures.

Fluctuation Assay and Rates

Fluctuation assays were carried out as previously described (Shah et al., 2012) and in Supplemental Experimental Procedures.

qRT-PCR and MNase-qPCR Analysis

RT-PCR, qRT-PCR, and MNase-qPCR analysis was carried out as described in Supplemental Experimental Procedures.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, one figure, and one table and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2014.10.048.

AUTHOR CONTRIBUTIONS

K.A.S. designed and performed experiments, analyzed data, and wrote the manuscript. R.J.M. and V.I.E. designed and performed experiments. S.M.M. supervised the project, analyzed data, and wrote the manuscript.

ACKNOWLEDGMENTS

We thank Steve Buratowski, Catherine Freudenreich, Mitch McVey, and Melissa Koch for their useful comments and helpful suggestions. This study was funded by NIH grants GM60987 and GM105473 to S.M.M. and supported by the Tufts Graduate Student Research Award to K.A.S.

Received: May 5, 2014
Revised: September 7, 2014
Accepted: October 19, 2014
Published: November 20, 2014

REFERENCES

Al-Mahdawi, S., Pinto, R.M., Ismail, O., Varshney, D., Lymperi, S., Sandi, C., Trabzuni, D., and Pook, M. (2008). The Friedreich ataxia GAA repeat expansion mutation induces comparable epigenetic changes in human and transgenic mouse brain and heart tissues. Hum. Mol. Genet. *17*, 735–746.

Angermayr, M., and Bandlow, W. (2003). Permanent nucleosome exclusion from the Gal4p-inducible yeast GCY1 promoter. J. Biol. Chem. *278*, 11026–11031.

Bidichandani, S.I., Ashizawa, T., and Patel, P.I. (1998). The GAA triplet-repeat expansion in Friedreich ataxia interferes with transcription and may be associated with an unusual DNA structure. Am. J. Hum. Genet. *62*, 111–121.

Bryant, G.O., Prabhu, V., Floer, M., Wang, X., Spagna, D., Schreiber, D., and Ptashne, M. (2008). Activator control of nucleosome occupancy in activation and repression of transcription. PLoS Biol. *6*, 2928–2939.

Campuzano, V., Montermini, L., Moltò, M.D., Pianese, L., Cossée, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A., et al. (1996). Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science *271*, 1423–1427.

Chapdelaine, P., Coulombe, Z., Chikh, A., Gérard, C., and Tremblay, J.P. (2013). A Potential New Therapeutic Approach for Friedreich Ataxia: Induction of Frataxin Expression With TALE Proteins. Mol. Ther. Nucleic Acids *2*, e119.

De Biase, I., Rasmussen, A., Endres, D., Al-Mahdawi, S., Monticelli, A., Cocozza, S., Pook, M., and Bidichandani, S.I. (2007). Progressive GAA expansions in dorsal root ganglia of Friedreich's ataxia patients. Ann. Neurol. *61*, 55–60.

Debacker, K., Frizzell, A., Gleeson, O., Kirkham-McCarthy, L., Mertz, T., and Lahue, R.S. (2012). Histone deacetylase complexes promote trinucleotide repeat expansions. PLoS Biol. *10*, e1001257.

Ditch, S., Sammarco, M.C., Banerjee, A., and Grabczyk, E. (2009). Progressive GAA.TTC repeat expansion in human cell lines. PLoS Genet. *5*, e1000704.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. Nature *489*, 101–108.

Dobi, K.C., and Winston, F. (2007). Analysis of transcriptional activation at a distance in Saccharomyces cerevisiae. Mol. Cell. Biol. *27*, 5575–5586.

Follonier, C., Oehler, J., Herrador, R., and Lopes, M. (2013). Friedreich's ataxia-associated GAA repeats induce replication-fork reversal and unusual molecular junctions. Nat. Struct. Mol. Biol. *20*, 486–494.

Gatchel, J.R., and Zoghbi, H.Y. (2005). Diseases of unstable repeat expansion: mechanisms and common principles. Nat. Rev. Genet. *6*, 743–755.

Giniger, E., Varnum, S.M., and Ptashne, M. (1985). Specific DNA binding of GAL4, a positive regulatory protein of yeast. Cell *40*, 767–774.

Grabczyk, E., and Usdin, K. (2000). The GAA*TTC triplet repeat expanded in Friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner. Nucleic Acids Res. *28*, 2815–2822.

Greene, E., Mahishi, L., Entezam, A., Kumari, D., and Usdin, K. (2007). Repeat-induced epigenetic changes in intron 1 of the frataxin gene and its consequences in Friedreich ataxia. Nucleic Acids Res. *35*, 3383–3390.

Hazelbaker, D.Z., Marquardt, S., Wlotzka, W., and Buratowski, S. (2013). Kinetic competition between RNA Polymerase II and Sen1-dependent transcription termination. Mol. Cell *49*, 55–66.

Herman, D., Jenssen, K., Burnett, R., Soragni, E., Perlman, S.L., and Gottesfeld, J.M. (2006). Histone deacetylase inhibitors reverse gene silencing in Friedreich's ataxia. Nat. Chem. Biol. *2*, 551–558.

House, N.C., Yang, J.H., Walsh, S.C., Moy, J.M., and Freudenreich, C.H. (2014). NuA4 initiates dynamic histone H4 acetylation to promote high-fidelity sister chromatid recombination at postreplication gaps. Mol. Cell *55*, 818–828.

Infante, J.J., Law, G.L., and Young, E.T. (2012). Analysis of nucleosome positioning using a nucleosome-scanning assay. Methods Mol. Biol. *833*, 63–87.

Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. Nat. Rev. Genet. *10*, 161–172.

Kerrest, A., Anand, R.P., Sundararajan, R., Bermejo, R., Liberi, G., Dujon, B., Freudenreich, C.H., and Richard, G.-F. (2009). SRS2 and SGS1 prevent chromosomal breaks and stabilize triplet repeats by restraining recombination. Nat. Struct. Mol. Biol. *16*, 159–167.

Ku, S., Soragni, E., Campau, E., Thomas, E.A., Altun, G., Laurent, L.C., Loring, J.F., Napierala, M., and Gottesfeld, J.M. (2010). Friedreich's ataxia induced pluripotent stem cells model intergenerational GAA·TTC triplet repeat instability. Cell Stem Cell *7*, 631–637.

Kumari, D., and Usdin, K. (2012). Is Friedreich ataxia an epigenetic disorder? Clin. Epigenetics *4*, 2.

Lin, Y., and Wilson, J.H. (2011). Transcription-induced DNA toxicity at trinucleotide repeats: double bubble is trouble. Cell Cycle *10*, 611–618.

Lin, Y., and Wilson, J.H. (2012). Nucleotide excision repair, mismatch repair, and R-loops modulate convergent transcription-induced cell death and repeat instability. PLoS ONE *7*, e46807.

Lin, Y., Dion, V., and Wilson, J.H. (2006). Transcription promotes contraction of CAG repeat tracts in human cells. Nat. Struct. Mol. Biol. *13*, 179–180.

Lin, Y., Hubert, L., Jr., and Wilson, J.H. (2009). Transcription destabilizes triplet repeats. Mol. Carcinog. *48*, 350–361.

Lohr, D., and Lopez, J. (1995). GAL4/GAL80-dependent nucleosome disruption/deposition on the upstream regions of the yeast GAL1-10 and GAL80 genes. J. Biol. Chem. *270*, 27671–27678.

Lohr, D., Venkov, P., and Zlatanova, J. (1995). Transcriptional regulation in the yeast GAL gene family: a complex genetic network. FASEB J. *9*, 777–787.

Marquardt, S., Hazelbaker, D.Z., and Buratowski, S. (2011). Distinct RNA degradation pathways and 3′ extensions of yeast non-coding RNA species. Transcription *2*, 145–154.

McMurray, C.T. (2010). Mechanisms of trinucleotide repeat instability during human development. Nat. Rev. Genet. *11*, 786–799.

Mirkin, S.M. (2007). Expandable DNA repeats and human disease. Nature *447*, 932–940.

Mirkin, S.M. (2010). Getting to the core of repeat expansions by cell reprogramming. Cell Stem Cell *7*, 545–546.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. Science *320*, 1344–1349.

Pearson, C.E., Nichol Edamura, K., and Cleary, J.D. (2005). Repeat instability: mechanisms of dynamic mutations. Nat. Rev. Genet. *6*, 729–742.

Reagan, M.S., and Majors, J.E. (1998). The chromatin structure of the GAL1 promoter forms independently of Reb1p in Saccharomyces cerevisiae. Mol. Gen. Genet. *259*, 142–149.

Shah, K.A., Shishkin, A.A., Voineagu, I., Pavlov, Y.I., Shcherbakova, P.V., and Mirkin, S.M. (2012). Role of DNA polymerases in repeat-mediated genome instability. Cell Reports *2*, 1088–1095.

Shishkin, A.A., Voineagu, I., Matera, R., Cherng, N., Chernet, B.T., Krasilnikova, M.M., Narayanan, V., Lobachev, K.S., and Mirkin, S.M. (2009). Large-scale expansions of Friedreich's ataxia GAA repeats in yeast. Mol. Cell *35*, 82–92.

Smith, D.J., and Whitehouse, I. (2012). Intrinsic coupling of lagging-strand synthesis to chromatin assembly. Nature *483*, 434–438.

St John, T.P., and Davis, R.W. (1981). The organization and transcription of the galactose gene cluster of Saccharomyces. J. Mol. Biol. *152*, 285–315.

Thompson, D.M., and Parker, R. (2007). Cytoplasmic decay of intergenic transcripts in Saccharomyces cerevisiae. Mol. Cell. Biol. *27*, 92–101.

Tomassini, B., Arcuri, G., Fortuni, S., Sandi, C., Ezzatizadeh, V., Casali, C., Condò, I., Malisan, F., Al-Mahdawi, S., Pook, M., and Testi, R. (2012). Interferon gamma upregulates frataxin and corrects the functional deficits in a Friedreich ataxia model. Hum. Mol. Genet. *21*, 2855–2861.

Tremblay, J.P., Chapdelaine, P., Coulombe, Z., and Rousseau, J. (2012). Transcription activator-like effector proteins induce the expression of the frataxin gene. Hum. Gene Ther. *23*, 883–890.

Yang, J.H., and Freudenreich, C.H. (2010). The Rtt109 histone acetyltransferase facilitates error-free replication to prevent CAG/CTG repeat contractions. DNA Repair (Amst.) *9*, 414–420.

Zhang, Y., Shishkin, A.A., Nishida, Y., Marcinkowski-Desmond, D., Saini, N., Volkov, K.V., Mirkin, S.M., and Lobachev, K.S. (2012). Genome-wide screen identifies pathways that govern GAA/TTC repeat fragility and expansions in dividing and nondividing yeast cells. Mol. Cell *48*, 254–265.