Role of transcript and interplay between transcription and replication in triplet-repeat instability in mammalian cells

Paul M. Rindler[1] and Sanjay I. Bidichandani[1,2,*]

[1]Department of Biochemistry and Molecular Biology and [2]Department of Pediatrics, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104, USA

ABSTRACT

Triplet-repeat expansions cause several inherited human diseases. Expanded triplet-repeats are unstable in somatic cells, and tissue-specific somatic instability contributes to disease pathogenesis. In mammalian cells instability of triplet-repeats is dependent on the location of the origin of replication relative to the repeat tract, supporting the 'fork-shift' model of repeat instability. Disease-causing triplet-repeats are transcribed, but how this influences instability remains unclear. We examined instability of the expanded $(GAA{\bullet}TTC)_n$ sequence in mammalian cells by analyzing individual replication events directed by the SV40 origin from five different locations, in the presence and absence of doxycycline-induced transcription. Depending on the location of the SV40 origin, either no instability was observed, instability was caused by replication with no further increase due to transcription, or instability required transcription. Whereas contractions accounted for most of the observed instability, one construct showed expansions upon induction of transcription. These expansions disappeared when transcript stability was reduced via removal or mutation of a spliceable intron. These results reveal a complex interrelationship of transcription and replication in the etiology of repeat instability. While both processes may not be sufficient for the initiation of instability, transcription and/or transcript stability seem to further modulate the fork-shift model of triplet-repeat instability.

INTRODUCTON

Several inherited neuromuscular diseases are caused by the expansion of a triplet-repeat sequence (1). These expanded sequences are unstable and frequently change in length during intergenerational transmission and within somatic cells. Somatic instability in human and mouse is tissue-specific and age-dependent. Typically, a bias for further expansion is observed in the tissues primarily affected in disease pathology. This phenomenon is seen in dorsal root ganglia (DRG) of Friedreich ataxia patients (2), in the striatum of Huntington disease patients (3,4), and in muscle and brain of myotonic dystrophy patients (5–8). Modeling of disease-related patterns of somatic instability in transgenic mouse models for various triplet-repeats has revealed the importance of genomic context in reproducing the tissue-specific patterns of somatic instability observed in patients. For instance, the expanded $(GAA{\bullet}TTC)_n$ sequence that causes Friedreich ataxia shows progressive expansions specifically in the DRG and cerebellum of transgenic mice when placed in the context of the human *FXN* gene (9,10), but not when 'knocked in' into the corresponding intronic location of the mouse *Fxn* gene (11). Indeed, differing levels of somatic instability were observed for $(GAA{\bullet}TTC)_n$ sequences located in unlinked human genomic loci (12). These data underscore the importance of the genomic context in determining instability, and provide evidence for the existence of *cis*-acting factors that regulate repeat instability.

An important *cis*-acting modifier of triplet-repeat instability in mammalian cells is the location of the origin of replication relative to the repeat tract (13). Replication in transfected mammalian cells via the SV40 origin of replication (*ori*) has been shown to modulate $(CTG{\bullet}CAG)_n$, $(CGG{\bullet}CCG)_n$ and $(GAA{\bullet}TTC)_n$ instability (12–14). The instability observed was dependent upon the orientation of replication as well as the distance between the SV40 *ori* and the repeat sequence. These observations support the fork-shift model for repeat instability, wherein the relationship of the repeat tract and the Okazaki initiation zone serve to determine whether, and what type of, instability will result (15). In this regard, the similarities in the response of the various triplet-repeat motifs also

*To whom correspondence should be addressed. Tel: +1 405 271 1360; Fax: +1 405 271 3910; Email: sanjay-bidichandani@ouhsc.edu

support the fork-shift model as a common underlying mechanism, and suggest that replication is a likely contributor to somatic instability observed in patients. However, the relationship of DNA replication and somatic instability is not straightforward since there is poor correlation between the proliferative potential of human and mouse tissues, and the level and type of somatic instability observed (2,3,9,16–26). Moreover, the age-dependent increase in somatic instability in post-mitotic tissues (2–8) also argues against a direct role for DNA replication.

Disease-causing triplet-repeat sequences are located within transcriptional units, opening up the possibility of a relationship between transcription and replication in mediating repeat instability. Transcription has been shown to increase $(CAG\bullet CTG)_n$ and $(GAA\bullet TTC)_n$ instability in human cell culture (27–29), and placing the $(CAG\bullet CTG)_n$ sequence within a transcriptional unit was required for repeat instability in a *Drosophila* model (30). Furthermore, replication fork stalling is observed when replication and transcription are in a 'head-on' versus co-linear orientation (31). The nascent RNA produced during transcription may also interact with replication-mediated non-B DNA structures (31), thus producing mutagenic intermediates. At present, it is unclear if these potential interactions between replication and transcription impact triplet-repeat instability.

We designed a defined replication assay that would also allow doxycycline-inducible transcription to examine effects on instability of the $(GAA\bullet TTC)_n$ sequence in mammalian cells. Replication from the SV40 *ori* was initiated at one of five different locations to examine the effect on repeat instability of replication and transcription in the head-on versus co-linear orientation. Depending on the location of the SV40 *ori,* induction of transcription either resulted in no instability, predominant contractions, or both expansions and contractions. Our data support a complex relationship between replication and transcription in the etiology of triplet-repeat instability, with transcription further modulating an important *cis*-acting determinant of repeat instability, and also suggest a potential role for the nascent transcript in triplet-repeat instability.

MATERIALS AND METHODS

Plasmid construction

The pTRE-tight plasmid (Clontech) containing a doxycycline inducible CMV promoter was used. In order to place the $(GAA\bullet TTC)_n$ sequence in the stable orientation with respect to the bacterial ColE1 *ori* (32) the expression cassette in pTRE-tight was amplified by PCR and flipped via cloning into Xho I and Eco0109 I to create pTRE-Rev. The expression cassette was amplified using the following forward and reverse primers, respectively:

5′-TTCGTCTTCACTCGAGTTTA-3′ and
5′-AGGCCCTCCATGGGCATGCGCAGTGAAAAAA
ATGCTTTA-3′ (note: the reverse primer contains

additional restriction sites for use in unrelated experiments).

A spliceable intron was amplified from pRL-SV40 (Promega) using the following forward and reverse primers, respectively:

5′-TCTAGACAGGTAAGTATCAAGGTTAC-3′ and
5′-ACCGGTCCTGTGGAGAGAAAGGCAAA-3′.

The spliceable intron was inserted into Xba I and Age I (the latter site was engineered upstream of the SV40 polyadenylation sequence), in order to generate pTRE-Rev-Int. Fragments containing the SV40 *ori* cloned at one of five different distances from the future location (Bam HI / Pst I) of the $(GAA\bullet TTC)_n$ sequence were amplified by PCR from previously described constructs pSEA-N, pSEA-K, pSEA-H, pSEA-S and pSEA-A (12) (after ablation of the EcoR I and Xba I sites by digestion and fill-in reactions) and cloned into the EcoR I and Xba I sites of pTRE-Rev-Int to generate pTRE-N, pTRE-K, pTRE-H, pTRE-S and pTRE-A. Finally, the $(GAA\bullet TTC)_{115}$ sequence was amplified and cloned into the Bam HI and Pst I sites of pTRE-N, pTRE-K, pTRE-H, pTRE-S and pTRE-A to generate pT115N, pT115K, pT115H, pT115S and pT115A, respectively.

The pT115H construct was further modified by removing the spliceable intron cloned between Xba I and Age I to create pT115HR. Additionally, pT115H was subjected to site-directed mutagenesis of the consensus splice donor (GT to CA) and acceptor (AG to TC) sites of the synthetic intron to generate pT115HM. The latter was accomplished by amplifying the spliceable intron using the following forward and reverse primers: 5′-TCTAGACAGCAAAGTATCAAGGTTA-3′ and 5′-ACCGGTCGAGTGGAGAGAAAGGC-3′ followed by digestion and cloning into Xba I and Age I. The length and purity of the repeat sequence in each construct were confirmed by DNA sequencing.

Cell transfection and induction of transcription

COS Tet-On cells were generated by stable transfection of COS1 cells with pTet-On Advanced (Clontech) according to the manufacturer's protocol. Several clones were isolated and screened for increased expression of luciferase reporter activity in the presence of 1 μg/ml doxycycline. A stable integrant was identified that displayed ∼100-fold increase in luciferase activity upon induction with doxycycline, and ∼9-fold increase in steady-state transcript levels of our repeat-containing constructs (data not shown). COS Tet-On cells were grown in DMEM (with 10% FBS and 200 μg/ml G418). Transfections (each condition in triplicate) were performed in six-well plates using 1 μg of plasmid DNA and 10 μl Lipotaxi (Stratagene) as per the manufacturer's recommendations. After 4 h, 750 μl of DMEM was added and cells were incubated for 16 h (10% FBS, 200 μg/ml G418, with or without 1 μg/ml doxycycline), after which the medium in each well was replaced with fresh DMEM (10% FBS, 200 μg/ml G418, with or without 1 μg/ml doxycycline). Cells were incubated

for 48 h and DNA was extracted using the DNeasy tissue kit (Qiagen).

Analysis of repeat instability

This was performed as previously described (12). Briefly, DNA extracted from transfected COS Tet-On cells was digested with Dpn I to remove all unreplicated plasmids used in the initial transfection. Dpn I resistant plasmids were transformed into *Escherichia coli* (DH5α) to isolate individual products of replication. Individual colonies were analyzed by colony PCR as previously described (12,32). Instability was measured as the ratio of plasmids with altered repeat lengths (mutant) to the total number of successful PCR reactions, and represented as a percentage. Background instability was determined for each construct by transformation of the same DNA sample used for transfection into DH5α and measured by colony PCR. Enhanced instability following transfection in COS Tet-On cells was determined by comparison with the background in DH5α and calculated by Chi-square analysis ($P < 0.05$ was used as the threshold for statistical significance).

Determination of plasmid replication efficiency

DNA extracted from transfected COS Tet-On cells was used to determine plasmid replication efficiency by transformation of equal quantities of Dpn I or mock digested products into *E. coli* (DH5α). Individual transformation colonies were counted to determine the number of replication products (Dpn I resistant plasmids) and the total number of plasmids (mock digestion) extracted from transfected COS Tet-On cells. Replication efficiency was measured as the ratio of replication products to total number of plasmids extracted and expressed as a percentage. Replication efficiency in COS Tet-On cells was determined in the absence or presence of doxycycline-induced transcription and compared by Chi-square analysis.

Real-time quantitative RT-PCR

Total RNA was extracted using Tripure (Roche) according to the manufacturer's protocol. cDNA generated with QuantiTect Reverse Transcription kit (Qiagen) was used for individual amplification reactions with specific forward and reverse primer pairs. The following pair of primers was used to measure steady-state levels of repeat-containing transcripts:

5′-GGTTATCCACAGAATCAGGG-3′ and 5′-GCTGA TTATGATCCACCGGT-3′.

The following sets of primers were used for quantitative measurement of levels of transcript at locations upstream (GAA up) and downstream (GAA down) of the triplet-repeat:

GAA up: 5′-GCACAGATGCGTAAGGAGAA-3′ and 5′-CGTCGTGACTGGGAAAACC-3′; GAA down: 5′-GCATAAAGTGTAAAGCCTGGG-3′ and 5′-CA ATACGCAAACCGCCTC-3′.

Real-time quantitative RT-PCR was performed in triplicate using Power SYBR Green (Applied Biosystems) and normalized either to endogenous levels of HPRT, or to luciferase transcript from a co-transfected luciferase expression vector (as indicated). HPRT and luciferase quantitative RT-PCR was performed using the following specific forward and reverse primers:

HPRT: 5′-AGCCAGATTTGCTTGTTTGG-3′ and 5′-T CCAGCAGGTCAGCAAAGAA-3′ Luc: 5′CCGCTG GAAGATGGAACC-3′ and 5′-GCAACTCCGATAA ATAACGC-3′.

Statistical analysis was performed on two independent transfection experiments, with each construct tested in triplicate, using the unpaired, two-tailed *t*-test.

Transcript stability

Transfection of COS Tet-On cells was essentially performed as above. Following the initial 16 h incubation in the presence of doxycycline, the medium was replaced with fresh DMEM without doxycycline, thus withdrawing the stimulus for transcriptional induction. Total RNA was extracted at hourly intervals and assayed by real-time quantitative RT-PCR to evaluate transcript degradation over time following removal of doxycycline. Real-time RT-PCR was performed as described above.

RESULTS

Design of constructs to test the effect of transcription and replication on (GAA•TTC)$_{115}$ instability in mammalian cells

To test the hypothesis that transcription modulates the fork-shift model of replication-induced repeat instability, with possible differences between head-on and co-linear orientations, we constructed a series of plasmids that would allow replication of a (GAA•TTC)$_{115}$ sequence and doxycycline-inducible transcription upon transfection in COS Tet-On cells (Figure 1). These plasmids contain a doxycycline-inducible CMV promoter, a consensus spliceable intron, and the SV40 polyadenylation sequence. The SV40 *ori* was cloned in five different locations so as to result in co-linear (pT115N and pT115K) or head-on (pT115H, pT115S and pT115A) transcription and replication with respect to the repeat sequence. Mutagenesis in *E. coli* was minimized by inserting the (GAA•TTC)$_{115}$ sequence in the most stable orientation with respect to the ColE1 *ori* (Figure 1) (32). Additionally, the significance of instability observed for any construct following transfection in COS Tet-On cells was determined by comparison with the background level of instability determined for each construct in *E. coli* (i.e. without transfection).

Transcription induced (GAA•TTC)$_{115}$ instability depends on the location of the SV40 *ori* relative to the repeat tract

All five constructs were individually transfected into COS Tet-On cells and allowed to replicate in the absence (replication alone) or presence of doxycycline (replication and
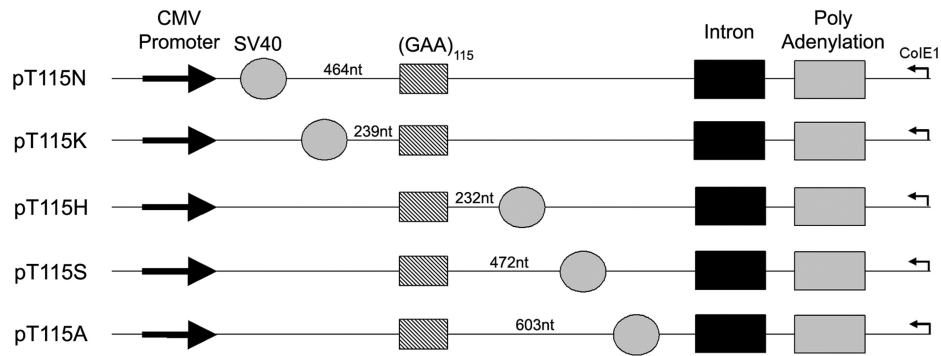
Figure 1. Design of constructs to test the effect of transcription and replication on $(GAA \bullet TTC)_n$ instability in mammalian cells. The expression cassette is essentially from pTRE-Tight (Clontech), which allows doxycycline-inducible transcription using the CMV promoter (large arrow). The SV40 origin of bidirectional replication (gray circles) was inserted at five different locations such that the orientation of transcription and replication would be co-linear (pT115N and pT115K) or head-on (pT115H, pT115S and pT115A) in relation to a patient-derived $(GAA \bullet TTC)_{115}$ sequence (hatched box). The distance in nucleotides of the SV40 origin of bidirectional replication from the proximal edge of the repeat tract is indicated for each construct. The repeat is in the stable orientation with respect to the ColE1 origin of replication (bent arrow) in order to minimize mutagenesis during propagation in *E. coli*. A spliceable intron (black box) was subcloned upstream from the SV40 polyadenylation sequence (gray box).

transcription). Following transfection, individual replication products were analyzed for repeat instability, with expansions and contractions evaluated and plotted separately (Figure 2A and B). Induction of transcription resulted in enhanced repeat instability, which was not observed with replication alone, using two constructs; pT115N showed enhanced frequency of contractions and pT115H showed an increase of both expansions and contractions. Using pT115K we observed a significant increase in contractions with replication alone, with no further increase in instability upon induction of transcription ($P = 0.38$). Replication and transcription together did not produce instability in two of the five constructs (pT115S and pT115A), indicating that both processes are insufficient for inducing repeat instability. Therefore, the location of the origin of replication determined if instability would occur, if replication alone was necessary for instability, or if both replication and transcription were required for instability. Indeed, the location of the origin also determined the pattern of instability seen; pT115N showed enhanced contractions, and pT115H showed both increased expansions and contractions. Moreover, while contractions accounted for most of the observed instability, an increase in the frequency of expansions was seen with pT115H. It should be noted that this is similar to the behavior of the expanded $(GAA \bullet TTC)_n$ sequence in tissues of Friedreich ataxia patients, most of which display contractions, with expansions observed much less commonly in specific regions of the nervous system (2,33).

A potential caveat of our experimental design is that the constructs in which replication and transcriptional elongation occur in the head-on orientation (pT115H, pT115S and pT115A) may experience replication fork stalling within, or even outside, the repeat tract thus effectively reducing the rate of replication of these constructs compared to those that have both processes in the co-linear orientation. However, measurement of replication efficiency in the presence or absence of transcription

showed no difference for either the constructs with co-linear (chi-square test $P = 0.17$) or head-on (chi-square test $P = 0.23$) orientations (Figure 3A). Whereas, we cannot rule out the possibility of transient replication fork slowing within the repeat in the head-on orientation, our data indicate that differences in the observed levels of instability is not likely due to any substantial difference in the replication efficiency of the head-on versus co-linear constructs.

We also considered the possibility that differences in instability may stem from differing rates of transcription through the repeat tract. For instance, transcriptional elongation through the expanded $(GAA \bullet TTC)_n$ sequence, which is already reduced when GAA is the non-template strand (34,35) as in our constructs, may be affected differentially based on the location of the origin of replication. We therefore measured transcript levels via real-time quantitative RT–PCR using primers both immediately upstream (GAA up) and downstream (GAA down) of the $(GAA \bullet TTC)_n$ sequence (Figure 3B; normalized to luciferase transcript from a co-transfected plasmid, to correct for both transfection efficiency and heterologous expression). No significant difference (using the unpaired two-tailed *t*-test) was seen for the amount of transcript upstream versus downstream of the $(GAA \bullet TTC)_n$ sequence for any of the five constructs (Figure 3B), indicating that in this experimental system the $(GAA \bullet TTC)_{115}$ sequence did not pose a significant hindrance to the progression of transcription for all five locations of the origin of replication. However, interestingly, we did note that all three constructs with a head-on configuration of replication and transcription had approximately half the steady-state levels of transcript compared with the co-linear constructs ($P < 0.001$; Figure 3B). Since the same reduction was seen whether GAA up or GAA down primer pairs were used, it suggests that the difference is not due to stalling within the $(GAA \bullet TTC)_n$ sequence, but likely due to interference with initiation or with RNA turnover. One could argue
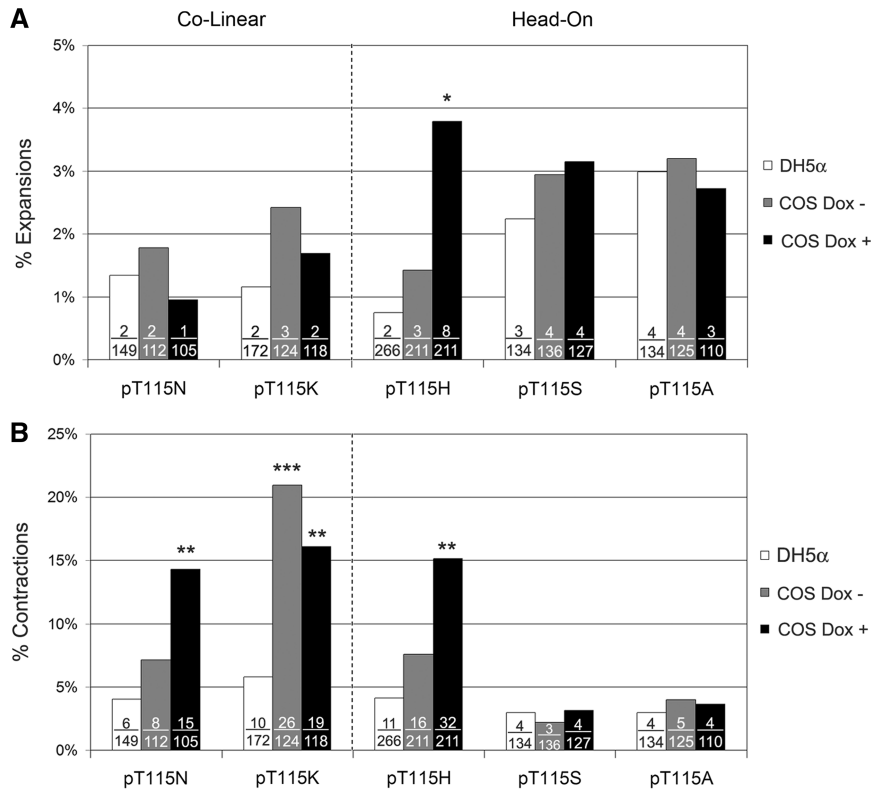
Figure 2. Transcription-induced (GAA•TTC)$_n$ instability is dependent on the relative location of the SV40 *ori*. Plasmids were replicated in COS Tet-On cells in the absence or presence of doxycycline-induced transcription. Repeat instability, calculated as the ratio of total number of length changes (numerator) and total number of successful PCR reactions (denominator), is displayed as percent values. Contractions and expansions are analyzed separately. White bars represent the background level of instability for each construct, i.e. without transfection in COS Tet-On cells. Gray and black bars represent instability following replication in COS Tet-On cells in the absence and presence of doxycycline-induced transcription, respectively. (**A**) Expansions induced following replication in COS Tet-On cells in the absence or presence of doxycycline-induced transcription. pT115H showed a significant increase in the frequency of expansions. (**B**) Contractions induced following replication in COS Tet-On cells in the absence or presence of doxycycline-induced transcription. pT115N and pT115H showed significantly increased frequency of contractions upon induction of transcription. pT115K showed increased frequency of contractions even without induction of transcription that was not enhanced further via transcription. *P*-values were derived by comparing instability to background using the Chi-square test (*$P < 0.05$, **$P < 0.01$ and ***$P < 0.001$).

that pT115S and pT115A, the two constructs that did not show any increase in instability upon replication and transcription, may be stable because they experienced half the level of transcription. However, the validity of this argument is questionable since they have the same level of transcript (and perhaps, also transcription) as pT115H, which showed enhanced frequency of both expansions and contractions (Figure 2).

Another possible source of instability may be transcription–transcription collisions; i.e. transcription arising from the CMV promoter and the SV40 early promoter, which is embedded within the same sequence as the bidirectional SV40 origin of replication. Given that the SV40 origin of replication is bidirectional, we minimized this possibility by inserting this sequence in all five locations around the (GAA•TTC)$_n$ sequence in such as way as to direct SV40-derived transcription away from the repeat tract. Furthermore, we excluded the SV40 enhancer sequence from the insert used to drive replication, which is known to reduce the efficiency of transcription from the early promoter by 10-fold in CV-1 cells (36).

We did not detect any expansions caused by replication alone, however it was possible to compare the size distribution of contractions observed with replication alone (pT115K) versus replication plus transcription (pT115N and pT115H). Replication plus transcription resulted in a uniform distribution of contraction sizes compared with replication alone, which produced significantly larger contractions (Mann–Whitney test, $P < 0.01$; Figure 4). The different size distribution of mutations via replication alone versus replication plus transcription suggests that the addition of transcription likely creates and/or resolves distinct mutagenic substrates. It is noteworthy that replication plus transcription with pT115K did not increase the frequency (Chi-square test, $P = 0.38$; Figure 2B) or magnitude (Mann–Whitney test, $P = 0.53$; data not shown) of contractions compared with what was observed with replication alone, supporting the notion that the contractions seen with pT115K are most likely due to replication alone (note also that the replication efficiency of pT115K did not change upon induction of transcription; Figure 3A).
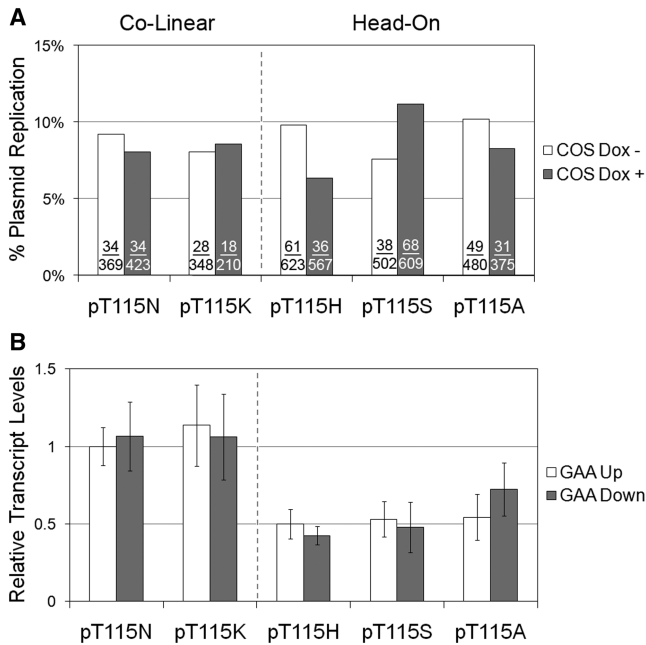
Figure 3. Transcriptional elongation is unaffected through the $(GAA \bullet TTC)_{115}$ sequence, and plasmid replication is unaffected by the induction of transcription. (**A**) Plasmid replication efficiency in transfected COS Tet-On cells was measured in the absence (white bars) or presence (gray bars) of doxycycline-induced transcription (see 'Materials and methods' section). The number of colony forming units was determined for products of replication (Dpn I resistant plasmids; numerator) and the total number of plasmids (mock digestion; denominator). Replication efficiency, calculated as the ratio of replication products and the total number of plasmids extracted, is displayed as percent values. Replication efficiency, in the presence or absence of transcription, showed no difference for either the constructs with co-linear (Chi-square test, $P = 0.17$) or head-on (chi-square test, $P = 0.23$) orientations. (**B**) Plasmids were replicated in COS Tet-On cells in the presence of doxycycline-induced transcription. Transcript levels were determined by real-time quantitative RT-PCR of cDNA generated from transcripts produced by all five constructs. Transcripts produced by each construct were amplified using primer pairs specific to sequences upstream (GAA up; white bars) or downstream (GAA down; gray bars) of the $(GAA \bullet TTC)_{115}$ sequence. Data are from triplicate assays of two independent transfections. Transcript levels were normalized to the level of luciferase transcript produced from co-transfected pTet-On Luc (Clontech) and are shown relative to levels obtained for pT115N upstream of the $(GAA \bullet TTC)_{115}$ sequence. Error bars represent the standard error of mean and data were analyzed using the unpaired, two-tailed t-test. No significant difference was found for the amount of transcript upstream versus downstream of the $(GAA \bullet TTC)_{115}$ sequence for any of the five constructs. However, all three constructs with a head-on configuration of replication and transcription had approximately half the steady-state levels of transcript compared with the co-linear constructs ($P < 0.001$).

Taken together, our data do not support a clear association of instability with the relative orientation (head-on or co-linear) of replication and transcriptional elongation. However, the addition of transcription seemed to modulate the *cis*-regulatory function afforded by the differential location of the origin of replication.

Transcription-induced instability of $(GAA \bullet TTC)_{115}$ is associated with transcript stability

Given that RNA-DNA hybrids akin to R-loop formation are known to form when the $(GAA \bullet TTC)_n$ sequence is
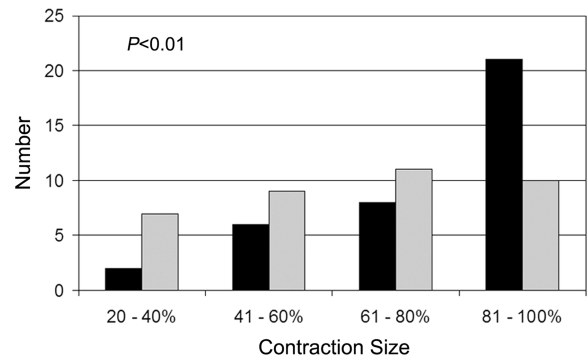

Figure 4. Differing size distribution of mutations associated with replication and transcription. Comparison of the size distribution of contractions caused by replication alone (pT115K; black bars; $n = 45$) versus transcription (pT115N and pT115H; gray bars; $n = 46$). The distribution of contractions >20% of the initial length is shown for both groups, subdivided equally into the indicated four size windows. The *y*-axis shows the number of contraction events. Contractions resulting from replication alone tended to be larger, whereas transcription produced a more uniform distribution across all size windows (Mann–Whitney test, $P < 0.01$).

transcribed (37), we decided to investigate if the triplet-repeat instability associated with induction of transcription was related to the stability of the transcript. A critical mediator of transcript stability in mammalian cells, as is well known for heterologous expression systems, is the presence of a spliceable intron (hence their consistent inclusion in commercial mammalian expression vectors). Using site-directed mutagenesis we created base-substitutions to mutate both the splice donor and acceptor sites of the synthetic intron in pT115H to create pT115HM, which would abolish splicing but not alter the size of the plasmid and thereby not add additional variables. It was decided to analyze pT115H because it showed both expansions and contractions. End-stage (not quantitative) RT–PCR showed that the splice site mutations in pT115HM completely abolished splicing normally seen with pT115H [Figure 5A; note that the synthetic intron in pT115K is inefficiently spliced ($\sim$30%) and this was confirmed in separate transfections; however, it should be noted that this is not an accurate estimate of the splicing efficiency since this RT-PCR is non-quantitative, and overexpression via doxycycline may overwhelm the splicing machinery]. In order to quantitatively determine if abolition of splicing had any effect on the stability of the mutant transcript, real-time RT-PCR was performed at hourly intervals following withdrawal of doxycycline (i.e. the source of transcriptional induction). As seen in Figure 5B, the mutant (unspliceable) pT115HM transcript was significantly unstable compared with pT115H, showing a precipitous fall in transcript level to $\sim$10% within 2 h. Whereas, despite the inefficient splicing of the synthetic intron in the pT115H transcript, it remained comparatively stable, remaining at $\sim$70% through 4 h [Figure 5B; it steadily decreased to $\sim$55% at 6 h and $\sim$20% at 12 h (data not shown)]. Transfection in COS Tet-On cells in the presence and absence of doxycycline showed that the mutant pT115HM construct
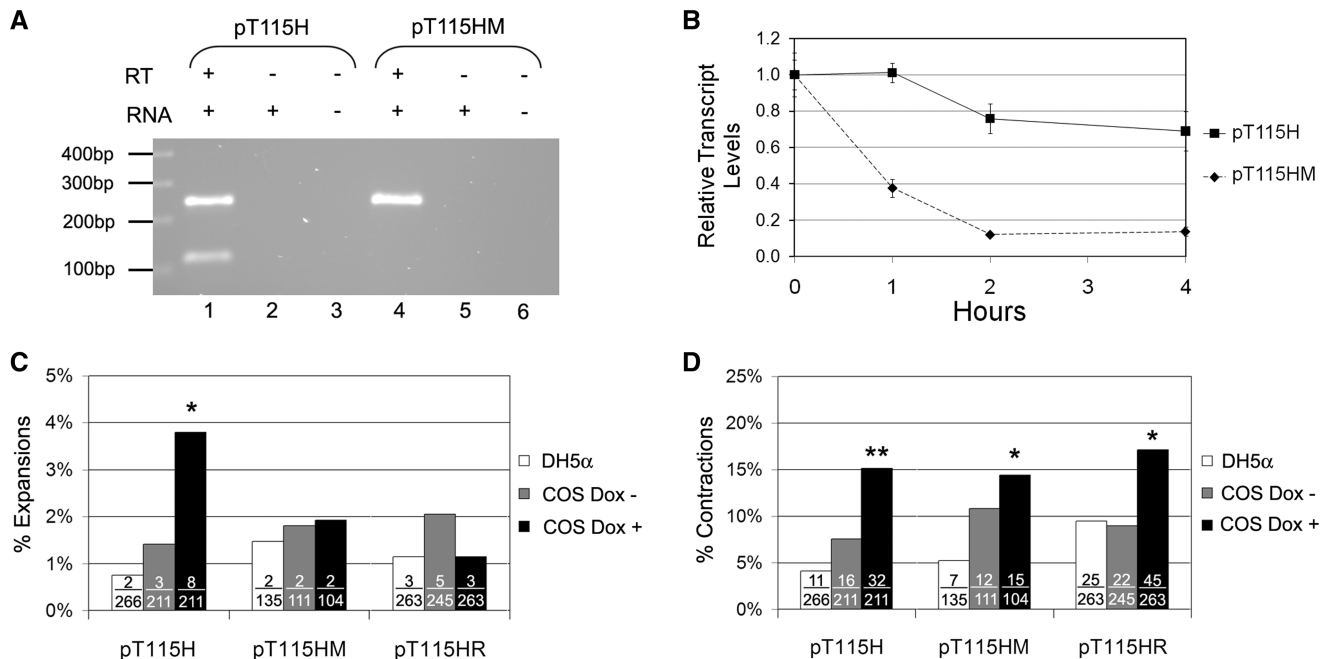
Figure 5. Transcription-induced instability of the $(GAA \bullet TTC)_n$ sequence is related to transcript stability. pT115H, which showed transcription-induced expansions and contractions, was modified to prevent splicing by introducing base substitutions in the splice donor and acceptor sites to generate pT115HM. Both constructs, transfected in COS Tet-On cells in the absence or presence of doxycycline-induced transcription, were analyzed for splicing efficiency, transcript stability, and mutational spectrum of the $(GAA \bullet TTC)_{115}$ sequence. (**A**) RT-PCR of RNA extracted from transfected COS Tet-On cells in the presence of doxycycline showed the predicted spliced product (115 bp) in pT115H but not in pT115HM (which only showed the unspliced product of 246 bp), indicating that the splice site mutations prevented splicing. Note that splicing of the synthetic intron in pT115H was inefficient. (**B**) Transcript stability was measured by estimating the amount of residual transcript upon withdrawal of doxycycline via real-time quantitative RT-PCR of RNA extracted at hourly intervals. pT115HM-derived transcript was significantly unstable compared with pT115H, showing a precipitous fall in transcript level to ~10% within 2 h. The pT115H transcript was comparatively stable, remaining at ~70% through 4 h. Error bars (SEM) are from triplicate RT-PCR assays of two separate experiments. (**C** and **D**) Expansions and contractions were analyzed by comparing the background instability prior to transfection (white bars) to instability following replication in the absence (gray bars) or presence (black bars) of doxycycline-induced transcription. pT115HM did not show the transcription-induced expansions seen with pT115H, but the contractions were seen with both constructs. The same result, i.e. increased contractions but abolition of expansions was also seen with an independent mutant of pT115H, made by removal of the synthetic intron (pT115HR). *P*-values are derived using Chi-square test (*$P < 0.05$ and **$P < 0.01$).

did not show any transcription-induced expansions (as was seen with pT115H), however, the transcription-induced contractions persisted (Figure 5C and D). These data tend to support the role of the transcript in the initiation of expansions rather than the process of transcriptional elongation *per se*. Furthermore, the pT115H construct was also mutated to remove the entire synthetic intron (pT115HR). The transcript generated by pT115HR was also unstable, decreasing to ~40% by 4 h and ~16% by 6 h (data not shown). Transfection of pT115HR in COS Tet-On cells with and without doxycycline-induced transcription showed the same result as pT115HM, i.e. abolition of expansions, but retention of contractions (Figure 5C and D), thus further supporting the role of the transcript in the expansions seen with pT115H.

DISCUSSION

There is considerable evidence for the role of transcription in either increasing or enabling instability of triplet-repeat sequences. For instance, transcription increased the instability of $(CAG \bullet CTG)_n$ and $(GAA \bullet TTC)_n$ sequences in cultured human cells (27,28), and was required for instability of $(CAG \bullet CTG)_n$ in a Drosophila model (30). Recently, it was shown that reduced transcription correlated with a reduced rate of expansion of the $(GAA \bullet TTC)_n$ sequence in cultured human cells (29). Consistent with these observations, we found in two of our five constructs (pT115N and pT115H) that transcription was required in order to initiate instability. However, since two other constructs were stable despite being transcribed and replicated (pT115S and pT115A), our data also indicate that neither replication alone, nor replication plus transcription are sufficient to induce triplet-repeat instability in mammalian cells.

It is also known that instability of triplet-repeats is modulated by *cis*-acting and tissue-specific factors; both play a role in determining whether instability would occur and the particular pattern of instability. For instance, we have previously shown that the expanded $(GAA \bullet TTC)_n$ sequence shows contractions in most tissues from Friedreich ataxia patients, but shows a tendency for expansion in specific regions of the nervous system (2,33). The tendency for tissue-specific expansion in the DRG and cerebellum was also seen in a mouse model containing

the expanded $(GAA\bullet TTC)_n$ sequence in the context of the entire human *FXN* gene (9,10). However, a similar $(GAA\bullet TTC)_n$ sequence 'knocked in' into the mouse *Fxn* locus did not show any instability (11), indicating that despite being replicated and transcribed strong *cis*-acting elements regulate triplet-repeat instability. A potent *cis*-acting modifier of instability in mammalian cells involves the distance of the origin of replication from the repeat tract (12–14). In the fork-shift model, replication-induced instability occurs as a function of the extent and the sequence of the repeat tract that maps within the Okazaki initiation zone (13,15). Our present experiments were designed to investigate if transcription plays a role in further modulating the fork-shift model in order to refine the understanding of the mechanism underlying locus-specific differences in triplet-repeat instability. Indeed, our data suggest a complex interplay between replication and transcription, where the distance between the origin of replication and the repeat tract also helped to determine if transcription would induce instability. This supports the role of transcription in further modulating the role of the *cis*-acting determinants of repeat instability.

It is believed that progressive expansion of the $(GAA\bullet TTC)_n$ sequence in the *FXN* gene in DRG of Friedreich ataxia patients plays a role in influencing the tissue-specific disease pathogenesis (2). Perhaps, it is also likely that progressive expansion of short 'borderline' alleles may help tip the balance and influence whether someone develops a mild, late-onset Friedreich ataxia (38). Therefore, using the same orientation of transcription seen in the endogenous *FXN* gene (GAA as the non-template strand), the transcription-dependent expansions seen with pT115H may have relevance to understanding a process that is important for disease pathogenesis. Our data, and other observations, tend to support a model that involves the formation of a R-loop (in order to explain the expansions seen with pT115H) such that the GAA-containing nascent transcript forms a hybrid with the template (TTC) strand. First, transcription of the $(GAA\bullet TTC)_n$ sequence, *in vitro* and in *E. coli*, is known to form a stable RNA–DNA hybrid of this type (37). Secondly, the repeat expansions seen with pT115H were dependent on the stability of the GAA-containing transcript, supporting the role for the nascent transcript rather than the process of transcriptional elongation *per se*. Thirdly, R-loops are known to be mutagenic, mostly via the transcription associated recombination (TAR) pathway (39). It is nevertheless puzzling that we see expansions with pT115H but not the other constructs, all of which should in theory also be predisposed to forming R-loops. Similar to the fork-shift model, it is possible that the precise location and extent of the R-loop may determine its mutagenic potential. In fact, R-loop formation itself is dependent of several factors such as the presence of a single-strand nick on the non-template strand, G-rich sequence in the non-template strand, and the distance from the promoter (40). Interestingly, TAR, just like somatic instability of the $(GAA\bullet TTC)_n$ sequence, is also subject to *cis*-acting controls, where the genomic context of the recombination substrate determines if TAR will occur (41). This differential ability of mutagenic R-loop formation and/or TAR may help explain, at least in part, the differences in the instability observed with the various constructs.

Replication fork stalling has been observed in bacteria when replication and transcription are in the head-on versus co-linear orientation (31). Indeed, we have found in *E. coli* that errors produced in absence of efficient replication restart mechanisms and during repair of double-strand breaks within the $(GAA\bullet TTC)_n$ sequence lead to high levels of instability of the $(GAA\bullet TTC)_n$ sequence (42,43). However, in mammalian cells this potential source of DNA instability caused by interference between replication and transcription is suppressed by Topoisomerase I and other factors that prevent assembly of mRNA-particle complexes (mRNPs) (44). Indeed, cells deficient in Topoisomerase I (and other mRNP assembly mutants) have enhanced R-loop formation which results in stalling of replication forks and DNA breaks. Our data showed equal replication efficiency in the absence or presence of transcription. However, despite these data, we cannot rule out the possibility that transient fork stalling or slowing down during a head-on collision (as may be possible in pT115H), likely involving only a small proportion of repeat-containing plasmids, could also play a role in the generation of mutagenic templates. However, the generality of this mechanism remains unclear since pT115S and pT115A showed no transcription-dependent increase in repeat instability, both of which have the same potential for head-on collision between transcription and replication.

In summary, our data show that replication and transcription work in concert to generate different patterns of triplet-repeat instability in mammalian cells. This has implications for the locus-specific, tissue-specific, and progressive somatic instability of triplet-repeats seen in human and mouse tissues. The locus-specific differences in instability of genomic triplet-repeats could be caused by the particular local configuration of replication and transcription. It is tempting to speculate that variable tissue levels of transcription and/or transcript stability could determine the tissue-specific variability in somatic instability of expanded triplet-repeats. Also, ongoing transcription in post-mitotic cells could be one of the factors underlying the progressive somatic instability seen in terminally differentiated tissues such as DRG in Friedreich ataxia patients.

REFERENCES

1. López Castel,A., Cleary,J.D. and Pearson,C.E. (2010) Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat. Rev. Mol. Cell Biol.*, **11**, 165–170.

2. 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.

3. Kennedy,L., Evans,E., Chen,C.M., Craven,L., Detloff,P.J., Ennis,M. and Shelbourne,P.F. (2003) Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. *Hum. Mol. Genet.*, **12**, 3359–3367.

4. Shelbourne,P.F., Keller-McGandy,C., Bi,W.L., Yoon,S.R., Dubeau,L., Veitch,N.J., Vonsattel,J.P., Wexler,N.S., US-Venezuela Collaborative Research Group. In Arnheim,N. *et al.* (2007) Triplet repeat mutation length gains correlate with cell-type specific vulnerability in Huntington disease brain. *Hum. Mol. Genet.*, **16**, 1133–1142.

5. Anvret,M., Ahlberg,G., Grandell,U., Hedberg,B., Johnson,K. and Edstrom,L. (1993) Larger expansions of the CTG repeat in muscle compared to lymphocytes from patients with myotonic dystrophy. *Hum. Mol. Genet.*, **2**, 1397–1400.

6. Ashizawa,T., Dubel,J.R. and Harati,Y. (1993) Somatic instability of CTG repeat in myotonic dystrophy. *Neurology*, **43**, 2674–2678.

7. Thornton,C.A., Johnson,K.J. and Moxley,R.T. (1994) Myotonic dystrophy patients have larger CTG expansions in skeletal muscle than in leukocytes. *Ann. Neurol.*, **35**, 104–107.

8. Monckton,D.G., Coolbaugh,M.I., Ashizawa,T., Siciliano,M.J. and Caskey,C.T. (1997) Hypermutable myotonic dystrophy CTG repeats in transgenic mice. *Nat. Genet.*, **15**, 193–196.

9. Clark,R.M., De Biase,I., Malykhina,A.P., Al-Mahdawi,S., Pook,M. and Bidichandani,S.I. (2007) The GAA triplet-repeat is unstable in the context of the human FXN locus and displays age-dependent expansions in cerebellum and DRG in a transgenic mouse model. *Hum. Genet.*, **120**, 633–640.

10. Al-Mahdawi,S., Pinto,R.M., Ruddle,P., Carroll,C., Webster,Z. and Pook,M. (2004) GAA repeat instability in Friedreich ataxia YAC transgenic mice. *Genomics*, **84**, 301–310.

11. Miranda,C.J., Santos,M.M., Ohshima,K., Smith,J., Li,L., Bunting,M., Cossee,M., Koenig,M., Sequeiros,J., Kaplan,J. *et al.* (2002) Frataxin knockin mouse. *FEBS Lett.*, **512**, 291–297.

12. Rindler,M., Clark,R.M., Pollard,L.M., De Biase,I. and Bidichandani,S.I. (2006) Replication in mammalian cells recapitulates the locus-specific differences in somatic instability of genomic GAA triplet-repeats. *Nucleic Acids Res.*, **34**, 6352–6361.

13. Cleary,J.D., Nichol,K., Wang,Y.H. and Pearson,C.E. (2002) Evidence of cis-acting factors in replication-mediated trinucleotide repeat instability in primate cells. *Nat. Genet.*, **31**, 37–46.

14. Nichol,E.K., Leonard,M.R. and Pearson,C.E. (2005) Role of replication and CpG methylation in fragile X syndrome CGG deletions in primate cells. *Am. J. Hum. Genet.*, **76**, 302–311.

15. Cleary,J.D. and Pearson,C.E. (2005) Replication fork dynamics and dynamic mutations: the fork-shift model of repeat instability. *Trends Genet.*, **21**, 272–280.

16. Monckton,D.G., Wong,L.J., Ashizawa,T. and Caskey,C.T. (1995) Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: small pool PCR analyses. *Hum. Mol. Genet.*, **4**, 1–8.

17. Mangiarini,L., Sathasivam,K., Mahal,A., Mott,R., Seller,M. and Bates,G.P. (1997) Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. *Nat. Genet.*, **15**, 197–200.

18. Lia,A.S., Seznec,H., Hofmann-Radvanyi,H., Radvanyi,F., Duros,C., Saquet,C., Blanche,M., Junien,C. and Gourdon,G. (1998) Somatic instability of the CTG repeat in mice transgenic for the myotonic dystrophy region is age dependent but not correlated to the relative intertissue transcription levels and proliferative capacities. *Hum. Mol. Genet.*, **7**, 1285–1291.

19. Wheeler,V.C., Auerbach,W., White,J.K., Srinidhi,J., Auerbach,A., Ryan,A., Duyao,M.P., Vrbanac,V., Weaver,M., Gusella,J.F. *et al.* (1999) Length-dependent gametic CAG repeat instability in the Huntington's disease knock-in mouse. *Hum. Mol. Genet.*, **8**, 115–122.

20. Kennedy,L. and Shelbourne,P.F. (2000) Dramatic mutation instability in HD mouse striatum: does polyglutamine load contribute to cell-specific vulnerability in Huntington's disease? *Hum. Mol. Genet.*, **9**, 2539–2544.

21. Fortune,M.T., Vassilopoulos,C., Coolbaugh,M.I., Siciliano,M.J. and Monckton,D.G. (2000) Dramatic, expansion-biased, age-dependent, tissue-specific somatic mosaicism in a transgenic mouse model of triplet repeat instability. *Hum. Mol. Genet.*, **9**, 439–445.

22. Lorenzetti,D., Watase,K., Xu,B., Matzuk,M.M., Orr,H.T. and Zoghbi,H.Y. (2000) Repeat instability and motor incoordination in mice with a targeted expanded CAG repeat in the Sca1 locus. *Hum. Mol. Genet.*, **9**, 779–785.

23. Seznec,H., Lia-Baldini,A.S., Duros,C., Fouquet,C., Lacroix,C., Hofmann-Radvanyi,H., Junien,C. and Gourdon,G. (2000) Transgenic mice carrying large human genomic sequences with expanded CTG repeat mimic closely the DM CTG repeat intergenerational and somatic instability. *Hum. Mol. Genet.*, **9**, 1185–1194.

24. Gomes-Pereira,M., Fortune,M.T. and Monckton,D.G. (2001) Mouse tissue culture models of unstable triplet repeats: in vitro selection for larger alleles, mutational expansion bias and tissue specificity, but no association with cell division rates. *Hum. Mol. Genet.*, **10**, 845–854.

25. van den Broek,W.J., Nelen,M.R., Wansink,D.G., Coerwinkel,M.M., te Riele,H., Groenen,P.J. and Wieringa,B. (2002) Somatic expansion behaviour of the (CTG)n repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch-repair proteins. *Hum. Mol. Genet.*, **11**, 191–198.

26. Watase,K., Venken,K.J., Sun,Y., Orr,H.T. and Zoghbi,H.Y. (2003) Regional differences of somatic CAG repeat instability do not account for selective neuronal vulnerability in a knock-in mouse model of SCA1. *Hum. Mol. Genet.*, **12**, 2789–2795.

27. 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.

28. Soragni,E., Herman,D., Dent,S.Y., Gottesfeld,J.M., Wells,R.D. and Napierala,M. (2008) Long intronic GAA*TTC repeats induce epigenetic changes and reporter gene silencing in a molecular model of Friedreich ataxia. *Nucleic Acids Res.*, **36**, 6056–6065.

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

30. Jung,J. and Bonini,N. (2007) CREB-binding protein modulates repeat instability in a Drosophila model for polyQ disease. *Science*, **315**, 1857–1859.

31. Mirkin,E.V. and Mirkin,S.M. (2005) Mechanisms of transcription-replication collisions in bacteria. *Mol. Cell. Biol.*, **25**, 888–895.

32. Pollard,L.M., Sharma,R., Gomez,M., Shah,S., Delatycki,M.B., Pianese,L., Monticelli,A., Keats,B.J. and Bidichandani,S.I. (2004) Replication-mediated instability of the GAA triplet repeat mutation in Friedreich ataxia. *Nucleic Acids Res.*, **32**, 5962–5971.

33. Sharma,R., Bhatti,S., Gómez,M., Clark,R., Murray,C., Ashizawa,T. and Bidichandani,S.I. (2002) The GAA triplet-repeat sequence in Friedreich ataxia shows a high level of somatic instability in vivo with a significant predilection for large contractions. *Hum. Mol. Genet.*, **11**, 2175–2187.

34. 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.

35. Ohshima,K., Montermini,L., Wells,R.D. and Pandolfo,M. (1998) Inhibitory effects of expanded GAA TTC triplet repeats from intron 1 of the Friedreich ataxia gene on transcription and replication in vivo. *J. Biol. Chem.*, **273**, 14588–14595.

36. Byrne,B.J., Davis,M.S., Yamaguchi,J., Bergsma,D.J. and Subramanian,K.N. (1983) Definition of the simian virus 40 early promoter region and demonstration of a host range bias in the enhancement effect of the simian virus 40 72-base-pair repeat. *Proc. Natl Acad. Sci. USA*, **80**, 721–725.

37. Grabczyk,E., Mancuso,M. and Sammarco,M.C. (2007) A persistent RNA.DNA hybrid formed by transcription of the Friedreich ataxia triplet repeat in live bacteria, and by T7 RNAP in vitro. *Nucleic Acids Res.*, **35**, 5351–5359.

38. Sharma,R., Gómez,M., De Biase,I., Ashizawa,T. and Bidichandani,S.I. (2004) Friedreich ataxia in carriers of

somatically unstable borderline GAA repeat alleles. *Ann. Neurol.*, **56**, 898–901.

39. Aguilera,A. and Gómez-González,B. (2008) Genome instability: a mechanistic view of its causes and consequences. *Nat. Rev. Genet.*, **9**, 204–217.

40. Roy,D., Zhang,Z., Lu,Z., Hsieh,C.L. and Lieber,M.R. (2010) Competition between the RNA transcript and the non-template DNA strand during R-loop formation in vitro: a nick can serve as a strong R-loop initiation site. *Mol. Cell. Biol.*, **30**, 146–159.

41. Gottipati,P., Cassel,T.N., Savolainen,L. and Helleday,T. (2008) Transcription-associated recombination is dependent on replication in mammalian cells. *Mol. Cell. Biol.*, **28**, 154–164.

42. Pollard,L.M., Chutake,Y., Rindler,P.M. and Bidichandani,S.I. (2007) Deficiency of RecA-dependent RecFOR and RecBCD pathways causes increased instability of the $(GAA\bullet TTC)_n$ sequence when GAA is the lagging strand template. *Nucleic Acids Res.*, **35**, 6884–6894.

43. Pollard,L.M., Bourn,R. and Bidichandani,S.I. (2008) Repair of double-strand breaks within the repeat tract enhances instability of the $(GAA\bullet TTC)_n$ sequence. *Nucleic Acids Res.*, **36**, 489–500.

44. Tuduri,S., Crabbé,L., Conti,C., Tourrière,H., Holtgreve-Grez,H., Jauch,A., Pantesco,V., De Vos,J., Thomas,A., Theillet,C. *et al.* (2009) Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nat. Cell Biol.*, **11**, 1315–1324.