Convergent transcription through microsatellite repeat tracts induces cell death

ARTICLE *in* MOLECULAR BIOLOGY REPORTS · JULY 2014

Impact Factor: 2.02 · DOI: 10.1007/s11033-014-3432-y · Source: PubMed

3 AUTHORS, INCLUDING:

Yunfu Lin
University of Texas MD Anderson Cancer Ce…

27 PUBLICATIONS **837** CITATIONS

Mol Biol Rep (2014) 41:5627–5634
DOI 10.1007/s11033-014-3432-y

Convergent transcription through microsatellite repeat tracts induces cell death

William Y. Lin · Yunfu Lin · John H. Wilson

Received: 18 September 2013 / Accepted: 24 May 2014 / Published online: 11 July 2014

Abstract Microsatellite sequences, composed of short tandem repeats and randomly distributed in human genome, can become unstable during various DNA metabolic processes. Expansions of CAG, GAA, CGG and CCTG repeats located in specific genes are responsible for several human disorders. It is known that a major percentage of human genes simultaneously express both sense and antisense transcripts. Recently, we reported that convergent transcription through a CAG_{95} tract in human cells leads to cell cycle arrest as well as robust apoptosis. In this study, we studied the effects of convergent transcription through other types of repeats, using cell lines that contain substrates with inducible sense and antisense transcription through CGG_{66}, GAA_{102}, or $CCTG_{134}$ tracts. We found that convergent transcription through all these repeats inhibits cell growth and induces cell death, though more moderately than convergent transcription through a CAG tract. These results suggest that convergent transcription through various types of tandem repeats represent a novel type of stress to cells.

Keywords Convergent transcription · Tandem repeats · Cell death · DNA toxicity

Introduction

Microsatellite sequences, or short tandem repeats (STRs), are composed of continuous runs of short repetitive units of 1–9 nucleotides. Tracts of such repeats are inherently unstable, becoming longer (expanding) or shorter (contracting) at high frequencies [1]. Previous studies [2–6] in bacteria, yeast, flies, mammalian cells, and mice have shown that repeat instability can occur during DNA replication, DNA repair, recombination, and transcription [2, 7, 8]. Each of these processes exposes single strands of DNA, allowing secondary structures such as hairpins, slipped-strand DNA duplexes, and triple helical DNA structures to form within the repeat tracts [9–12], which interferes with the cells' ability to convert the secondary structures back to a normal DNA duplex with same length repeat [13, 14].

In most places in the human genome, instability of microsatellite sequences, especially the more common small changes of one or a few repeat units, does not cause detectable pathology. Certain types of repeats in specific genes, however, have a tendency to expand beyond critical length thresholds, which causes human neurodegenerative disorders characterized by loss or death of neurons in disease-specific brain regions [15–17]. For example, expansions of CAG tracts are responsible for Huntington disease (HD), myotonic dystrophy 1 (DM1), and several spinocerebellar ataxias (SCA) [16]. Expansions of CGG, GAA, and CCTG repeat tracts cause fragile × mental retardation syndrome (FMR1), Friedriech ataxia (FRDA), and myotonic dystrophy 2 (DM2), respectively [15]. Studies over the past two decades have focused on the molecular basis for expansion-biased repeat instability in germline and somatic cells and on the means by which expanded repeats cause disease [3, 4, 6, 15, 18]. The mechanisms of repeat instability and disease-associated neuronal cell death are not clearly defined [4, 16, 19].

Transcription through CAG and GAA tracts promotes instability [20–22], which is likely induced by secondary structures that arrest RNA polymerase II (RNAPII) and

W. Y. Lin · Y. Lin (✉) · J. H. Wilson
Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX 77030, USA
e-mail: yunful@bcm.edu

trigger transcription-coupled DNA repair [6, 23, 24]. Transcription-induced changes in length of the repeat tract require transcription-coupled nucleotide excision repair (TC-NER), which removes DNA lesions that block transcription [23, 25], and are modulated by several factors including mismatch repair (MMR) components, formation of RNA/DNA hybrids in the repeat tract, the single strand break repair pathway, epigenetic modifications, and stress responses [26–30]. Moreover, antisense transcription also promotes repeat instability [31, 32] and simultaneous sense and antisense expression—convergent transcription—stimulates repeat instability synergistically [31]. These studies have demonstrated that transcription and antisense transcription play important roles in causing repeat instability.

Beyond its role in promoting repeat instability, convergent transcription through a CAG_{95} tract in the genome triggers cell cycle arrest and massive cell death via apoptosis [31], via signaling pathways that resemble the cellular response to DNA damage [33, 34]. Convergent-transcription-induced cell death is likely caused by the difficulty of removing transcription roadblocks on both strands, which activates ATR and its downstream targets cell cycle checkpoint kinase 1 (CHK1) and p53 [35]. We hypothesized that other structure-inducing repeats would also cause cell death when subjected to convergent transcription. In this study, we tested our hypothesis using human cell lines that contain one copy of substrates with inducible convergent transcription through CGG_{66}, GAA_{102}, or $CCTG_{134}$ repeat tracts. We find that convergent transcription through each of these repeats induces cell cycle arrest and cell death. These results, combined with our previous findings, suggest that convergent transcription through repeat tracts represents a novel type of stress for mammalian cells.

Materials and methods

Cell lines and cell culture

All cell lines used in this study were constructed using RS11 cells, which are derived from the human fibrosarcoma HT1080 cell line. The construction of RS11 cells was described previously [31]. Briefly, RS11 cells contain a gene expressing the rtTA protein, a fusion of the HSV VP16 activation domain and the reverse tetracycline repressor protein, to drive expression of pTRE-CMV^{mini} promoter in the presence of the inducer doxycycline. RS11 cells also contain the genes for RheoReceptor-1 and RheoActivator, which drive expression of the pNERB-X1 promoter in the presence of the inducer RSL1.

Plasmids pDIT11 (CGG_{66}), pDIT21 (GAA_{102}), and pDIT31 ($CCTG_{134}$) were derived from plasmid pDIT1, which contains an *HPRT* minigene with a CAG tract, flanked by two *Not*I sites, in its intron [31]. pDIT11 was constructed by PCR amplifying the CGG_{66} tract from pRW3691 (Dr. M. Napierala, MD Anderson Cancer Center) and inserting it into pDIT1 in place of the CAG tract. pDIT21 and pDIT31 were constructed in the same way, using the GAA_{102} tract from pRW4212 (Dr. M. Napierala) and the $CCTG_{134}$ tract from pRW5114 (Dr. M. Napierala). The repeat tracts from plasmids pRW3691, pRW4212, and pRW5114 included a few hundred base pairs of sequences flanking the repeats, which were originally from the genomes of human patients.

To construct the DIT11 (CGG_{66}), DIT21 (GAA_{102}) and DIT31 ($CCTG_{134}$) cell lines, we electroporated five million RS11 cells with 1 μg of the appropriate linearized plasmid, either pDIT11, pDIT21, or pDIT31. After electroporation, cells were plated in the presence of puromycin (2.5 μg/mL) for 2 weeks. Individual puromycin-resistant colonies were analyzed by Southern blotting to identify clones with a single integrated copy of the *HPRT* minigene that was responsive to both inducers. Cell lines were normally grown at 37 °C with 5 % CO_2 in DMEM/F-12 medium supplemented with 10 % fetal bovine serum and 1 % MEM nonessential amino acids. To starve the cells, cells were grown in DMEM/F-12 medium supplemented with 1 % fetal bovine serum.

Induction of transcription

Cells were maintained in the absence of transcription inducers. In experiments, the time at which inducers were added to the cells was defined as day 0. Sense transcription of *HPRT* minigene was induced by adding doxycycline (dissolved in water) at a final concentration of 2.0 μg/mL. Because the half-life of doxycycline in medium is about 24 h, we added 1.0 μg/mL of doxycycline directly into the medium each day until the treatment was completed. Antisense transcription of *HPRT* minigene was induced by adding RSL1 (dissolved in DMSO) to the medium at a final concentration of 500 nM. No additional RSL1 was required for treatments up to 7 days.

Viable cell measurements

In accord with our previous studies, adherent cells (attached to the plate) were operationally defined as viable, and nonadherent cells (present in the medium) were operationally defined as dead [31]. We have previously shown that adherent cells contain fewer than 4 % of cells that stain with propidium iodide, indicating that >96 % of adherent cells are viable [31]. The number of dead cells was determined by counting the nonadherent cells in the medium. After the nonadherent cells were removed by washing, the number of

viable cells was by detaching the adherent cells from the dish by trypsin treatment and counting them. The percentage of dead cells was calculated as the number of nonadherent cells divided by the total number of cells, which is the sum of adherent and nonadherent cells.

Real-time RT-PCR

Total RNA was extracted from about one million cells, using RNeasy mini kits (Qiagen). Induction of *HPRT* sense or antisense transcription was measured on RNA extracted one day after addition of doxycycline or RSL1 to the medium. The levels of sense and antisense transcripts were measured separately using strand-specific real-time RT-PCR, as described previously [31]. For sense transcripts, we synthesized cDNA from the sense transcript, using reverse transcriptase and a primer (SSP-H8) that links the *HPRT* sense-strand specific H8 sequence to the SSP universal primer, whose sequence is not present in the genome. Reverse transcriptase was then inactivated by incubation at 95 °C for 10 min, the cDNA was amplified using the *HPRT*-specific H7 primer and the SSP universal primer, and amounts were quantified by real-time PCR. For antisense transcripts, we synthesized cDNA from the antisense transcript, using reverse transcriptase and a primer (SSP-H7) that links the *HPRT* antisense-strand specific H7 sequence to the SSP universal primer. The antisense cDNA was then amplified using a mixture of the H7 and SSP primers and quantified by real-time PCR. Sequences of the primers are listed in Table 1. For strand-specific real-

Table 1 Sequences of RT-PCR primers used in this study

SSP-H7	CGATGCTTGGACAGCCTGACCAGTCAACAGGGGAC
SSP-H8	CGATGCTTGGACAGCCTGCGTGGGGTCCTTTTCACC
SSP	CGATGCTTGGACAGCCTG
H7	ACCAGTCAACAGGGGAC
H8	CGTGGGGTCCTTTTCACC

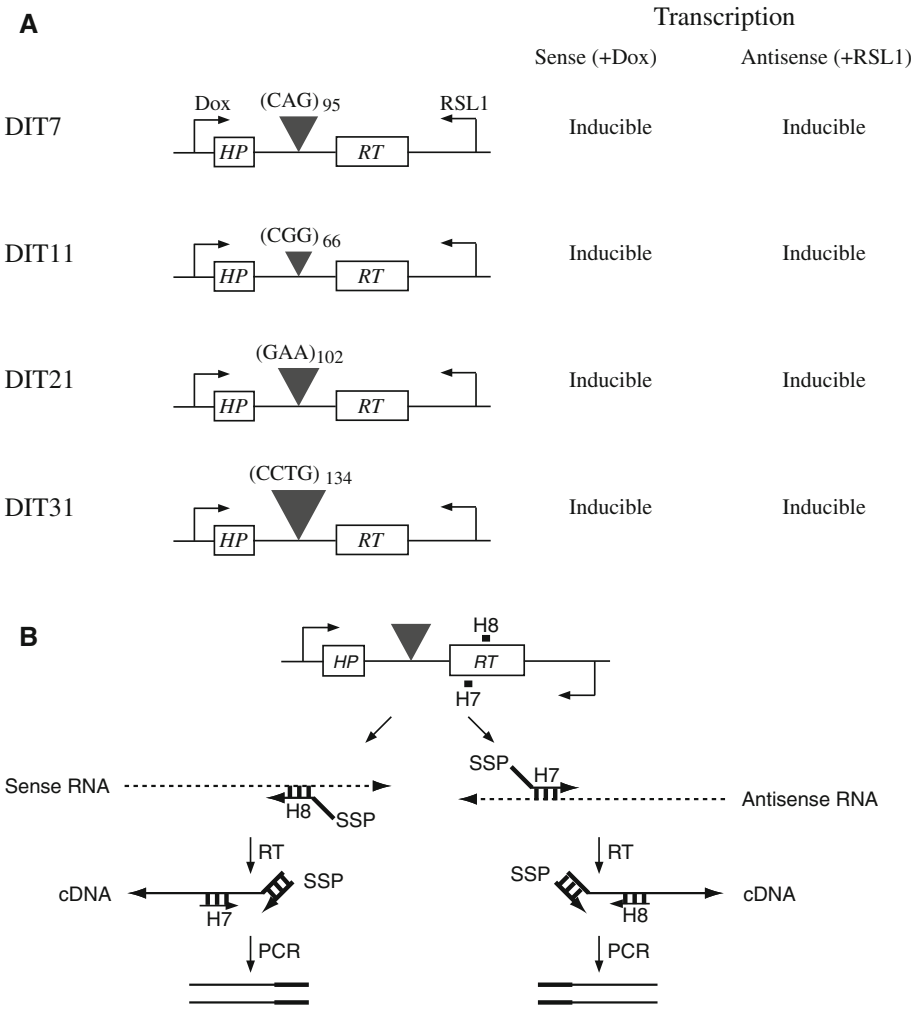

Fig. 1 Cell lines for testing effects of convergent transcription. **a** In all cells, the repeat tracts are centered in the 2.1-kb intron in the randomly integrated single copy of *HPRT* minigene, and the repeat is 1.6 kb downstream of the sense promoter and 2.5 kb upstream of the antisense promoter. Sense transcription is driven by the inducible promoter pCMVmini, which responds to doxycycline (Dox); and antisense transcription is driven by the inducible promoter pNEBR-X1, which responds to RSL1. **b** Schematic strategy for strand-specific real-time RT-PCR to quantitatively measure sense and antisense transcripts from the *HPRT* minigene

time RT-PCR, total RNA (50 ng per reaction) was reverse transcribed at 50 °C for 30 min and assayed using the SYBR Green PCR kit (Qiagen). Results were normalized to the concentration of β-actin RNA, which was determined the same way. Conditions for real-time PCR were 95 °C for 15 min, followed by 45 cycles of 94 °C for 15 s, 50 °C for 30 s, and 72 °C for 30 s.

Statistics

Statistical analyses of significance were conducted using Student's *t* test to compare the means and standard deviations derived from multiple experiments.

Results

Cell lines with convergently transcribed repeat tracts

To test the effects of convergent transcription through various repeat sequences, we constructed several human cell lines that each carried a CGG_{66}, GAA_{102}, or $CCTG_{134}$ repeat tract inserted into the intron of a randomly integrated single copy of the *HPRT* minigene (Fig. 1a). These cell lines were constructed by integrating repeat-containing plasmids in the genome of cell line RS11, which constitutively expresses the protein regulators required for both the Tet-ON and the Rheo-Switch inducible expression systems [31]. Cell lines DIT11-7 and DIT11-9 were independently derived and each carries a CGG_{66} tract. The independently derived DIT21-11 and DIT21-18 cell lines carry a GAA_{104} repeat tract, and the independently derived DIT31-1 and DIT31-26 cell lines contain a $CCTG_{134}$ tract. In each of these cell lines, sense transcription could be induced by addition of doxycycline and antisense transcription could be induced by RSL1. Sense transcription was induced 15–25 fold by addition of doxycycline, and antisense transcription was induced 8–15 fold by addition of RSL1 (Table 2), as determined by strand-specific quantitative RT-PCR (Fig. 1b).

Table 2 Induced sense and antisense transcript levels of *HPRT* minigene

Cell line	Dox + RSL1	
	Sense*	Antisense*
DIT11-7	15	8
DIT11-9	15	9
DIT21-11	17	10
DIT21-18	24	15
DIT31-1	19	12
DIT31-26	23	10

* The levels of sense and antisense transcripts in corresponding cells without the presence of inducers were arbitrarily defined as 1

Convergent transcription induces cell death

Previously, we showed that convergent transcription through a CAG_{95} tract in human cells induced massive apoptosis-associated cell death, with only about 50 % of cells alive after 4-day treatment [31]. Convergent

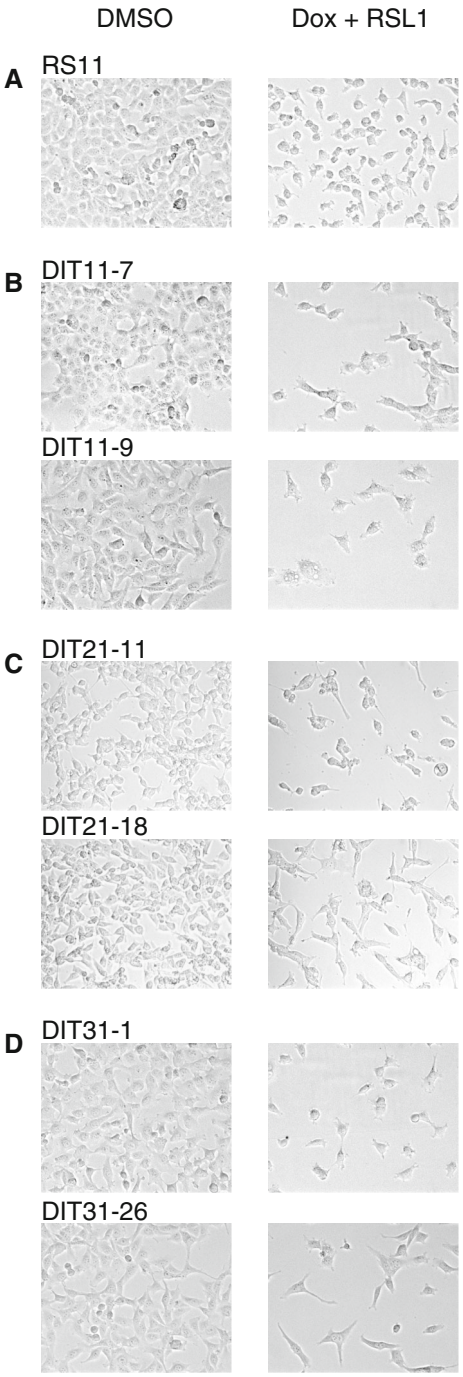

Fig. 2 Convergent transcription reduces viable cells. Cells were plated at the same initial density and photographed after 7-day treatment with inducers doxycycline and RSL1. **a** RS11 cells. **b** DIT11 cells. **c** DIT21 cells. **d** DIT31 cells

transcription also induced cell death in DIT11, DIT21 and DIT31 cells; however, cell death was apparent only after 7 days of treatment with doxycycline plus RSL1. As a control for the nonspecific effects of transcription inducers, we used RS11 cells, which are the parental cells for the DIT11, DIT21, and DIT31 cell lines. After 7 days of treatment with doxycycline plus RSL1, the number of viable RS11 cells was lower than in control experiments with DMSO, suggesting doxycycline and RSL1 may have some nonspecific effects, but the viable cells look healthy (Fig. 2a). Compared with RS11 cells, convergent transcription leads to relatively fewer viable cells for DIT11 cells (Fig. 2b), DIT21 cells (Fig. 2c), and DIT31 cells (Fig. 2d). In addition, the viable DIT11, DIT21, and DIT31 cells display abnormal morphological features, which were also observed in convergently transcribed DIT7 cells with a CAG_{95} tract [31].

To quantify the effects of convergent transcription on cell death, we counted the numbers of adherent (viable) and nonadherent (dead) cells and calculated the percentage of dead cells. Without inducers, dead cells ranged from 3 to 6 % for all cell lines and did not differ significantly (Fig. 3). By contrast, in the presence of inducers the percentages of dead cells significantly increased in all cell lines that carried repeat tracts (Fig. 3; Table 3). Thus, convergent transcription through CGG_{66} tracts (DIT11-7 and DIT11-9), GAA_{102} tracts (DIT21-11 and DIT21-18), and $CCTG_{134}$ tracts (DIT31-1 and DIT31-26) can induce cell death. Induction of sense or antisense transcription alone did not cause significant cell death in these cell lines in 7 days (Table 3). We also monitored the cell death after 3 or 5 days of induction of convergent transcription, no cell death was detected in 3 days, while slight and insignificant cell death was observed in 5 days (Table 3).

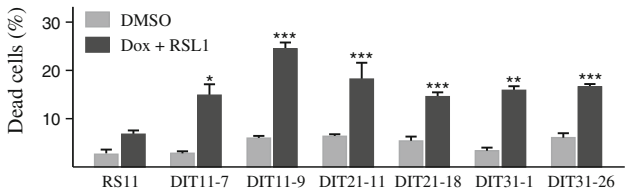

Fig. 3 Convergent transcription causes cell death. The percentage of dead cells, which is determined by the number of dead cells divided by the addition of both viable and dead cells, was assessed after 7-day treatment of transcription inducers. The percentages of dead cells are 6.7 % for RS11, and increased to 15 % for DIT11-7, 24 % for DIT11-9, 19 % for DIT21-11 16 % for DIT21-18, 16 % for DIT31-1, and 17 % for DIT31-26. Statistical significance relative to RS11 cells treated with doxycycline and RSL1 is indicated: $*p < 0.01$; $**p < 0.001$; $***p < 0.0001$

Convergent transcription suppresses growth of starved cells

In our previous studies we showed that convergent transcription through a CAG_{95} tract suppressed cell growth and induced cell death more quickly in starved cells [31]. To test for a similar effect of convergent transcription through other types of repeats, we maintained DIT11, DIT21 and DIT31 cells in 1 % FBS medium for 3 days in the presence or absence of inducers and then counted the number of viable cells. For RS11 cells, the number of viable cells was the same in the presence and absence of inducers (Fig. 4a), whereas fewer viable cells were observed in the presence of inducers in DIT11 cells (Fig. 4b), DIT21 cells (Fig. 4c), and DIT31 cells (Fig. 4d). As shown in Fig. 5, the relative numbers of viable cells was 71 % for DIT11-7, 73 % for DIT11-9, 75 % for DIT21-11, 80 % for DIT21-18, 75 % for DIT31-1, and 70 % for DIT31-26, significantly lower than RS11 cells. We also measured the number of dead cells, which was not significantly changed over the 3-day treatment period (data not shown). These results suggest that cell growth is reduced by convergent transcription through CGG_{66} tracts (DIT11-7 and DIT11-9), GAA_{102} tracts (DIT21-11 and DIT21-18), and $CCTG_{134}$ tracts (DIT31-1 and DIT31-26).

Discussion

An appreciable percentage of human genes are transcribed in both the sense and antisense directions, suggesting that

Table 3 Cell death induced by convergent transcription

Cell line	7 days	3 days	5 days	7 days		
	DMSO	Dox + RSL1	Dox + RSL1	Dox	RSL1	Dox + RSL1
RS11	2.3 ± 0.5	2.4 ± 0.3	2.5 ± 0.2	2.9 ± 0.3	2.4 ± 0.4	6.7 ± 0.8
DIT7	3.0 ± 0.8	19 ± 4.5	67 ± 3.9			91 ± 8.6
DIT11-7	3.1 ± 0.2	2.9 ± 0.4	3.9 ± 0.7	6.5 ± 0.6	3.7 ± 0.7	15 ± 2.2
DIT11-9	5.2 ± 0.4	4.8 ± 0.6	5.9 ± 0.9	7.8 ± 1.2	6.0 ± 0.5	24 ± 1.8
DIT21-11	5.7 ± 0.3	5.3 ± 0.3	6.2 ± 0.8	8.2 ± 0.8	6.2 ± 1.2	19 ± 3.7
DIT21-18	4.9 ± 0.7	4.6 ± 0.9	6.1 ± 0.7	8.4 ± 1.6	5.8 ± 0.6	16 ± 1.4
DIT31-1	3.5 ± 0.5	3.8 ± 0.6	4.7 ± 0.4	7.1 ± 0.7	4.6 ± 0.4	16 ± 1.3
DIT31-26	5.4 ± 0.7	5.1 ± 0.7	7.8 ± 0.6	8.1 ± 1.8	6.2 ± 0.9	17 ± 0.7

Shown are the percentages of dead cells calculated by the number of dead (nonadherent) cells divided by total number of cells, which is the addition of dead cells and viable (adherent) cells. The data are averaged from at least 6 independent measurements

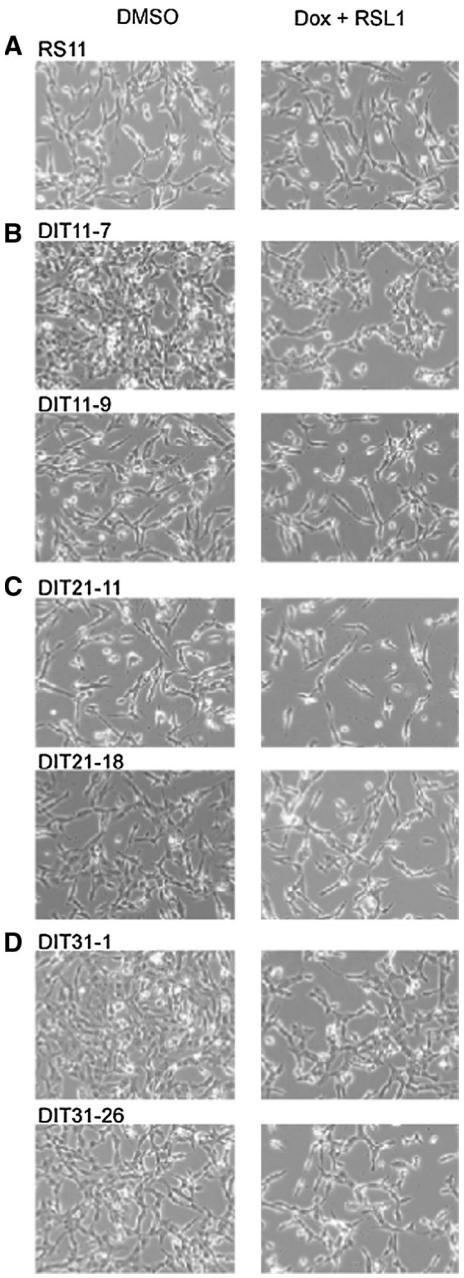

Fig. 4 Convergent transcription in starved cells. Cells were plated at the same initial density for individual cell lines in medium with 1 % FBS and photographed after 3 day treatment with inducers doxycycline and RSL1. **a** RS11 cells. **b** DIT11 cells. **c** DIT21 cells. **d** DIT31 cells

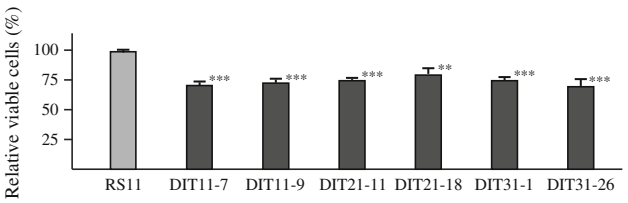

Fig. 5 Convergent transcription suppresses growth of starved cells. Shown are the relative numbers of viable cells, which were calculated by dividing the number of viable cells in treated population with that in untreated population. Same number of cells was plated for each cell line either in the presence or absence of doxycyline plus RSL1. Statistical significance relative to RS11 cells treated with doxycycline and RSL1 is indicated: $^{**}p < 0.001$; $^{***}p < 0.0001$

convergent transcription may be common in the human genome [36]. Antisense transcription has also been documented in an increasing number of those genes in which expansions of repeat sequences such as CAG, CGG, and GAA repeats cause disease [37–45]. We previously reported that convergent transcription through a CAG tract not only promotes repeat instability, but also causes cell-cycle arrest and induces cell death via apoptosis, potentially linking the mechanisms that underlie repeat instability and disease pathology, which typically involves the death or loss of neurons [31]. In this study, we extend our previous studies and reported that convergent transcription through CGG_{66}, GAA_{102}, and $CCTG_{134}$ repeat tracts can also suppress cell growth and induce cell death. Collectively, these studies suggest convergent transcription through repeat tracts creates a novel type of cell stress that can lead to cell death if unresolved.

The underlying mechanism for convergent transcription-induced cell stress has not been fully defined. It has been proposed, however, that during convergent transcription RNAPII complexes stall at a repeat tract on both DNA template strands, which makes it difficult for TC-NER to remove the structures that block the polymerase [23]. For CAG repeats, the blocking structures are likely to be CAG and CTG hairpins stabilized by MSH2/3 [46]. Consistent with this, depletion of XPA, a key component for TC-NER, increased convergent transcription-induced cell death, while depletion of MSH2 reduced cell death [47]. Whatever the specific structural problem, it activates the ATR pathway, subsequently inducing cell cycle arrest and apoptosis [31], similar to the effects of global interference of transcription [48–50]. The key for this mechanism is that a repeat tract can form structures that block transcription. Consistent with this idea, several in vitro studies have demonstrated that CAG, CGG and GAA repeats can block transcription [51–54].

Results from this study, combined with our previous results [31], demonstrate that cell growth arrest and cell death can be induced by convergent transcription through CAG_{95}, CGG_{66}, GAA_{102}, and $CCTG_{134}$ repeat tracts. Convergent transcription through a CAG tract, however, has much more dramatic effects than convergent transcription through CGG, GAA, or CCTG tracts. Convergent transcription through a CAG_{95} tract killed 50 % of cells after just 4 day of treatment with doxycycline and RSL13 [1], whereas convergent transcription through CGG_{66}, GAA_{102}, or $CCTG_{134}$ tracts caused cell death in only about

20 % of cells, even after 7 days of treatment. The general agreement between independently derived cell lines, suggests that the observed differences between the repeat tracts reflect an inherent property of the repeats, as opposed to, for example, integration at different sites in the genome. Based on the similar lengths of these repeats, we postulate that the different effects of convergent transcription on cell death are due to the efficiency with which they form structures capable of blocking transcription.

The ability of convergent transcription through a microsatellite sequence at a single genome locus to cause cell death suggest a novel type of cell stress, one that arises from events directly at the level of DNA, without participation of a toxic protein or a toxic RNA. If related to disease pathology, "DNA toxicity" would be expected to be associated with dominant diseases, since a single locus can cause cell death. The dominant repeat diseases include most of the disorders caused by expansion of a CAG tract, the repeat that causes the most dramatic effect in our assay. By contrast, CGG and GAA repeats, which are associated with recessive diseases, have a milder effect in cells. For CCTG repeats, although its expansion causes a dominant disease that can be severe in patients with more than 1,000 repeat units, it causes a much milder form of the disease at the relatively shorter lengths tested here. As we have shown, convergent transcription-induced cell death is proportional to the length of the CAG tract, and this tract length dependence presumably applies to CCTG repeats as well. It is still too early to say whether convergent transcription-induced cell death contributes to the cause or pathogenesis of repeat diseases. This may be a difficult answer to pin down. A recent study using transgenic fly showed that convergent transcription through a CAG_{100} repeat tract leads to pathological development, but the authors explained their results as toxic double-stranded RNA [55]. At this moment, it is unclear whether convergent transcription-induced cell death is due to DNA toxicity, RNA toxicity, or both.

Acknowledgments We thank members of the Wilson lab for helpful discussions. This work was supported by a grant from the NIH (GM38219) to J.H.W.

References

1. Richard GF, Kerrest A, Dujon B (2008) Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. Microbiol Mol Biol Rev 72:686–727
2. Lin Y, Dion V, Wilson JH (2006) Transcription and triplet repeat instability. In: Wells R, Ashizawa T (eds) Genetic instability and neurological diseases. Elsevier, Amsterdam, pp 691–704
3. Lopez Castel A, Cleary JD, Pearson CE (2010) Repeat instability as the basis for human diseases and as a potential target for therapy. Nat Rev Mol Cell Biol 11:165–170
4. Pearson CE, Edamura KN, Cleary JD (2005) Repeat instability: mechanisms of dynamic mutations. Nat Rev Genet 6:729–742
5. Wells RD, Dere R, Hebert ML, Napierala M, Son LS (2005) Advances in mechanisms of genetic instability related to hereditary neurological diseases. Nucleic Acids Res 33:3785–3798
6. Lin Y, Hubert L Jr, Wilson JH (2009) Transcription destabilizes triplet repeats. Mol Carcinog 48:350–361
7. Cleary JD, Pearson CE (2003) The contribution of cis-elements to disease-associated repeat instability: clinical and experimental evidence. Cytogenet Genome Res 100:25–55
8. Mirkin SM (2007) Expandable DNA repeats and human disease. Nature 447:932–940
9. Wells RD (2007) Non-B DNA conformations, mutagenesis and disease. Trends Biochem Sci 32:271–278
10. Pearson CE, Sinden RR (1996) Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. Biochemistry 35:5041–5053
11. Gacy AM, Goellner G, Juranic N, Macura S, McMurray CT (1995) Trinucleotide repeats that expand in human disease form hairpin structures in vitro. Cell 81:533–540
12. Wang G, Vasquez KM (2006) Non-B DNA structure-induced genetic instability. Mutat Res 598:103–119
13. Bacolla A, Larson JE, Collins JR, Li J, Milosavljevic A, Stenson PD et al (2008) Abundance and length of simple repeats in vertebrate genomes are determined by their structural properties. Genome Res 18:1545–1553
14. Bacolla A, Wells RD (2009) Non-B DNA conformations as determinants of mutagenesis and human disease. Mol Carcinog 48:273–285
15. La Spada AR, Taylor JP (2010) Repeat expansion disease: progress and puzzles in disease pathogenesis. Nat Rev Genet 11:247–258
16. Orr HT, Zoghbi HY (2007) Trinucleotide repeat disorders. Annu Rev Neurosci 30:575–621
17. Gatchel JR, Zoghbi HY (2005) Diseases of unstable repeat expansion: mechanisms and common principles. Nat Rev Genet 6:743–755
18. Li LB, Yu Z, Teng X, Bonini NM (2008) RNA toxicity is a component of ataxin-3 degeneration in Drosophila. Nature 453:1107–1111
19. Riley BE, Orr HT (2006) Polyglutamine neurodegenerative diseases and regulation of transcription: assembling the puzzle. Genes Dev 20:2183–2192
20. Ditch S, Sammarco MC, Banerjee A, Grabczyk E (2009) Progressive GAA.TTC repeat expansion in human cell lines. PLoS Genet 5:e1000704
21. Jung J, Bonini N (2007) CREB-binding protein modulates repeat instability in a Drosophila model for polyQ disease. Science 315:1857–1859
22. Lin Y, Dion V, Wilson JH (2006) Transcription promotes contraction of CAG repeat tracts in human cells. Nat Struct Mol Biol 13:179–180
23. Hanawalt PC, Spivak G (2008) Transcription-coupled DNA repair: two decades of progress and surprises. Nat Rev Mol Cell Biol 9:958–970
24. Kim N, Jinks-Robertson S (2012) Transcription as a source of genome instability. Nat Rev Genet 13:204–214
25. Lin Y, Wilson JH (2007) Transcription-induced CAG repeat contraction in human cells is mediated in part by transcription-coupled nucleotide excision repair. Mol Cell Biol 27:6209–6217
26. Dion V, Lin Y, Hubert L Jr, Waterland RA, Wilson JH (2008) Dnmt1 deficiency promotes CAG repeat expansion in the mouse germline. Hum Mol Genet 17:1306–1317
27. Hubert L Jr, Lin Y, Dion V, Wilson JH (2011) Topoisomerase 1 and single-strand break repair modulate transcription-induced

CAG repeat contraction in human cells. Mol Cell Biol 31:3105–3112
28. Lin Y, Dent SY, Wilson JH, Wells RD, Napierala M (2010) R loops stimulate genetic instability of CTG.CAG repeats. Proc Natl Acad Sci USA 107:692–697
29. Lin Y, Wilson JH (2009) Diverse effects of individual mismatch repair components on transcription-induced CAG repeat instability in human cells. DNA Repair (Amst) 8:878–885
30. Mittelman D, Sykoudis K, Hersh M, Lin Y, Wilson JH (2010) Hsp90 modulates CAG repeat instability in human cells. Cell Stress Chaperones 15:753–759
31. Lin Y, Leng M, Wan M, Wilson JH (2010) Convergent transcription through a long CAG tract destabilizes repeats and induces apoptosis. Mol Cell Biol 30:4435–4451
32. Nakamori M, Pearson CE, Thornton CA (2011) Bidirectional transcription stimulates expansion and contraction of expanded (CTG)*(CAG) repeats. Hum Mol Genet 20:580–588
33. Harper JW, Elledge SJ (2007) The DNA damage response: ten years after. Mol Cell 28:739–745
34. Ciccia A, Elledge SJ. The DNA damage response: making it safe to play with knives. Mol Cell 40:179–204
35. Lin Y, Wilson JH (2011) Transcription-induced DNA toxicity at trinucleotide repeats: double bubble is trouble. Cell Cycle 10:611–618
36. Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M et al (2005) Antisense transcription in the mammalian transcriptome. Science 309:1564–1566
37. Moseley ML, Zu T, Ikeda Y, Gao W, Mosemiller AK, Daughters RS et al (2006) Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. Nat Genet 38:758–769
38. Ladd PD, Smith LE, Rabaia NA, Moore JM, Georges SA, Hansen RS et al (2007) An antisense transcript spanning the CGG repeat region of FMR1 is upregulated in premutation carriers but silenced in full mutation individuals. Hum Mol Genet 16:3174–3187
39. Cho DH, Thienes CP, Mahoney SE, Analau E, Filippova GN, Tapscott SJ (2005) Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF. Mol Cell 20:483–489
40. Chen WL, Lin JW, Huang HJ, Wang SM, Su MT, Lee-Chen GJ et al (2008) SCA8 mRNA expression suggests an antisense regulation of KLHL1 and correlates to SCA8 pathology. Brain Res 1233:176–184
41. He Y, Vogelstein B, Velculescu VE, Papadopoulos N, Kinzler KW (2008) The antisense transcriptomes of human cells. Science 322:1855–1857
42. Wilburn B, Rudnicki DD, Zhao J, Weitz TM, Cheng Y, Gu X, et al. An antisense CAG repeat transcript at JPH3 locus mediates expanded polyglutamine protein toxicity in Huntington's disease-like 2 mice. Neuron 70:427–440
43. Chung DW, Rudnicki DD, Yu L, Margolis RL. A natural antisense transcript at the Huntington's disease repeat locus regulates HTT expression. Hum Mol Genet
44. Sopher BL, Ladd PD, Pineda VV, Libby RT, Sunkin SM, Hurley JB, et al. CTCF Regulates ataxin-7 expression through promotion of a convergently transcribed, antisense noncoding RNA. Neuron 70:1071–1084
45. Hubert L Jr, Lin Y, Dion V, Wilson JH (2011) Xpa deficiency reduces CAG trinucleotide repeat instability in neuronal tissues in a mouse model of SCA1. Hum Mol Genet 20:4822–4830
46. Owen BA, Yang Z, Lai M, Gajek M, Badger JD 2nd, Hayes JJ et al (2005) (CAG)(n)-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition. Nat Struct Mol Biol 12:663–670
47. Lin Y, Wilson JH (2012) Nucleotide excision repair, mismatch repair, and R-loops modulate convergent transcription-induced cell death and repeat instability. PLoS ONE 7:e46807
48. Ljungman M (2005) Activation of DNA damage signaling. Mutat Res 577:203–216
49. Ljungman M (2007) The transcription stress response. Cell Cycle 6:2252–2257
50. Ljungman M, Zhang F (1996) Blockage of RNA polymerase as a possible trigger for u.v. light-induced apoptosis. Oncogene 13:823–831
51. Belotserkovskii BP, Liu R, Tornaletti S, Krasilnikova MM, Mirkin SM, Hanawalt PC. Mechanisms and implications of transcription blockage by guanine-rich DNA sequences. Proc Natl Acad Sci USA 107:12816–12821
52. Belotserkovskii BP, De Silva E, Tornaletti S, Wang G, Vasquez KM, Hanawalt PC (2007) A triplex-forming sequence from the human c-MYC promoter interferes with DNA transcription. J Biol Chem 282:32433–32441
53. Parsons MA, Sinden RR, Izban MG (1998) Transcriptional properties of RNA polymerase II within triplet repeat-containing DNA from the human myotonic dystrophy and fragile X loci. J Biol Chem 273:26998–27008
54. Salinas-Rios V, Belotserkovskii BP, Hanawalt PC. DNA slip-outs cause RNA polymerase II arrest in vitro: potential implications for genetic instability. Nucleic Acids Res
55. Lawlor KT, O'Keefe LV, Samaraweera SE, van Eyk CL, McLeod CJ, Maloney CA, et al (2011) Double-stranded RNA is pathogenic in Drosophila models of expanded repeat neurodegenerative diseases. Hum Mol Genet