

# Processing of double-R-loops in (CAG)·(CTG) and *C9orf72* (GGGGCC)·(GGCCCC) repeats causes instability

Kaalak Reddy<sup>1,2</sup>, Monika H.M. Schmidt<sup>1,2</sup>, Jaimie M. Geist<sup>1,3</sup>, Neha P. Thakkar<sup>1,2</sup>, Gagan B. Panigrahi<sup>1,2</sup>, Yuh-Hwa Wang<sup>4</sup> and Christopher E. Pearson<sup>1,2,\*</sup>

<sup>1</sup>Department of Genetics, The Hospital for Sick Children, Peter Gilgan Centre for Research & Learning, 686 Bay Street, Toronto, Ontario M5G 0A4, Canada, <sup>2</sup>Program of Molecular Genetics, University of Toronto, Toronto, Ontario M5G 0A4, Canada, <sup>3</sup>Department of Biology, Laurentian University, Sudbury, Ontario P3E 2C6, Canada and <sup>4</sup>Department of Biochemistry & Molecular Genetics, University of Virginia School of Medicine, Charlottesville, VA 22908, USA

Received November 12, 2013; Revised July 08, 2014; Accepted July 9, 2014

## ABSTRACT

R-loops, transcriptionally-induced RNA:DNA hybrids, occurring at repeat tracts (CTG)<sub>n</sub>, (CAG)<sub>n</sub>, (CGG)<sub>n</sub>, (CCG)<sub>n</sub> and (GAA)<sub>n</sub>, are associated with diseases including myotonic dystrophy, Huntington's disease, fragile X and Friedreich's ataxia. Many of these repeats are bidirectionally transcribed, allowing for single- and double-R-loop configurations, where either or both DNA strands may be RNA-bound. R-loops can trigger repeat instability at (CTG)·(CAG) repeats, but the mechanism of this is unclear. We demonstrate R-loop-mediated instability through processing of R-loops by HeLa and human neuron-like cell extracts. Double-R-loops induced greater instability than single-R-loops. Pre-treatment with RNase H only partially suppressed instability, supporting a model in which R-loops directly generate instability by aberrant processing, or via slipped-DNA formation upon RNA removal and its subsequent aberrant processing. Slipped-DNAs were observed to form following removal of the RNA from R-loops. Since transcriptionally-induced R-loops can occur in the absence of DNA replication, R-loop processing may be a source of repeat instability in the brain. Double-R-loop formation and processing to instability was extended to the expanded *C9orf72* (GGGGCC)·(GGCCCC) repeats, known to cause amyotrophic lateral sclerosis and frontotemporal dementia, providing the first suggestion through which these repeats may become unstable. These findings

provide a mechanistic basis for R-loop-mediated instability at disease-associated repeats.

## INTRODUCTION

R-loops are thermodynamically stable, RNA:DNA structures that may form during transcription. As nascent RNA is produced within the transcription bubble, short stretches of RNA:DNA hybrids are transiently formed with the parental DNA strand and lost with subsequent ejection from the transcribing RNA polymerase (1). At certain GC-rich DNA sequences, stable R-loops are formed and persist following passage of the RNA polymerase along the transcribed DNA template due to the high stability of rG-dC base pairs between RNA and DNA, preferentially occurring when G clusters are present in the transcript (2,3). Biological roles for R-loops have been described in DNA replication initiation at mitochondrial and prokaryotic origins of replication, in class switch recombination at immunoglobulin genes, and in DNA methylation regulation at CpG island sequences (4–7). In these cases, the formation of R-loops is a necessary event regulating essential downstream activities. However, R-loops can also act as mutagenic intermediates and can cause either gene-specific or genome-wide instability. For example, mutations in the THO/TREX complex, which is required for proper coupling of transcription and mRNA export, cause wide-scale co-transcriptional R-loop formation triggering aberrant recombination leading to genome-wide instability (8–10). In contrast, gene-specific R-loop-mediated mutagenesis may have specific sequence requirements, as described below.

Certain sequences prone to R-loop formation, such as at expanded (CTG)·(CAG) repeat tracts, can trigger instability in a gene-specific manner (11,12). Gene-specific instabil-

\*To whom correspondence should be addressed. Tel: +1 416-813-8256; Fax: +1 416-813-8170; Email: cepearson.sickkids@gmail.com

Present address:

Kaalak Reddy, MRC Human Genetics Unit, MRC IGMM, University of Edinburgh, Edinburgh, UK.

ity of trinucleotide repeat sequences is the causative mutation for various neurological, neuromuscular and neurodegenerative diseases (13). At least 14 diseases are caused by gene-specific expansions of (CAG)·(CTG) repeats including Huntington's disease, myotonic dystrophy type 1 and a series of spinocerebellar ataxias. Others include fragile X mental retardation involving (CGG)·(CCG) repeats and Friedreich's ataxia involving a (GAA)·(TTC) repeat (14). Recently, the leading genetic cause of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) has been linked to an expanded (GGGGCC)·(GGCCCC) repeat in the *C9orf72* gene which also shows length variation between tissues of a given individual (15). In many instances, these disease-associated repeats, including the *C9orf72* tract, are transcribed in either one or both directions (16,17). Evidence suggests that transcription across the repeat tract drives repeat instability (18–21). Notably, simultaneous bidirectional transcription can exacerbate genetic instability of a (CAG)·(CTG) repeat tract in human cells (21–23). Transcription across CAG, CTG, CGG, CCG and GAA repeats can lead to the formation of single-R-loops, where the RNA transcript is retained in a stable hybrid with the DNA and the non-template DNA strand is rendered single-stranded (3,11,24). We previously demonstrated the formation of novel double-R-loop structures in which both strands of a convergent bidirectionally transcribed (CAG)·(CTG) repeat tract or a (CGG)·(CCG) repeat tract contained RNA:DNA hybrids on both strands, termed double-R-loops (3). Double-R-loops may be associated with the increased instability observed during bidirectional transcription (22,23).

Transcriptionally-induced R-loops on expanded CAG repeat tracts was coincident with increased levels of trinucleotide repeat contractions. Knockdown of RNase H1 and H2 enzymes led to even greater contraction frequency in the repeat tract, further supporting the role of R-loops in instability (11). Interestingly, we found that upon removal of the RNA with RNase H, from both single- and double-R-loops, a portion of the DNA template remained single-stranded (3). Single-stranded regions of the repeat tract may assume slipped-DNA (S-DNA) structures formed from misalignment during re-annealing of the separated DNA strands, which may trigger instability, as previously hypothesized (11,25).

Although previous findings established a connection between transcription, R-loops and instability, there are important unanswered questions regarding the mechanistic basis of R-loop-mediated instability. For example, can R-loops be directly processed post-transcriptionally in the absence of DNA replication? This is an important question because instability occurs in various non-proliferative cells such as neurons (14). Also, since many (CAG)·(CTG) repeat tracts are transcribed in both directions, does the configuration of the R-loop (rCAG-containing R-loop, rCUG-containing R-loop or rCAG+rCUG-containing double-R-loops) influence the instability of the repeat tract? Finally, might R-loop-mediated instability involve an S-DNA intermediate produced through R-loop degradation by RNase H elimination of the RNA portion? Addressing these questions regarding R-loop processing has proven difficult because there are currently no suitable assays to di-

rectly elucidate the mechanism of R-loop processing upon (CAG)·(CTG) repeat instability.

In this study, we establish an *in vitro* R-loop processing assay using human cell extracts and demonstrate that R-loops and double-R-loops formed at an expanded (CAG)<sub>79</sub>·(CTG)<sub>79</sub> repeat tract can lead to repeat instability post-transcriptionally and in the absence of DNA replication. Similarly, we show that a (GGGGCC)<sub>n</sub>·(GGCCCC)<sub>n</sub> repeat tract from the *C9orf72* gene with expansion lengths ranging from 4 to 94 repeats was able to form R- and double-R-loops. We show that double-R-loops arising from simultaneous convergent bidirectional transcription, cause the greatest instability relative to either single-R-loop alone. Finally, R-loop degradation by RNase H only partially suppresses the observed instability likely due to the formation of S-DNA structures within the repeat tract. In support of this model, we also reveal that removal of the RNA entity from R-loops by RNase H can induce S-DNA formation. These findings shed light on the mechanistic basis for R-loop-mediated instability at disease-associated repeat tracts.

## MATERIALS AND METHODS

### *In vitro* transcription and RNase treatments to generate R-loop templates

Plasmids bearing an expanded (CAG)<sub>79</sub>·(CTG)<sub>79</sub> repeat tract with convergent T3 and T7 RNA polymerase promoters are described in detail previously (26). Large-scale plasmid preparations were prepared from dam+ *Escherichia coli* cells as described previously (25). Briefly, cells were harvested and lysed with lysozyme (Invitrogen) and a detergent solution of 1% Brij 58 (Sigma) and 0.4% deoxycholate (Sigma). Plasmids were treated with RNase A and T1 (Sigma), phenol-extracted and purified twice by cesium chloride/ethidium bromide centrifugation.

(GGGGCC)<sub>n</sub>·(GGCCCC)<sub>n</sub> clones, where  $n = 13, 21, 40$  or 60 repeats (provided by Tao Zu and Laura Ranum as previously described (27), were prepared and propagated in DH5α cells and prepared as above. Requests for these clones must be directed to Dr Ranum.

Transcription reactions were performed as described (28,29). Briefly, 500 ng of template DNA in 1x transcription buffer (Roche) and 1x bovine serum albumin (NEB) were mixed in a final volume of 100 μl for 1 h with 20 U of the appropriate RNA polymerase: T7, T3 or T7+T3 (Roche). Total nucleic acid material was subsequently extracted with phenol/chloroform, then chloroform extraction followed by precipitation with 100% ethanol and 3 mM sodium acetate. Samples were resuspended in 10 μl 1x TE for subsequent RNase treatments. For incorporation of radioactive nucleotides, each transcription reaction was additionally carried out in the presence of 3.5 μCi of [α-<sup>32</sup>P]rCTP. Samples were run through sephadex G-50 columns (GE Healthcare) prior to precipitation.

Samples were treated with either 1 μg of RNase A (Roche) alone or with 1 μg of RNase A (Roche) and 1 U of *E. coli* RNase H (Roche) in a final volume of 10 μl containing 1x NEB buffer #2 at room temperature for 30 min. Nucleic acid material was subsequently extracted as described above.

All *in vitro* transcription reaction products were analyzed on 1% agarose gels run in 1x Tris–Borate–ethylenediaminetetraacetic acid (EDTA) buffer at 80 V for 3 h. Gels were subsequently stained with ethidium bromide (0.5 mg/ml) to allow visualization of total nucleic acid under ultraviolet (UV) light. For samples containing radioactive isotopes, gels were dried and exposed to X-ray film (Kodak BioMax XAR).

### Human cell extract treatment

HeLa S3 cell line was purchased from National Cell Culture Center, National Center for Research Resources, National Institutes of Health (Bethesda). The human SH-SY5Y neuroblastoma cells were cultured and neuronally differentiated with 10  $\mu$ M all-*trans* retinoic acid (Sigma), with medium changes every 3 days as described (30,31). Whole cell extracts were prepared as described previously (32,33). Cell extracts prepared in this manner are functional in *in vitro* DNA replication, mismatch repair, excision repair (psoralens cross-links, cisplatin adducts, UV dimers, and pyrimidine dimers), double-strand break repair, homologous recombination, triplex-mediated recombination, processing of slipped-strand DNAs formed by CAG/CTG repeats (31,34–35), and are capable of inducing replication-mediated CAG/CTG expansions and contractions (26,36–37). R-loop templates prepared from *in vitro* transcription and RNase H treatments (described above) were incubated with  $\sim$ 32  $\mu$ g of extract in a 50  $\mu$ l volume containing 100  $\mu$ M each of deoxyadenosine triphosphate (dATP), dGTP, deoxythymidine triphosphate (dTTP), and deoxycytidine triphosphate (dCTP), guanosine triphosphate GTP, 200  $\mu$ M each of uridine triphosphate (UTP), and cytidine triphosphate (CTP), 4 mM of adenosine triphosphate (ATP), 40 mM of creatine phosphate (Roche Molecular Biochemicals) and 100  $\mu$ g/ml of creatine kinase (Roche Molecular Biochemicals). Reactions were carried out at 37°C for 4 h and subsequently stopped by adding 50  $\mu$ l of stop solution (2  $\mu$ g/ $\mu$ l proteinase K, 2% sodium dodecyl sulphate and 50 mM EDTA, pH 8.0) with further incubation for 60 min at 37°C. Nucleic acid material was subsequently extracted as described above. Samples were further purified using QIAquick enzyme clean-up kits as per manufacturer's instructions prior to transforming into bacteria for Stability of Trinucleotide Repeats by analysis of Individual Products (STRIP) analysis (see below). For *in vitro* DNA replication analysis, cell extract treatment was carried out in the same way as described above, but with the addition of 1  $\mu$ g of SV40 T-antigen (Chimerx) and 0.1 Ci of [ $\alpha$ -<sup>32</sup>P]dCTP.

### STRIP

The STRIP assay is described in detail previously (26,38). Briefly, products of human cell extract processing were transformed into *E. coli* XL1-MutS (Agilent) (Invitrogen). Individual bacterial colonies (each representing one processed template) were picked and cultured for a limited growth period (maximum of 6 h, 4–6 generations). During colony propagation, bacterial contribution was minimized for all clones by (i) having the (CTG)-(CAG) repeat in the stable orientation relative to the unidirectional bacterial

ColE1 origin of replication; (ii) the limited colony growth time ( $\leq$ 6 h); and (iii) bacterial strain selection. Bacterial culture conditions including choice of bacterial strain have been thoroughly investigated in order to minimize background instability during the STRIP assay and have been described in detail previously (26,38). Miniprep DNA was analyzed for changes in repeat length by analysis of the repeat-containing fragment on 4% polyacrylamide gels. Repeat lengths (CAG/CTG) were classed into one of three categories: '<79 repeats', '79 repeats' and '>79 repeats' representing contraction, stable and expansion products, respectively. Similar procedures were used for the *C9orf72* clones. The magnitudes of repeat length changes were determined by electrophoretic sizing of the repeat-containing fragments on 4% polyacrylamide gels relative to the starting length material and a known set of size markers.

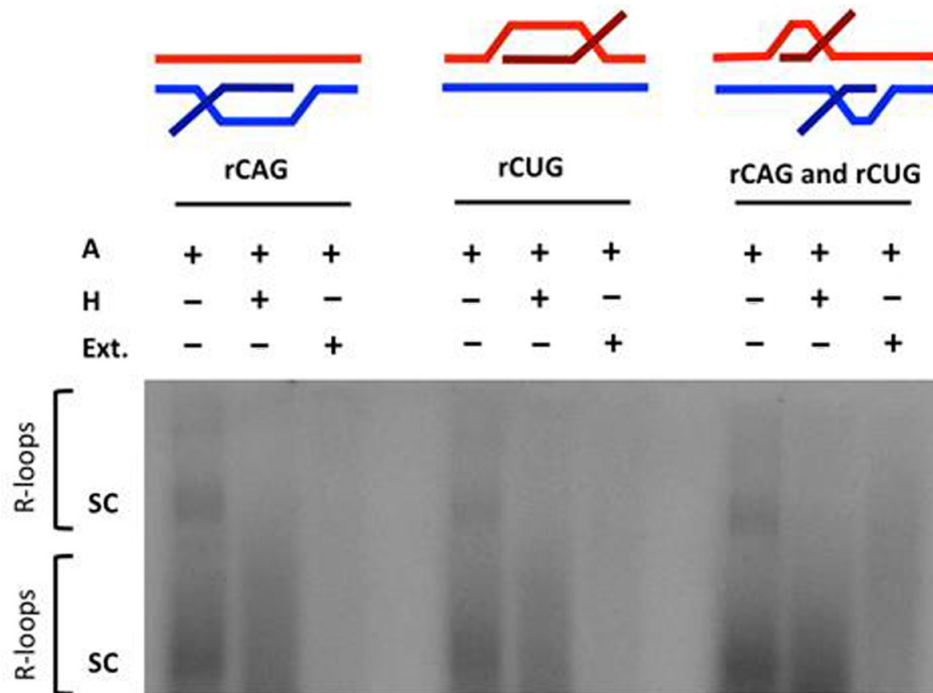
### Electron microscopy

RNase A and RNase H-treated transcription reaction products were de-proteinized and analyzed by electron microscopy (EM) as described previously (3). Briefly, binding reactions with bacterial single-strand binding protein (SSB) were carried out in a 50  $\mu$ l reaction mixture containing 8 mM NaCl, 20 mM HEPES (hydroxy ethyl piperazine ethanesulfonic acid) (pH 7.5) and 300 ng SSB for 10 min at room temperature. Complexes were fixed with 0.6% glutaraldehyde (v/v) for 10 min at room temperature, followed by filtration through a 2-ml column of Bio-Gel A5m (Bio-Rad) to remove excess glutaraldehyde and free proteins. Fractions containing DNA–SSB protein complexes were prepared for EM. Briefly, the indicated SSB–DNA complexes were mixed in a buffer containing 2 mM spermidine, adsorbed to glow-charged carbon-coated grids, washed with a water/graded ethanol series and rotary shadow cast with tungsten. Samples were examined using a Philips 420 electron microscope. Micrographs are shown in reverse contrast.

## RESULTS

### R-loops can be processed by human cell extracts

To investigate the role of R-loops in (CAG)-(CTG) repeat instability, we utilized human whole cell extracts (HeLa) that are proficient in *in vitro* DNA repair in an *in vitro* processing assay which we modified from previous studies of repeat instability in the lab (26,31,34–35). To assess the status of the R-loop following human cell extract treatment, we first generated R-loops through *in vitro* transcription with [ $\alpha$ -<sup>32</sup>P]rCTP-incorporation into the nascent RNA. Following treatment with RNase A (a ribonuclease that digests unbound, single-stranded RNA), there was a resistant autoradiographic signal consistent with R-loop formation as demonstrated previously (3) (Figure 1). Treatment of these products with both RNase A and RNase H (a ribonuclease that specifically degrades RNA that is bound to DNA as would occur in an R-loop) led to a reduction in the RNase A-resistant autoradiographic signal in the supercoiled forms of the DNA template consistent with digestion of R-loop material as demonstrated previously (3) (Figure 1; sc). However, we saw some signal persists indicat-



**Figure 1.** R-loop processing by human cell extract. *In vitro* transcription of a (CAG)<sub>79</sub>·(CTG)<sub>79</sub> repeat-containing plasmid with [ $\alpha$ -<sup>32</sup>P]rCTP was performed followed by RNase A treatment (to cleave single-stranded RNA); labeled 'A' or RNase H treatment (to also cleave RNA:DNA hybrids of the R-loop); labeled 'H' or human cell extract treatment; labeled 'Ext.' as indicated. The configuration of the R-loop generated is schematically represented above the gel. Autoradiographic signal in the gel represents R-loop formation. The position of supercoiled plasmid in dimer and monomer form is indicated by 'sc' where the top 'sc' represents linked dimers and bottom 'sc' represents monomers.

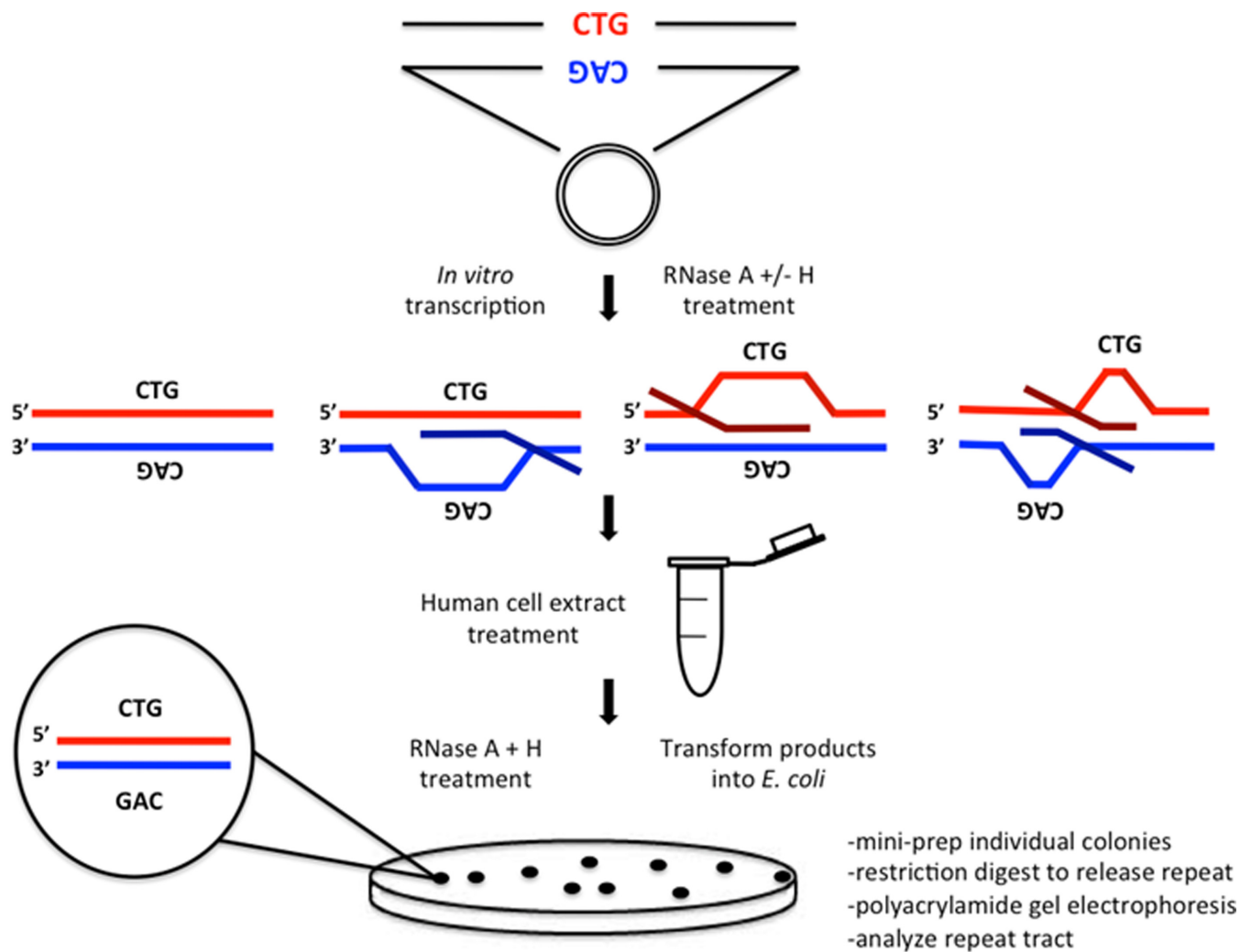
ing incomplete digestion of the products, which is consistent with the inability of RNase H to recognize and cleave RNA in an RNA:DNA hybrid shorter than four base pairs in length (39) as well as the extremely high sensitivity of autoradiography relative to ethidium bromide staining (compare Figures 1–3; A and A+H treatments). While R-loops are known to be favorably retained in supercoiled state (3), interestingly, this material that is partially resistant to the purified RNase H, is more stable in the monomeric rather than the concatameric forms. Upon treatment of the R-loops with HeLa cell extract, the autoradiographic RNA signal was largely reduced, suggesting that the HeLa extract is able to more completely process R-loop substrates, likely due to the presence of other factors (Figure 1).

To test whether the R-loops were acting as initiation sites for replication, where the RNA portion may serve as a primer for initiation by the HeLa cell extract, we analyzed dNTP radioincorporation into the R-loop-containing templates in the presence of HeLa cell extract (Supplementary Figure S1). Radioincorporation was compared to templates that did not contain R-loops, but were incubated with HeLa cell extract in the presence of T-antigen, which is able to initiate *in vitro* DNA replication from an SV40 origin of replication present in our repeat-containing templates (Supplementary Figure S1). We only observed high levels of dNTP radioincorporation that was *DpnI*-resistant, indicative of DNA replication, in the presence of T-antigen and not when R-loop templates were incubated with HeLa cell extract (Supplementary Figure S1). Thus, the R-loops do

not act as an initiation site for DNA replication in our system.

### Establishing a (CAG)·(CTG) repeat R-loop processing system

To assess whether R-loop processing might alter the DNA lengths of the (CAG)·(CTG) repeat tracts, we carried out a modified version of our previously reported STRIP protocol (Figure 2) (26,36–38). Essentially, the repeat tract lengths of single molecules were assessed in R-loop preparations that had been processed by HeLa cell extracts and compared to the same DNA templates devoid of R-loops that had also been processed by HeLa cell extracts. R-loop substrates for HeLa cell extract processing were generated through *in vitro* transcription across the repeat, initiated by T3 and/or T7 phage RNA polymerases, whose promoters flank the repeats, as previously described (Figure 3) (3). Single-stranded RNA was eliminated from R-loop preparations by treating transcription products with RNase A. Controls where the RNA portion was eliminated from the R-loop, were prepared by treatment with both RNase A+H (Figure 3). This preparation potentially contains aberrant S-DNAs upon R-loop removal, as noted above (see also (3)). A DNA template devoid of R-loops that had not been transcribed served as a non-R-loop and non-S-DNA control (Figure 3). These substrates were all treated with HeLa cell extract and repeat lengths were analyzed by STRIP (Figure 2).



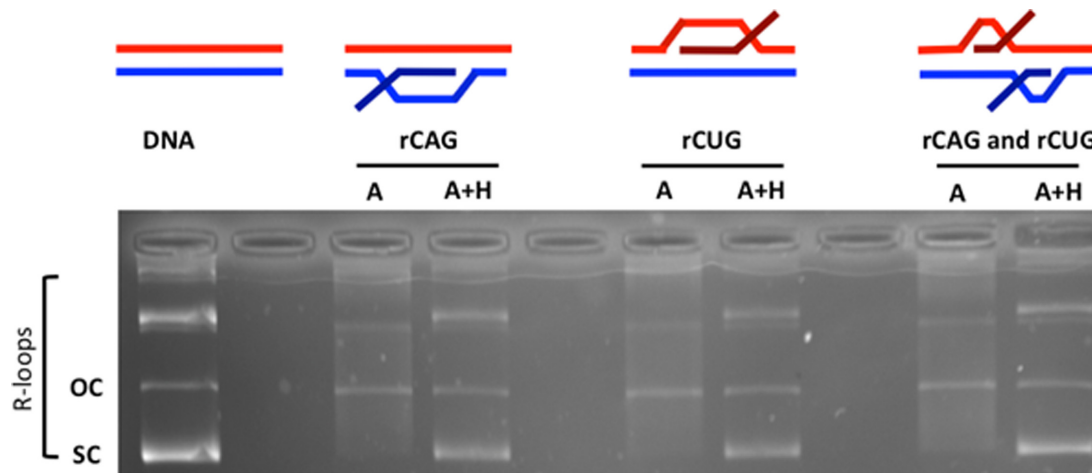
**Figure 2.** R-loop instability assay. Plasmids bearing an expanded  $(CAG)_{79} \cdot (CTG)_{79}$  repeat tract were in vitro transcribed and treated with the appropriate RNase to produce R-loop templates of each configuration as schematically depicted. R-loop templates were subsequently treated with human (HeLa) cell extract to allow processing to occur. Nucleic acid products were extracted and subjected to a final RNase A+H treatment to remove any residual R-loop products and transformed into *E. coli* bacteria and plated overnight. Individual colonies were picked (representing individual products of R-loop processing) and minimally cultured (see Materials and Methods). DNA was then extracted and restriction digested to release the repeat tract. Products were electrophoresed alongside known size markers to determine repeat length.

### Processing of double-R-loops generates increased instability

Varying levels of repeat instability were evident by electrophoretic resolution of repeat-containing fragments from individual molecules, exhibited by slower-migrating repeat-containing fragments (representing repeat tract expansions) as well as by faster-migrating fragments (representing repeat tract contractions) (Figure 4, each lane represents one individual molecule following HeLa extract treatment). The levels of instability appeared to depend upon the substrate configuration. An aliquot of the starting parental template with  $(CAG)_{79} \cdot (CTG)_{79}$  repeats that had not been transcribed nor subjected to cell extract treatment or STRIP, was loaded in the first lane of each gel to serve as a marker to compare products of instability against (Figure 4, indicated by arrow). The non-transcribed DNA template that was treated with human cell extract and subjected to STRIP served as the cell extract processing control to assess basal

levels of length heterogeneity present in the starting material (instability during preparation in bacteria), as well as any instability that may be incurred by exposure of the fully-paired DNA repeat to the HeLa extract (Figure 4). Some level of length heterogeneity is expected for the DNA template due to its unstable length of 79 repeat units, resulting from processing of endogenous DNA damage (potentially including single strand breaks, oxidative damage, nucleotide mismatches, etc.) by human cell extract, repeat length heterogeneity present in the starting plasmid, as well as from bacterial culturing during the STRIP procedure (26,38). The tract length heterogeneity in this starting template serves as the background level of tract length instability above which any potential R-loop processing instability must rise.

The various R-loop substrates generated increased levels of repeat instability following HeLa extract treatment relative to the control DNA template (Figure 4 shows rep-



**Figure 3.** R-loop formation in  $(CAG)_{79} \cdot (CTG)_{79}$  templates. Templates were *in vitro* transcribed with T3 and/or T7 RNA polymerases and treated with RNase A (which digests single-stranded RNA) to form each R-loop configuration indicated schematically above the gel. The presence of R-loops forces the plasmid into a more open configuration, thus reducing electrophoretic migration within the gel. Treatment of the R-loop with RNase H cleaves RNA that is base-paired to DNA (the RNA:DNA hybrid) and thus collapses the R-loop, returning DNA to supercoiled form. The position of supercoiled plasmid is indicated as 'sc' and open circular plasmid as 'oc'. Products above these are catenated multimers, which also form R-loops.

representative gels). We quantified the differences in instability resulting from each R-loop processing reaction by assessing the repeat sizes of a large set of single molecules, compared to known size markers. Products were categorized as 'stable' (having 79 repeats), 'contractions' (having fewer than 79 repeats) or 'expansions' (having greater than 79 repeats) (Figure 5). There was a significant increase in instability from processing of R-loops formed by simultaneous convergent transcription relative to the DNA control template that had not been transcribed (Figure 5A,  $P = 0.0345$  by  $\chi^2$  analysis). There was a modest but consistent increase in instability arising from HeLa cell extract processing of R-loops from either direction alone, but the differences were not significant compared to DNA. It is noteworthy that instability arising from rCAG R-loop processing was consistently higher than instability arising from rCUG R-loop processing (Figure 5A). This observation reflects upon previous findings that rCAG R-loops exhibit increased formation/stability than rCUG R-loops (3,11).

To ensure that the differences in instability we observed were a result of HeLa extract processing and not from culturing in bacteria, we performed STRIP analysis as described above, but transformed the products of *in vitro* transcription following RNase A+H treatment into the bacteria without HeLa cell extract treatment (Supplementary Figure S2). There were no significant differences in instability between DNA control and R-loop templates without HeLa extract treatment (Supplementary Figure S2).

### Distribution of products from R-loop processing

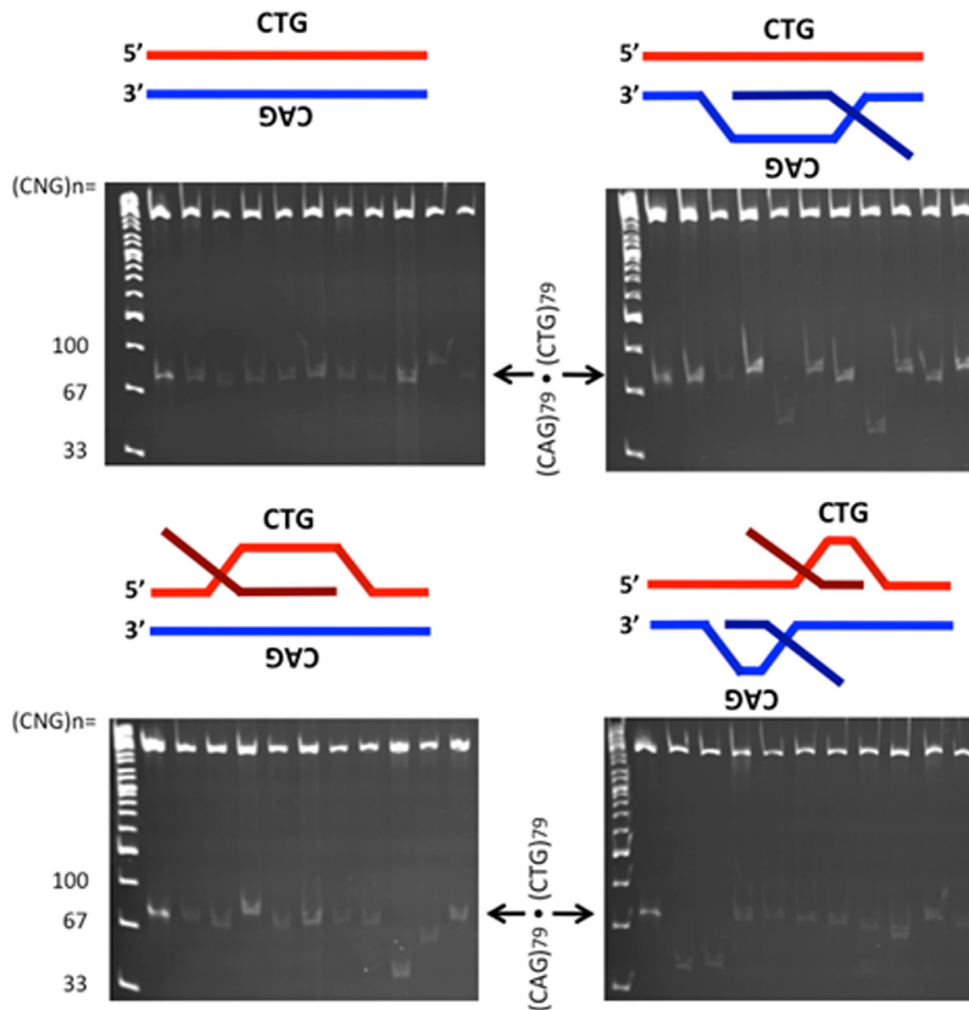
Repeat instability can vary both in mutation direction (expansion versus contraction) and magnitude of change. The unstable products of HeLa cell extract processing from the various R-loop configurations showed increases in expansions and contractions compared to the DNA control (Figure 5B). Processing of rCAG+rCUG double-R-loops resulted in a significant increase in contractions compared

to control DNA (Figure 5B;  $P = 0.0125$  by  $\chi^2$  analysis). Processing of rCUG and rCAG+rCUG double-R-loops yielded more contractions than expansions (Figure 5B). On the other hand, processing of rCAG R-loops yielded more expansions than contractions (Figure 5B). Thus, R-loop configuration can influence the direction of instability (expansion versus contraction).

The magnitudes of repeat length changes for each R-loop processing are plotted in Figure 5C. Most of the mutant products showed small changes, relative to the starting template size distribution in the  $(CAG)_{79} \cdot (CTG)_{79}$  preparation. There were fewer products with repeat numbers increasing or decreasing from this starting parental length (Figure 5C). The products from DNA control processing reactions displayed a range of repeat sizes from 31 to 100 units. The range of products in the R-loop-processed reactions was broader, from as few as 8 repeats (rCAG+rCUG processing) to as many as 127 repeats (rCUG processing). Repeat sizes of all products from human cell extract processing are listed in Supplementary Table S1.

### Neuron-like extracts yield instability through double-R-loop processing

Somatic expansions occur in the central nervous system of patients suffering CAG/CTG expansions and mouse models of these diseases (14,40). These ongoing mutations may exacerbate age-of-onset, disease severity and progression. Neural differentiation is associated with changes in repair capacity (41). Using extracts of the non-proliferating human SH-SY5Y cells differentiated into neuron-like cells (Figure 6A) (30,31), we assessed whether R-loop processing could alter repeat lengths. The products of SH-SY5Y extract-processed double-R-loops showed moderate, yet significantly increased levels of instability compared to the mock-treated DNA (Figure 6B,  $P = 0.00276$ , by  $\chi^2$  analysis). Processing of double-R-loops that have had the RNA component removed (RNase A+H) did not alter levels of



**Figure 4.** Products of R-loop processing by gel analysis. R-loop templates were formed by *in vitro* transcription; the DNA control template was not transcribed, while the other templates were transcribed and treated with RNase A to remove single-stranded RNA but retain the R-loop configuration as indicated schematically above each gel. These templates were treated with HeLa cell extract. Products of HeLa extract processing were extracted, treated with RNase A+H to remove residual R-loops and transformed into *E. coli*. Following minimal culturing, products of each R-loop configuration were resolved on polyacrylamide gels alongside known size markers. The product immediately adjacent to the ladder lane in each gel (indicated by arrow) is the untranscribed, unprocessed parental DNA template that serves as a size marker containing 79 (CAG)<sub>79</sub>·(CTG)<sub>79</sub> repeats.

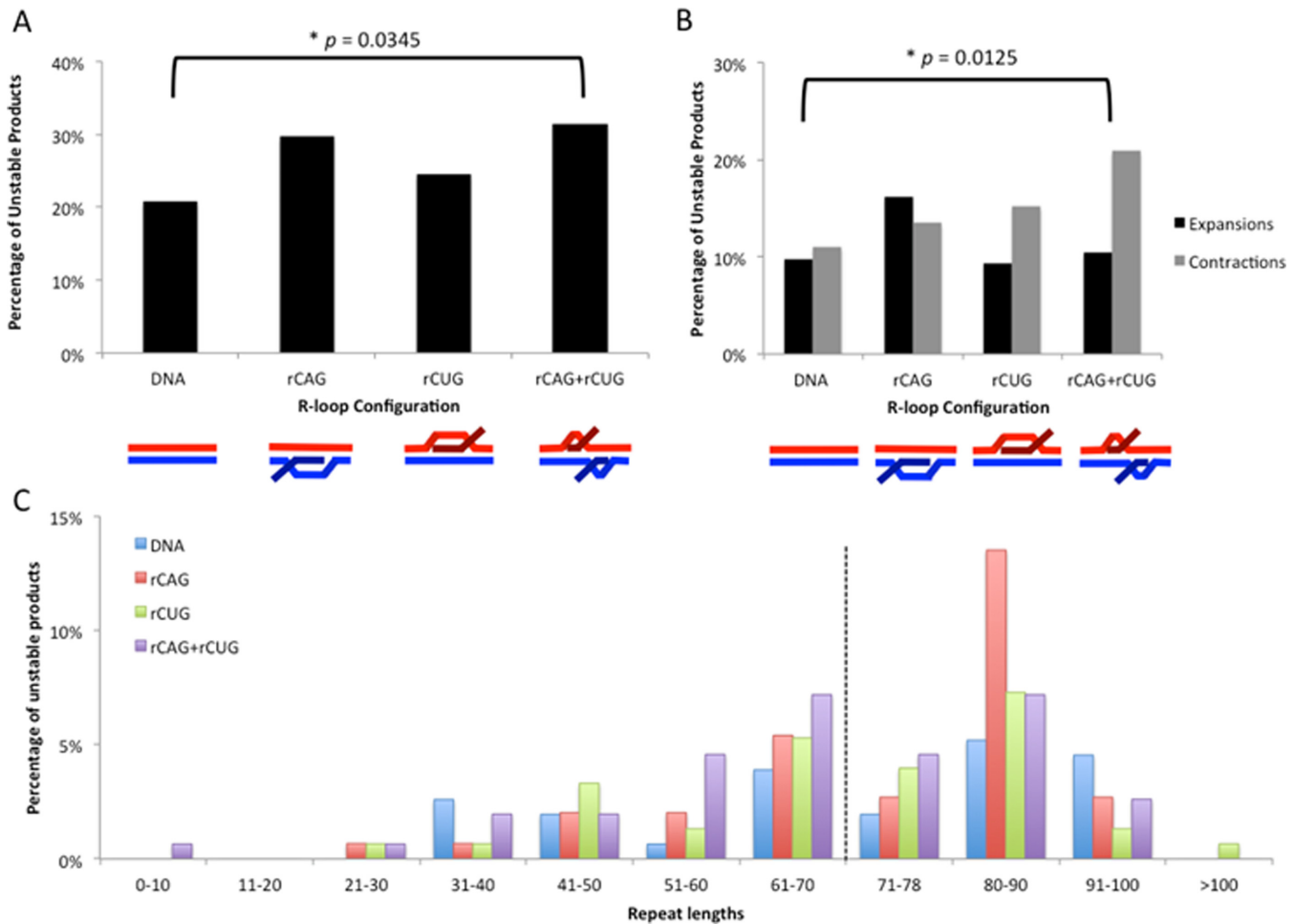
instability compared to the control DNA, but was significantly different from processing of the intact double-R-loop (Figure 6A,  $P = 0.691$ ,  $P = 0.00229$ , respectively by  $\chi^2$  analysis). Processing of intact double-R-loops resulted in a significant increase in contractions compared to control DNA (Figure 6C;  $P = 0.000682052$  by  $\chi^2$  analysis). Thus, double-R-loop processing by a neuron-like cell extract can lead to repeat instability.

The magnitudes of repeat length changes for double-R-loop processing are plotted in Figure 6D, see also Supplementary Table S3. Most of the mutant products showed small changes, relative to the starting template size distribution in the (CAG)<sub>79</sub>·(CTG)<sub>79</sub> preparation. The products from DNA control SH-SY5Y-processing reactions displayed a wide range of repeat sizes, from 20 to 131 repeat units, with a distribution similar to the starting material. The size distribution of the DNA products of SH-SY5Y-treated double-R-loops was significantly different prior to removal of the RNA portion (RNase A,  $P = 0.00046$ ,

Wilcoxon signed rank test), and to a lesser degree following its removal (RNase A+H,  $P = 0.02642$ , Wilcoxon signed rank test). These results further support the suggestion double-R-loop processing by a neuron-like cell extract can lead to repeat instability.

#### RNase H treatment partially suppresses increased instability from R-loop processing

Removal of the RNA component of the R-loop can alter the structure of the remaining DNA template, such that homo-duplex S-DNAs may arise following out-of-register reannealing of the complementary DNA strands (3). Using EM, we analyzed individual molecules following R-loop removal with RNase H (Figure 7A). There was an increase in the number of molecules bound by SSB following R-loop removal compared to DNA control samples that were not transcribed (Figure 7A). Levels of S-DNAs were similar for each R-loop configuration. The binding of SSB to



**Figure 5.** Instability analysis of products from R-loop processing by human cell extract. (A) Percentage of total unstable products following processing. Products were characterized as either stable (having 79 repeats) or unstable (having fewer than or greater than 79 repeats), based on electrophoretic migration and plotted. Data are derived from three independent *in vitro* transcription and human cell extract processing reactions with ~150 colonies (~50 colonies per replicate) representing 150 individual products of cell extract treatment analyzed for each R-loop configuration. Individual experiments were compared with each other within a triplicate using the  $\chi^2$  test to ensure there were no significant differences between experiments and then data were pooled for each experimental condition. Specific colony numbers are as follows: DNA-154, rCAG-148, rCUG-151, rCAG+rCUG-153. Products of R-loop processing were compared to the DNA control processing products using the  $\chi^2$  test. (B) Percentage of contractions and expansions from processing. Unstable products were further separated into contractions (fewer than 79 repeats) and expansions (greater than 79 repeats) and plotted. The distribution of contractions and expansions were compared between R-loop products and DNA control products using the  $\chi^2$  test. (C) Distribution of unstable products of R-loop processing. Sizes were estimated for each unstable product of processing from electrophoretic migration position relative to known size markers as previously described (26) and plotted. Only unstable products are shown; the stable repeat size of 79 is indicated by the dashed vertical line.

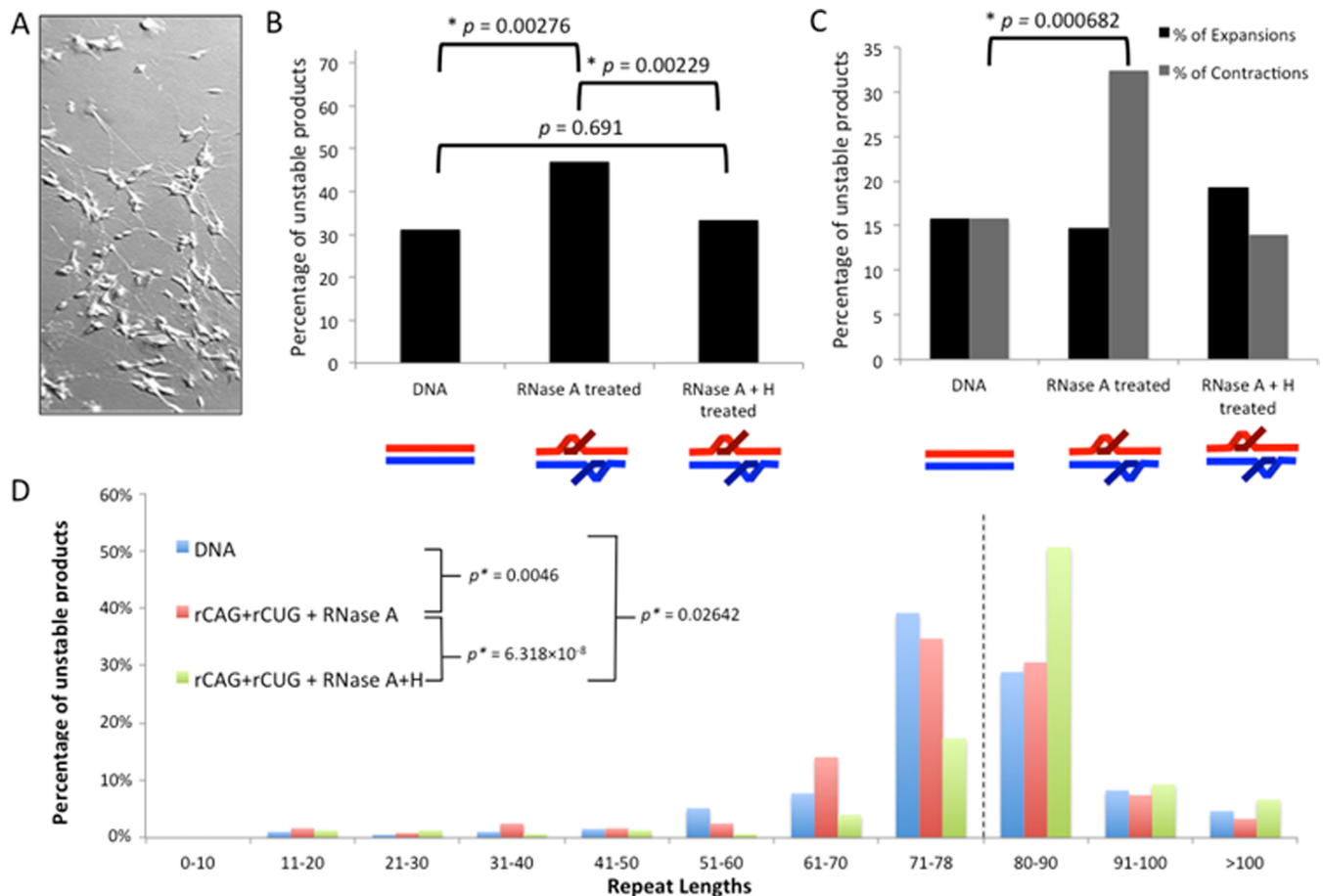
DNA templates following R-loop removal supports the formation of S-DNA structures that are unpaired from their complementary DNA strand and contain regions of single-strandedness detectable by SSB binding. As previously reported, slip-out sizes were too small to detect in the absence of SSB (42).

To further support the formation of S-DNA structures following RNase H removal, templates were additionally treated with mung bean nuclease, which degrades single-stranded DNA, as would be present in slipped-structures following RNA:DNA hybrid removal with RNase H (Supplementary Figure S3). Visualization by EM revealed that there was an increase in the number of molecules bound by SSB following R-loop removal with RNase H compared to a DNA template control and a reduction in the number of molecules bound by SSB following mung bean nuclease

treatment as would be expected due to the formation of S-DNA structures following R-loop removal with RNase H (Supplementary Figure S3).

We next assessed whether the DNA remaining after RNA elimination from the R-loop can lead to instability through processing by HeLa cell extracts. We analyzed the instability arising from R-loops that were treated with RNase H prior to treatment with HeLa cell extract (Figure 7B). For each R-loop configuration, treatment with RNase H prior to HeLa extract treatment consistently reduced the overall instability (Figure 7B). Interestingly, the RNase H-induced reduction in instability was only partial relative to DNA control and instability was not significantly less than R-loop preparations that were not treated with RNase H prior to processing by HeLa extract (Figure 7B). Furthermore, analysis of the individual unstable products resulting



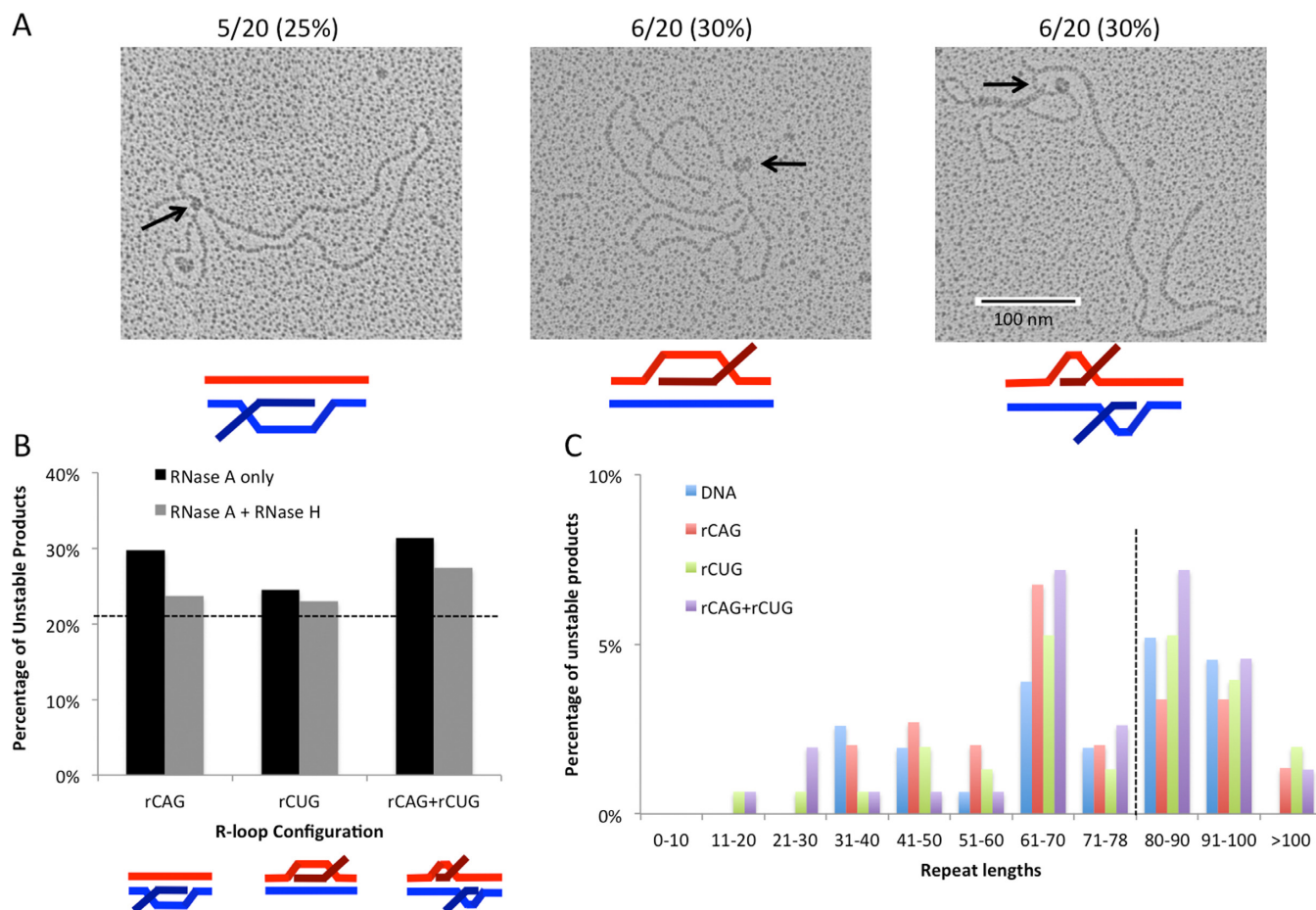


**Figure 6.** Instability analysis of products from double-R-loop processing by neuron cell extract. (A) Retinoic acid-differentiated SH-SY5Y cells. (B) Percentage of total unstable products following processing. Products were characterized as either stable (having 79 repeats) or unstable (having fewer than or greater than 79 repeats), based on electrophoretic migration and plotted. Data are derived from three independent *in vitro* transcription and retinoic acid human SH-SY5Y cell extract processing reactions with ~150 colonies (~50 colonies per replicate) representing 150 individual products of cell extract treatment analyzed for each R-loop configuration. Individual experiments were compared with each other within a triplicate using the  $\chi^2$  test to ensure there were no significant differences between experiments and then data were pooled for each experimental condition. Specific colony numbers are as follows: DNA-194; rCAG+rCUG+RNase A-121; rCAG+rCUG+RNase A+H-149. Products of R-loop processing were compared to the DNA control processing products using the  $\chi^2$  test. (C) Percentage of contractions and expansions from processing. Unstable products were further separated into contractions (fewer than 79 repeats) and expansions (greater than 79 repeats) and plotted. The distribution of contractions and expansions were compared between R-loop products and DNA control products using the  $\chi^2$  test. (D) Distribution of unstable products of R-loop processing. Sizes were estimated for each unstable product of processing from electrophoretic migration position relative to known size markers as previously described (26) and plotted. Only unstable products are shown; the stable repeat size of 79 is indicated by the dashed vertical line.

from RNase H-treated samples revealed a similar distribution as observed with processing of R-loop samples without RNase H-treatment (compare Figures 5C–7C, and Supplementary Tables S1–S2). These products also showed a broader length distribution than products from the DNA control, which did not contain repeats shorter than 31 units or longer than 100 units (Figure 7C). These observations were most notable for double-R-loop processing following R-loop removal with RNase H (Figure 7). Thus, RNA degradation from the R-loop partially suppresses instability evident from the modest but consistent decrease in instability upon RNase H treatment; however, R-loop removal with RNase H still triggers instability, which may occur through the formation of S-DNA structures (as evidenced by EM) through misaligned reannealing of the repeat tract.

### R-loop formation in *C9orf72* repeats, processing and instability

To assess whether R-loops may form at the expanded *C9orf72* (GGGGCC)<sub>n</sub>-(GGCCCC)<sub>n</sub> repeats, we used a series of clones with tract lengths ranging from 13 to 60 repeats. Transcription of a tract with 60 repeats in either direction producing r(GGCCCC) or r(GGGGCC) leads to R-loop formation, resistant to RNase A and sensitive to RNase H (Figure 8A), consistent with R-loop formation (28,29). Transcription through the C-rich template strand produced slightly more hybrids compared to the G-rich template strand. Simultaneous transcription across both strands also led to double-R-loops. To assess a potential length effect, we used tracts with 13, 21 and 60 repeats; these too formed R-loops and double-R-loops, with little variation in amounts of R-loops formed (Figure 8B). Thus, this

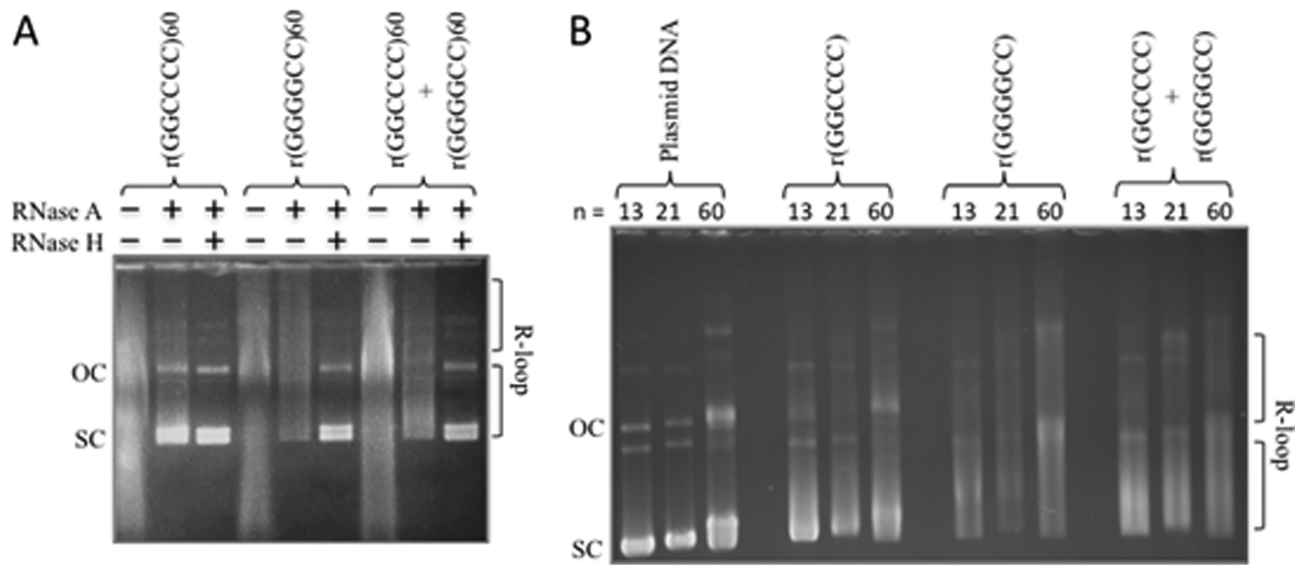


**Figure 7.** Instability analysis following R-loop removal. (A) Detection of S-DNA structures following R-loop removal with RNase H. DNA templates were transcribed to generate r(CAG), r(CUG) or r(CAG)+r(CUG) R-loops and then treated with RNase A and RNase H to remove all single-stranded RNA and RNA:DNA hybrids. EM was performed on individual molecules in the presence of bacterial SSB to detect unpaired DNA strands that exist in S-DNA structures (see ‘Materials and Methods’). A total of 20 molecules were analyzed for each sample type. DNA controls that were not transcribed contained 2/20 (10%) molecules bound by SSB. The number of samples bound by SSB at a single position following R-loop removal is indicated above each image and expressed as a percentage. (B) Percentage of unstable products following processing of RNase H-treated R-loops. R-loop products of each configuration were treated with RNase H prior to cell extract processing and assessed for instability through STRIP analysis as in (A). Products from (A) (RNase A only) were compared to RNase H-treated R-loops (RNase A+H) using the  $\chi^2$  test. Data for RNase H-treated R-loop processing are derived from three independent *in vitro* transcription and human cell extract processing reactions with  $\sim 150$  colonies representing 150 individual products of cell extract treatment for each RNase H-treated R-loop configuration. Specific colony numbers are as follows: rCAG-148, rCUG-152, rCAG+rCUG-153. Dashed line indicates DNA control level of instability (21%) for comparison. (C) Distribution of unstable products of HeLa extract processing following RNase-H-mediated R-loop removal. Sizes were determined for each unstable product of processing from electrophoretic migration position relative to known size markers as previously described (26) and plotted. Only unstable products are shown; the stable repeat size of 79 is indicated by the dashed vertical line.

ALS/FTD disease-associated hexanucleotide repeat tract can also form biophysically stable single- and double-R-loop structures.

To test whether processing of these *C9orf72* R-loops might alter repeat tract lengths, we prepared double-R-loops from a plasmid with 40 repeats, treated these with RNase A or RNase A+H and subjected these to SH-SY5Y cell extracts. These repeats were extremely unstable in bacterial preparations, showing extremely high levels of unavoidable length heterogeneity, all of which are contractions of the starting length (Figure 9A, see titrated DNA concentration in first four lanes of each gel). Due to this high level of variation in the starting preparation, we assessed length distributions as a measure of instability. A sample of the individual products is shown in Figure 9A (see also Supplemen-

tary Table S4). The products from DNA control SH-SY5Y-processing reactions displayed a broad range of repeat sizes from 4 to 64 units, similar to the starting material. Products in the R-loop-processed reactions ranged from 5 to 69 repeats. However, the size distribution of the DNA products of SH-SY5Y-treated double-R-loops was significantly different prior to removal of the RNA portion [DNA versus RNase A,  $P = 0.00715$ , confidence interval (CI) 95%, Wilcoxon signed rank test]. Interestingly, following removal of the RNA from the double-R-loop, treatment with SH-SY5Y extracts led to a significantly different size distribution of the DNA products (DNA versus RNase A+H,  $P = 0.000362$ , CI 95%, Wilcoxon signed rank test), which differed significantly from the intact double-R-loop (RNase A+H versus RNase A,  $P = 0.00631$ , CI 95%, Wilcoxon



**Figure 8.** *C9orf72* repeat forms R-loops and double-R-loops. (A) Templates with (GGGGCC)<sub>60</sub>·(GGCCCC)<sub>60</sub> were *in vitro* transcribed with T3 and/or T7 RNA polymerases and treated with RNase A (which digests single-stranded RNA) to form each R-loop configuration indicated schematically above the gel. The presence of R-loops forces the plasmid into a more open configuration, thus reducing electrophoretic migration within the gel. Treatment of the R-loop with RNase H cleaves RNA that is base-paired to DNA (the RNA:DNA hybrid) and thus collapses the R-loop, returning DNA to supercoiled form. The slower migrating products above these are catenated multimers, which also form R-loops. (B) Templates with 13, 21 or 60 *C9orf72* repeats were transcribed as in panel (A) to reveal single and double-R-loop formation.

signed rank test). These results provide the first possible mechanism through which the *C9orf72* repeats may be unstable and further support the suggestion double-R-loop processing by a neuron-like cell extract can lead to repeat instability. These results also suggest that processing of a *C9orf72* double-R-loop that has had its RNA component removed, can also lead to instability.

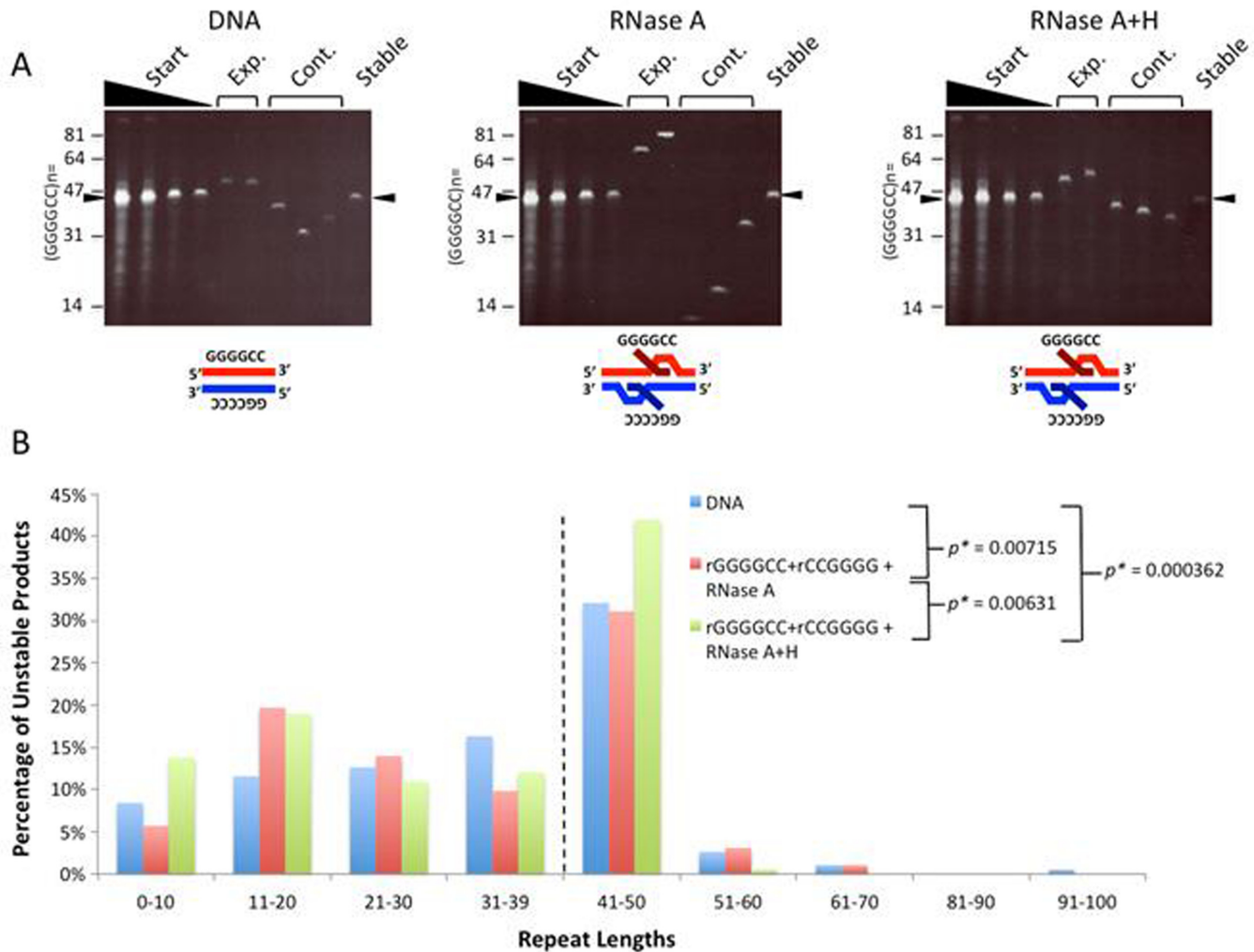
## DISCUSSION

In this study, we have developed a repeat R-loop processing assay to assess how such processing might contribute to repeat instability. We show that R-loops generated by *in vitro* transcription of an expanded (CAG)<sub>79</sub>·(CTG)<sub>79</sub> DNA repeat tract as well as an expanded (GGGGCC)<sub>40</sub>·(GGCCCC)<sub>40</sub> can be processed by human cell extract to increase instability of the repeat tract. Specifically, we demonstrate that (i) there is a significant increase in instability ( $P < 0.05$ ) following double-R-loop processing, but only a modest increase following single-R-loop processing, which was not found to be statistically significant; (ii) the range of repeat length changes is broader as a result of R-loop processing compared to DNA controls; (iii) RNase H treatment only partially suppresses R-loop-mediated instability; (iv) slipped-DNAs can be formed following degradation of the RNA from R-loops; and (v) double-R-loop processing by extracts of a non-proliferating neuronally-differentiated cell line can also lead to instability.

The mechanism through which R-loops cause disease-associated repeat instability is poorly understood. Some studies on the role of R-loops in genome-wide instability support a mechanism involving DNA replication fork stalling in the presence of R-loops leading to aberrant recombination (43,44). In the case of (CAG)·(CTG) instabil-

ity however, transcription is able to generate instability independent of DNA replication (19), suggesting direct engagement of repair factors by secondary structures generated during or following transcription (3,11). This model of aberrant DNA repair is consistent with instability arising in post-mitotic cells like neurons, which in many diseases display high levels of repeat instability (14). However, the idea that aberrant R-loop processing could lead to repeat instability had not been previously tested. In this study, we demonstrate that transcriptionally formed R-loop structures can directly trigger instability through human cell extract processing. Thus, genome maintenance alone may account for R-loop-mediated instability in non-replicating tissues. However, it should be noted that although not required, DNA replication may further exacerbate R-loop-associated (CAG)·(CTG) instability, e.g. if transcription and replication machinery were to collide at expanded (CAG)·(CTG) repeat tracts. A recent study found that many long transcribed human genes were susceptible to DNA breaks leading to genetic instability at common fragile site sequences where transcription and DNA replication machinery overlap (45). RNase H1 was found to suppress this common fragile site instability supporting a role for R-loops in the instability process (45). A similar situation may arise during simultaneous transcription and DNA replication of expanded (CAG)·(CTG) repeats, which are prone to R-loop formation. While beyond the scope of the current study, the effects of concomitant DNA replication and transcription in (CAG)·(CTG) repeat instability warrant further investigation.

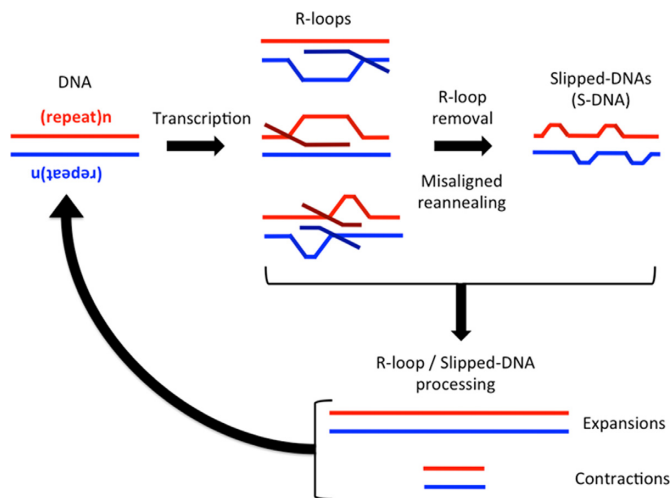
Expansion of the *C9orf72* (GGGGCC)·(GGCCCC) repeat has recently been demonstrated to be the leading cause of ALS/FTD. Individuals with *C9orf72* expansions show length variation between tissues of a given individual (15).



**Figure 9.** *C9orf72* double-R-loops when processed by neuron cell extracts lead to instability. (A) Products of double-R-loop processing by gel analysis. DNA control template with (GGGGCC)<sub>40</sub>·(GGCCCC)<sub>40</sub> was not transcribed, while the other templates were transcribed and treated with RNase A to remove single-stranded RNA, but retain the R-loop configuration as indicated schematically above each gel. These templates were treated with SH-SY5Y cell extract. Products of SH-SY5Y extract processing were extracted, treated with RNase A+H to remove residual R-loops and transformed into *E. coli*. Following minimal culturing, products of each R-loop configuration were resolved on polyacrylamide gels alongside known size markers. For each gel, the first four lanes show a titration of DNA concentration of the starting DNA, to reveal repeat length heterogeneity. A sample of products with lengths larger, smaller or similar to the starting lengths are shown. Specific colony numbers are as follows: DNA-168; rGGGGCC+rGGCCCC+RNase A-167; rGGGGCC+rGGCCCC+RNase A+H-157. The arrowhead indicates the size of the major starting length of 40 repeats. (B) Distribution of unstable products of R-loop processing. Sizes were estimated for each unstable product of processing from electrophoretic migration position relative to known size markers as previously described (26) and plotted. Only unstable products are shown; the starting major repeat size of 40 is indicated by the dashed vertical line.

The mechanism of *C9orf72* repeat expansions has yet to be studied. Interestingly, the *C9orf72* tract, is transcribed in either one or both directions (17). In all cases of repeat instability, an unusual DNA structure is believed to be involved. Here we present the novel finding that the *C9orf72* repeats can form both single- and double-R-loop structures. We also show that processing of the double-R-loop structure can lead to repeat instability, providing the first suggestion of a possible mechanism of *C9orf72* instability. The *C9orf72* tract is bidirectionally transcribed (17). We, and others recently demonstrated that the G-rich RNA strand can assume a G-quadruplex structure (28,46–47). Preliminary experiments suggest that the G-rich and C-rich strands of the *C9orf72* repeat can assume G-quadruplex and *i-*

*motif* structures (Zamiri B, Macgregor and Pearson, unpublished data). Interestingly, herein we show that transcription across the G-rich strand yields more R-loops than across the C-rich strand (Figure 8). This may be due to R-loop stabilization by a potential G-quadruplex structure formed by the non-transcribe G-rich DNA strand (48), and/or due to the increased biophysical stability of R-loops formed with G-clusters in the transcript (2). We previously observed R-loop formation following transcription across either the (CGG)<sub>n</sub> or the (CCG)<sub>n</sub> repeats of the FRAXA locus, as well as double-R-loops for both(26). It is noteworthy that we found production of a rCGG transcript yielded more R-loops than production of the rCCG transcript, similar to our finding herein that the rGGGGCC transcript led to



**Figure 10.** Proposed model for R-loop-mediated instability. Expanded (CAG)<sub>n</sub>-(CTG)<sub>n</sub> repeats are transcribed in a bidirectional manner generating rCAG-containing R-loops, rCUG-containing R-loops or double-R-loops. These products are then either directly processed by various repair proteins to generate instability or are first processed to remove the R-loop. R-loop removal may cause misaligned reannealing of the two DNA strands to generate S-DNA, which is then further processed to generate instability. This process of transcription-induced R-loop formation, R-loop processing to instability can cycle back again and again, in non-proliferating tissues, thereby continually amplifying the instability.

more R-loops than for the rGGCCCC transcript. Our results suggest that transcription may drive *C9orf72* repeat instability as it does for trinucleotide repeats (21–23).

Repeat expansions associated with numerous neurological and neuromuscular diseases show high levels of expansions in non-proliferating tissues including specific brain regions (14,40). We found that R-loop-processing-mediated repeat instability was independent of DNA replication. Since R-loop formation occurs via transcription, which can arise in the absence of DNA replication, it is likely that instability mediated by R-loop processing can lead to instability in non-proliferating tissues and cells, such as the brain (Figure 10). That slipped S-DNAs can be an intermediate of R-loop processing via the degradation of the RNA component, suggests that human patient tissues displaying high levels of repeat instability might harbor S-DNAs at the expanded repeats. In fact, recent evidence revealed the presence of slipped-DNAs at the expanded DM1 locus in various DM1 patient tissues, where the levels of S-DNAs correlated with the levels of instability (49). Multiple cycles of transcription, R-loop formation and processing could lead to the accumulation of high levels of instability (Figure 10).

Many disease-associated repeat loci are known to be transcribed in both directions across the repeat, including the DM1, HD, SCA8, FRAXA, FRAXE and *C9orf72* loci (16). The effect of transcription direction on R-loop formation may thus be an important determinant of somatic instability based on tissue-specific differences in transcription levels and directions. Previous studies reported higher levels of instability during simultaneous convergent bidirectional transcription relative to either direction alone potentially through R-loop formation (21–23). Although the basis for R-loop-mediated instability was poorly understood,

the detection of double-R-loop structures from simultaneous bidirectionally transcribed repeat templates supports a mechanism in which direct processing of these double-R-loops may generate the instability (3). We observed the highest levels of instability following processing of double-R-loops with human neuron-like cell extract relative to single-R-loops. Double-R-loop structures may be susceptible to reduced correct repair leading to instability potentially through error-prone repair as previously demonstrated for S-DNAs with repeat slip-outs (31,34). In the case of S-DNAs, there was asymmetry in repair efficiency depending on whether the slip-out sequence was CTG or CAG (where CAG slip-outs were repaired more efficiently than CTG slip-outs) (31). We and others have observed asymmetry in the formation of R-loops where rCAG R-loops form at a greater proportion than rCUG R-loops (3,11). In this study, we observe a corresponding asymmetry in R-loop-mediated instability, where rCAG R-loops trigger more instability (with a modest but consistent bias for expansions) than rCUG R-loops (with a modest but consistent bias for contractions) and rCAG+rCUG R-loops generate the greatest instability (with a significant bias for contractions). Thus, aberrant error-prone repair leading to expansions or contractions may be influenced by the structure of the slipped-out DNA formed in the R-loop, which is in turn dependent upon the direction in which the sequence is transcribed and whether it is simultaneously, bidirectionally transcribed.

It is important to determine which proteins respond to and interact with R-loops and how this generates repeat instability. One cellular factor known to be involved in trinucleotide repeat R-loop processing is RNase H (11). Lin *et al.* (11) demonstrated an increase in transcription-mediated CAG instability in human cells following shRNA knockdown of RNase H1 and H2 (the major RNase H enzymes in eukaryotes). Our results further support this finding as treatment of R-loops with RNase H prior to HeLa cell extract treatment reduced the levels of instability. Interestingly, this reduction in instability was partial relative to a non-R-loop DNA control. We found that R-loop removal with RNase H can enhance the formation of single-stranded regions in the DNA that have not properly reannealed to their complementary strand. These regions can form slipped S-DNA structures that have the potential to induce aberrant processing leading to instability, a suggestion further supported by our detection of S-DNAs following degradation of the RNA component of R-loops. Our observation is similar to findings of DNA misalignment following RNase H-mediated R-loop removal at human immunoglobulin class switch sequences (50). Subsequent aberrant processing of these S-DNAs with DNA repair complexes such as MutS $\beta$  may be what induces the partial instability observed following RNase H-mediated R-loop removal from the (CAG)<sub>n</sub>-(CTG)<sub>n</sub> repeat tract (34).

Transcription-mediated instability of (CAG)<sub>n</sub>-(CTG)<sub>n</sub> repeat tracts involves various nucleotide excision repair and mismatch repair proteins (19–20,23,51–52). How these proteins generate instability is poorly understood. An interaction between nucleotide excision repair proteins XPF-ERCC1 and XPG and R-loops was established previously for immunoglobulin class switch sequences (53). Both pro-

teins were found to cleave R-loops formed in class switch sequences, which is proposed to be a necessary step during class switch recombination (53). We similarly demonstrated that XPF-ERCC1 and XPG proteins recognize and cleave CAG and CTG slipped structures 5' and 3' of the S-DNA junction, respectively (54). Subsequent error-prone DNA repair synthesis may generate instability. Binding of mismatch repair proteins and/or other repair proteins to S-DNA in the R-loop may also trigger instability through aberrant repair processes. The assessment of individual protein factors upon R-loop processing warrants further analysis.

Lin *et al.* (11) proposed a model in which R-loop formation stabilizes non-B-DNA structures in the displaced, non-template DNA strand of the R-loop, signaling various repair factors such as those of transcription-coupled repair, nucleotide excision repair and/or mismatch repair to act upon the slipped-structures thus generating aberrant repair-mediated instability. This is similar to the model proposed by Pearson *et al.* (25) in which unwinding and re-annealing of (CAG)-(CTG) repeats during DNA metabolism (including transcription) leads to the formation of slipped-, S-DNAs as a mutagenic intermediate of instability. Based on these combined findings, we propose a model in which convergent bidirectional transcription generates R-loops in various configurations that are either directly processed to trigger instability, or are first processed into S-DNA through RNA removal, which is then further processed triggering instability (Figure 10). While the level of instability we observe is modest, it represents the possible outcome of only a single round of transcription, R-loop formation and processing. Such events may occur frequently (possibly for every round of transcription) and the accumulation of such changes may amount to the high levels of instability in non-proliferating patient tissues.

In summary, we have established a novel R-loop processing assay utilizing human cell extracts. We demonstrate that double-R-loop processing generates instability of an expanded disease-associated (CAG)-(CTG) and a (GGGGCCG)-(GGCCCC) repeat tract post-transcriptionally and in the absence of DNA replication. Some of the double-R-loop-mediated instability may arise from the processing of S-DNAs that arise from the removal of RNA from R-loops.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGMENTS

The authors thank Drs. Tao Zu and Laura Ranum for providing (GGGGCC)-(GGCCCC) expansion clones containing various repeat lengths. They thank Dr. Meghan Slean for comments on this manuscript. We also thank Ms. Jodie Simard for technical assistance and Dr. David I. McIntosh for assistance in data processing.

## FUNDING

Canadian Institutes of Health Research [MOP-97896 to C.E.P.]; Amyotrophic Lateral Sclerosis Association

(ALSA), U.S.A. [to C.E.P.]; National Institutes of Health [R01GM101192 to Y.-H.W.]. Funding for open access charge: Canadian Institutes of Health Research [MOP-97896 to C.E.P.]; National Institutes of Health [R01GM101192 to Y.-H.W.].

*Conflict of interest statement.* None declared.

## REFERENCES

1. Temiakov, D., Montesana, P.E., Ma, K., Mustaev, A., Borukhov, S. and McAllister, W.T. (2000) The specificity loop of T7 RNA polymerase interacts first with the promoter and then with the elongating transcript, suggesting a mechanism for promoter clearance. *Proc. Natl Acad. Sci. U.S.A.*, **97**, 14109–14114.
2. Roy, D. and Lieber, M.R. (2009) G clustering is important for the initiation of transcription-induced R-loops in vitro, whereas high G density without clustering is sufficient thereafter. *Mol. Cell Biol.*, **29**, 3124–3133.
3. Reddy, K., Tam, M., Bowater, R.P., Barber, M., Tomlinson, M., Nichol Edamura, K., Wang, Y.H. and Pearson, C.E. (2011) Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. *Nucleic Acids Res.*, **39**, 1749–1762.
4. Masukata, H. and Tomizawa, J. (1990) A mechanism of formation of a persistent hybrid between elongating RNA and template DNA. *Cell*, **62**, 331–338.
5. Xu, B. and Clayton, D.A. (1995) A persistent RNA-DNA hybrid is formed during transcription at a phylogenetically conserved mitochondrial DNA sequence. *Mol. Cell Biol.*, **15**, 580–589.
6. Daniels, G.A. and Lieber, M.R. (1995) RNA:DNA complex formation upon transcription of immunoglobulin switch regions: implications for the mechanism and regulation of class switch recombination. *Nucleic Acids Res.*, **23**, 5006–5011.
7. Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I. and Chedin, F. (2012) R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell*, **45**, 814–825.
8. Huertas, P. and Aguilera, A. (2003) Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol. Cell*, **12**, 711–721.
9. Luna, R., Jimeno, S., Marin, M., Huertas, P., Garcia-Rubio, M. and Aguilera, A. (2005) Interdependence between transcription and mRNP processing and export, and its impact on genetic stability. *Mol. Cell*, **18**, 711–722.
10. Dominguez-Sanchez, M.S., Barroso, S., Gomez-Gonzalez, B., Luna, R. and Aguilera, A. (2011) Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. *PLoS Genet.*, **7**, e1002386.
11. Lin, Y., Dent, S.Y., Wilson, J.H., Wells, R.D. and Napierala, M. (2010) R loops stimulate genetic instability of CTG.CAG repeats. *Proc. Natl Acad. Sci. U.S.A.*, **107**, 692–697.
12. Nakamori, M., Gourdon, G. and Thornton, C.A. (2011) Stabilization of expanded (CTG)<sup>n</sup>(CAG)<sup>n</sup> repeats by antisense oligonucleotides. *Mol. Ther.*, **19**, 2222–2227.
13. Lopez 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.
14. Pearson, C.E., Nichol Edamura, K. and Cleary, J.D. (2005) Repeat instability: mechanisms of dynamic mutations. *Nat. Rev. Genet.*, **6**, 729–742.
15. van Blitterswijk, M., DeJesus-Hernandez, M., Niemantsverdriet, E., Murray, M.E., Heckman, M.G., Diehl, N.N., Brown, P.H., Baker, M.C., Finch, N.A., Bauer, P.O. *et al.* (2013) Association between repeat sizes and clinical and pathological characteristics in carriers of C9ORF72 repeat expansions (Xpansize-72): a cross-sectional cohort study. *Lancet Neurol.*, **12**, 978–988.
16. Batra, R., Charizanis, K. and Swanson, M.S. (2010) Partners in crime: bidirectional transcription in unstable microsatellite disease. *Hum. Mol. Genet.*, **19**, R77–R82.
17. Xu, Z., Poidevin, M., Li, X., Li, Y., Shu, L., Nelson, D.L., Li, H., Hales, C.M., Gearing, M., Wingo, T.S. *et al.* (2013) Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. *Proc. Natl Acad. Sci. U.S.A.*, **110**, 7778–7783.

18. Bowater, R.P., Jaworski, A., Larson, J.E., Parniewski, P. and Wells, R.D. (1997) Transcription increases the deletion frequency of long CTG-CAG triplet repeats from plasmids in *Escherichia coli*. *Nucleic Acids Res.*, **25**, 2861–2868.
19. 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.
20. Jung, J. and Bonini, N. (2007) CREB-binding protein modulates repeat instability in a *Drosophila* model for polyQ disease. *Science*, **315**, 1857–1859.
21. Nakamori, M., Pearson, C.E. and Thornton, C.A. (2011) Bidirectional transcription stimulates expansion and contraction of expanded (CTG)<sub>n</sub>(CAG) repeats. *Hum. Mol. Genet.*, **20**, 580–588.
22. Lin, Y., Leng, M., Wan, M. and Wilson, J.H. (2010) Convergent transcription through a long CAG tract destabilizes repeats and induces apoptosis. *Mol. Cell Biol.*, **30**, 4435–4451.
23. 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.
24. 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.
25. Pearson, C.E. and Sinden, R.R. (1996) Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. *Biochemistry*, **35**, 5041–5053.
26. Panigrahi, G.B., Cleary, J.D. and Pearson, C.E. (2002) In vitro (CTG)<sub>n</sub>(CAG) expansions and deletions by human cell extracts. *J. Biol. Chem.*, **277**, 13926–13934.
27. Zu, T., Liu, Y., Bañez-Coronel, M., Reid, T., Pletnikova, O., Lewis, J., Miller, T.M., Harms, M.B., Falchook, A.E., Subramony, S.H. *et al.* (2013) RAN proteins and RNA foci from antisense transcripts in C9orf72 ALS and frontotemporal dementia. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, E4968–E4977.
28. Reddy, K., Zamiri, B., Stanley, S.Y., Macgregor, R.B. Jr and Pearson, C.E. (2013) The disease-associated r(GGGGCC)<sub>n</sub> repeat from the C9orf72 gene forms tract length-dependent uni- and multimolecular RNA G-quadruplex structures. *J. Biol. Chem.*, **288**, 9860–9866.
29. Haeusler, A.R., Donnelly, C.J., Periz, G., Simko, E.A., Shaw, P.G., Kim, M.S., Maragakis, N.J., Troncoso, J.C., Pandey, A., Sattler, R. *et al.* (2014) C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature*, **507**, 195–200.
30. Pahlman, S., Mamaeva, S., Meyerson, G., Mattsson, M.E., Bjelfman, C., Ortoft, E. and Hammerling, U. (1990) Human neuroblastoma cells in culture: a model for neuronal cell differentiation and function. *Acta Physiol. Scand. Suppl.*, **592**, 25–37.
31. Panigrahi, G.B., Lau, R., Montgomery, S.E., Leonard, M.R. and Pearson, C.E. (2005) Slipped (CTG)<sub>n</sub>(CAG) repeats can be correctly repaired, escape repair or undergo error-prone repair. *Nat. Struct. Mol. Biol.*, **12**, 654–662.
32. Li, J.J. and Kelly, T.J. (1985) Simian virus 40 DNA replication in vitro: specificity of initiation and evidence for bidirectional replication. *Mol. Cell Biol.*, **5**, 1238–1246.
33. Roberts, J.D. and Kunkel, T.A. (1988) Fidelity of a human cell DNA replication complex. *Proc. Natl Acad. Sci. U.S.A.*, **85**, 7064–7068.
34. Panigrahi, G.B., Slean, M.M., Simard, J.P., Gileadi, O. and Pearson, C.E. (2010) Isolated short CTG/CAG DNA slip-outs are repaired efficiently by hMutSbeta, but clustered slip-outs are poorly repaired. *Proc. Natl Acad. Sci. U.S.A.*, **107**, 12593–12598.
35. Panigrahi, G.B., Slean, M.M., Simard, J.P. and Pearson, C.E. (2012) Human mismatch repair protein hMutLalpha is required to repair short slipped-DNAs of trinucleotide repeats. *J. Biol. Chem.*, **287**, 41844–41850.
36. Lopez Castel, A., Tomkinson, A.E. and Pearson, C.E. (2009) CTG/CAG repeat instability is modulated by the levels of human DNA ligase I and its interaction with proliferating cell nuclear antigen: a distinction between replication and slipped-DNA repair. *J. Biol. Chem.*, **284**, 26631–26645.
37. Tome, S., Panigrahi, G.B., Lopez Castel, A., Foiry, L., Melton, D.W., Gourdon, G. and Pearson, C.E. (2011) Maternal germline-specific effect of DNA ligase I on CTG/CAG instability. *Hum. Mol. Genet.*, **20**, 2131–2143.
38. 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.
39. Nowotny, M., Gaidamakov, S.A., Ghirlando, R., Cerritelli, S.M., Crouch, R.J. and Yang, W. (2007) Structure of human RNase H1 complexed with an RNA/DNA hybrid: insight into HIV reverse transcription. *Mol. Cell*, **28**, 264–276.
40. 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.
41. Nospikel, T. and Hanawalt, P.C. (2002) DNA repair in terminally differentiated cells. *DNA Repair*, **1**, 59–75.
42. Pearson, C.E., Wang, Y.H., Griffith, J.D. and Sinden, R.R. (1998) Structural analysis of slipped-strand DNA (S-DNA) formed in (CTG)<sub>n</sub>(CAG)<sub>n</sub> repeats from the myotonic dystrophy locus. *Nucleic Acids Res.*, **26**, 816–823.
43. Gan, W., Guan, Z., Liu, J., Gui, T., Shen, K., Manley, J.L. and Li, X. (2011) R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev.*, **25**, 2041–2056.
44. Gomez-Gonzalez, B., Garcia-Rubio, M., Bermejo, R., Gaillard, H., Shirahige, K., Marin, A., Foiani, M. and Aguilera, A. (2011) Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. *EMBO J.*, **30**, 3106–3119.
45. Helmrich, A., Ballarino, M. and Tora, L. (2011) Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Mol. Cell*, **44**, 966–977.
46. Zamiri, B., Reddy, K., Macgregor, R.B. Jr and Pearson, C.E. (2014) TMPyP4 porphyrin distorts RNA G-quadruplex structures of the disease-associated r(GGGGCC)<sub>n</sub> repeat of the C9orf72 gene and blocks interaction of RNA-binding proteins. *J. Biol. Chem.*, **289**, 4653–4659.
47. Fratta, P., Mizielinska, S., Nicoll, A.J., Zloh, M., Fisher, E.M., Parkinson, G. and Isaacs, A.M. (2012) C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. *Sci. Rep.*, **2**, 1016.
48. Duquette, M.L., Handa, P., Vincent, J.A., Taylor, A.F. and Maizels, N. (2004) Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev.*, **18**, 1618–1629.
49. Axford, M.M., Wang, Y., Nakamori, M., Zannis-Hadjopoulos, M., Thornton, C.A. and Pearson, C.E. (2013) Detection of slipped-DNAs at the trinucleotide repeats of the myotonic dystrophy type I disease locus in patient tissues. *PLoS Genet.*, **9**(12), e1003866.
50. Yu, K., Chedin, F., Hsieh, C.L., Wilson, T.E. and Lieber, M.R. (2003) R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat. Immunol.*, **4**, 442–451.
51. Lin, Y. and Wilson, J.H. (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.
52. Lin, Y. and Wilson, J.H. (2009) Diverse effects of individual mismatch repair components on transcription-induced CAG repeat instability in human cells. *DNA Repair*, **8**, 878–885.
53. Tian, M. and Alt, F.W. (2000) Transcription-induced cleavage of immunoglobulin switch regions by nucleotide excision repair nucleases in vitro. *J. Biol. Chem.*, **275**, 24163–24172.
54. Slean, M.M., Reddy, K., Wu, B., Nichol, E., Edamura, K., Kekis, M., Nelissen, F.H., Aspers, R.L., Tessari, M., Schärer, O.D., Wijmenga, S.S. *et al.* (2013) Interconverting conformations of slipped-DNA junctions formed by trinucleotide repeats affect repair outcome. *Biochemistry*, **52**, 773–785.