

Cas9 nickase-mediated contractions of CAG/CTG repeats are transcription-dependent and replication-independent

Meghan Larin¹, Florence Gidney¹, Lorène Aeschbach², Laura Heraty¹, Emma L. Randall¹, Aeverie E.R. Heuchan¹, Marcela Buřičová¹, Melvin Bérard² and Vincent Dion^{1,3,*}

¹UK Dementia Research Institute at Cardiff University, Hadyn Ellis Building, Maindy Road, Cardiff CF24 4HQ, UK

²Centre for Integrative Genomics, University of Lausanne, Lausanne CH-1015, Switzerland

³Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, Maindy Road, Cardiff CF24 4HQ, UK

*To whom correspondence should be addressed. Tel: +44 29225 10893; Email: dionv@cardiff.ac.uk

Present addresses:

Meghan Larin, AstraZeneca R&D, 1 Francis Crick Avenue, Cambridge CB2 0AA, UK.

Florence Gidney, Huntington's Disease Centre, Department of Neurodegenerative Disease, UCL Queen Square Institute of Neurology, University College London, London WC1N 3BG, UK.

Lorène Aeschbach, AC Immune SA EPFL, Innovation Park Building B, 1015 Lausanne, Switzerland.

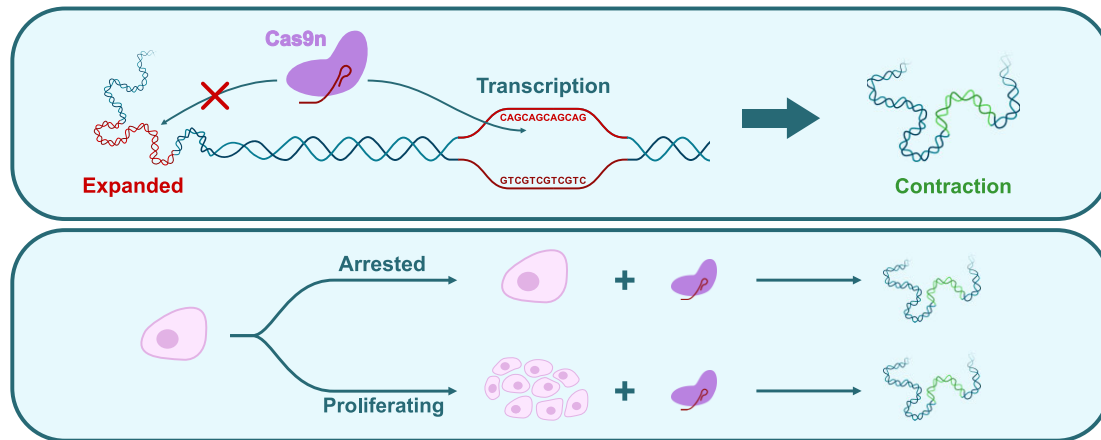
Marcela Buřičová, Revvity, Cambridge Research Park, 8100 Beach Dr, Waterbeach, Cambridge CB259TL, UK.

Melvin Bérard, Department of Cellular and Molecular Biology, University of Geneva, Quai Ansermet 30, 1205 Geneva, Switzerland.

Abstract

There is currently no disease-modifying treatment for expanded CAG/CTG repeat disorders. Given that longer repeat tracts lead to an earlier age of disease onset and faster progression, contracting them is expected to improve symptoms and/or delay onset. We have previously demonstrated that the Cas9 D10A nickase can effectively contract CAG/CTG repeats when targeted to the repeat tract itself. However, the mechanism remains unclear. Here, we tested whether nickase-mediated contractions depend on transcription or on replication using human cell models. We find that transcription promotes contractions and that they occur independently of the rate of cell division. These results support the therapeutic potential of this approach in non-dividing cells.

Graphical abstract



Introduction

The expansion of CAG/CTG repeats causes at least 15 different neurodegenerative and neuromuscular disorders, including Huntington's disease (HD), myotonic dystrophy type 1 (DM1), and several spinocerebellar ataxias (1). There are currently no disease-modifying treatments available for any of these disorders. Importantly, in HD, repeat tract length accounts for about 60% of the variation in the age at disease onset, with longer repeats leading to an earlier appearance of

the symptoms (2). This inverse correlation between repeat size and age at disease onset is also apparent in several other expanded CAG/CTG repeat disorders, suggesting that reducing the rate of expansion would slow the progression, if not prevent it in multiple diseases (3). This model is further substantiated by the identification of the mismatch repair pathway as a major modifier of the age at disease onset (4–6). Indeed, mismatch repair has a central role in the rate of repeat expansion in somatic tissues (7), providing a potential link to disease

Received: June 13, 2024. Revised: September 10, 2024. Editorial Decision: September 11, 2024. Accepted: September 21, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of NAR Molecular Medicine.

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

progression. Furthermore, contracting the repeat tract may be more effective at provoking a phenotypic change than merely preventing somatic expansions (8).

To date, three methods for contracting CAG/CTG repeats have been used (9–12). One is via a small molecule, naphthyridine-azaquinolone (NA), which binds to secondary structures formed by CAG/CTG repeats (10). NA induced contractions in HD-derived fibroblasts, HT1080 cells containing 850 CAG/CTG repeats at an ectopic locus, and in two mouse models (10,13). The current hypothesis to account for NA promoting contractions is centred on its ability to stabilize secondary DNA structures produced by transcription through the repeat tract, which are then processed by the mismatch repair complex MutS β (10). These data suggest that contractions in this case occur via a mechanism related to that of somatic expansion.

A second approach is to induce double-strand breaks within the repeat tract with programmable nucleases, including zinc-finger nucleases (9,11), TALENs (12) and CRISPR-Cas9⁹. In these cases, the double-strand break(s) created within the repeat are thought to be repaired via homology directed repair, either in a Rad51-dependent manner (11), or through single-strand annealing (12). In our previous study, we found that both zinc-finger nucleases and CRISPR-Cas9 induced expansions and contractions at comparable rates (9). The risk, therefore, is that some cells within a population show further expansions, which would exacerbate disease phenotypes.

We have taken a third approach that avoids double-strand break-inducing nucleases. Instead, we used the Cas9 D10A nickase. When directed to the repeat tract itself using a single-guide RNA (sgRNA), we found that the Cas9 D10A nickase induces contractions efficiently in HEK293-derived cells (9), in astrocytes and neurons from patient-derived induced pluripotent stem cells (14), as well as in a mouse model of HD (14). We found that nickase-mediated contractions occurred only at expanded alleles larger than 41 repeats, and that off-target mutations, if present, were below detection limits (9,14). How contractions occur, however, is unclear. In a HEK293-derived model, contractions were unaffected by the knockdown of MSH2, a central player in mismatch repair (9). This suggests that the mechanism of contractions caused by the Cas9 nickase may be different to that of NA- and programmable nuclease-induced contractions.

Understanding how the Cas9 nickase induces contractions is critical for the translation of this approach into the clinic. Here we address two outstanding mechanistic questions about Cas9 nickase-mediated contractions: (i) are Cas9 nickase-mediated contractions dependent on transcription through the repeat tract? and, (ii) are they driven by replication? Our work indicates that contractions are promoted by transcription through the repeat tract while being uncoupled from replication rates. These results support the use of the Cas9 nickase as a therapy as contractions will be expected in non-dividing neurons and other terminally differentiated cells that express the mutant gene.

Materials and methods

Cell lines and culture conditions

The cell lines used in this study and their construction are found in Table S1. GFP(CAG)₁₀₁¹⁵ cells were maintained at 37°C with 5% CO₂ in Dulbecco's modified Eagle's medium

(DMEM) glutamax, supplemented with 10% dialysed foetal bovine serum (FBS, Gibco), 100 U ml⁻¹ penicillin, 100 μ g ml⁻¹ streptomycin, 15 μ g ml⁻¹ blasticidin and 150 μ g ml⁻¹ hygromycin. GNickx and GNickSx cells were cultured under the same conditions, but with the addition of G418 (400 μ g ml⁻¹) and puromycin (1 μ g ml⁻¹). The FLAH25 cells (15) were maintained at 37°C with 5% CO₂ in DMEM F-12, 10% FBS (Gibco), 100 U ml⁻¹ penicillin, 100 μ g ml⁻¹ streptomycin, G418 (500 μ g ml⁻¹) and puromycin (2.5 μ g ml⁻¹). FNick102 cells, which stably express Cas9 D10A together with the blasticidin resistance gene (BSD), were further supplemented with 10 μ g ml⁻¹ blasticidin. Cells were arrested using 0.5 μ M palbociclib. The media was refreshed every 2–3 days.

Both GFP(CAG)₁₀₁ and FLAH25 cells were genotyped using the Cell Line Authentication service from Eurofins. They were confirmed to be HEK293.2sus and HT1080, respectively. Mycoplasma testing was performed by Eurofins and using in-house testing regularly throughout the study and were found to be negative. The repeat tract length for each cell line was determined using PCR for the region around the repeat tract and Sanger sequencing or long-read sequencing. The primers used for PCR are found in Table S2.

Transient transfections of GFP(CAG)₁₀₁ cells were conducted as described (16). GNickX and GNickSX cells were cultured for up to 42 days in the presence or absence of 2 μ g ml⁻¹ of doxycycline (dox). Cells were collected and genomic DNA was used to amplify the repeat tract using primers oVIN-0016 and oVIN-0487 (See Table S2). The plasmids used are found in Table S3. Specifically, lentiCas9n(D10A)-Blast (17) was a gift from Feng Zhang (Addgene plasmid # 63593) and hCas9_D10A (18) was a gift from George Church (Addgene plasmid # 41816).

Cell sorting

FNick102 cells were transduced at a multiplicity of infection of 5 with viruses containing sgCTG and a GFP cassette (Table S3). Two days after transduction, cells were treated with or without doxycycline. At the end of the 21- or 42-day treatment, the cells were harvested with PBS + 10mM EDTA and sorted using a BD FACSAria II cell sorter based on their GFP intensity. Cells that had GFP fluorescence comparable to that of untransduced cells were kept and analysed (GFP⁻). Cells that were above the background fluorescence were further divided into two populations with the bottom 50% of GFP-expressing cells being referred to as GFP⁺ and the top 50% as GFP⁺⁺.

Small pool PCR

Small pool PCR was performed as before (19) with the following exceptions. Cells were harvested 3 weeks or 6 weeks after doxycycline addition, and DNA was extracted and then quantified using the Quant-iT Qubit dsDNA HS Assay Kit (Fisher Scientific). Blinded counts were completed for each membrane based on allele frequency calculated using a Poisson distribution. Differences between conditions were determined statistically by Mann–Whitney *U* test using GraphPad Prism software (v10.2.2).

Long read sequencing

Single-molecule real time (SMRT) sequencing was done using an amplicon-based approach. Briefly, the repeat region in GNickS12, GNickS115 and GNick102 were amplified using

primers containing the barcodes (see Table S2) together with Thermo™ Phusion II High Fidelity polymerase. Multiple independent PCRs were performed for each sample and pooled together before loading. Equimolar pools of the samples were loaded at a concentration between 10 and 12 μM . The sequencing was done using a Sequel IIe at Cardiff University School of Medicine. Repeat size was determined from the resulting CCS using Repeat Detector v1.0 as described (20).

BrdU incorporation and cell cycle analysis

For BrdU staining, FLAH25 cells had their medium supplemented with BrdU ($10 \mu\text{g ml}^{-1}$) for 1 h before being fixed, stained and imaged. The protocol used was as described (14) except that DNA was denatured with 2M HCl, followed by neutralization with 0.1M sodium borate. Cells were stained using the BU1/75 (ICR1) antibody (Abcam) with a dilution of 1:1000 and counterstained with DAPI. The images were taken using an Opera Phenix and analysed using FIJI (21) to count the total number of nuclei per frame. Cell cycle analysis was done using a 5-minute cell cycle assay from Chemotec and measured using a NucleoCounter NC-250 Automated Cell Analyzer. The data was then analysed using FlowJo (v10.0.8r1).

Results

Transcription through the repeat tract promotes Cas9 nickase-mediated contractions

To address whether transcription promotes contractions, we used HEK293-derived GFP(CAG)₁₀₁ cells (9,16,22). This line contains a doxycycline-inducible promoter driving transcription through 101 CAG repeats located in the intron of a GFP mini gene (Figure 1A). The repeats interfere with splicing in a length-dependent manner, with longer repeats resulting in lower GFP expression and shorter repeats increasing GFP levels, providing a fluorescent proxy for repeat size (9,22). This line has a single copy of the GFP minigene integrated on chromosome 12 (23). It recapitulates the instability pattern seen in other systems, including an expansion bias and HDAC3-dependency (24–28).

To determine whether transcription through the repeat tract promotes contractions, we transiently transfected the Cas9 nickase together with either a sgRNA against the repeat tract (sgCTG) or an empty vector. We either induced transcription through the repeat tract by adding doxycycline to the cultures for 168 h or kept them without dox and thus without transcription (Figure 1B). We then added doxycycline to all cultures for 72 h to read out steady state GFP levels. We found that transfection of both the Cas9 nickase and the sgCTG led to a significant increase in the GFP⁺ cells in the presence of doxycycline compared to the control conditions transfected with an empty plasmid and without dox ($P = 0.0094$, using a Kruskal–Wallis test with Dunn's multiple comparison post-hoc test). The GFP⁺ population is enriched with contractions (9) and contains cells with GFP intensities that lie within the brightest 1% of the cells transfected with the empty sgRNA vector, which is the negative control in these experiments (16). Importantly, we found no significant increase in GFP⁺ cells in the absence of doxycycline ($P = 1.0$), sgCTG ($P = 1.0$), or the nickase ($P = 0.39$; Figure 1C), suggesting that transcription is essential for contractions in HEK293-derived cells.

Production of stable cell lines constitutively expressing the Cas9 nickase and sgCTG

Our previous results suggested that non-expanded (≤ 42) repeats are not targeted for contraction (9,14). This observation, together with the result that transcription is essential for Cas9 nickase-mediated contractions, suggested that we could construct a cell line that stably and continuously expresses both the Cas9 nickase and the sgCTG. This cell line should be viable and only contract the ectopic repeat when there is transcription through the repeat tract. To do this, we constructed GFP(CAG)₁₁₅ cells that constitutively express the Cas9 nickase together with sgCTG (GNickS115, Table S1). The addition of doxycycline to these cells led to contractions, including in the non-expanded range as measured by PCR (Figure 1D). By contrast, control lines containing 102 CAGs and an empty sgRNA vector (GNick102) or lines with a non-pathogenic 12 CAGs (GNick12) both remained stable. We noted that there is a low molecular band that appears in the GNickS115 lines, which is not seen in the absence of the sgRNA (GNick102) or when the repeat tract is short (GNick12). We wondered whether this was a contraction of the expanded repeat tract in a fraction of the cells in the clone containing the sgRNA, the Cas9 nickase, and the large repeat. To address this, we used Single-Molecule Real-Time (SMRT) long-read sequencing on the same experiments and found that indeed they were contracted alleles. Further, we confirmed that contractions accumulated in GNickS115 at a higher frequency at day 26 when the cells were treated with dox. We found that the main contracted band seen on the gel had a mode of three CAG repeats with intact flanking sequences (Supplementary Figure S1). Together, these results show that transcription promotes nickase-mediated contractions and emphasises that constitutive expression of the Cas9 nickase is not toxic despite the presence of endogenous, non-expanded repeats.

To test whether this effect of transcription is unique to HEK293-derived cells, we used a separate human cell line: the HT1080-derived FLAH25 (29). HT1080 cells have been used extensively to investigate repeat instability (11,15,29–33), making them a convenient model. Similar to GFP(CAG)₁₀₁, FLAH25 cells harbour 102 CAGs in the intron of an ectopic *HPRT* minigene under the control of doxycycline (15). We transduced these cells with a lentivirus expressing the Cas9 nickase and isolated single clones to create a stable FLAH25 Nickase with 102 CAGs (FNick102 – Table S1). We then delivered the sgCTG using a lentiviral vector that also contained a GFP cassette, generating FNickS102 cells (Figure 2A). Unlike the GNickS115 cells, FNickS102 cells are a population with the sgCTG/GFP lentivirus delivered *de novo* for each experiment. We incubated FNickS102 cells with or without doxycycline and then sorted the GFP negative cells and the dimmest (GFP⁺) and brightest (GFP⁺⁺) halves of the GFP-expressing cells for DNA extraction and analysis by small-pool PCR (34) (Figure 2B, C). The expectation was that cells with a large amount of GFP also express more sgCTG but the transcription through the repeat tract in the *HPRT* minigene would remain the same between the two populations. GFP⁺ cells did not show a statistically significant shift when comparing with and without doxycycline treatment at day 42 (Figure 2D – $P = 0.08$). We also found no significant increase in contraction in the GFP⁺⁺ cells exposed to doxycycline after 21 days compared to those without doxycycline (Supplementary Figure S2 – $P = 0.44$). However, there was

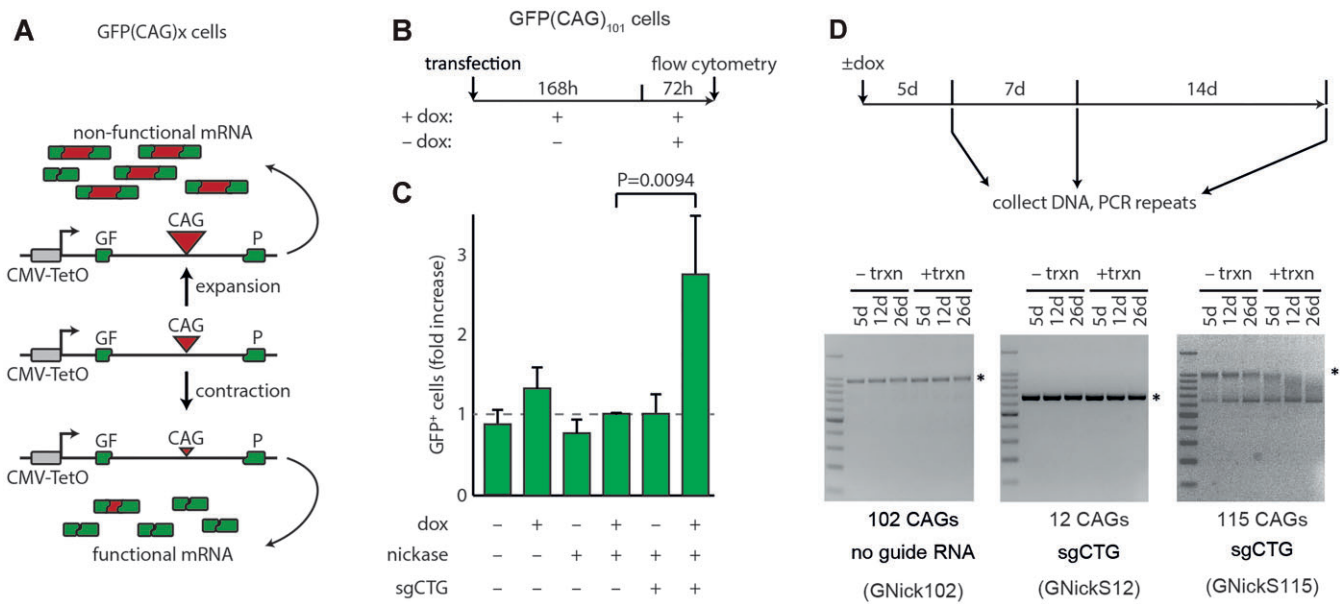


Figure 1. Transcription is essential for Cas9 nickase-mediated contractions in HEK293-derived cells. **(A)** Schematic of the GFP minigene reporter. Longer repeats are spliced into the mRNA along with 38 nucleotides downstream of the repeat tract, creating a frameshift and the degradation of the mRNA (42). Conversely, shorter repeats allow for efficient splicing of the reporter and robust GFP expression. The promoter, CMV-TetO, is induced by doxycycline (dox). **(B)** Timeline of experiments. **(C)** Fold change of GFP⁺ cells, which are enriched for contractions, over the control condition where dox is added and the Cas9 nickase is transfected, but there is no sgRNA expression. $N \geq 4$ for each condition. **(D)** Timeline of experiments (top) for stable GNick and GNickS cells. After 5, 12 and 26 days of doxycycline addition, the cells were harvested, and a PCR of the repeat region was run on agarose gels. The asterisk indicates the modal repeat size in each line at the start of the experiment.

a clear effect of doxycycline treatment in increasing contractions after 42 days compared to cells cultured without doxycycline (Figure 2D – $P < 0.0001$ for GFP⁺⁺ cells). These results suggest that, like in the GNickS115 cells, transcription through the repeat tract promotes contractions in HT1080-derived cells.

Cas9-mediated contractions occur independently of proliferation rates

One advantage of the HT1080 cells is that they can be efficiently arrested using the CDK4/6 inhibitor palbociclib (35). This allowed us to test whether contractions occur in cells whose proliferation is compromised, modelling disease-relevant cells, such as neurons, that are terminally differentiated. We found that FLA25 cell counts plateaued over 6 days of palbociclib treatment (Supplementary Figure S3A). The cells arrested as measured by the lack of BrdU incorporation and DAPI staining (Supplementary Figure S3A). Given these observations, we arrested FNick102 cells using palbociclib 6 days prior to transduction with the sgCTG/GFP lentivirus. Two days after transduction, we induced transcription through the repeat tract by adding doxycycline to the culture media (Figure 3A). As before, we sorted cells based on the GFP expression at 21 and 42 days and performed small-pool PCR to determine repeat size distribution (Figure 3B, Supplementary Figure S3B). We found that contractions occurred readily in palbociclib-arrested cells (Figure 3B, C, Supplementary Figure S3C), suggesting that population doubling does not correlate with frequencies of nickase-mediated contractions.

To test whether contractions in palbociclib-arrested cells are also dependent on transcription, we compared cells arrested both with and without doxycycline treatment. We

found that, similar to replicating cells, palbociclib-arrested FNickS102 cells showed enhanced contraction frequencies when exposed to doxycycline after 21 and 42 days (Supplementary Figure S3C; Figure 3C, $P < 0.0001$), suggesting that transcription through the repeat tract promotes nickase-mediated contractions in non-replicating FNickS102 cells.

Discussion

Here we found that transcription promotes contractions when the Cas9 D10A nickase is paired with a sgRNA against CAG repeats in two independent cell lines. Moreover, nickase-induced contractions did not scale with the number of cell divisions, suggesting that contractions can occur without a major role for replication. This adds to our previous work showing that the same treatment induced contractions in patient-derived induced pluripotent stem cells differentiated into astrocytes and neurons (14).

In FNickS102 and GNickS115 cells, we found that simply culturing the cells without doxycycline for 6 weeks led to an increase in contractions compared to cells that were untransduced (Figures 1D and 2D). We found previously that in GFP(CAG)_x cells, there is some transcription through the locus in the absence of doxycycline as evidenced by low but detectable background GFP expression (9). We speculate that this leaky transcription may be responsible for the contractions observed under these conditions. Another potential model is that the large amount of sgCTG masks any effect of transcription, perhaps because the sgCTG is rate-limiting. This is supported by the population of GFP⁺ FNickS102 cells, which contains less of the sgCTG/GFP lentivirus, yielding lower levels of contractions compared to the GFP⁺⁺ population (Figure 2D). The prediction is that transducing FNick102

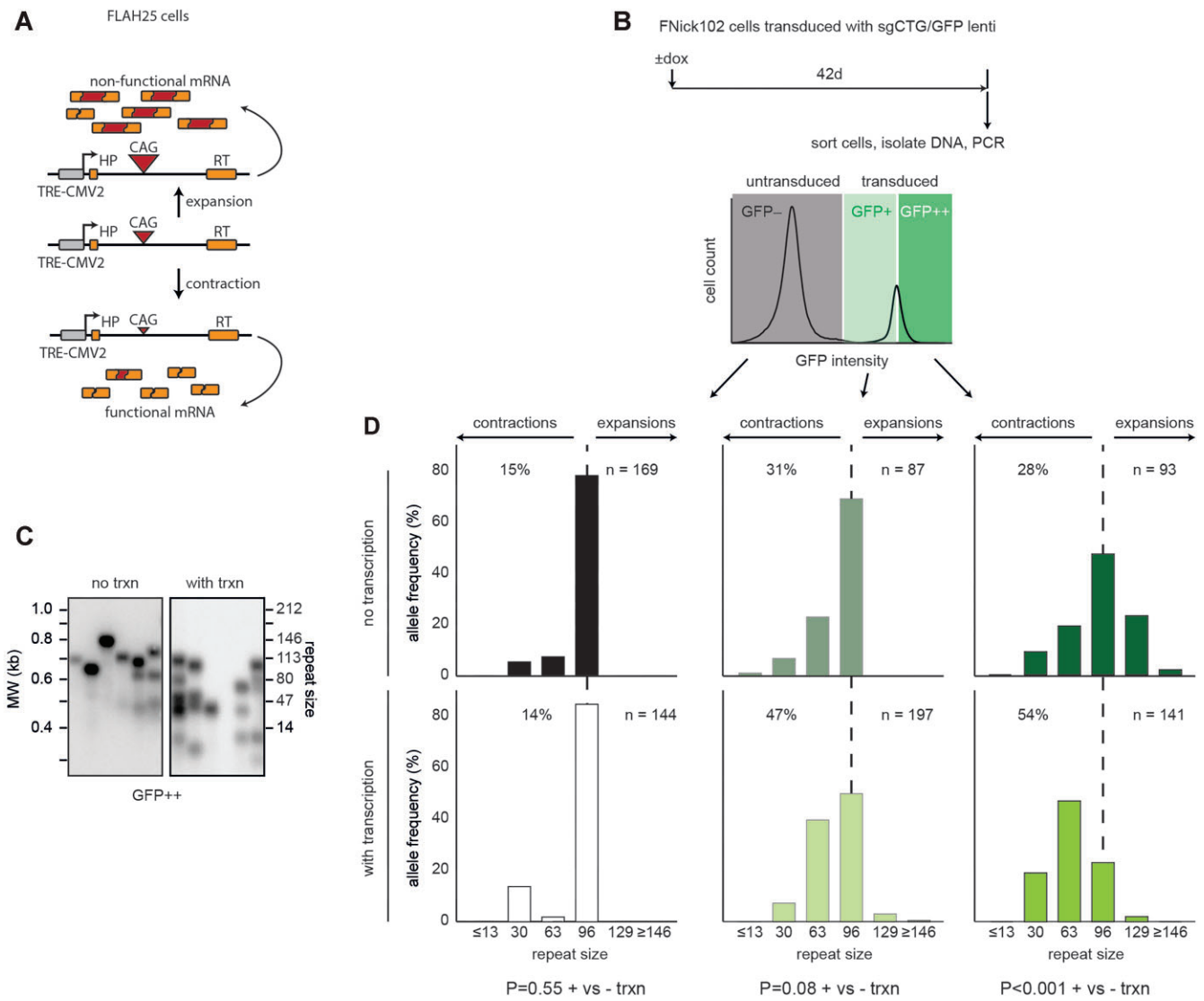


Figure 2. Transcription promotes nickase-mediated CAG repeat contraction in HT1080-derived cells. **(A)** FLAH25 cells contain a doxycycline inducible promoter driving transcription through an *HPRT* minigene in which 102 CAG repeats were inserted. As in the GFP reporter in GNickS lines, the presence of the repeats interferes with *HPRT* splicing. Here, however, we did not use the reporter. **(B)** FNick102 cells were transduced with a lentivirus that expresses both sgCTG and GFP and allowed to proliferate for 6 weeks. Then the cells were sorted into GFP⁺ cells, which are the dimmest 50% of the GFP-expressing cells, GFP⁺⁺ cells, which express the largest amount of GFP or GFP⁻ cells, which did not express GFP above background at the end of the experiments. All three populations are grown in the same well. **(C)** Representative small pool PCR of GFP⁺⁺ FNickS102 cells treated with or without doxycycline. **(D)** Quantification of the GFP⁺, GFP⁺⁺ and GFP⁻ cells in the presence or absence of doxycycline. The % in each graph represents the fraction of contractions. The number of repeats on the x-axes correspond to the median repeat size of a given interval, except for the ≤ 13 and ≥ 146 intervals. The *P*-values were calculated with Mann-Whitney *U* test between the GFP⁺ cells with or without doxycycline.

cells with a larger amount of sgCTG may lead to further contractions even in the absence of transcription. It is also unclear why there are no significant contractions until 6 weeks in the FNickS102 GFP⁺⁺ cells. The simplest explanation for this lag is that contractions are slow to accumulate and do not reach significance within 3 weeks. A non-mutually exclusive hypothesis is that the contractions occur via small changes in repeat size that we cannot resolve on the small-pool PCR gel after only 3 weeks. Regardless of the exact reason for these results, our conclusion still holds: transcription through the repeat tract promotes Cas9 nickase-mediated contractions.

Transcription has been shown to promote spontaneous and NA-mediated contractions in HT1080-derived cells (10,15). In these cases, the prevailing model is that transcription pro-

duces R-loops that are then processed, possibly via the mismatch repair pathway, to lead to contractions (10,30). In the case of the Cas9 nickase, there are further possibilities. For example, Cas9 is inhibited by the presence of nucleosomes (36,37). Indeed, CAG/CTG repeat tracts are strong nucleosome-positioning sequences (38–41) and transcription may act in displacing nucleosomes such that Cas9 can access its target sequence. This would also be in line with our previous work, which showed that knockdown of MSH2 and XPA, essential for mismatch repair and transcription-coupled nucleotide excision repair, respectively, had no effect on Cas9 nickase-induced contractions (9). Another possibility is that the R-loop is provided by Cas9/sgRNA binding to DNA.

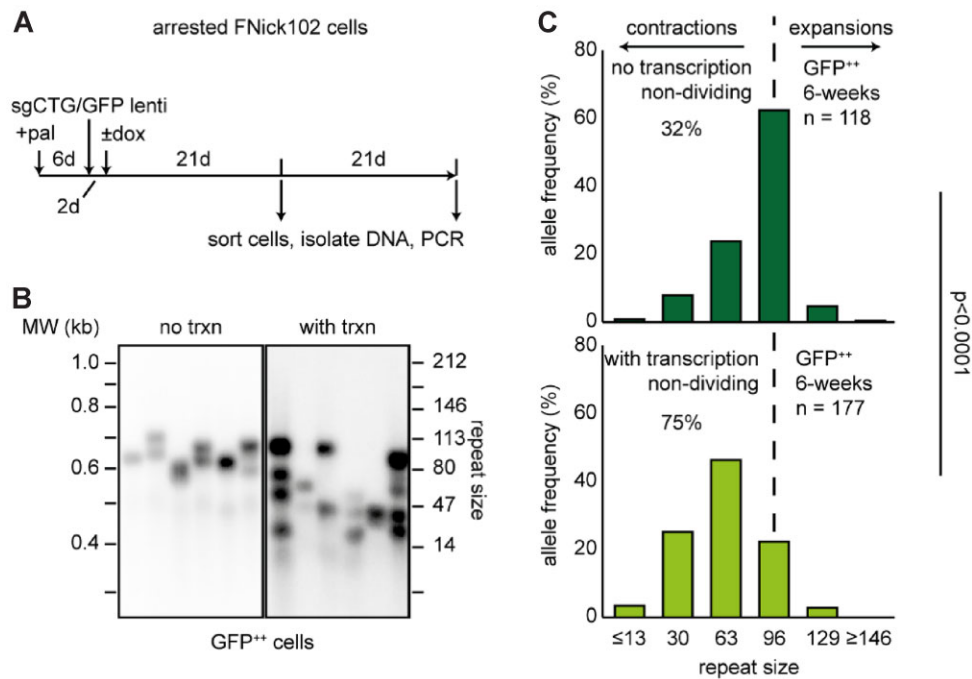


Figure 3. Transcription promotes nickase-mediated CAG repeat contraction independently of cell proliferation. **(A)** Timeline of the experiments (pal = palbociclib, dox = doxycycline). **(B)** Representative small pool PCR of GFP⁺⁺ cells after 42 days. **(C)** Quantification of the small pool PCRs for GFP⁺⁺ cells at 42 days in culture. The % in each graph represents the fraction of contractions. The number of repeats on the x-axes correspond to the median repeat size of a given interval, except for the ≤13 and ≥146 intervals. The *P*-value was calculated with Mann–Whitney *U* test between the GFP⁺⁺ cells with or without doxycycline.

We have developed two different tools that will allow us to probe the mechanism of nickase-induced contractions further. Specifically, we have created stable cell lines that constitutively express both the Cas9 nickase and the sgCTG. Given the ease of genetic engineering in immortalized cell lines, it will be possible to make genetic modifications to test specific models of repeat contractions. Consistent with our previous findings that short (≤42) repeats do not appear to be mutated by the Cas9 nickase (9,14), the viability of these stable lines suggests that there is little overt toxicity caused by the Cas9 nickase even in the presence of a sgRNA against the repeat tract. The results presented here help further demonstrate that the Cas9 nickase constitutes a potential therapeutic intervention for expanded repeat disorders. Continuing to elucidate the underlying mechanism will help improve both the efficiency and safety of the approach.

Data availability

All cell lines and raw data generated herein are available upon request from the corresponding author. The sequencing data is available publicly via Sequence Read Archive (SRA) (ID: PRJNA1153810) - <https://www.ncbi.nlm.nih.gov/bioproject/1153810>.

Supplementary data

[Supplementary Data](#) are available at NARMME Online.

Acknowledgements

We thank the members of the Dion lab for critical reading of the manuscript and for fruitful discussions. Christopher Smith helped with the BrdU experiments, Alvaro Murillo

with blindly counting small pool PCR membranes during revisions, and Ruban Rex Durairaj with uploading the sequencing datasets to SRA. Johanne Morgan assisted with running and maintaining the Sequel IIe. We acknowledge the support of the Supercomputing Wales project, which is part-funded by the European Regional Development Fund (ERDF) via the Welsh Government.

Author contributions: M.L.: Conceptualization, Methodology, Investigation, Visualization, Formal Analysis, Writing – original draft, Writing – review & editing. FG: Investigation, Methodology, Formal Analysis, Writing – review & editing. L.A.: Investigation, Methodology, Writing – review & editing. L.H.: Investigation, Writing – review & editing. E.R.: Investigation, Methodology, Writing – review & editing. AERH: Investigation, Formal analysis, Writing – review & editing. M Buřičová: Investigation, Writing – review & editing. M Bédard: Investigation, Writing – review & editing. V.D.: Conceptualization, Visualization, Formal Analysis, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition.

Funding

Dion lab is funded by a Professorship from the Academy of Medical Sciences [AMSPR1\1014]; UK Dementia Research Institute [UK DRI-3006] through UK DRI Ltd, principally funded by the UK Medical Research Council.

Conflict of interest statement

V.D. declares that he had a research contract with Pfizer Inc., which ended within the past year. It was unrelated to this work.

References

- Khristich,A.N. and Mirkin,S.M. (2020) On the wrong DNA track: molecular mechanisms of repeat-mediated genome instability. *J. Biol. Chem.*, **295**, 4134–4170.
- Mcallister,B., Gusella,J.F., Landwehrmeyer,G.B., Lee,J.-M., Macdonald,M.E., Orth,M., Rosser,A.E., Williams,N.M., Holmans,P., Jones,L., *et al.* (2021) Timing and impact of psychiatric, cognitive, and motor abnormalities in Huntington disease. *Neurology*, **96**, e2395–e2406.
- 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.
- Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium (2019) CAG repeat not polyglutamine length determines timing of Huntington's disease onset. *Cell*, **178**, 887–900.
- Mcallister,B., Donaldson,J., Binda,C.S., Powell,S., Chughtai,U., Edwards,G., Stone,J., Lobanov,S., Elliston,L., Schuhmacher,L.-N., *et al.* (2022) Exome sequencing of individuals with Huntington's disease implicates FAN1 nuclease activity in slowing CAG expansion and disease onset. *Nat. Neurosci.*, **25**, 446–457.
- Bettencourt,C., Hensman-Moss,D., Flower,M., Wiethoff,S., Brice,A., Goizet,C., Stevanin,G., Koutsis,G., Karadima,G., Panas,M., *et al.* (2016) DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. *Ann. Neurol.*, **79**, 983–990.
- Schmidt,M.H.M. and Pearson,C.E. (2016) Disease-associated repeat instability and mismatch repair. *DNA Repair (Amst.)*, **38**, 117–126.
- Wheeler,V.C. and Dion,V. (2021) Modifiers of CAG/CTG repeat instability: insights from mammalian models. *J. Huntingt. Dis.*, **10**, 123–148.
- Cinesi,C., Aeschbach,L., Yang,B. and Dion,V. (2016) Contracting CAG/CTG repeats using the CRISPR-Cas9 nickase. *Nat. Commun.*, **7**, 13272.
- Nakamori,M., Panigrahi,G.B., Lanni,S., Gall-Duncan,T., Hayakawa,H., Tanaka,H., Luo,J., Otabe,T., Li,J., Sakata,A., *et al.* (2020) A slipped-CAG DNA-binding small molecule induces trinucleotide-repeat contractions in vivo. *Nat. Genet.*, **52**, 146–159.
- Mittelman,D., Moye,C., Morton,J., Sykoudis,K., Lin,Y., Carroll,D. and Wilson,J.H. (2009) Zinc-finger directed double-strand breaks within CAG repeat tracts promote repeat instability in human cells. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 9607–9612.
- Bétemps,L., Descorps-Declère,S., Frenoy,O., Poggi,L., Mosbach,V., Tomé,S., Viterbo,D., Klein,A., Ma,L., Lameiras,S., *et al.* (2024) TALEN-induced contraction of CTG trinucleotide repeats in myotonic dystrophy type 1 cells. bioRxiv doi: <https://doi.org/10.1101/2023.10.14.562330>, 03 January 2024, preprint: not peer reviewed.
- Hasuike,Y., Tanaka,H., Gall-Duncan,T., Mehkary,M., Nakatani,K., Pearson,C.E., Tsuji,S., Mochizuki,H. and Nakamori,M. (2022) CAG repeat-binding small molecule improves motor coordination impairment in a mouse model of Dentatorubral-pallidoluysian atrophy. *Neurobiol. Dis.*, **163**, 105604.
- Murillo,A., Alpaugh,M., Larin,M., Randall,E.L., Heraty,L., Durairaj,R.R., Aston,A.N., Taylor,A.S., Monteyes,A.M., Stöberl,N., *et al.* (2024) Cas9 nickase-mediated contraction of CAG/CTG repeats at multiple disease loci. bioRxiv doi: <https://doi.org/10.1101/2024.02.19.580669>, 21 February 2024, preprint: not peer reviewed.
- 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.
- Cinesi,C., Yang,B. and Dion,V. (2020) GFP reporters to monitor instability and expression of expanded CAG/CTG repeats. *Methods Mol. Biol. Clifton NJ*, **2056**, 255–268.
- Sanjana,N.E., Shalem,O. and Zhang,F. (2014) Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods*, **11**, 783–784.
- Mali,P., Yang,L., Esvelt,K.M., Aach,J., Guell,M., Dicarlo,J.E., Norville,J.E. and Church,G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823–826.
- Aeschbach,L. and Dion,V. (2017) Minimizing carry-over PCR contamination in expanded CAG/CTG repeat instability applications. *Sci. Rep.*, **7**, 18026.
- Taylor,A.S., Barros,D., Gobet,N., Schuepbach,T., Mcallister,B., Aeschbach,L., Randall,E.L., Trofimenko,E., Heuchan,E.R., Barszcz,P., *et al.* (2022) Repeat Detector: versatile sizing of expanded tandem repeats and identification of interrupted alleles from targeted DNA sequencing. *NAR Genomics Bioinforma*, **4**, lqac089.
- Schindelin,J., Arganda-Carreras,I., Frise,E., Kaynig,V., Longair,M., Pietzsch,T., Preibisch,S., Rueden,C., Saalfeld,S., Schmid,B., *et al.* (2012) Fiji: an open-source platform for biological-image analysis. *Nat. Methods*, **9**, 676–682.
- Santillan,B.A., Moye,C., Mittelman,D. and Wilson,J.H. (2014) GFP-based fluorescence assay for CAG repeat instability in cultured human cells. *PLoS One*, **9**, e113952.
- Ruiz Buendía,G.A., Leleu,M., Marzetta,F., Vanzan,L., Tan,J.Y., Ythier,V., Randall,E.L., Marques,A.C., Baubec,T., Murr,R., *et al.* (2020) Three-dimensional chromatin interactions remain stable upon CAG/CTG repeat expansion. *Sci. Adv.*, **6**, eaaz4012.
- Yang,B., Borgeaud,A.C., Buřičová,M., Aeschbach,L., Rodríguez-Lima,O., Ruiz Buendía,G.A., Cinesi,C., Taylor,A.S., Baubec,T. and Dion,V. (2022) Expanded CAG/CTG repeats resist gene silencing mediated by targeted epigenome editing. *Hum. Mol. Genet.*, **31**, 386–398.
- Debacker,K., Frizzell,A., Gleeson,O., Kirkham-Mccarthy,L., Mertz,T. and Lahue,R.S. (2012) Histone deacetylase complexes promote trinucleotide repeat expansions. *PLoS Biol.*, **10**, e1001257.
- Gannon,A.-M.M., Frizzell,A., Healy,E. and Lahue,R.S. (2012) MutSβ and histone deacetylase complexes promote expansions of trinucleotide repeats in human cells. *Nucleic Acids Res*, **40**, 10324–10333.
- Suelves,N., Kirkham-McCarthy,L., Lahue,R.S. and Ginés,S. (2017) A selective inhibitor of histone deacetylase 3 prevents cognitive deficits and suppresses striatal CAG repeat expansions in Huntington's disease mice. *Sci. Rep.*, **7**, 6082.
- Williams,G.M., Paschalis,V., Ortega,J., Muskett,F.W., Hodgkinson,J.T., Li,G.-M., Schwabe,J.W.R. and Lahue,R.S. (2020) HDAC3 deacetylates the DNA mismatch repair factor MutSβ to stimulate triplet repeat expansions. *Proc. Natl. Acad. Sci.*, **117**, 23597–23605.
- Nakatani,R., Nakamori,M., Fujimura,H., Mochizuki,H. and Takahashi,M.P. (2015) Large expansion of CTG•CAG repeats is exacerbated by MutSβ in human cells. *Sci. Rep.*, **5**, 11020.
- Lin,Y., Dent,S.Y.R., 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.
- Dion,V., Lin,Y., Hubert,L., Waterland,R.A. and Wilson,J.H. (2008) Dnmt1 deficiency promotes CAG repeat expansion in the mouse germline. *Hum. Mol. Genet.*, **17**, 1306–1317.
- 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 (Amst.)*, **8**, 878–885.
- 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.
- Monckton,D.G., Wong,L.J., Ashizawa,T. and Caskey,C.T. (1995) Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: small pool PCR analyses. *Hum. Mol. Genet.*, **4**, 1–8.

35. Leontieva, O.V. and Blagosklonny, M.V. (2013) CDK4/6-inhibiting drug substitutes for p21 and p16 in senescence. *Cell Cycle*, **12**, 3063–3069.
36. Yarrington, R.M., Verma, S., Schwartz, S., Trautman, J.K. and Carroll, D. (2018) Nucleosomes inhibit target cleavage by CRISPR-Cas9 in vivo. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, 9351–9358.
37. Horlbeck, M.A., Witkowsky, L.B., Guglielmi, B., Replogle, J.M., Gilbert, L.A., Villalta, J.E., Torigoe, S.E., Tjian, R. and Weissman, J.S. (2016) Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *eLife*, **5**, e12677.
38. Godde, J.S. and Wolffe, A.P. (1996) Nucleosome assembly on CTG triplet repeats. *J. Biol. Chem.*, **271**, 15222–15229.
39. Wang, Y.-H. and Griffith, J. (1995) Expanded CTG triplet blocks from the myotonic dystrophy gene create the strongest known natural nucleosome positioning elements. *Genomics*, **25**, 570–573.
40. Volle, C.B. and Delaney, S. (2012) CAG/CTG repeats alter the affinity for the histone core and the positioning of DNA in the nucleosome. *Biochemistry*, **51**, 9814–9825.
41. Wang, Y.-H., Amirhaeri, S., Kang, S., Wells, R.D. and Griffith, J.D. (1994) Preferential nucleosome assembly at DNA triplet repeats from the myotonic dystrophy gene. *Science*, **265**, 669–671.
42. Gorbunova, V., Seluanov, A., Dion, V., Sandor, Z., Meservy, J.L. and Wilson, J.H. (2003) Selectable System for Monitoring the Instability of CTG/CAG Triplet Repeats in Mammalian Cells. *Mol. Cell. Biol.*, **23**, 4485–4493.