

ARTICLE

Received 15 Jun 2016 | Accepted 12 Sep 2016 | Published 9 Nov 2016

DOI: 10.1038/ncomms13272

OPEN

# Contracting CAG/CTG repeats using the CRISPR-Cas9 nickase

Cinzia Cinesi<sup>1</sup>, Lorène Aeschbach<sup>1</sup>, Bin Yang<sup>1</sup> & Vincent Dion<sup>1</sup>

CAG/CTG repeat expansions cause over 13 neurological diseases that remain without a cure. Because longer tracts cause more severe phenotypes, contracting them may provide a therapeutic avenue. No currently known agent can specifically generate contractions. Using a GFP-based chromosomal reporter that monitors expansions and contractions in the same cell population, here we find that inducing double-strand breaks within the repeat tract causes instability in both directions. In contrast, the CRISPR-Cas9 D10A nickase induces mainly contractions independently of single-strand break repair. Nickase-induced contractions depend on the DNA damage response kinase ATM, whereas ATR inhibition increases both expansions and contractions in a MSH2- and XPA-dependent manner. We propose that DNA gaps lead to contractions and that the type of DNA damage present within the repeat tract dictates the levels and the direction of CAG repeat instability. Our study paves the way towards deliberate induction of CAG/CTG repeat contractions *in vivo*.

<sup>1</sup>Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland. Correspondence and requests for materials should be addressed to V.D. (email: Vincent.dion@unil.ch).

Repetitive DNA sequences are hotspots for genome instability because they pose a particular challenge to the DNA repair and replication machineries. Their mutation often leads to disease<sup>1</sup>. For example, tracts of CAG/CTG triplets (henceforth referred to as CAG repeats) longer than about 35 units cause at least 14 different currently incurable neurological and neuromuscular diseases<sup>2</sup>. In addition, when CAG repeats expand to pathological lengths, they become highly dynamic and their length changes at high frequencies in both somatic and germ cells throughout the lifetime of an individual<sup>3–6</sup>.

The molecular mechanisms governing CAG repeat instability revolve around the ability of these sequences to fold into non-B-DNA structures when exposed as single-stranded DNA<sup>7–9</sup>. These unusual structures are mistaken for damaged DNA whether or not they contain lesions. The subsequent repair is error-prone due to the repetitive nature of the sequences and their structure-forming ability<sup>3</sup>. Another non-mutually exclusive model suggests that DNA damage within the repeat tract triggers repair, which is, in turn, error-prone due to secondary structures formed by these sequences<sup>5</sup>. In support of these models, several DNA repair pathways promote the instability of expanded CAG repeats, including mismatch repair<sup>10</sup>, double-strand break (DSB) repair<sup>11–14</sup>, transcription-coupled nucleotide excision repair<sup>15,16</sup>, base excision repair (BER)<sup>17,18</sup>, as well as DNA replication<sup>19</sup>. In contrast, single-strand break (SSB) repair (SSBR)<sup>20</sup> and signalling via the DNA damage response (DDR)<sup>21</sup> antagonize CAG repeat instability. Therefore, changes in repeat length provide an opportunity to understand the interaction and interdependence of several different DNA repair pathways at naturally occurring sequences.

Importantly, repeat length determines in large part the severity of the diseases caused by expanded repeats<sup>4</sup>. It has therefore been proposed that contracting the repeat tract would be beneficial in reducing phenotype expression. Repeat expansion, on the other hand, would further exacerbate the disease symptoms<sup>4,22,23</sup>. Currently, there is no treatment that specifically shrinks CAG repeats. This is, in part, because the assays used to measure repeat instability are tedious, slow and/or can only survey instability in one direction. Consequently, the understanding of the mechanism of CAG repeat instability remains poor. Elucidating how contractions can be induced, without also provoking expansions, is critical in designing therapeutic avenues.

Here we present a green fluorescent protein (GFP)-based chromosomal reporter assay that can monitor both CAG repeat expansions and contractions in the same human cell population. We combined this assay with gene-editing tools, namely zinc finger nucleases (ZFNs) and the CRISPR-Cas9 technology. We found that DSBs induced within the expanded repeat tract either by the Cas9 nuclease or a ZFN led to both expansions and contractions. Remarkably, the Cas9 D10A mutant (referred to as the Cas9 nickase) induces instability with a marked bias towards contractions and no detectable off-target mutations. We implicate the DDR kinases ataxia telangiectasia mutated (ATM) and ataxia telangiectasia and Rad3 related (ATR) in promoting contractions and preventing instability, respectively. Moreover, it is not dependent on the SSBR factors XRCC1 and PARylation. Cas9 nickase-induced repeat contraction appears to occur via a pathway different from SSBR- or BER-induced instability. We propose that DNA gaps may be the crucial mutagenic intermediate during nickase-induced contractions. Our results have important implications for gene editing in expanded trinucleotide repeat diseases.

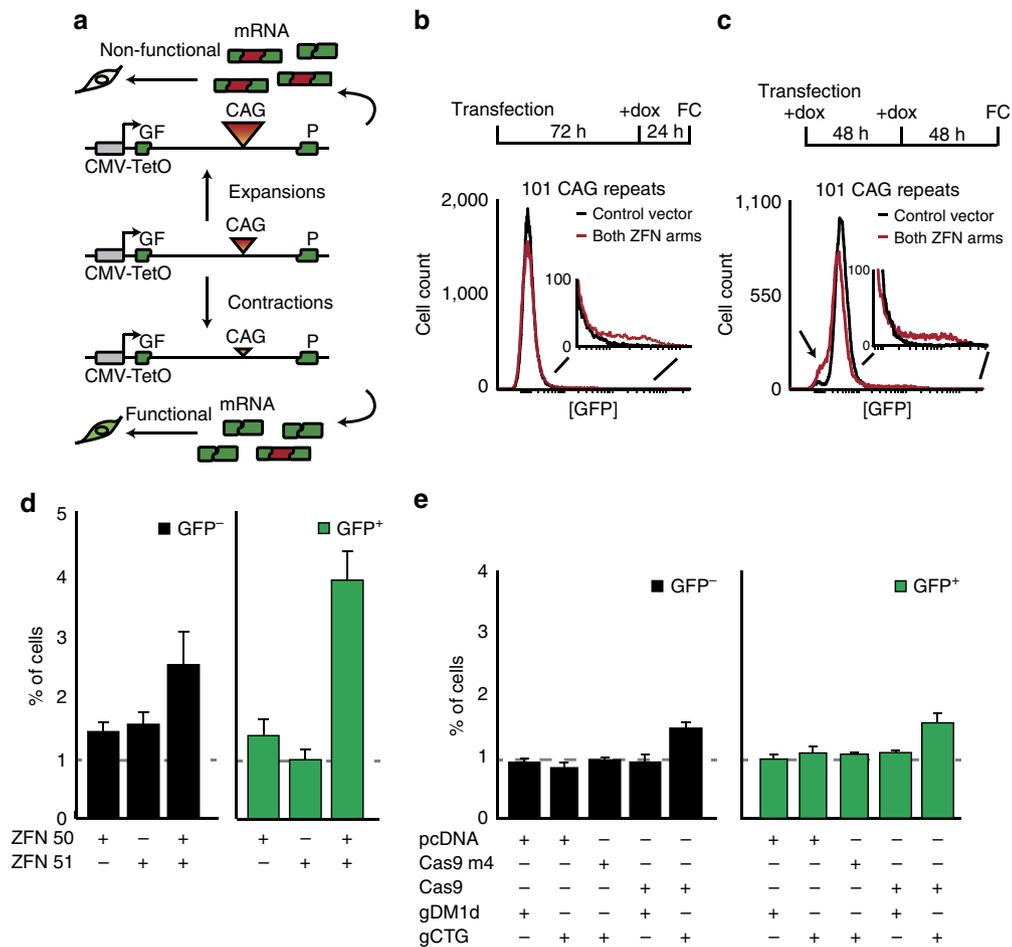
## Results

**A GFP-based assay to detect CAG repeat instability.** We made use of a recently described GFP-based assay capable of detecting

contractions in human cells<sup>24</sup> (Fig. 1a). In this assay, CAG repeats within the intron of a GFP mini-gene interfere with splicing in a repeat length-dependent manner, with longer repeats diminishing GFP production. Thus, GFP intensities, measured by flow cytometry, serve as a proxy for the length of the repeat tract (Supplementary Fig. 1A,B). The reporter is present as a single copy integrated in the genome of human HEK293 T-Rex Flp-In cells. Its transcription is driven by a doxycycline (dox)-inducible promoter. A second isogenic cell line, GFP(CAG)<sub>0</sub>, harbours the same reporter at the same genomic location but is devoid of a CAG repeat. Santillan *et al.*<sup>24</sup> validated the assay by expressing a ZFN that cuts the CAG repeat tract. This treatment increased the number of cells with higher GFP intensities (GFP<sup>+</sup>) in a reporter cell line with 89 repeats (GFP(CAG)<sub>89</sub>) by about 3.5-folds, suggesting that the ZFN treatment induced contractions. They did not report testing for expansions.

To determine whether we could monitor expansions using this assay, we sorted GFP<sup>-</sup> and GFP<sup>+</sup> cells from a population of GFP(CAG)<sub>101</sub> cells using fluorescence-activated cell sorting (FACS). We defined GFP<sup>-</sup> cells as those within the 1% of the cells in the population expressing the least amount of GFP. Similarly, GFP<sup>+</sup> cells are the brightest 1% in the population. From the GFP<sup>-</sup> population, we isolated 19 clones with expansions reaching up to 258 CAGs (Supplementary Fig. 1C). Of the 12 GFP<sup>+</sup> clones isolated, 11 had contractions, the largest of which shrank the repeat tract down to 33 CAGs. The allele sizes in GFP<sup>-</sup> and GFP<sup>+</sup> cells were significantly different ( $P = 1.0 \times 10^{-5}$ , using a Wilcoxon *U*-test), demonstrating that repeat size differences can be detected with this assay. Sequencing the region flanking the CAG repeats also uncovered deletions in five single clones with contractions (Supplementary Fig. 1D). With the exception of one clone that contained a complex rearrangement, the clones with deletions included 2 bp of microhomology at the junction, suggesting that a minor CAG repeat instability pathway is due to an error-prone alternative end-joining mechanism, as suggested recently<sup>25</sup>. Similar results were obtained after sorting cells from populations that were kept in culture for 6 months with or without dox (Supplementary Fig. 1E–H). These results demonstrate that the assay can detect contractions as well as expansions that nearly triple the size of the repeat tract.

**DSBs induce both contractions and expansions.** To determine whether ZFN-induced expansions in addition to the contractions reported by Santillan *et al.*<sup>24</sup>, we first repeated the same experiment. Here we defined GFP<sup>-</sup> and GFP<sup>+</sup> cells as those with GFP intensities in the brightest and dimmest 1% after transfection with the control vector, (Supplementary Fig. 2A—see Methods). We reproduced their results: ZFN expression increased the frequency of GFP<sup>+</sup> cells by 3.2-folds, but had no effect on the number of GFP<sup>-</sup> cells (Fig. 1b). While optimizing the assay, we noted that GFP intensities increased on the addition of dox for 72 h before reaching a steady-state level (Supplementary Fig. 2B). This is in contrast to the 24 h previously reported<sup>24</sup>. Increasing the time of GFP induction raised the overall apparent average intensity of GFP and unmasked an additional GFP<sup>-</sup> cell population only in the sample transfected with both ZFN arms (Fig. 1c). This approach revealed 2.5- and 3.9-fold increases in the proportion of GFP<sup>-</sup> and GFP<sup>+</sup> cells, respectively, on expression of both ZFN arms compared with transfecting an empty vector control (Fig. 1d). Expressing either ZFN arm individually led to only small changes in GFP levels: between 1.3- and 1.4-fold increases in the number of GFP<sup>-</sup> cells and between 0.9- and 1.5-fold for GFP<sup>+</sup> cells (Fig. 1d). Expressing both ZFN arms in the GFP(CAG)<sub>0</sub> cell line had no effect on GFP intensities (Supplementary Fig. 2C), confirming



**Figure 1 | DSBs within CAG repeats lead to expansions and contractions.** (a) GFP-based assay to detect changes in repeat length. (b) Representative flow cytometry profiles after expression of a ZFN in GFP(CAG)<sub>101</sub> using the protocol from ref. 24. (c) Representative flow cytometry profiles with increased dox induction time uncovering an increase in GFP<sup>-</sup> cells on ZFN expression in GFP(CAG)<sub>101</sub> cells (arrow). (d) Quantification of the ZFN experiments in c revealed that ZFN induces the appearance of GFP<sup>-</sup> and GFP<sup>+</sup> cells. ZFNs are composed of two different ZFN arms, each fused to a FokI nuclease that must dimerize to be active. ZFN 50 and ZFN 51 are individual ZFN arms<sup>24</sup>. The dashed line represents the number of cells present in gates set to include the dimmest (GFP<sup>-</sup>) or brightest (GFP<sup>+</sup>) 1% of the cells when a control vector, pcDNA3.1 Zeo, is transfected. Error bars are s.e.m. from 15 replicates for experiments with both ZFN arms, 12 for the single ZFN transfections. (e) Quantification of GFP<sup>-</sup> and GFP<sup>+</sup> cells obtained after expression of the indicated vectors. Dashed line: dimmest (GFP<sup>-</sup>) or brightest (GFP<sup>+</sup>) 1% of the cells transfected with the Cas9 nuclease vector and the empty gRNA plasmid, pPN10. The error bars are s.e.m. Number of replicates per treatment: pcDNA + gDM1d,  $n=3$ ; pcDNA + gCTG,  $n=5$ ; Cas9 m4 + gCTG,  $n=4$ ; Cas9 + gDM1d,  $n=3$ ; Cas9 + gCTG,  $n=7$ . FC, flow cytometry; dox, doxycycline.

that the presence of the repeat tract is necessary. We confirmed that GFP<sup>-</sup> cells contained expansions and GFP<sup>+</sup> cells harboured contractions by sorting cells exposed to both ZFN arms. Of the 9 GFP<sup>-</sup> clones analysed, 8 revealed an expansion (Supplementary Fig. 2D,E). None of them contained deletions and were therefore not GFP<sup>-</sup> because they had lost the GFP reporter. Of the 13 GFP<sup>+</sup> clones, 11 had contractions. Of those, 3 had deletions in the flanking sequences, which is similar to the findings of a previous study constrained to measuring only contractions and using a different ZFN<sup>26</sup>. Here again, the size of GFP<sup>-</sup> and GFP<sup>+</sup> cells in the recovered clones were significantly different ( $P=5 \times 10^{-4}$ , using a Wilcoxon  $U$ -test). These results demonstrate that GFP<sup>-</sup> and GFP<sup>+</sup> cells accurately reflect the presence of expansions and contractions, making this assay especially well suited to detect expansions and contractions quickly within a chromosomal environment.

To confirm that DSBs within the repeat tract lead to both expansions and contractions, we used a second type of programmable nuclease: CRISPR-Cas9. This bacterial nuclease

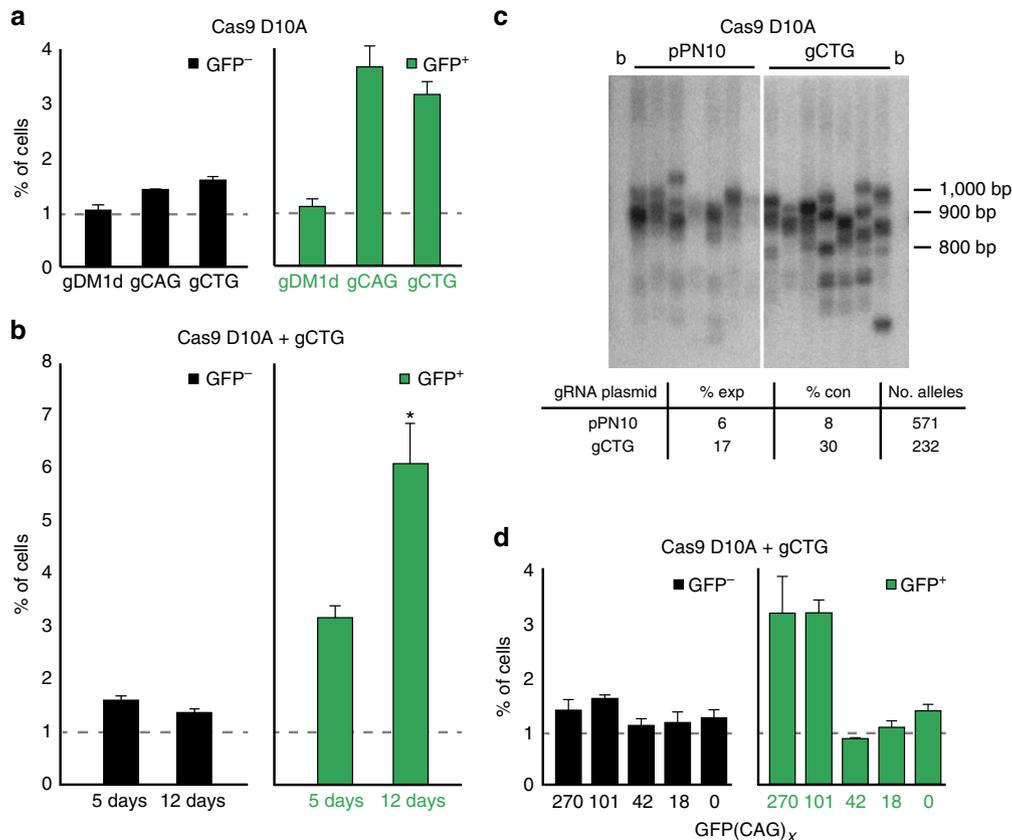
is guided to virtually any sequence of interest by a guide RNA (gRNA) molecule, where it induces blunt-ended DSBs, making it a highly effective gene-editing tool<sup>27-29</sup>. Transfection of a vector expressing a gRNA that targets the unrelated *DMPK* locus (gDM1d) together with Cas9 did not affect GFP expression (Fig. 1e). Similarly, expressing a gRNA containing six CTGs as the target sequence (gCTG) alone, expressing the Cas9 nuclease plus the gCTG in GFP(CAG)<sub>0</sub> cells, or the gCTG together with a catalytically inactive version of Cas9 (Cas9m4) did not change GFP expression significantly (Fig. 1e and Supplementary Fig. 2F). However, expressing the Cas9 nuclease together with gCTG resulted in a meek 1.4- and 1.5-fold induction of GFP<sup>-</sup> and GFP<sup>+</sup> cells, respectively, compared with co-transfecting the Cas9 expression vector with the empty gRNA vector (Fig. 1e). This low efficiency may reflect that the protospacer adjacent motif next to the target sequence of gCTG is not the canonical NGG. We conclude that DSBs induced within the repeat tract by a ZFN or the Cas9 nuclease provoke nearly as many expansions as contractions.

**The Cas9 nickase induces mainly CAG repeat contractions.** The use of the Cas9 enzyme allowed us to test whether the type of DNA damage present within the repeat tract influences CAG repeat instability. The Cas9 D10A mutant can be used with the same gRNA to introduce DNA nicks on the strand complementary to the gRNA<sup>30</sup>. DNA nicks are important intermediates in repeat instability *in vitro*<sup>31,32</sup>. We therefore asked whether inducing DNA nicks with the Cas9 nickase could influence CAG repeat instability.

We found that expressing the Cas9 nickase together with gCTG in GFP(CAG)<sub>101</sub> cells increased the number of GFP<sup>-</sup> cells by 1.6-fold and GFP<sup>+</sup> cells by 3.2-folds compared with cells expressing only the nickase (Fig. 2a). Transfecting the Cas9 nickase with gCAG, which cuts the opposite strand compared with gCTG, had a similar effect, leading to increases of 1.4- and 3.7-folds in GFP<sup>-</sup> and GFP<sup>+</sup> cells, respectively (Fig. 2a). To control for potential indirect effects on GFP expression, we expressed the Cas9 nickase along with gDM1d. This had no effect on GFP expression (Fig. 2a). In addition, the gCTG alone did not increase either GFP<sup>-</sup> or GFP<sup>+</sup> cells, similar to expressing the gCTG together with the Cas9m4 mutant (Fig. 1e), suggesting that the activity of the nickase is necessary. Increasing the number of

transfections to three in the span of 12 days further increased the number of GFP<sup>+</sup> cells to 6.2-folds, without a concomitant change in GFP<sup>-</sup> cells (Fig. 2b and Supplementary Fig. 2G,H;  $P=0.32$  and  $0.001$  for GFP<sup>-</sup> and GFP<sup>+</sup> cells, respectively, using a Wilcoxon  $U$ -test). The Cas9 nickase did not increase the number of dead cells, which could skew the quantification of GFP<sup>-</sup> and GFP<sup>+</sup> cells (Supplementary Table 1). Also, the difference in the number of GFP<sup>+</sup> cells induced between the nuclease and the nickase was not due to differences in expression levels of the Cas9 enzyme (Supplementary Fig. 3A,B). We further confirmed that the way we quantified the data did not induce a bias against expansions (Supplementary Fig. 3C,D). These observations suggest that the Cas9 nickase leads to instability with a bias towards contractions.

To confirm this effect using an assay that is independent of the GFP reporter, we isolated DNA after expressing the Cas9 nickase and gCTG together and performed small-pool PCR (SP-PCR). This method bypasses the inherent advantage in amplification efficiency that smaller alleles have by setting up a larger number of reactions, each with only a few genomes as templates<sup>33</sup>. Using samples treated according to our 12-day regimen, we could detect larger and more frequent contractions in cells exposed to both the



**Figure 2 | The Cas9 nickase causes CAG repeat contraction.** (a) Quantification of the effect of Cas9 nickase expression in GFP(CAG)<sub>101</sub> cells. Number of replicates per treatment: gCTG,  $n=37$ ; gCAG,  $n=3$ ; gDM1d,  $n=3$ . Quantification of gCTG experiments also include results from Cas9 nickase expression treated with DMSO and siRNAs against vimentin. Dashed line: dimmest (GFP<sup>-</sup>) or brightest (GFP<sup>+</sup>) 1% of the cells transfected with the Cas9 nickase vector and the empty gRNA plasmid. (b) Quantification of GFP<sup>-</sup> and GFP<sup>+</sup> cells after 5 days (1 transfection) or 12 days (3 transfections) of expressing the Cas9 nickase along with gCTG in GFP(CAG)<sub>101</sub> cells. Number of replicates per treatment: 5 days,  $n=37$  (same as in a); 12 days,  $n=4$  ( $*P=0.001$ , using a Wilcoxon  $U$ -test, compared with 5-day treatment). Dashed line: dimmest (GFP<sup>-</sup>) or brightest (GFP<sup>+</sup>) 1% of the cells transfected with the Cas9 nickase vector and the empty gRNA plasmid over the indicated period of time. (c) SP-PCR (top) and its quantification (bottom) of DNA isolated from GFP(CAG)<sub>101</sub> cells after Cas9 nickase expression with or without gCTG (shown: 100 pg DNA per PCR) or pPN10 (50 pg DNA per PCR) for 12 days. b, no DNA blank. (d) The ability of the Cas9 nickase to induce contractions is repeat-length dependent. Dashed line: dimmest (GFP<sup>-</sup>) or brightest (GFP<sup>+</sup>) 1% of the cells transfected with the Cas9 nickase vector and the empty gRNA plasmid in the indicated cell line. For each cell line the 1% threshold is determined independently. The error bars are s.e.m. Number of replicates per treatment: GFP(CAG)<sub>270</sub>,  $n=3$ ; GFP(CAG)<sub>101</sub>,  $n=37$  (same as in a); GFP(CAG)<sub>42</sub>,  $n=3$ ; GFP(CAG)<sub>18</sub>,  $n=2$ ; GFP(CAG)<sub>0</sub>,  $n=13$ .

**Table 1 | Effect of the Cas9 nickase targeted by gCTG at CAG/CTG sites in the genome.**

Locus	No. of repeats <sup>†</sup>		No. of alleles sequenced	No. with changes
	Allele 1	Allele 2		
AR	20 + 5	21 + 5	18	0
ATN1	15	16	18	0
ATXN1	12 + 11	12 + 12	18	0
DMPK	5	5	18	0
PPP2R2B	10	10	18	0
TBP	9 + 18	9 + 19	18	0
TCF4	14	17	18	0

See Supplementary Table 1 for sequence composition at these loci.

<sup>†</sup>Alleles from GFP<sup>+</sup> cells sorted from GFP(CAG)<sub>101</sub> cells transfected with the Cas9 nickase and gCTG expressing plasmids.

nickase and gCTG (Fig. 2c). The number of contractions accounted for nearly a third of the total alleles compared with only 8% when cells were transfected with the Cas9 nickase-expressing vector alone (Fig. 2c,  $P < 0.0001$ , using a Fisher's exact test). On nickase expression there was also an increase in the number of expansions, but there were fewer of them and the changes in size were smaller than for the contractions. We conclude that the Cas9 nickase targeted by gCAG or gCTG leads to a marked bias towards contractions, which is in sharp contrast to the results we obtained with the ZFN and the Cas9 nuclease.

We next examined the effect of repeat length on Cas9 nickase-induced contractions. To do so, we used GFP(CAG)<sub>x</sub> cell lines with repeat sizes ranging from 0 to 270 CAGs. We detected slight increases of 1.2- to 1.6-fold in GFP<sup>-</sup> cells on expression of both the Cas9 nickase and gCTG. This effect was largely independent of the repeat size, suggesting that this slight increase in GFP<sup>-</sup> cells seen in GFP(CAG)<sub>101</sub> is only partly caused by changes in repeat length (Fig. 2d). By contrast, the same treatment increased the proportion of GFP<sup>+</sup> cells in GFP(CAG)<sub>270</sub> and GFP(CAG)<sub>101</sub> cells, but not in GFP(CAG)<sub>42</sub>, GFP(CAG)<sub>18</sub> nor GFP(CAG)<sub>0</sub> (Fig. 2d). These observations suggest that normal-length repeats are not prone to instability on action of the Cas9-nickase. We further substantiated this claim by examining the extent of the Cas9-induced changes at seven different loci in the genome harbouring repeats of normal sizes (Supplementary Table 2). We used nine GFP<sup>+</sup> clones with contractions within the GFP reporter caused by the action of the Cas9 nickase guided by gCTG. Of the 126 alleles sequenced, we found that they all remained mutation-free (Table 1), suggesting that the frequency of off-target mutations caused by the nickase is low. Together, these results argue that expanded CAG repeats are targets of the Cas9 nickase, leading predominantly to contractions.

#### SSBR is not involved in Cas9 nickase-induced contractions.

Our results suggest that the type of damage induced within the repeat tract influences instability. It was therefore important to confirm the mutagenic intermediate created by the Cas9 nickase. The simplest hypothesis is that DNA nicks are themselves mutagenic. We therefore tested the effect of X-Ray Repair Cross-Complementing Protein 1 (XRCC1) and Poly (ADP-ribose) polymerase (PARP) activities on nickase-induced instability. The XRCC1-PARP1 complex works as a nick sensor and is involved in their repair<sup>34</sup>. In addition, XRCC1 interacts with a number of DNA glycosylases and is required for the repair of single-nucleotide gaps, i.e., SSBs that arise during BER<sup>35</sup>. This is highly relevant because BER causes expansion in a Huntington disease mouse model<sup>17,18</sup>, and both XRCC1 and PARP1 protect against contractions in a mammalian-based assay that is blind to expansions<sup>20</sup>. Therefore, the prediction was that the knockdown of XRCC1 or the inhibition of PARP using Olaparib would

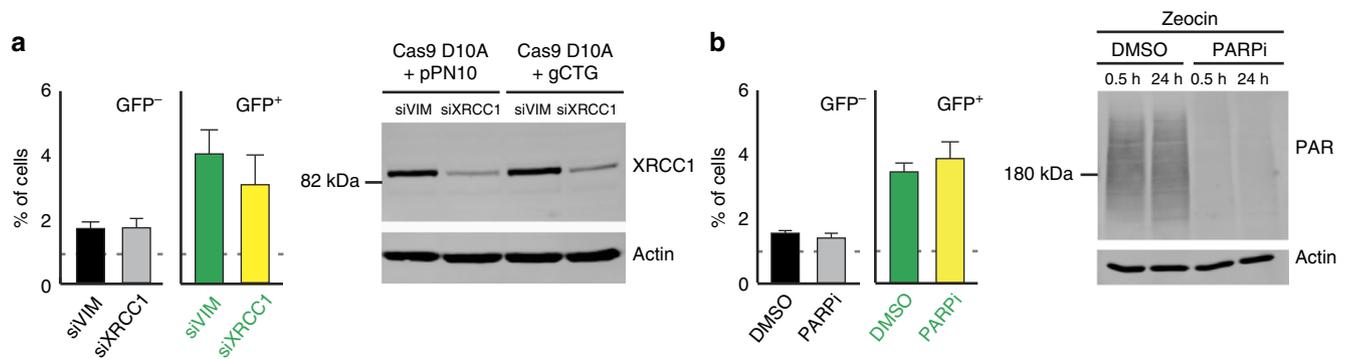
significantly affect the contraction frequencies caused by the Cas9 nickase if DNA nicks or SSBs are mutagenic. This prediction was not confirmed: neither the knockdown of XRCC1 nor the chemical inhibition of PARP activity changed the frequency of GFP<sup>+</sup> cells compared with controls (Fig. 3a,b and Supplementary Fig. 4A,B). We confirmed that the XRCC1 protein levels were substantially reduced and that the Olaparib concentration used led to an accumulation of cells in G2 and that it inhibited PARylation in response to Zeocin assault (Supplementary Table 3 and Fig. 3a,b). These observations suggest that the mutagenic intermediate is neither a DNA nick nor a SSB, and imply that nickase-induced contractions occur through a mechanism that is distinct from spontaneous and BER-dependent CAG repeat instability. We posit instead that DNA gaps larger than a single nucleotide may lead to nickase-induced contractions.

#### ATR and ATM in Cas9 nickase-induced CAG repeat instability.

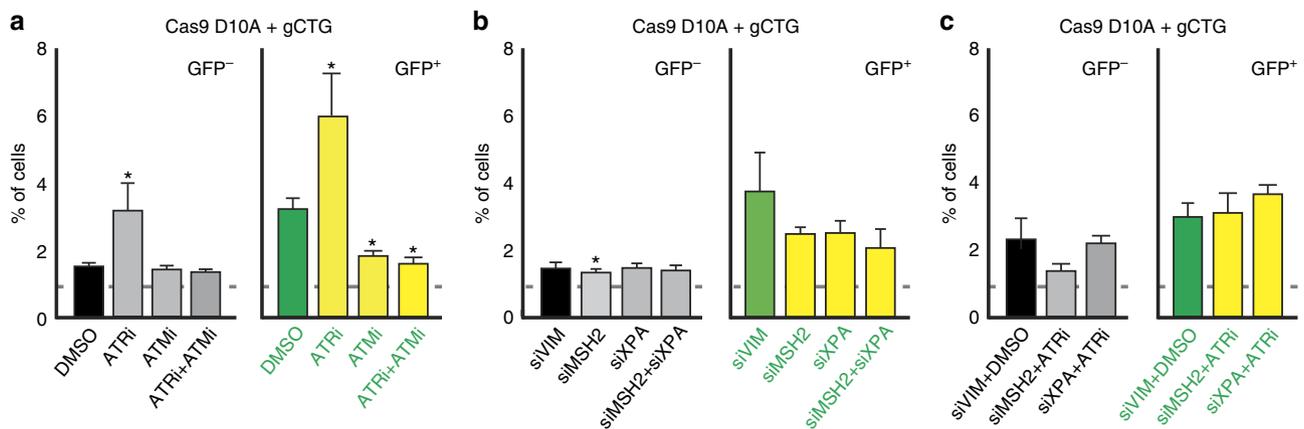
DNA gaps, for example, those induced by ultraviolet light during G1 of the cell cycle, activate ATR<sup>36</sup>. We therefore tested the effect of inhibiting this DDR kinase on nickase-induced instability using the small molecule VE-821 (ref. 37). We found that this inhibitor led to a 3.1- and 5.9-fold increase in GFP<sup>-</sup> and GFP<sup>+</sup> cells, respectively, when used in combination with the Cas9-nickase and gCTG (Fig. 4a,  $P = 0.03$  compared with dimethylsulphoxide (DMSO) treated cells, using a Wilcoxon U-test). This treatment did not affect GFP expression in GFP(CAG)<sub>0</sub> (Supplementary Fig. 4B), confirming that the effect depends on the Cas9 nickase activity within the expanded repeat tract. These data suggest that ATR prevents CAG repeat instability at Cas9 nickase-induced damage.

ATM is a related DDR kinase that is partially redundant with ATR<sup>38</sup>. Thus, we wanted to know what effect ATM might have on nickase-induced contractions. KU60019, a specific inhibitor of ATM<sup>39</sup>, led to a nearly two-fold reduction in the frequency of GFP<sup>+</sup> cells compared to DMSO-treated cells (Fig. 4a,  $P = 0.01$ , using a Wilcoxon U-test). To test whether the effect of ATR was dependent on the activity of ATM, we treated the cells with both inhibitors simultaneously. This double treatment reduced the number of contractions induced by the Cas9 nickase compared with DMSO-treated cells ( $P = 0.03$ , using a Wilcoxon U-test), to a level similar to using the ATM inhibitor alone (Fig. 4a,  $P = 0.57$ , using a Wilcoxon U-test). These observations suggest that the activity of ATM is required to cause nickase-induced instability in the absence of ATR.

**A role for MSH2 and XPA in the absence of ATR activity.** We next aimed to further define how the Cas9 nickase leads to a contraction bias. A central player in CAG repeat instability is MutS Homolog 2 (MSH2), which is essential for mismatch repair. MSH2 knockout in mouse models and its knockdown in human cell-based assays nearly eliminates expansions<sup>10,40,41</sup>. Its role in



**Figure 3 | SSB repair is not involved in Cas9-nickase-induced repeat instability.** (a) Left: XRCC1 knockdown ( $n = 4$ ) did not affect the GFP expression in GFP(CAG)<sub>101</sub> cells ( $P = 0.7$  for both GFP<sup>-</sup> and GFP<sup>+</sup> cells, using a Wilcoxon  $U$ -test). Dashed line: dimmest (GFP<sup>-</sup>) or brightest (GFP<sup>+</sup>) 1% of the cells transfected with the Cas9 nickase vector; the empty gRNA plasmid; and the indicated siRNAs. Right: western blot showing knockdown efficiency. (b) Left: same as a, but cells treated with the PARP inhibitor Olaparib ( $n = 4$ ,  $P = 0.5$  for GFP<sup>-</sup> cells and  $P = 0.4$  for GFP<sup>+</sup> cells compared with DMSO-treated cells, using a Wilcoxon  $U$ -test). Dashed line: dimmest (GFP<sup>-</sup>) or brightest (GFP<sup>+</sup>) 1% of the cells transfected with the Cas9 nickase vector together with the empty gRNA plasmid, and treated with either DMSO or Olaparib. Right: PAR levels 30 min and 24 h after treatment with 100  $\mu\text{g ml}^{-1}$  of Zeocin. The error bars represent the s.e.m.



**Figure 4 | Mechanism of Cas9-nickase-induced repeat instability.** (a) Quantification of GFP<sup>-</sup> and GFP<sup>+</sup> cells on treatment with DMSO ( $n = 20$ , which includes the amount of DMSO for treatment with one or two inhibitors in GFP(CAG)<sub>101</sub> cells; these controls were indistinguishable from each other), an ATR inhibitor (VE-821,  $n = 5$ ), or an ATM inhibitor (KU60019,  $n = 5$ ), or both ( $n = 3$ ). Dashed line: dimmest (GFP<sup>-</sup>) or brightest (GFP<sup>+</sup>) 1% of the cells transfected with the Cas9 nickase vector together with the empty gRNA plasmid and treated with the DMSO or the indicated inhibitor. (b) Quantification of the GFP<sup>-</sup> and GFP<sup>+</sup> cells on knockdown with siRNAs (siVIM,  $n = 13$ ; siMSH2,  $n = 14$ ; siXPA,  $n = 9$ ; siMSH2 + siXPA,  $n = 4$ ). siVIM quantifications include results from knockdown of vimentin with both 10 and 20 nM of siRNAs as the results were indistinguishable from each other. Dashed line: dimmest (GFP<sup>-</sup>) or brightest (GFP<sup>+</sup>) 1% of the cells transfected with the Cas9 nickase vector; the empty gRNA plasmid; and the indicated siRNAs. (c) Quantification of the GFP<sup>-</sup> and GFP<sup>+</sup> cells on combinatorial knockdown of the indicated siRNAs and the ATR inhibitor VE-821 (siVIM + DMSO,  $n = 6$ ; siMSH2 + ATRi,  $n = 3$ ; siXPA + ATRi,  $n = 6$ ). Dashed line: dimmest (GFP<sup>-</sup>) or brightest (GFP<sup>+</sup>) 1% of the cells transfected with the Cas9 nickase vector together with the empty gRNA plasmid, and treated with the indicated inhibitors and siRNAs. The error bars are s.e.m. \* $P \leq 0.05$ .

contraction, however, is more controversial. In mouse models, knocking out *MSH2* either promoted or had no effect on contractions<sup>10</sup>. In human cells *MSH2* downregulation promotes contractions or instability in both directions, depending on the model system used<sup>15,40,42,43</sup>. We found that *MSH2* knockdown did not consistently reduce the number of Cas9 nickase-induced GFP<sup>+</sup> cells compared with a control knockdown of vimentin (Fig. 4b,  $P = 0.14$ , using a Wilcoxon  $U$ -test). *MSH2* promotes CAG repeat contractions together with the NER factor, Xeroderma Pigmentosum, Complementation Group A (*XPA*)<sup>15</sup>, in a human cell-based assay. *XPA* is also required for CAG repeat instability in mouse neuronal tissues<sup>44</sup> and for contractions in a human cell-based assay<sup>15</sup>. It was therefore not surprising that the knockdown of *XPA* alone or in combination with *MSH2* knockdown did not significantly reduce the frequency of nickase-induced GFP<sup>+</sup> cells (Fig. 4b,  $P = 0.18$ , using a

Wilcoxon  $U$ -test, for comparing *XPA* and vimentin (*VIM*) knockdowns; and  $P = 0.07$ , using a Wilcoxon  $U$ -test, when comparing double knockdown to vimentin knockdown). These results argue that neither *MSH2* nor *XPA* are involved in generating contractions at Cas9 nickase-induced lesions.

We reasoned that ATR inhibition may be increasing the number of expansions and contractions because DSB intermediates may form under these conditions. We therefore tested whether the NER pathway, which is known to generate DSBs on ultraviolet damage<sup>45</sup> and at short inverted repeats<sup>46</sup>, could contribute to repeat instability in the absence of ATR activity. Knockdown of *XPA* in cells treated with VE-821 led to results indistinguishable from those obtained when treating cells with DMSO together with a control siRNA (Fig. 4c,  $P = 0.70$ , using a Wilcoxon  $U$ -test). Similarly, the effect of VE-821 treatment was suppressed by *MSH2* knockdown (Fig. 4c,  $P = 0.71$  compared

with control DMSO and vimentin siRNA treatments, using a Wilcoxon *U*-test). These results suggest that expansions and contractions induced by the inhibition of ATR occur because of a XPA- and MSH2-dependent activity that may eventually generates DSBs.

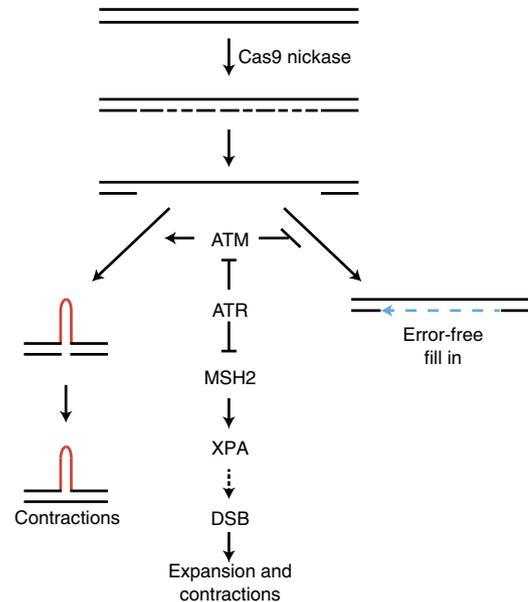
## Discussion

Many assays have been used with great success to dissect the mechanisms of repeat instability. Unfortunately, they are often slow, labour-intensive and/or cannot probe both expansions and contractions at once<sup>15,24,33,47–49</sup>. Here we have adapted a chromosomal-based reporter assay such that it can monitor instability in both directions within only 5 days. We show that the assay can be coupled to pharmaceutical treatments, siRNAs and cDNA overexpression, making it highly versatile and well suited for screening.

DNA nicks appear to be repaired by distinct and still poorly understood mechanisms. For example, they stimulate homology-directed repair in a human cell-based assay<sup>50</sup>. Intriguingly, this process is suppressed by RAD51 and BRCA2, which are required for homologous recombination at DSBs<sup>50</sup>. PARP inhibition also stimulates nick-induced homology-directed repair<sup>51</sup>. Our observation that the same PARP inhibitor has no effect on nickase-induced contraction is suggestive of a different pathway being used at CAG repeats and that DNA nicks are not the mutagenic intermediates leading to nickase-induced contractions. Furthermore, the lack of an effect when knocking down XRCC1 or inhibiting PARP1 implies that the Cas9 nickase leads to contraction via a pathway different from that of BER-generated SSBs. We cannot rule out that DNA nicks lead to contractions independently of the known pathways leading to spontaneous instability. Instead, however, we offer a model (Fig. 5) whereby the Cas9 nickase induces several nicks on the same strand within the repeat tract, thereby generating DNA gaps. This hypothesis is attractive because it provides an explanation for the repeat-length dependency of nickase-induced contractions: shorter repeats have fewer gCTG-binding sites and thus DNA gaps are not created as readily, leading to a stable tract. Together, these observations suggest that different types of DNA lesions found within the repeat tract are repaired by different pathways, which may dictate the direction of repeat instability.

DNA gaps are important intermediates in CAG repeat instability in model systems as varied as yeast and mice<sup>52–54</sup>. How they lead to contraction, however, has remained unclear. In our model (Fig. 5), we propose that DNA gaps caused by the Cas9 nickase are converted to contractions via an ATM-dependent mechanism—perhaps by promoting ligation of single-stranded DNA ends across a hairpin. This intermediate could be further processed or simply replicated in the following cell cycle to create a contraction. DNA gap filling, promoted by ATR, would prevent the involvement of ATM, providing an explanation for the apparent role of ATR in antagonizing ATM. When ATR signalling is compromised, an intermediate, possibly stabilized by MSH2 (ref. 55) and/or XPA<sup>56</sup>, lingers and is processed more often by an XPA-dependent recruitment of downstream nucleases. The resulting DSB is further repaired via the same error-prone pathway that processes ZFN and Cas9-induced DSBs.

The yeast homologue of ATR, Mec1, prevents the appearance of contractions, most likely by preventing DSB formation at expanded CAG repeats<sup>21</sup>. Tel1, the ATM homologue, had no effect on CAG repeat instability<sup>14</sup>. Admittedly, budding yeast displays a bias towards contractions in wild-type cells and may therefore process CAG repeats differently than human cells. Nevertheless, it is unclear why the roles that we have uncovered here should be different than the ones uncovered in yeast. One



**Figure 5 | Model for Cas9-nickase-induced repeat contraction.**

possibility is that the mutagenic intermediates that Mec1 and Tel1 are sensing in the yeast studies were different than those involved here.

ATR and ATM heterozygosities also have distinct effects on the instability of CGG/CCG repeats in mice. In agreement with data presented here, ATR prevents the expansion of CGG/CCG repeats both in somatic tissues as well as in non-replicating prophase I-arrested mouse oocytes<sup>57</sup>. The effect of ATR on contractions was not reported. *Atm*<sup>+/-</sup> animals, by contrast, did not display an overt somatic instability phenotype. Instead, they showed markedly increased frequencies of expansions in the male germlines<sup>58</sup>. The effect on contractions was not reported. The reason for these differences is currently unclear but may include the very different nature of the trinucleotide repeats studied (CAG/CTG versus CGG/CCG), and/or the difference between the human cells used here and the *in vivo* mouse model used in both previous studies. More work is required to resolve this issue.

Our results have profound implications for somatic gene editing of expanded CAG diseases. Programmable nucleases were proposed to provide a tool to shorten repeat tracts and a much needed cure<sup>59</sup>. Some attempts have been made to test this hypothesis using ZFNs or TALENs<sup>24,26,60,61</sup>. Our data caution that inducing DSBs within the repeat tract in an attempt to shrink them would also lead to repeat expansion. This would be a problem because the expansions are likely to exacerbate the disease phenotype<sup>22,23</sup>. An alternative may be to induce two DSBs in regions immediately flanking, but not within, the repeat tract. This approach would be prone to off target effects<sup>62–64</sup>, may lead to the mutation of the wild-type allele, and gRNAs would have to be designed and tested for each disease locus. Our approach would be simpler, using a single gRNA that could target any of the disease loci. In addition, our data argue that only longer, pathogenic, repeat tracts are targeted for contractions; an ideal scenario as it leaves the normal allele intact.

For CAG repeat contraction to be a viable therapeutic avenue, the disease phenotypes must be reversible. There is some evidence that this is the case. Indeed, the myotonia and cardiac symptoms of a myotonic dystrophy mouse model were reversible on shutting off the expression of the pathogenic transgene<sup>65</sup>. Similarly, halting the expression of a spinocerebellar ataxia type 1 allele with 82 CAGs markedly improved the pathological

phenotype of Purkinje cells and reversed motor dysfunction<sup>66</sup>. Homology-directed replacement of an expanded CAG repeat in induced pluripotent stem cells (iPSC) derived from Huntington disease patients improved susceptibility to cell death and mitochondrial defects<sup>67</sup>. Removing a CGG/CCG repeat tract along with flanking sequences from the *FMRI* gene with CRISPR-Cas9 nuclease reactivated the expression of FMRP in a few iPSC clones<sup>68</sup>. Finally, excising expanded GAA/TTC repeats in Friedreich Ataxia fibroblasts reactivated the expression of frataxin, improved the activity of the Fe-S-containing Aconitase, and increased cellular ATP levels<sup>69</sup>. Together with our results, these studies offer great hope that Cas9 nickase-mediated shrinkage of expanded repeat tracts in somatic tissues may alleviate disease symptoms in patients.

## Methods

**Cell culture.** The GFP(CAG)<sub>0</sub> and GFP(CAG)<sub>101</sub> cells lines were a kind gift from John H. Wilson<sup>24</sup>. The cells tested negative for mycoplasma using the MycoAlert detection kit (Lonza) at the start of our experiments and during the revisions of this manuscript. The GFP(CAG)<sub>15</sub>, GFP(CAG)<sub>18</sub>, GFP(CAG)<sub>42</sub>, GFP(CAG)<sub>50</sub> and GFP(CAG)<sub>270</sub> were isolated from populations grown for 6 months unperturbed or after transfection with the ZFN. They did not contain mutations in the region flanking the repeat tract. The cells were maintained at 37 °C with 5% CO<sub>2</sub> in Dulbecco's modified Eagle's medium (DMEM) glutamax, supplemented with 10% fetal bovine serum (FBS), 100 U ml<sup>-1</sup> penicillin (pen), 100 µg ml<sup>-1</sup> streptomycin (strep), 15 µg ml<sup>-1</sup> blasticidine and 150 µg ml<sup>-1</sup> hygromycin. When the cells were destined for flow cytometry, they were kept in DMEM glutamax, with 10% of dialysed calf serum, along with pen-strep. During the long-term culturing, the untransfected and unperturbed cells were split one to five twice a week, and the medium was supplemented with blasticidine and hygromycin to ensure continued expression of the TetR and GFP transgenes.

**Plasmids and siRNA transfections.** The plasmids used in this study are found in Supplementary Table 4. They are available on request. cDNA transfections were performed using 6 × 10<sup>5</sup> cells per well in 12-well plates using a total of 1 µg of DNA and Lipofectamine 2000 (Life Technologies) per well. The culture medium was replaced 6 h after transfection and 2 µg ml<sup>-1</sup> of dox, diluted in DMSO, was added. Controls without dox were treated with DMSO alone. Forty-eight hours later, the medium was replaced and dox was freshly added. Flow cytometry, protein extraction and/or DNA extraction were performed after another 48 h of incubation.

The siRNAs used in this study are found in Supplementary Table 5. When transfecting with both a cDNA and a siRNA, 8 × 10<sup>5</sup> cells per well were used along with 1 µg of DNA and 20 nM of siRNAs using Lipofectamine 2000. The medium was replaced 6 h later and dox was added. Forty-eight hours after the first transfection, we performed a second siRNA transfection with RNAiMax (Life Technologies) using half of the cells present and 20 nM of siRNA. We collected the cells to assess knockdown efficiency or GFP fluorescence analysis 48 h later. When transfecting two siRNAs, we used a final siRNA concentration of 40 nM, where 20 nM of each individual siRNA were used. We found that single knockdowns at 20 nM were no different from those also containing 20 nM of the vimentin siRNA and were pooled for the statistical analyses and in the presented figures.

**Pharmacological inhibitors.** When using small-molecule inhibitors (Supplementary Table 6), the cells were treated as above. The medium, along with the dox and the inhibitors, was replaced after 48 h and for another 48 h of treatment. Cell cycle analysis was performed after 96 h of treatment. Briefly, the cells were fixed with 100% ethanol and treated with RNaseA (50 µg ml<sup>-1</sup>) before adding propidium iodide (50 µg ml<sup>-1</sup>). Flow cytometry analysis was performed as described below.

**Flow cytometer and cell sorting.** In preparation for flow cytometry analysis, cells were re-suspended in phosphate-buffered saline (PBS) with 1 mM EDTA to a concentration of about 10<sup>6</sup> cells per ml. For each condition, we measured at least 2 × 10<sup>5</sup> events using a LSRII from BD. Data analysis was done using Flowing II. FACS was performed using a FACS Aria II (BD) or MoFlo Astrios (Beckman Coulter). For single-clone analyses, we re-suspended the cells to a concentration of 2 × 10<sup>6</sup> cells per ml and sorted the GFP<sup>-</sup> and GFP<sup>+</sup> cells. The cells were then expanded in DMEM glutamax supplemented with pen-strep, blasticidine, hygromycin, 5% FBS and 5% dialysed calf serum. For viability tests, cells were treated as described above except that 96 h after the first transfection they were collected in PBS with 1 mM EDTA, and 1 µM of TO-PRO-3 was added as a dead cell marker.

**Quantification of GFP<sup>-</sup> and GFP<sup>+</sup> cells.** To quantify the fold increase in the number of GFP<sup>-</sup> or GFP<sup>+</sup> cells, we first established gates that contained the top or bottom 1% of GFP-expressing cells in the control treatment, for example, the nickase plasmid transfected together with an empty gRNA vector (pPN10). For each treatment or cell line, therefore, the top and bottom 1% were adjusted to take

any shift in GFP expression into account. In some cases, we adjusted the voltage of the flow cytometer laser to accommodate samples with very high or very low GFP expression. This adjustment did not interfere with the quantification (Supplementary Fig. 3C,D). Once the GFP gates were established, we calculated the percentage of cells from the test population (for example, expressing both the Cas9 nickase and the gCTG) falling within these same gates. In cases where inhibitors or siRNAs were used, the control population expressed the Cas9 nickase, pPN10 and the inhibitor or siRNA. The 1% cutoffs were used to keep a balance between having enough cells for robust statistics and detecting significant fold changes<sup>24</sup>. This method probably underestimates the frequencies of change compared with SP-PCR (Fig. 2).

**Repeat length determination and SP-PCR.** To determine the repeat length of each sorted clone, we isolated DNA using the PeqGold MicroSpin Tissue DNA kit (PeqLab). The DNA was then amplified with primers oVIN-0437 and oVIN-0459 (Supplementary Table 7). Several PCR reactions were set-up with MangoTaq and the products were gel-extracted, pooled and sent for sequencing with the same primers used for the amplification. The repeat size was determined from at least two different amplification and sequencing reactions. The longest repeat size determined was used in the rare cases where the repeat length was not identical between the runs. SP-PCR was done based on the protocol described in ref. 70. Briefly, primers oVIN-0459 and oVIN-0460 were used for the amplification along with between 50 and 100 pg of genomic DNA per PCR. The products were then run on an agarose gel and transferred into a membrane. The probe was derived from a PCR product amplified with the same primers from a plasmid containing 40 repeats. The primers used to amplify the off-target loci are found in Supplementary Table 7.

**Antibodies and western blotting.** Protein extraction was done using RIPA buffer and proteinase inhibitor cocktail tablets (Roche, Germany) and at least 10 µg of proteins were loaded onto 6 or 10% Tris/glycine SDS polyacrylamide gels and transferred onto nitrocellulose membranes. The antibodies used in this study are found in Supplementary Table 8. An Odyssey Infrared Imager (Licor) was used for signal detection. All uncropped western blots are found in Supplementary Fig. 5.

**Statistics.** When determining whether there were differences in the frequency of GFP<sup>-</sup> and GFP<sup>+</sup> cells between treatments, we were unable to guarantee that the data were normally distributed using a two-tailed Kolmogorov-Smirnov test. We therefore used a two-tailed Wilcoxon *U*-test as it is non-parametric. We also performed two-tailed Student's *t*-tests, which gave similar results as the *U*-tests. The same was true when comparing length of the repeat tracts in clones sorted from different populations. We used a Poisson distribution to evaluate the total number of alleles amplified in our SP-PCR experiments based on the proportion of PCRs that did not yield a detectable product. Fisher's exact tests were used to determine whether there were changes in the number of contractions and expansions seen in the SP-PCR experiment. All statistical analyses were done using R Studio version 0.99.441. We concluded that a significant difference existed when *P* < 0.05.

**Data availability.** The data presented in this study are available from the corresponding author.

## References

- Wang, G. & Vazquez, K. M. Impact of alternative DNA structures on DNA damage, DNA repair, and genetic instability. *DNA Repair (Amst.)* **19**, 143–151 (2014).
- Orr, H. T. & Zoghbi, H. Y. Trinucleotide repeat disorders. *Annu. Rev. Neurosci.* **30**, 575–621 (2007).
- Dion, V. Tissue specificity in DNA repair: lessons from trinucleotide repeat instability. *Trends Genet.* **30**, 220–229 (2014).
- Lopez Castel, A., Cleary, J. D. & Pearson, C. E. Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat. Rev. Mol. Cell Biol.* **11**, 165–170 (2010).
- McMurray, C. T. Mechanisms of trinucleotide repeat instability during human development. *Nat. Rev. Genet.* **11**, 786–799 (2010).
- Usdin, K., House, N. C. & Freudenreich, C. H. Repeat instability during DNA repair: insights from model systems. *Crit. Rev. Biochem. Mol. Biol.* **50**, 142–167 (2015).
- Lee, D. Y. & McMurray, C. T. Trinucleotide expansion in disease: why is there a length threshold? *Curr. Opin. Genet. Dev.* **26**, 131–140 (2014).
- Mirkin, S. M. Expandable DNA repeats and human disease. *Nature* **447**, 932–940 (2007).
- Axford, M. M. *et al.* Detection of slipped-DNAs at the trinucleotide repeats of the myotonic dystrophy type I disease locus in patient tissues. *PLoS Genet.* **9**, e1003866 (2013).
- Schmidt, M. H. & Pearson, C. E. Disease-associated repeat instability and mismatch repair. *DNA Repair (Amst)* **38**, 117–126 (2016).

11. Su, X. A., Dion, V., Gasser, S. M. & Freudenreich, C. H. Regulation of recombination at yeast nuclear pores controls repair and triplet repeat stability. *Genes Dev.* **29**, 1006–1017 (2015).
12. Marcadier, J. L. & Pearson, C. E. Fidelity of primate cell repair of a double-strand break within a (CTG)<sub>n</sub>(CAG) tract. Effect of slipped DNA structures. *J. Biol. Chem.* **278**, 33848–33856 (2003).
13. Richard, G. F., Dujon, B. & Haber, J. E. Double-strand break repair can lead to high frequencies of deletions within short CAG/CTG trinucleotide repeats. *Mol. Gen. Genet.* **261**, 871–882 (1999).
14. Sundararajan, R., Gellon, L., Zunder, R. M. & Freudenreich, C. H. Double-strand break repair pathways protect against CAG/CTG repeat expansions, contractions and repeat-mediated chromosomal fragility in *Saccharomyces cerevisiae*. *Genetics* **184**, 65–77 (2010).
15. Lin, Y., Dion, V. & Wilson, J. H. Transcription promotes contraction of CAG repeat tracts in human cells. *Nat. Struct. Mol. Biol.* **13**, 179–180 (2006).
16. Jung, J. & Bonini, N. CREB-binding protein modulates repeat instability in a *Drosophila* model for polyQ disease. *Science* **315**, 1857–1859 (2007).
17. Kovtun, I. V. *et al.* OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells. *Nature* **447**, 447–452 (2007).
18. Mollersen, L. *et al.* Neil1 is a genetic modifier of somatic and germline CAG trinucleotide repeat instability in R6/1 mice. *Hum. Mol. Genet.* **21**, 4939–4947 (2012).
19. Yang, Z., Lau, R., Marcadier, J. L., Chitayat, D. & Pearson, C. E. Replication inhibitors modulate instability of an expanded trinucleotide repeat at the myotonic dystrophy type 1 disease locus in human cells. *Am. J. Hum. Genet.* **73**, 1092–1105 (2003).
20. Hubert, Jr L., Lin, Y., Dion, V. & Wilson, J. H. Topoisomerase 1 and single-strand break repair modulate transcription-induced CAG repeat contraction in human cells. *Mol. Cell. Biol.* **31**, 3105–3112 (2011).
21. Lahiri, M., Gustafson, T. L., Majors, E. R. & Freudenreich, C. H. Expanded CAG repeats activate the DNA damage checkpoint pathway. *Mol. Cell* **15**, 287–293 (2004).
22. Budworth, H. *et al.* Suppression of somatic expansion delays the onset of pathophysiology in a mouse model of Huntington's disease. *PLoS Genet.* **11**, e1005267 (2015).
23. Wheeler, V. C. *et al.* Mismatch repair gene Msh2 modifies the timing of early disease in Hdh(Q111) striatum. *Hum. Mol. Genet.* **12**, 273–281 (2003).
24. Santillan, B. A., Moye, C., Mittelman, D. & Wilson, J. H. GFP-based fluorescence assay for CAG repeat instability in cultured human cells. *PLoS ONE* **9**, e113952 (2014).
25. Chatterjee, N., Lin, Y., Yotnda, P. & Wilson, J. H. Environmental stress induces trinucleotide repeat mutagenesis in human cells by Alt-nonhomologous end joining repair. *J. Mol. Biol.* **428**, 2978–2980 (2016).
26. Mittelman, D. *et al.* Zinc-finger directed double-strand breaks within CAG repeat tracts promote repeat instability in human cells. *Proc. Natl Acad. Sci. USA* **106**, 9607–9612 (2009).
27. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
28. Sternberg, S. H. & Doudna, J. A. Expanding the biologist's toolkit with CRISPR-Cas9. *Mol. Cell* **58**, 568–574 (2015).
29. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
30. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
31. Panigrahi, G. B., Lau, R., Montgomery, S. E., Leonard, M. R. & Pearson, C. E. 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 (2005).
32. Hou, C., Chan, N. L., Gu, L. & Li, G. M. Incision-dependent and error-free repair of (CAG)<sub>n</sub>/(CTG)<sub>n</sub> hairpins in human cell extracts. *Nat. Struct. Mol. Biol.* **16**, 869–875 (2009).
33. Monckton, D. G., Wong, L. J., Ashizawa, T. & Caskey, C. T. Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: small pool PCR analyses. *Hum. Mol. Genet.* **4**, 1–8 (1995).
34. Caldecott, K. W., Aoufouchi, S., Johnson, P. & Shall, S. XRCC1 polypeptide interacts with DNA polymerase beta and possibly poly (ADP-ribose) polymerase, and DNA ligase III is a novel molecular 'nick-sensor' in vitro. *Nucleic Acids Res.* **24**, 4387–4394 (1996).
35. Horton, J. K. *et al.* XRCC1 and DNA polymerase beta in cellular protection against cytotoxic DNA single-strand breaks. *Cell Res.* **18**, 48–63 (2008).
36. Costanzo, V. *et al.* An ATR- and Cdc7-dependent DNA damage checkpoint that inhibits initiation of DNA replication. *Mol. Cell* **11**, 203–213 (2003).
37. Prevo, R. *et al.* The novel ATR inhibitor VE-821 increases sensitivity of pancreatic cancer cells to radiation and chemotherapy. *Cancer Biol. Ther.* **13**, 1072–1081 (2012).
38. Marechal, A. & Zou, L. DNA damage sensing by the ATM and ATR kinases. *Cold Spring Harb. Perspect. Biol.* **5**, a012716 (2013).
39. Golding, S. E. *et al.* Improved ATM kinase inhibitor KU-60019 radiosensitizes glioma cells, compromises insulin, AKT and ERK prosurvival signaling, and inhibits migration and invasion. *Mol. Cancer Ther.* **8**, 2894–2902 (2009).
40. Du, J., Campau, E., Soragni, E., Jespersen, C. & Gottesfeld, J. M. Length-dependent CTG.CAG triplet-repeat expansion in myotonic dystrophy patient-derived induced pluripotent stem cells. *Hum. Mol. Genet.* **22**, 5276–5287 (2013).
41. Manley, K., Shirley, T. L., Flaherty, L. & Messer, A. Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nat. Genet.* **23**, 471–473 (1999).
42. Nakatani, R., Nakamori, M., Fujimura, H., Mochizuki, H. & Takahashi, M. P. Large expansion of CTG<sup>\*</sup>CAG repeats is exacerbated by MutSbeta in human cells. *Sci. Rep.* **5**, 11020 (2015).
43. Seriola, A. *et al.* Huntington's and myotonic dystrophy hESCs: down-regulated trinucleotide repeat instability and mismatch repair machinery expression upon differentiation. *Hum. Mol. Genet.* **20**, 176–185 (2011).
44. Hubert, Jr L., Lin, Y., Dion, V. & Wilson, J. H. Xpa deficiency reduces CAG trinucleotide repeat instability in neuronal tissues in a mouse model of SCA1. *Hum. Mol. Genet.* **20**, 4822–4830 (2011).
45. Bradley, M. O. & Taylor, V. I. DNA double-strand breaks induced in normal human cells during the repair of ultraviolet light damage. *Proc. Natl Acad. Sci. USA* **78**, 3619–3623 (1981).
46. Lu, S. *et al.* Short inverted repeats are hotspots for genetic instability: relevance to cancer genomes. *Cell Rep.* **10**, 1674–1680 (2015).
47. Cleary, J. D., Nichol, K., Wang, Y. H. & Pearson, C. E. Evidence of cis-acting factors in replication-mediated trinucleotide repeat instability in primate cells. *Nat. Genet.* **31**, 37–46 (2002).
48. Claassen, D. A. & Lahue, R. S. Expansions of CAG.CTG repeats in immortalized human astrocytes. *Hum. Mol. Genet.* **16**, 3088–3096 (2007).
49. Pelletier, R., Farrell, B. T., Miret, J. J. & Lahue, R. S. Mechanistic features of CAG<sup>\*</sup>CTG repeat contractions in cultured cells revealed by a novel genetic assay. *Nucleic Acids Res.* **33**, 5667–5676 (2005).
50. Davis, L. & Maizels, N. Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. *Proc. Natl Acad. Sci. USA* **111**, E924–E932 (2014).
51. Metzger, M. J., Stoddard, B. L. & Monnat, Jr R. J. PARP-mediated repair, homologous recombination, and back-up non-homologous end joining-like repair of single-strand nicks. *DNA Repair (Amst.)* **12**, 529–534 (2013).
52. House, N. C., Yang, J. H., Walsh, S. C., Moy, J. M. & Freudenreich, C. H. NuA4 initiates dynamic histone H4 acetylation to promote high-fidelity sister chromatid recombination at postreplication gaps. *Mol. Cell* **55**, 818–828 (2014).
53. Dae, D. L., Mertz, T. & Lahue, R. S. Postreplication repair inhibits CAG.CTG repeat expansions in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **27**, 102–110 (2007).
54. Kovtun, I. V. & McMurray, C. T. Trinucleotide expansion in haploid germ cells by gap repair. *Nat. Genet.* **27**, 407–411 (2001).
55. Pearson, C. E., Ewel, A., Acharya, S., Fishel, R. A. & Sinden, R. R. Human MSH2 binds to trinucleotide repeat DNA structures associated with neurodegenerative diseases. *Hum. Mol. Genet.* **6**, 1117–1123 (1997).
56. Lin, Y. & Wilson, J. H. Nucleotide excision repair, mismatch repair, and R-loops modulate convergent transcription-induced cell death and repeat instability. *PLoS ONE* **7**, e46807 (2012).
57. Entezam, A. & Usdin, K. ATR protects the genome against CGG.CCG-repeat expansion in Fragile X premutation mice. *Nucleic Acids Res.* **36**, 1050–1056 (2008).
58. Entezam, A. & Usdin, K. ATM and ATR protect the genome against two different types of tandem repeat instability in Fragile X premutation mice. *Nucleic Acids Res.* **37**, 6371–6377 (2009).
59. Richard, G. F. Shortening trinucleotide repeats using highly specific endonucleases: a possible approach to gene therapy? *Trends Genet.* **31**, 177–186 (2015).
60. Liu, G., Chen, X., Bissler, J. J., Sinden, R. R. & Leffak, M. Replication-dependent instability at (CTG)<sub>n</sub> x (CAG) repeat hairpins in human cells. *Nat. Chem. Biol.* **6**, 652–659 (2010).
61. Richard, G. F. *et al.* Highly specific contractions of a single CAG/CTG trinucleotide repeat by TALEN in yeast. *PLoS ONE* **9**, e95611 (2014).
62. Cho, S. W. *et al.* Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* **24**, 132–141 (2014).
63. Kucsu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat. Biotechnol.* **32**, 677–683 (2014).
64. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).
65. Mahadevan, M. S. *et al.* Reversible model of RNA toxicity and cardiac conduction defects in myotonic dystrophy. *Nat. Genet.* **38**, 1066–1070 (2006).

66. Zu, T. *et al.* Recovery from polyglutamine-induced neurodegeneration in conditional SCA1 transgenic mice. *J. Neurosci.* **24**, 8853–8861 (2004).
67. An, M. C. *et al.* Genetic correction of Huntington's disease phenotypes in induced pluripotent stem cells. *Cell Stem Cell* **11**, 253–263 (2012).
68. Park, C. Y. *et al.* Reversion of FMR1 methylation and silencing by editing the triplet repeats in fragile X iPSC-derived neurons. *Cell Rep.* **13**, 234–241 (2015).
69. Li, Y. *et al.* Excision of expanded gaa repeats alleviates the molecular phenotype of Friedreich's ataxia. *Mol. Ther.* **23**, 1055–1065 (2015).
70. Dion, V., Lin, Y., Hubert, Jr L., Waterland, R. A. & Wilson, J. H. Dnmt1 deficiency promotes CAG repeat expansion in the mouse germline. *Hum. Mol. Genet.* **17**, 1306–1317 (2008).

### Acknowledgements

We thank F. Hamaratoglu, J.H. Wilson and members of the Dion laboratory for helpful discussions. H. Ferreira, G. Gourdon, F. Hamaratoglu, J. E. Martin, M. Op, A. Orioli and G.-F. Richard critically read the manuscript. We thank J.H. Wilson and J. Lingner for reagents; P. Nunes and A. Feola for cloning some of the gRNA vectors; and O. Rodriguez Lima for blindly quantifying the SP-PCR blots. This work is supported by a Swiss National Science Foundation professorship (#144789) and by Gebert RUF Stiftung and UNISCIENTIA STIFTUNG within the programme «Rare Diseases – New Approaches» (GRS-060/14) to V.D.

### Author contributions

Everyone contributed to performing the experiments and analysing the results. C.C. and V.D. designed the experiments and wrote the paper.

### Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** Procedure ongoing for European Patent application number EP16165203.7.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Cinesi, C. *et al.* Contracting CAG/CTG repeats using the CRISPR-Cas9 nickase. *Nat. Commun.* **7**, 13272 doi: 10.1038/ncomms13272 (2016).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016