Cas9 nickase-mediated contraction of CAG/CTG repeats at multiple disease loci

Alvaro Murillo[1], Melanie Alpaugh[2], Meghan Larin[1,3], Emma L. Randall[1], Laura Heraty[1], Ruban Rex Durairaj[1], Alys N. Aston[1], Alysha S. Taylor[1], Alex Mas Monteys[4], Nina Stöberl[5], Aeverie E. R. Heuchan[1], Pascale Aeschlimann[1], Soumyasree Bhattacharyya[1], Nicholas D. Allen[5], Jack Puymirat[6], Beverly L. Davidson[4], Francesca Cicchetti[7], Mariah Lelos[5], and Vincent Dion[1,8*]

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

2: Department of Molecular and Cellular Biology, University of Guelph, Guelph, Ontario, Canada N1G 2W1

3: Current address: AstraZeneca R&D 1 Francis Crick Avenue Cambridge, UK CB2 0AA.

4: Raymond G. Perelman Center for Cellular and Molecular Therapeutics, Children's Hospital of Philadelphia, Philadelphia, 19104, USA.

5: School of Biosciences, Cardiff University, The Sir Martin Evans Building, Museum Avenue, Cardiff, CF10 3AX

6: LOEX, CHU de Québec-Université Laval Research Center, Quebec City, QC, Canada.

7: Centre de Recherche du CHU de Québec, Axe Neurosciences; Département de Psychiatrie et Neurosciences, Université Laval, Québec, QC, Canada.

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

*: corresponding author: dionv@cardiff.ac.uk

Keywords: CRISPR; Cas9 nickase, gene editing, off-target mutations, Huntington's disease, myotonic dystrophy type 1.

Abstract:

Expanded CAG/CTG repeats cause at least 15 different neurodegenerative and neuromuscular diseases that all remain without an effective disease modifying treatment. Because the size of the repeat tract accounts for the majority of the variation in disease severity, contracting them presents an attractive therapeutic avenue. Here, we show that the CRISPR-Cas9 nickase targeting the CAG/CTG repeat itself leads to efficient contractions in Huntington's disease patient-derived neurons and astrocytes, as well as in myotonic dystrophy type 1 patient-derived neurons. Using single-cell DNA sequencing, PCR-free whole genome sequencing, and targeted long-read sequencing of the *HTT* locus, we found no off-target mutations above background in neurons and astrocytes. Furthermore, we delivered the Cas9 nickase and sgRNA stereotactically to a mouse model of Huntington's disease using adeno-associated viruses, and found contractions accumulating in over half of the infected cells over a period of 5 months. We also found that the Cas9 nickase was prone to silencing, further improving the safety of the approach. Our results provide the proof of concept for using the Cas9 nickase to contract the repeat tract safely in multiple cell types and diseases.

Introduction

The heterozygous expansion of CAG/CTG repeats at 15 different loci in the genome causes clinically distinct neurodegenerative and neuromuscular disorders, including Huntington's disease (HD) and myotonic dystrophy type 1 (DM1)[1,2]. They are all currently without a disease modifying treatment[3,4]. The size of the expanded repeat tract explains up to 60% of the variation in the age at disease onset in HD[5], with longer repeats leading to more severe phenotypes. This is compounded by repeat expansions accumulating over time in somatic tissues, especially in those cell types that are particularly affected by the disease[6–8]. Moreover, DNA repair factors involved in somatic expansion have been identified as disease modifiers[9–12]. These observations suggest that somatic expansion is a major driver of disease pathogenesis and that contractions in somatic tissues may provide a much needed therapeutic avenue.

Several studies have used gene editing *in vivo* at the *DMPK* and *HTT* loci, where a repeat expansion causes DM1 and HD, respectively. Most of these approaches aimed to excise the repeat locus[13–21]. They used the Cas9 nuclease or the Cas9D10A nickase targeted with multiple single guide RNAs (sgRNAs) to the sequences flanking the repeat to induce double-strand breaks. This led to editing efficiencies ranging from 2% to 54%. Critically, given the error-prone nature of double-strand break repair[22,23], it is not possible to control the exact insertion and deletion events that are introduced. Consequently, edited cells have a range of different repair products, even when a template is provided for homology-directed repair[24]. Experiments in cellular models have uncovered examples of frequent on-target rearrangements, including repeat expansions and inversions upon double-strand break-based editing[25–27]. Off-target mutations are also concerning for gene editing[28,29]. Most *in vivo* studies that excised the repeats reported no off-target mutations, save one where editing efficiencies at a predicted off-target site was as high as at the on-target locus[19]. This was observed despite the use of a double nicking approach to create the double-strand break, which is thought to be less prone to off-target mutations[30]. Taken together, these results show that double-strand breaks often cause unintended mutations both on and off target. They should be avoided. Choi *et al* addressed this by using a cytosine deaminase fused to the Cas9D10A nickase[31]. Targeted to the repeat tract, this fusion produced CAA

interruptions within the *HTT* CAG tract and reduced repeat expansion, rather than removing or contracting the repeat tract. The hope is that by preventing expansions, disease severity will be alleviated in the long term.

Despite patients having a heterozygous expansion, typical gene editing approaches mutate both alleles, which may have undesirable consequences[3,32,33]. Three studies have circumvented this problem by designing sgRNAs that target a specific single-nucleotide polymorphism distinguishing the two *HTT* alleles[14,20,21]. However, this approach reduces the total number of individuals who stand to benefit since the most common variant is found in only 30% of HD individuals of European ancestry[21].

Altogether, these results highlight the need to refine gene editing approaches such that they are highly efficient, allele-selective, do not induce mutagenic double-strand breaks, or on- and off-target mutations, whilst being applicable to as many individuals as possible, regardless of the disease.

We have previously shown that the Cas9D10A nickase could induce contractions of an ectopic CAG repeat in HEK293-derived cells when targeted directly to the repeat tract using a single sgRNA[34,35]. The single-strand breaks, or nicks, induced contractions in over 30% of the cells within 2 weeks, whereas introducing double-strand breaks with the Cas9 nuclease provoked both expansions and contractions. Cas9D10A-induced contractions were allele-length specific with repeat sizes of 42 or fewer remaining stable, and those with 101 repeats or more contracted, often to below the pathogenic threshold, which is 35 CAG repeats for HD and about 50 CTG repeats for DM1. The results suggested that contractions only occurred on the expanded allele and that off-target mutations were at a minimum[34].

Here we provide the proof of concept that using the Cas9D10A nickase targeted to the repeat tract induces contractions in disease-relevant cells and *in vivo*. We found this system to be free of unwanted on- and off-target mutations, contracting the repeat tract in an allele selective manner, and working at both the *HTT* and *DMPK* loci.

Results

Nickase-induced contractions in patient-derived cells.

We first investigated the possibility of contracting CAG/CTG repeats in disease-relevant cells. Specifically, we used HD astrocytes derived from the induced pluripotent stem cell (iPSC) line CS09iHD109-n1, which harbours 130 to 140 CAG/CTG repeats at the *HTT* locus[36]. Astrocytes are highly relevant to HD pathogenesis as they display changes in glutamate uptake leading to excitotoxicity[37,38], aggregate formation[39], and neuronal death[40,41]. We differentiated HD iPSCs to astrocyte progenitors and then to mature S100β$^+$ astrocytes (Fig. 1a). The cells were sorted using CD44, ensuring that they were all of an astrocytic lineage. We found little replication in these cultures over the 6 week experiments as assessed by Ki67 staining, a measure of proliferation potential (Supplementary Fig. 1a). We transduced differentiated astrocytes with a lentivirus (Supplementary Table 1) expressing Cas9D10A and Blasticidin S deaminase, which we used to select transduced cells. A second virus expressed the sgRNA against the repeat tract (sgCTG) and contained a GFP cassette to monitor silencing of the lentiviral vector. We found that GFP expression was sustained over the 6 week time course of the experiment (Supplementary Fig. 1b). Similarly, Cas9D10A was expressed 3 and 6 weeks post-transduction (Supplementary Fig. 1c). Six weeks after transduction, we assessed repeat size distribution using small-pool PCR (Fig. 1b), the gold standard for determining repeat size mosaicism[42]. We observed a 6-fold increase in the number of alleles shorter than 102 CAGs when sgCTG was present together with Cas9D10A (27% of alleles ≤101 CAGs) compared to cells transduced with Cas9D10A only (4.3% of alleles ≤101 CAGs - Fig. 1c). Importantly, we also found a decrease in the number of large alleles (>134 CAGs) in these cultures (10%) compared to populations transduced with Cas9D10A only (23%; P≤0.0001). Overall, 6 weeks post-transduction, the cells treated with both Cas9D10A and sgCTG had lost on average 20 repeats compared to those that were treated with Cas9D10A only. These results suggest that HD iPSCs S100β$^+$ astrocyte cultures undergo robust contractions when exposed to both the nickase and sgCTG and are accompanied by a decrease in expansions.

In HD, the apparent driving factor of pathogenesis is neuronal death, which occurs first in the striatum with medium spiny neurons, followed by neurons in the cortex[43]. We differentiated the HD iPSCs into cortical neurons (Fig. 1d) as the protocol is well established and can be done robustly. Indeed, these neurons contained 99% of

□-tubulin positive cultures (Supplementary Fig. 1d). We exposed them to the nickase and sgCTG for 3 or 6 weeks in the presence of blasticidin to select for transduced cells and performed small-pool PCRs (Supplementary Fig. 1e, Fig. 1f). We found a 2-fold increase in the frequency of alleles shorter than 127 CAGs in iPSC-derived neurons comparing Cas9D10A + sgCTG transduced cells (14%) to the population without sgCTG (6.7%) after 3 weeks (Supplementary Fig. 1f, $P<0.023$). The longer alleles changed little during that time with 4.5% of non-transduced cells containing alleles of 160 CAGs or more compared to 6.2% in Cas9D10A plus sgCTG-treated cells. A different set of cultures analysed at 6 weeks showed a 7-fold increase in the frequency of alleles with fewer than 127 CAGs and a 3-fold decrease in the frequency of alleles above 160 repeats when comparing neuronal cultures transduced with Cas9D10A only and ones transduced with both Cas9D10A and sgCTG (Fig. 1e-f, $P\leq0.0003$). We confirmed this increase in contraction frequencies using single-molecule real-time (SMRT) long read sequencing (Supplementary Fig. 2a-d). These results suggest that the Cas9 nickase can induce contractions of the repeat tract in HD iPSC-derived cortical neurons.

The advantage of our design is that the repeat tract itself is targeted and, therefore, it should induce contractions at multiple disease loci. To test this directly, we obtained DM1 iPSCs and differentiated them into cortical neurons (Fig. 1g). Although DM1 displays debilitating myotonia and muscle wasting, the brain is extensively affected, including the cortex[44]. In these cultures, over 98% of the cells stained for □-tubulin III (Supplementary Fig. 1g). As expected for repeat expansion sizes in DM1, this line was found to have a mode of 858 CTGs at the *DMPK* locus. However, due to rarer alleles reaching over 2500 CTGs in the cultures (as measured by small-pool PCRs), the average repeat size was 996 CTGs (Fig. 1h). Over the span of 6 weeks, we found that the number of alleles with ≤580 repeats increased from 9.1% in the DM1 neurons treated with Cas9D10A only to 24% in cells treated with both Cas9D10A and sgCTG (Fig. 1i, $P=0.0003$,). This was further accompanied by the loss of larger alleles (>913) from 17% to 11%, primarily coming from alleles with >2500 CTGs, suggesting a size-specific effect. Although this was not sufficient to lead to repeat sizes in the non-pathogenic range, the cells lost on average 166 repeats over the course of 6 weeks. We conclude that the Cas9D10A nickase can induce contractions

in at least two distinct disease loci without changing the Cas9 nickase construct or the sgRNA.

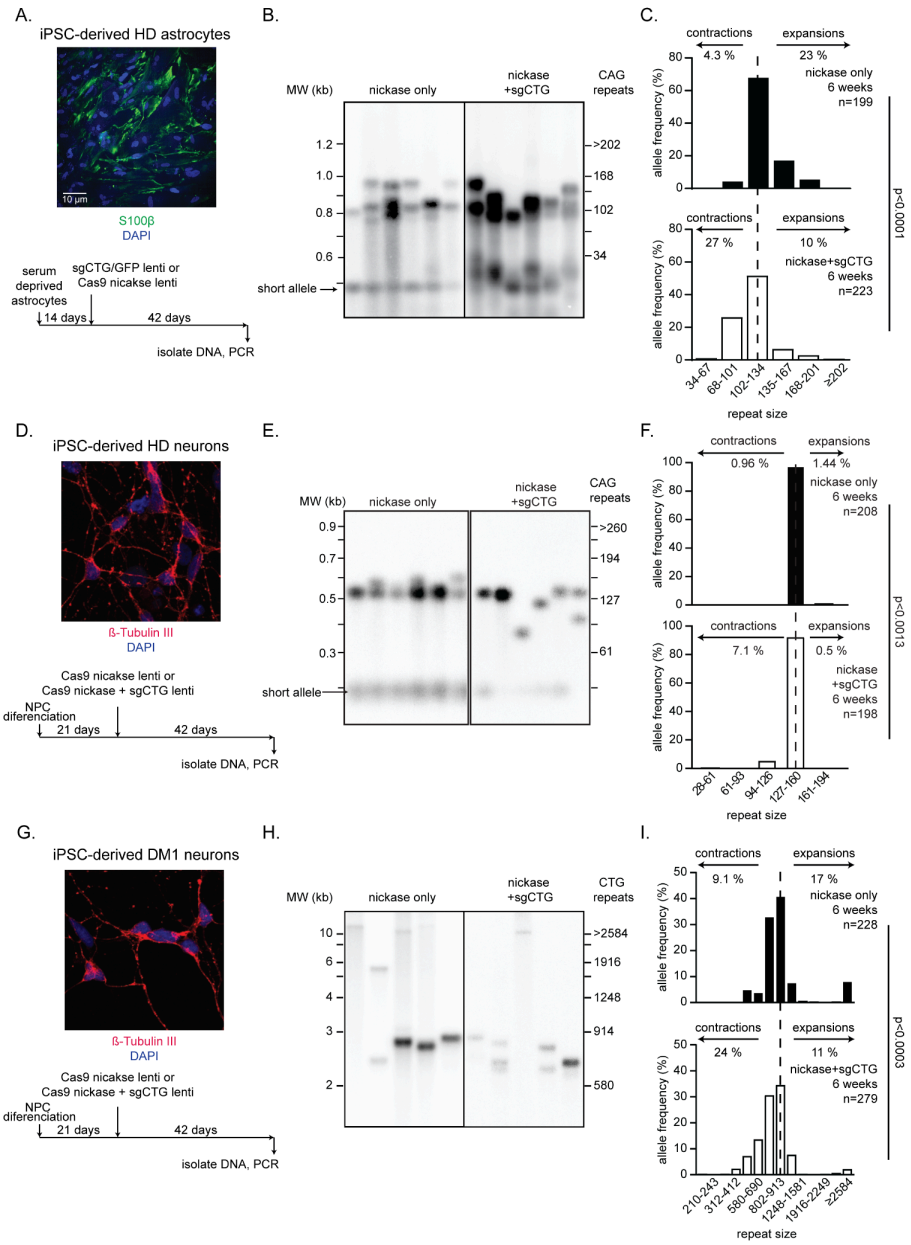


Figure 1. ***Cas9D10A nickase-induced contractions in HD and DM1 iPSC-derived cells.*** *A) Top: Representative confocal image of HD iPSC-derived astrocytes stained with S100β and DAPI. Bottom: timeline of the experiments. B) Representative small-pool PCR blot showing contractions in HD iPSC-derived astrocytes that were either transduced with the Cas9D10A only or both the Cas9D10A nickase and the sgCTG for 6 weeks. C) Quantification of the small-pool PCR blots for HD iPSC-derived astrocytes. D) Top: Representative confocal image of HD iPSC-derived cortical neurons stained with β-Tubulin III and DAPI. Bottom: timeline of the experiments. E) Representative small-pool PCR blot showing contractions in HD iPSC-derived cortical neurons that were transduced with Cas9D10A only or both Cas9D10A and the sgCTG for 6 weeks. F) Quantification of the small-pool PCR blots for HD iPSC-derived cortical neurons. G) Top: Representative confocal image of DM1 iPSC-derived cortical neurons stained with β-Tubulin III and DAPI. Bottom: timeline of the experiments. H) Representative small-pool PCR blot showing contractions in HD iPSC-derived*

cortical neurons that were transduced with Cas9D10A only or both Cas9D10A and the sgCTG for 6 weeks. I) Quantification of the small-pool PCR blots for DM1 iPSC-derived cortical neurons. 'n' is the number of alleles counted in the small-pool PCR experiments. Scale bar = 10 µm.

Off-target mutations remain at background levels

An important safety concern in the context of gene editing is the generation of off-target mutations[29]. Most approaches to mitigate this issue include carefully designing the sgRNA, including its target sequence. Here, this is not possible as we target the repeat tract itself. Therefore, it was important to determine whether this approach could cause changes to the genome at any significant rate. To address this, we used three different approaches. First, we used the HD iPSC-derived neurons and performed single-cell DNA sequencing of individual cells that were either not transduced (n = 48) or transduced with both Cas9D10A and sgCTG (n = 48) for 3 weeks. As in Supplementary Fig. 1ef, the transduced cells were selected for the presence of the transgene using blasticidin. We determined the number of single and double base pair substitutions with LoFreq[45]. We found that the number of cells with at least one base substitution was not significantly different between the non-transduced and the Cas9D10A + sgCTG transduced cells (Table 1, P=0.40). We conclude that the levels of off-target mutations in these cells is below the detection limit of this experiment.

Table 1: Analysis of mutations in single-cell DNA sequencing from HD iPSC-derived cortical neurons

Single-cell DNA sequencing	Number of cells with at least one mutation*
Non-transduced	33/48
Cas9D10A + sgCTG	28/48

**: Each cell was analysed using LoFreq and the number includes the single and double base pair substitutions. When present, the average number of mutations per cell was 2.1 for the non-transduced and 1.8 for the transduced cells. There is no statistical difference using a Fisher's exact test (P=0.40).*

Upon closer inspection, we found that the single-cell DNA sequencing datasets did not contain any of the 111 loci found in the reference genome (GRCh38) to have at least 6 CAG/CTG repeats next to a NGG or NAG PAM identified using Cas-OFFinder[46]. This list includes 80 genes, with the other sites being intergenic (Supplementary Table 2). These are the most likely off-target loci to be mutated given the sgCTG target sequence is CUGs. We therefore used a second approach

to look for off-targets: PCR-free whole genome sequencing using Illumina short-read sequencing. We used the same HD iPSC-derived cortical neurons treated for 3 weeks. In these experiments, we had a coverage at these sites of 6,010 alleles for the transduced cells and 5,045 for the non-transduced cells. LoFreq identified 5 single base substitutions genome-wide in the two cultures treated with both Cas9D10A and sgCTG compared to 3 mutations in the two non-transduced cultures. Thus, the mutation frequencies were 1 in 1,202 for transduced cells, and 1 in 1,682 for non-transduced cells. Of the 8 base substitutions identified, the closest to one to one of the likely off-target sites was 4.3kb away. Our results suggest that a 3 week treatment with both Cas9D10A and sgCTG does not induce mutations above background.

We wondered whether off-target mutations increased with longer expression of the Cas9D10A and sgCTG. We therefore analysed mutations in the 80 genes most likely to be off-targets 6 weeks after transduction with either Cas9D10A only or Cas9D10A and sgCTG. This was done under selection such that all the cells contained the lentivirus. In this experiment, we sequenced 1488 alleles from 4 independent cultures. LoFreq identified no base pair substitutions at these sites (Table 2). We conclude that after 6 weeks off-target mutations are below 1 in 1,488.

Table 2: Mutation analysis in PCR-free whole genome sequencing of HD iPSC-derived cortical neurons and astrocytes.

Cell type	Treatment (6 weeks)	No. of genes with potentially targeted CAG/CTG repeat	No. of base pair substitutions*
HD iPSC-derived cortical neurons	Cas9D10A only	80	0
	Cas9 D10A + sgCTG	80	0
HD iPSC-derived astrocytes	Not transduced	80	0
	Cas9D10A + sgCTG	80	0

**: Each sample was analysed using LoFreq and the number includes the single and double base pair substitutions.*

Given that Cas9 accessibility to DNA is influenced by chromatin structure[47], we considered the possibility that off-target mutations may be cell-type specific. We therefore tested whether HD iPSC-derived astrocytes contained off-target mutations

in the selected 80 genes. The total number of alleles sequenced was 384 for the cultures transduced with Cas9D10A plus sgCTG and 555 for those transduced with Cas9D10A alone. As was the case with the HD iPSC-derived cortical neurons, we found no mutations in these cells after 6 weeks of treatment. We conclude that off-target mutations in HD iPSC-derived astrocytes are below our current limit of detection for this experiment.

The third approach was to look for rearrangements at the non-expanded *HTT* allele 6 weeks after treatment in HD iPSC-derived neuronal cultures using our long-read sequencing dataset (Supplementary Fig. 2a-d). We first sorted unaligned circular consensus sequences (CCSs) with 6 to 20 repeats as determined by Repeat Detector[48]. Then we looked for rearrangements using NGMLR alignment, Picard sort and samtools indexing, followed by Sniffles2[49] (see methods). This analysis revealed that one sample transduced with Cas9D10A and sgCTG had an insertion downstream of the repeat tract at a frequency of 0.107. However, we could not identify this rearrangement with the Integrative Genomics Viewer[50] or by manually looking for the rearranged alleles in CCSs. Moreover, we found no rearrangements in the other 3 samples treated with both Cas9D10A and sgCTG, which came from the same neuronal epithelium, but differentiated from neuronal progenitor cells to cortical neurons separately. We therefore conclude that there is no *bona fide* rearranged allele in this sample. Altogether, we examined 90,240 CCSs for the samples transduced with both Cas9D10A and sgCTG and 69,493 CCSs for the cultures transduced with Cas9D10A alone, without finding any rearranged allele. Our results suggest that mutations are rare in the non-expanded alleles upon a 6 week exposure to Cas9D10A and sgCTG.

Long-read sequencing detects somatic expansion in the mouse striatum

Next, we aimed to determine whether the Cas9 nickase could induce contractions *in vivo*. We first sought to validate the SMRT sequencing approach, specifically asking whether we could detect somatic expansions. For this, we turned to an HD mouse model, R6/1[51], which contains a hemizygous transgene expressing exon 1 of human *HTT* with an expanded CAG repeat. It has been extensively used for studying repeat instability[52]. We PCR-amplified the repeat tract over time, sequenced the products, and determined repeat size mosaicism with Repeat Detector. The pattern of repeat

size distribution was as expected[53]. Specifically, there was a small but progressive shift in the modal repeat size towards expansion accompanied by a bimodality beyond 10 weeks of age (Supplementary Fig. 2e), which is likely due to different cell types having different rates of expansion. Together with the results in HD iPSC-derived cortical neuron cultures, this observation suggests that SMRT sequencing is robustly detecting both contractions and expansions, in line with previous work[48,54].

Expression of AAV-delivered GFP and Cas9D10A in vivo

Next, we established whether we could deliver Cas9D10A and sgCTG directly to the striatum of mice using recombinant adeno-associated virus serotype 9 (AAV9). Cas9 is too large to be packaged into a single AAV and therefore we used a two vector system with one AAV harbouring Cas9D10A on a miniCMV promoter and the other expressing both sgCTG and GFP (Fig. 2ab, Supplementary Table 1).

When injected in R6/1 mice two days after birth (P2), GFP was detectable throughout the striatum along with some expression in the cortex (Supplementary Fig. 3a). Expression was still present 3 months post injection without a significant drop in expression over time (Supplementary Fig. 3b). We found that in the striatum, AAV9 exhibited strong tropism to $NeuN^+$ neurons and expressed GFP well (Supplementary Fig. 3c). Most of the GFP expressing cells in the cortex had an astrocytic morphology and did not colocalize with NeuN (Supplementary Fig. 3d). Furthermore, using qPCR, we found that there were 26 ± 6 copies of the sgCTG/GFP AAV genome per cell on average 1 month after injection (Supplementary Fig. 3e). These data demonstrate that AAV9 can infect striatal neurons in R6/1 mice and that GFP is robustly expressed over several months.

Cas9D10A expression, however, proved more challenging. We tested several AAV constructs previously shown to induce edits *in vivo*[20,55,56] using the Cas9 nuclease version and introduced the D10A mutation. We could not detect Cas9D10A expression using immunofluorescence in the striatum 28 days post injection in adult wild type mice (Supplementary Fig. 4a-d). However, Cas9D10A expression was detected for 3 of the constructs, using western blotting, 15 days post injection in adult wild type mice (Supplementary Fig. 4e). We tested 2 of these constructs in

adult R6/1 mice and found that one was silenced 1 month post injection (Supplementary Fig. 4f). The last construct, containing a miniCMV promoter, was still expressed 2 months after injection, albeit at lower levels (Supplementary Fig. 4g). Similarly, when injected in 2-day-old (P2) animals, we could detect Cas9D10A for up to 5 months after injections, but the expression markedly decreased over time (Supplementary Fig. 4h). These data suggest that Cas9D10A expression wanes over time at a rate heavily dependent on the exact construct used.

To have an idea of the multiplicity of infection, we measured the average copy number of Cas9D10A 1 month after injection of R6/1 mice at P2. We found them to be lower than for the sgCTG/GFP AAV, at 0.3 ±0.15 copies per cell on average (Supplementary Fig. 4i). We therefore expected the efficiency of contractions to be capped at 26% on average, as this is the predicted number of cells with at least one copy of the Cas9D10A AAV, assuming a Poisson distribution of viral genomes per cell.

Large contractions seen after delivery of Cas9D10A and sgCTG to neonatal mice

Measuring repeat size in P2-injected mice showed little to no change in the hemisphere that was injected with both Cas9D10A and the sgCTG cassette after a month compared to the hemisphere injected only with the Cas9D10A AAV (Fig. 2d). This, however, progressed over a period of 5 months with the appearance of short alleles of 12 to 17 CAGS, which is well within the non-pathogenic range (Supplementary Fig. 5; Fig. 2e). These large contractions were seen in all 3 mice that were sequenced at 5 months post injection. The instability index[57], calculated from the sequencing data, shows a slower increase over time in the hemisphere injected with both Cas9D10A and sgCTG compared to Cas9D10A alone (Fig. 2f, P=0.0004), which was accompanied by a more negative contraction index (Fig. 2g, P=0.0002), and a lower expansion index (Fig. 2h, P=0.005). These results suggest that there were more contractions and fewer expansions in the hemispheres injected with both Cas9D10A and sgCTG. Of note, because of the way these indices are calculated, they do not include the larger contractions seen 5 months post-injection. We examined the CCSs containing the large contractions further, aligning them to the *HTT* locus and visually noting the contractions in the fasta files. We further found that the average repeat size was reduced compared to Cas9D10A-injected mice,

especially after months 4 and 5 (Fig. 2i, P<0.001). This appeared to be due to fewer large expansions and more large contractions. To measure the percentage of the alleles that are shorter than expected, we took the difference in the area under the curve between the Cas9D10A-injected and the Cas9D10A and sgCTG-injected hemispheres (Supplementary Fig. 6). We observed a rise in the percentage of short alleles in the samples treated with both Cas9D10A and sgCTG from 5.1 ± 1.5% 1 month post-injection to 15.7 ± 7.6 % 5 months after the injections (Fig. 2j, P=0.035). Given that the expected maximum was 26%, we conclude that most cells that express Cas9D10A together with sgCTG in the striatum of R6/1 mice harbour contractions after 5 months.

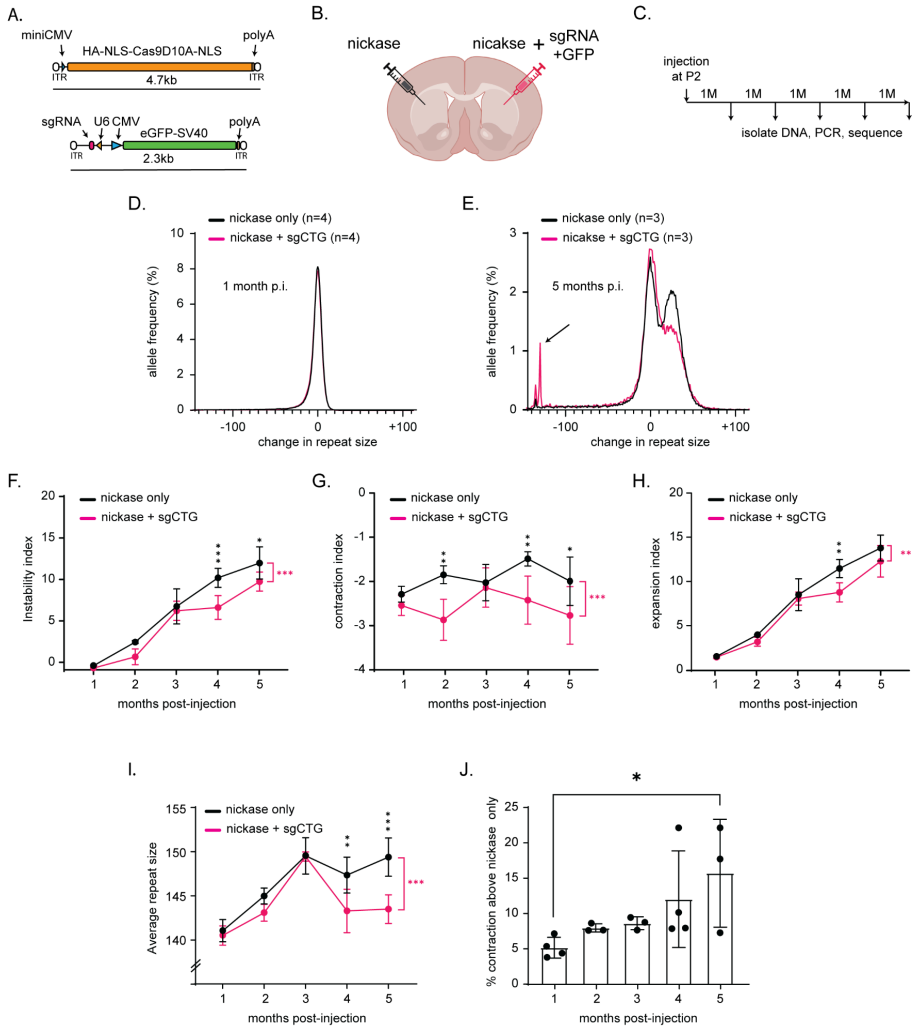


Figure 2. *CAG repeat contraction in P2-injected R6/1 animals. A) Cartoon of the AAV constructs used. B) Schematic of the experimental approach. C) Timeline of the experiment. P2 = 2 days after birth. D) Aggregated graphs of repeat size distribution in hemispheres of R6/1 mice injected with Cas9D10A only (black) versus Cas9D10A + sgCTG (magenta). Note that the graphs are normalised*

*to the modal peak size found in each hemisphere (see methods). The mice used in these experiments had a modal repeat size ranging from 138 to 151 CAGs. E) Same as D but 5 months after injection. Arrow points to large contractions down to 17 CAGs. F) Instability index of the mice treated with both Cas9D10A and sgCTG (magenta) versus Cas9D10A only (black). A two-way anova showed an effect of the treatment (P<0.001, magenta asterisks), with a post-hoc Tukey's multiple comparison test showing significant differences 4 and 5 months post injection (*P<0.05, ***P<0.001, black asterisks). G) The same data as in F, but showing only the contraction index (more negative values indicating more contractions). *P<0.05, **P<0.01, ***P<0.001. H) Same as in F but only showing the expansion index.**P<0.01. I) The average repeat size over time. **P<0.01, ***P<0.001 using a two-way ANOVA. I) Difference in the area under the curve between the Cas9D10A + sgCTG treatment and Cas9D10A alone (*P<0.05 using a one-way ANOVA). Error bars represent ± standard deviation. Number of animals: Cas9D10A nickase only; 1 month n=4, 2 months n=3; 3 months n=3; 4 months n=4; 5 months n=3; Cas9D10A nickase + sgCTG; 1 month n=4, 2 months n=3; 3 months n=3; 4 months n=4; 5 months n=3).*

Contractions in the R6/1 adult striatum upon Cas9D10A and sgCTG delivery

Next, we asked whether contractions also occurred in adult mice upon treatment with the Cas9D10A AAV. Therefore, we repeated the experiments, injecting some 3.5 month old adult R6/1 mice with both Cas9D10A and sgCTG and comparing them with age-matched un-injected controls. We tracked the instability 1 and 2 months post-injections (Fig. 3a-c). We did not go beyond 2 months in these experiments because the R6/1 mice become markedly unwell beyond this age and to continue would have inflicted unnecessary suffering.

In contrast to the injections at P2, we found over 400 copies of each AAV per cell on average in the adult mice 1 month post-injection (Supplementary Fig. 4jk). Consequently, we found that we could get contractions *in vivo* after 1 and 2 months with the instability index being lower and the contraction index decreasing over time, indicative of an increase in the number and/or size of the short alleles (Fig. 3def). We conclude that our gene editing regimen induces contractions in adult R6/1 mice.

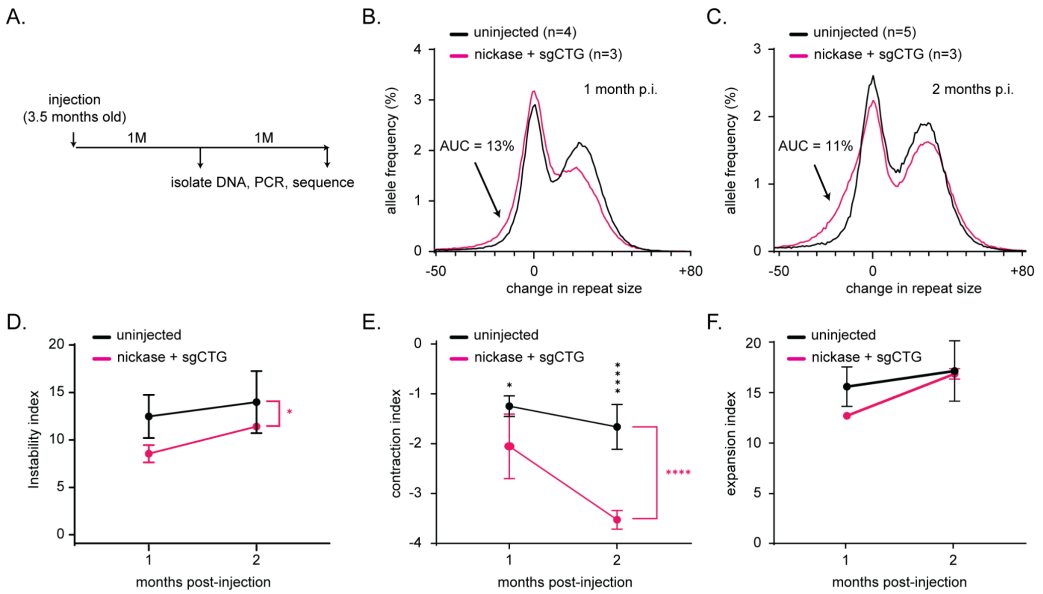


Figure 3: *AAV-based delivery of Cas9D10A and sgCTG leads to contractions in the R6/1 adult striata. A) Timeline of the experiments. B) Aggregated graphs of repeat size distribution in the striatum of uninjected adult R6/1 mice (black) compared to striata injected with Cas9D10A + sgCTG (magenta) 1 month post-injection. Note that the graphs are normalised to the modal peak size found in each mouse (see methods). The mice used in these experiments had a modal repeat size ranging from 138 to 145. C) Same as B, but 2 months after injection. D) Instability index of the mice treated with both Cas9D10A and sgCTG (magenta) versus uninjected (black). A two-way anova showed an effect of the treatment (P<0.05). E) The same data as D, but showing only the contraction index (more negative values indicating more/larger contractions). Magenta asterisks: Two way anova, ****P<0.0001, post-hoc Tukey's multiple comparison test at individual time points (black asterisks) *P<0.05, ****P<0.0001. F) Same as in E but only showing the expansion index.*

Discussion

Here we provide evidence that the Cas9D10A nickase can induce contractions when targeted to expanded CAG/CTG repeats in HD iPSC-derived neurons and astrocytes and in DM1 iPSC-derived neurons. These contractions were not accompanied by an increase in off-target mutations and they were allele-selective. Using AAVs to deliver Cas9D10A and sgCTG, we could induce contractions in an HD mouse model. In all cases, the frequency of contractions increased over time with expression of the Cas9D10A. These data demonstrate that multiple loci and cell types can be targeted with this approach. Our data suggest that the approach does not yield a significant number of off-target mutations and implies that it is safe. Moreover, Cas9D10A was silenced over time, further improving the safety of the approach.

In P2-injected mice, 5 months post-injection, we found an average editing efficiency of up to 15%. This was in the context of having an average cap on the editing efficiency of 26%, which is the predicted portion of the cells with at least one copy of Cas9D10A. Therefore, editing efficiency in the infected cells was on average 58%. Although work is needed to improve the infection efficiency in P2 injected mice, the efficiency of contractions is comparable to what has been seen by others using sgRNAs flanking the repeats. By contrast, here, we recovered repeat contractions rather than a collection of uncontrolled mutations[15,16,24,25]. Thus, our approach is as efficient as excising the repeat tract, while remaining more precise and avoiding both on-target and off-target mutations.

In vitro, interestingly, the size and frequency of the contractions appear to be much smaller in the HD neuronal cultures than in the astrocyte cultures, despite having similar starting repeat sizes. The differences could be due to differential expression

of the Cas9D10A nickase, the sgRNA and/or of DNA repair enzymes that process the contractions, however these hypotheses need further investigation. We also considered that replication may play a role, but we found markers of proliferation to be low in both cultures. *In vivo,* the pattern of repeat contraction was also markedly different between the P2- and 3.5 month-injected animals. At 1 and 2 months after injection, adult-injected animals showed a more apparent accumulation of alleles 20 to 40 repeats shorter than the modal size. Despite these differences, one pattern has emerged whereby longer repeats appear to be better targets for contractions in all systems investigated here. It may also be that in addition to the shorter repeats, there is a prevention of somatic expansions when Cas9D10A and sgCTG are both present. These results suggest that the Cas9D10A nickase will be more efficient in affected cells that have larger expansions, which occur in more affected individuals and later in life.

Importantly, contraction efficiency improved over time *in vivo*, suggesting that short alleles will accumulate for as long as expression of both Cas9D10A and sgCTG is maintained. Thus, the duration of the expression is expected to be a rate limiting factor in the editing efficiency. However, the propensity of Cas9D10A to be silenced *in vivo* may bypass the need to remove the enzyme artificially. More work is required to understand this silencing event and to gain control over Cas9 expression *in vivo*.

One limitation of our work is the use of two different AAVs to deliver Cas9D10A and sgCTG. We used the most common Cas9 orthologue from *S. pyogenes*, which is too large to fit within an AAV together with a sgRNA cassette[58]. This is an issue because greater quantities of viral vectors must be injected to ensure that both AAVs infect the same cells. One way around this is to use smaller Cas9 orthologues, for example the Cas9 from *S. aureus*[59], as has been done *in vivo* for the *DMPK* locus[17,18]. Unexpectedly, this did not lead to higher levels of editing compared to a two-vector system[17,18]. Therefore, the size of the Cas9 enzyme and its delivery in the brain remain important challenges.

A gene editing approach would have several advantages over more traditional approaches, including antisense oligonucleotides and microRNAs, which reduce the amount of *HTT* or *DMPK* mRNA[4,60]. Specifically, our gene editing approach would

permanently improve the most proximal cause of the disease by reducing the size of the expanded CAG/CTG repeat. Recently, a small molecule that targets DNA secondary structures formed by expanded CAG/CTG repeats was shown to induce contractions *in vivo*[61,62]. However multiple stereotactic injections or continuous supply of the drug over 16 weeks were required for the effect to be significant. A gene editing approach would potentially require a single dose. Crucially, from a patient benefit standpoint, it is imperative that we try multiple approaches to maximise the chances of making it to the clinic.

To our knowledge, the data presented here is the first example of a gene editing approach using a single Cas9 and sgRNA pair capable of editing the mutation that causes several clinically distinct disorders in a safe and efficient manner.

Methods

iPSC cultures and differentiation conditions

For the HD iPSC-derived astrocytes, we obtained CS09iHD109-n1[36] iPSCs from the Cedars-Sinai Medical Center's David and Janet Polak Foundation Stem Cell Core Laboratory. They originally had a heterozygous expansion of the CAG repeat in the *HTT* gene. We found them to have expanded to around 135 CAG repeats. IPSCs were grown in plates coated with Matrigel in E8 Flex Medium at 37°C, 5% CO_2. For astrocyte differentiation iPSCs grown in E8 medium were supplemented with Rho-associated protein kinase inhibitor (RI) at approximately 60% confluency for 24 h. Neural progenitors (NPCs) were first derived by culture in SLI medium (advanced DMEM F-12 supplemented with 1x PenStrep, 1x Glutamax, 1% Neurobrew (w/o) vitamin A, 10 μM SB-431542, 200 nM LDN-193189, 1.5 μM IWR-1) for 8 days. On day 8, cells were treated with RI for 1 h and split 1:4 into NB medium (advanced DMEM F-12 1x PenStrep, 1x Glutamax, 2% Neurobrew (w/o) vitamin A) supplemented with RI for 24 h. From day 14 to 16, 20 μM PluriSIn was added to the medium to eliminate any undifferentiated iPSCs. On day 16, differentiated NPCs were frozen. Astrocyte progenitor cells (APCs) were differentiated from these NPCs as previously described[63]. Six million NPCs were thawed into NB medium containing 20 μM PluriSIn. After 2 days, the medium was changed to NF (advanced DMEM

F-12 containing 1x PenStrep, 1x Glutamax, 2% Neurobrew with vitamin A, 20 ng ml^{-1} fibroblast growth factor 2 "Improved Sequence" (FGF-2 IS) and cells were allowed to reach confluency. They were then passaged three times in NEL medium (advanced DMEM F-12 with 1x PenStrep, 1x Glutamax, 2% Neurobrew with vitamin A, 20 ng ml^{-1} epidermal growth factor 20 ng ml^{-1} leukaemia inhibitory factor). CD44 positive cells were collected using the DB FACSAria Fusion, and maintained into NEF medium (advanced DMEM F-12 supplemented with 1x PenStrep, 1x Glutamax, 2% Neurobrew with vitamin A, 20 ng ml^{-1} epidermal growth factor, 20 ng ml^{-1} FGF-2 IS). To mature astrocytes, APCs were seeded onto Matrigel coated 6 well plates at 200,000 cells per well with STEMdiff™ Astrocyte Maturation medium (STEMCELL Technologies). After 2 weeks, the media was changed to advanced DMEM F-12 with 2% FBS for another 2 weeks. Once mature, astrocytes were plated in advanced DMEM F-12 medium without FBS for treatment with Cas9D10A.

CS09iHD109-n1 and DM1 iPSCs (GM24559, obtained from the Coriell Institute) were differentiated to cortical neurons[64] by first bringing the cells to confluency and maintaining them in neuronal induction medium (0.5x advanced DMEM F-12 and 0.5x Neurobasal medium supplemented with 25 U ml^{-1} PenStrep, 2.5 µg ml^{-1} insulin solution human, 50 µM 2-Mercaptoethanol, 1x MEM non-essential amino acids solution, 500 µM sodium pyruvate solution, 1x N2 supplement, 1x B27 supplement, 1mM L-Glutamine, 10 µM SB-431542, 1 µM dorsomorphin dihydrochloride) for 12 days. The resulting neuroepithelium was disaggregated and plated in a new precoated 2x Matrigel well with neuronal maintenance media (NMM; 50x advanced DMEM F-12 and 50x Neurobasal medium supplemented with 25 U ml^{-1} PenStrep, 2.5µg ml^{-1} insulin, 50 µM 2-Mercaptoethanol, 1x MEM non-essential amino acids, 500 µM sodium pyruvate solution, 1x N2 supplement, 1x B27 supplement, and 1 mM L-Glutamine) also containing 20 ng ml^{-1} human FGF-2 IS. Once neuronal rosettes were visible, Neural Rosette Selection Reagent (Stem Cell Tech) was added to the culture for 1 h at 37°C for NPC selection. NPC in suspension were collected and isolated cells were plated. NPC were maintained in NMM+FGF-2 IS. 50,000 live cells cm^{-2} and 30,000 live cells cm^{-2} of HD and DM1 NPC, respectively, were plated for neuronal differentiation in wells pre-coated with Laminin from Engelbreth-Holm-Swarm murine sarcoma basement membrane (Merck). Cells were

plated in NMM and half the media was replaced twice a week. Neuronal cultures were kept for 3 weeks for maturation before being used for lentiviral transduction.

Mature neuronal and astrocytic cultures were transduced with lentiviral vectors at a multiplicity of infection of 5. The lentiviruses drive the expression of the Cas9D10A nickase together with a blasticidin resistance gene, BSD, alone or together with the sgCTG cassette (Supplementary Table 1). Transduced cells were selected using blasticidin (5 µg ml^{-1}) for 7 days. Cells were collected with scrapers 3 and 6 weeks after the infection to isolate DNA and measure repeat size.

Immunostaining

Mature astrocytes were plated onto CELLSTAR 96-well tissue culture plates (Greiner Bio-One Ltd.), fixed with 4% paraformaldehyde (PFA), washed with PBS, permeabilized with 0.4% Triton-X in PBS, washed with PBS, and then blocked for 1 h at RT with 0.2% Triton-X in PBS + 1% normal goat serum (NGS, Merck). Antibodies used and their concentration is found in Supplementary Table 4. Cells were counterstained with DAPI (1 µg ml^{-1}) before imaging with the Opera Phenix High Content Screening System (Perkin Elmer).

NPCs were plated on 13 mm coverslips for neuronal differentiation. Neurons were fixed in 4% of PFA. Cells were washed in PBS, blocked, and permeabilized for 1 h in 1% BSA, 0.1% Triton X-100, and 4% NGS in PBS, and incubated with primary antibody in 1% BSA, 0.1% Triton X-100, and 1% NGS in PBS overnight at 4°C. After washing with PBS, cells were incubated with appropriate Alexa-conjugated secondary antibody in 1% BSA, 0.1% Triton X-100 in PBS for 1 h at room temperature. Cells were counterstained with DAPI Staining Reagent (1 µg ml^{-1}; Chemometec) and after several washes in PBS, coverslips were mounted onto SuperFrost Plus slides (Thermo Fisher Scientific) with Fluoromount aqueous mounting medium (Merck) and imaged with the Leica SP8 confocal microscope. Images were captured using LAS X software and were analysed with Fiji software.

Flow cytometry

Mature cultures were resuspended by Accutase (Fisher Scientific) and fixed with PFA 2%, Saponin 0,1%, in PBS. Fixation step was inactivated using 1.25 M glycine. Cells were washed/blocked with 500 µl of blocking solution (Saponin 0,5%; BSA 0,5%; in PBS) and incubated with primary antibody in 300 µl Saponin 0,5%; BSA 0,5%; NGS 1% in PBS for 30 min at 4°C. After washing with a cleaning buffer (Saponin 0,1%; BSA 0,2%; in PBS), cells were incubated with appropriate Alexa-conjugated secondary antibody in the cleaning buffer for 30 min at 4°C. Cells were counterstained with DAPI Staining Reagent (1 ng ml^{-1}; Fisher Scientific). Finally, cells were resuspended in BSA 0,5% and EDTA 5 mM in PBS. Samples were run on the DBFACSAria Fusion and analysed using FlowJo v10 software.

Small-pool PCR

SP-PCR was done as described before[34,35], but using primers oVIN-1345 and oVIN-1347 (See Supplementary Table 3 for the sequence of all primers used) for the amplification of *HTT* exon 1, and oVIN-1251 and oVIN-1252 for *DMPK*. Genomic DNA was extracted with the NucleoSpin Tissue DNA extraction kit (Macherey-Nagel). PCRs were run using between 10-500 pg of DNA measured by Quant-iT Qubit dsDNA HS Assay Kit (ThermoFisher) to find the amount of amplifiable alleles for each sample. The products were run on 1.5% agarose gels in 1X TAE buffer and transferred onto a GeneScreen Hybridization Transfer Membrane. The hybridization was done at 48°C (for *HTT* blots) or 51°C (for *DMPK* blots) in ULTRAhyb Ultrasensitive Hybridization buffer (Fisher Scientific) with a ^{32}P-labelled oligo containing 10 CAG repeats (oVIN-100). The membranes were exposed to an Imaging K phosphor screen and scanned using Bio-Rad PMI Personal Molecular Imager. Images were captured using Quantity One 1-D Analysis Software (Basic V4.6.6) and analysed with Fiji. Blinded allele counts were completed for each membrane based on allele frequency calculated by Poisson distribution as described[35]. Each band was sized using the ladder run on either side of the gel. For the *HTT* locus in astrocytes, we excluded repeats below 34 repeats so as to not conflate contractions with the non-expanded allele. For the *DMPK* locus, small-pool PCRs, sizing was further divided between fragment sizes of 2 kb and 3 kb by fitting an exponential decay of the distance travelled by the ladder bands. This allowed us

to subdivide this region of over 580 repeats into three bins, where most of the contractions were observed. Differences between conditions were determined using a Mann-Whitney U Test calculated by GraphPad Prism software (v.10.0.0).

WGS and single cell DNA sequencing library preparation

Genomic DNA was extracted from cells using the Macherey-Nagel NucleoSpin Tissue DNA mini kit and made into dual-indexed single stranded DNA libraries using 'Illumina DNA PCR-free Prep kit, Tagmentation' (Illumina). Quantification of the libraries was done using the KAPA Illumina Universal Library Quant kit (Roche). The sequencing was performed on an Illumina NovaSeq 6000 using the v1.5 S4 reagent kit, 300 cycles using XP 4-lane splitter kit.

For the whole genome single-cell DNA sequencing, single cells were sorted using a BD FACSAria Fusion into 96-well plates suspended in 5 µl TE. Dual-indexed DNA libraries were prepared using Takara Bio's PicoPLEX Gold Single cell DNA-Seq Kit according to the kit guidelines. Quantification was done using Qubit HS dsDNA kit (Thermo Fisher Scientific). Libraries were sequenced using Illumina NovaSeq 6000 on a NovaSeq SP flow cell with 300 cycles.

Whole genome sequencing analysis

Whole Genome Sequencing (WGS) data was demultiplexed and obtained as fastq files. Paired-end sequences were aligned to GRCh38 with the Burrows-Wheeler Aligner (BWA)[65] version bwa/0.7.17 and the mem algorithm. Subsequently, the Sequence Alignment/Map (SAM)[66] files produced were sorted via the Picard (picard/2.27.5 - https://broadinstitute.github.io/picard/). These files were then converted to binary alignment/map (BAM) format and indexed using SAMtools (samtools/1.9).

To assess coverage of off-target regions, Mosdepth[67] was employed. Read groups were attributed to the BAM files with the Picard tool, which also facilitated the subsequent analysis for single nucleotide polymorphisms using LoFreq (LoFreq_star-2.1.5)[45], referencing the entire genome. The variant call format (VCF) files generated were scrutinised for variants overlapping with the set of 111 most likely off-targets (Supplementary Table 2), provided in a bed file.

Long read sequencing of repeats

Expanded repeats were sequenced as before using SMRT HiFi sequencing[48]. In brief, primers oVIN-1321 and oVIN-3490 were used to amplify the human *HTT* locus, including the CAG repeat from iPSC-derived neurons and R6/1 mouse genomes. Mice samples were mechanically disaggregated in Nucleospin Tissue DNA extraction T1 buffer using a mini hand-held homogenizer (ThermoFisher). Genomic DNA was extracted with the NucleoSpin Tissue DNA extraction kit (Macherey-Nagel). PCR products were cleaned with AMPure PB beads using the manufacturer's instructions at a ratio of 0.6X. Library preparation was done according to the manufacturer's instructions. Barcoded sample pool purity was analysed on the 5200 Fragment Analyser (Agilent). Sample pools were sequenced using the Pacific Biosciences Sequel IIe.

Precise CAG/CTG repeat sizes were determined using Repeat Detector[47] (version 1.0.15eb445). Unaligned reads were assessed using the restrictive profile with a repeat size range of [0–250]. For each analysis, the -–with-revcomp option was enabled and data was output to a histogram (-o histogram option). Obtained histograms were plotted using GraphPad Prism software (v.10.0.0). As the sequences are unaligned when we determine repeat size, the CCSs containing between 0 and 5 CAG/CTGs are removed from analysis as they include a large proportion of sequences that do not align to the *HTT* locus. The CCSs are available from SRA (https://www.ncbi.nlm.nih.gov/sra) PRJNA1077893.

Alignment of long reads

For aligning specific PacBio sequencing repeat regions, the repeat regions were filtered using custom python scripts to process fasta files based on Repeat Detector-generated log files. Repeat Detector was run to confirm exact repeat sizes have been filtered. These repeat regions were aligned with the NGMLR aligner using the *HTT* region from GRCh38 containing 19 CAG repeats as reference sequence. The resulting SAM files were converted into BAM and .bai index files using Picard and Samtools respectively. The bam files were checked for percentage of mapped reads using Samtools flagstat command. The indexed BAM files were then

visualised using the Integrative Genomics Viewer (IGV 2.16.2) to assess the alignment quality and to confirm the presence of the expected CAG repeat count. Sniffles2 was run on the aligned BAM files for identifying rearrangements.

Area Under the Curve

To calculate the difference in the area under the curve (Supplementary Figure 6), we used the histograms generated by Repeat Detector, removing the reads with 5 or fewer CAGs as they are low quality reads that do not align with the *HTT* locus. The total read counts were normalised for each sample and adjusted such that the mode is set to 0. Then we subtracted the frequency of reads of the treated samples from those in the control conditions for each repeat size. The positive values were then added together to produce the difference in the area under the curve. It represents the frequency of alleles that have changed beyond what is found in the control samples.

Instability, expansion and contraction indices

The instability index was calculated as previously described for images[57], but using normalised read frequencies and a 5% threshold. The expansion and contraction indices were calculated such that expansion index + contraction index = instability index.

Mouse housing

R6/1 mice[51] were maintained as hemizygotes with *ad libitum* access to food and water and maintained in a temperature-controlled environment on a 12 h dark/light cycle. All experimental procedures done in Cardiff followed protocols in accordance with the United Kingdom Animals (Scientific Procedures) Act of 1986. All experimental procedures performed on R6/1 mice were approved by Cardiff University Animal Welfare and Ethical Review Board and carried out under Home Office Licenses P49E8C976 and PP7595333. Stereotactic injections done in wild type mice presented in Supplementary Figure 4a-d were done at the Center de Recherche du CHU de Québec (Québec, QC, Canada). All procedures on these animals were completed in accordance with the guidelines of the Canadian Council

on Animal Care and were approved by the Comité de Protection des Animaux du CRCHUQ-UL under protocol number 2019-49, CHU-17-106.

Stereotactic Injections

The R6/1 mice used here had between 138 and 151 CAG/CTG repeats. Mice were anaesthetised with isoflurane and placed in a stereotaxic frame. In P2 animals, a syringe was used to penetrate the skin and skull. In adult (3.5 months old) mice, the scalp was shaved, a longitudinal incision was made to expose the skull surface, and 2 burr holes were drilled above the infusion sites. A 4 µl viral suspension ($4.8x10^{10}$vg of Cas9D10A and sgCTG/GFP AAVs in 1:1 ratio) was stereotaxically injected into the striatum of adult mice according to the Paxinos and Franklin mouse brain atlas (AP +0.8; ML +-1.8; DV -2.2 mm from bregma). P2 mice were stereotaxically injected with 1 µl ($1.44x10^{10}$vg of Cas9D10A and sgCTG/GFP AAVs in 1:1 ratio) virus suspension into the striatum (AP +2.3; ML +-1.4; DL 1.8 mm from lambda). Hamilton syringes 701 N 10 µl and 5 µl FIX NDL (26S/51/3) were used for the infusion in adult and post-natal mice respectively. The infusion rate was 200 nL min^{-1}, and the needle remained in place for 5 min after infusion for vector absorption before removal of the syringe. In adult mice, the injection site was closed with stitches, and mice recovered in incubators. For the wild type animals injected with the different constructs found in Supplementary Figure 4a-d, each construct was injected together with the sgCTG AAV9 into the cortex and the striatum (AP: +/- 1.6, MD: +/- 1.4, DV -0.75) using a total of $6x10^{10}$ vg, at a ratio of Cas9D10A to sgCTG of 1:1, divided equally between the two hemispheres. Animals were between 8 and 9 months of age.

Perfusion and immunostaining of mouse tissues

R6/1 animals were deeply anaesthetised with Dolethal (Vetoquino) and transcardially perfused with 4% PFA in PBS. Brains were removed, 2 h post-fixed, and stored in 30% sucrose in PBS at 4°C. Wild type animals presented in Supplementary Figure 4a-d were anaesthetised the same way but the perfusion was done with PBS. 30 µm coronal sections were cut with a freezing sliding microtome and stored at −20°C in cryoprotectant solution (30% ethylene glycol, 30% glycerol, 20 mM PBS) until processing. Free-floating brain sections were washed in PBS, blocked and

permeabilized for 1 h in 1% BSA, 0.2% Triton X-100, and 4% NGS in PBS, and incubated with primary antibody in 1% BSA, 0.1% Triton X-100, and 1% NGS in PBS overnight at 4°C. After washing with PBS, sections were incubated with appropriate Alexa-conjugated secondary antibody in 1% BSA, 0.1% Triton X-100 in PBS for 1 h at room temperature. Tissues were stained with DAPI and after several washes in PBS, sections were mounted onto SuperFrost Plus slides (Thermo Fisher Scientific), air-dried, and coverslipped with Fluoromount aqueous mounting medium (Merck). Leica SP8 confocal (LAS X software) and Evos FL Auto 2 (Invitrogen EVOS FL Auto 2.0 Imaging System) microscopes were used to capture images.

Western blotting

To assess Cas9 levels in astrocytic cultures, proteins were extracted using RIPA buffer supplemented with cOmplete™EDTA-free protease inhibitors. Mouse brain tissues were mechanically disaggregated in RIPA buffer supplemented with cOmplete™EDTA-free protease inhibitors using a mini hand-held homogenizer (ThermoFisher). Samples were incubated for 30 min in a rotator at 4°C and soluble fractions were collected by centrifugation (30 min; 13000g at 4°C). Protein concentrations were determined using Pierce BCA protein assay (ThermoFisher). 10 to 30 µg protein was loaded on 4-12% Bis-Tris Plus precast gels (ThermoFisher) and run in 1X MES buffer (ThermoFisher) alongside Bio-Rad Kaleidoscope molecular weight marker. Proteins were transferred onto a nitrocellulose membrane using the Bolt system (ThermoFisher). The results were imaged on the Odyssey Imager (LI-COR Biosciences).

q-PCR

Quantitative real-time PCR (qPCR) was performed using the Applied Biosystems QuantStudio 7 Flex Real-Time PCR system and analysed with the QuantStudio Real-Time PCR Software (v1.7 Thermo Fisher). Briefly, qPCR was assayed in a total volume of 10 µl reaction mixture containing the ready-to-use FastStart Universal SYBR Green Master (Rox) (Thermo Fisher) and 0.5 µM of forward and reverse mix appropriate primers (Supplementary Table 3). All qPCR reactions were run in triplicates. Mean cycle threshold (Ct) values for each reaction were recorded and the relative DNA copies were calculated and normalised to Actin (2^(ActinCt-geneCt)).

Statistics

Total number of alleles amplified by SP-PCR was calculated using Poisson distribution based on the proportion of PCRs that did not yield a detectable product as described[35]. We determined P-values for small-pool PCR using a Mann-Whitney U nonparametric test comparing control and Cas9D10A + sgCTG conditions. We determined statistical significance between the Cas9D10A only and Cas9D10A + sgCTG HD iPSC-derived neuronal populations shown in Supplementary Fig. 2cd using a Student's t-test. Furthermore, differences in instability indices over time were calculated by one-way ANOVA with a post-hoc Tukey's multiple comparison test. We used two-way ANOVA to estimate differences on treatment and time, followed by a post-hoc Tukey's multiple comparison test. Statistical analyses were done using GraphPad Prism software version 10.0.0.

Acknowledgements

We would like to thank Andrew Jefferson for training and for keeping the fluorescence microscopes in excellent working order and Joanne Morgan and the core team at the Centre for Neuropsychiatric Genetics and Genomics at Cardiff University for support with sequencing. We thank Vanessa Wheeler and Geneviève Gourdon for critical reading of the manuscript. This research was undertaken using the supercomputing facilities at Cardiff University operated by Advanced Research Computing at Cardiff (ARCCA) on behalf of the Cardiff Supercomputing Facility and the HPC Wales and Supercomputing Wales (SCW) projects. We acknowledge the support of the latter, which is part-funded by the European Regional Development Fund (ERDF) via the Welsh Government. We also thank the University of Lausanne for providing some plasmids used in this study.

Funding

This work was supported by the UK Dementia Research Institute (DRI-TAP2022FA2, DRI-TAP2022FA14, and UK DRI-3006) which receives its funding from UK DRI Ltd, funded by the UK Medical Research Council, Alzheimer's Society and Alzheimer's Research UK and by the Neurogenetics Therapeutic Programme. VD is supported by a professorship from the Academy of Medical Sciences (AMSPR1\1014) and the

Moondance Foundation Laboratories. This project is part of the UCL Neurogenetic Therapies Programme, generously funded by The Sigrid Rausing Trust. It was partially supported by a grant from Huntington Society of Canada (to FC, VD, and Vanessa Wheeler) and by the Myotonic Dystrophy Foundation (to VD, JP, and Geneviève Gourdon). NS was funded by European Union's Horizon 2020 research and innovation programme awarded to NDA under the Marie Skłodowska-Curie Grant Agreement No. 813851. This project was further supported by the Future Leaders in Neuroscience Research Award by Neuroscience and Mental Health Innovation Institute at Cardiff University to A.M.

Author contributions

AM: Conceptualization, Formal Analysis, Investigation, Methodology, Supervision, Visualization, Writing – original draft, Writing – review & editing
MA: Conceptualization, Formal Analysis, Investigation, Methodology, Visualization, Writing – review & editing
MLarin: Conceptualization, Formal Analysis, Investigation, Methodology, Visualization, Writing – review & editing
ER: Methodology, Writing – review & editing
LH: Methodology, Writing – review & editing
RRD: Formal Analysis, Visualization, Writing – review & editing
ANA: Methodology, Writing – review & editing
AST: Formal Analysis, Visualization, Writing – review & editing
AMM: Resources, Writing – review & editing
NS: Investigation, Resources, Writing – review & editing
AERH: Formal Analysis, Writing – review & editing
PA: Methodology, Writing – review & editing
SB: Methodology, Writing – review & editing
NDA: Resources, Funding acquisition, Supervision, Writing – review & editing
JP: Conceptualization, Supervision, Funding Acquisition, Writing – review & editing
BLD: Resources, Supervision, Writing – review & editing
FC: Conceptualization, Funding acquisition, Supervision, Writing – review & editing
MLelos: Conceptualization, Funding acquisition, Supervision, Writing – review & editing
VD: Conceptualization, Funding acquisition, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing

Conflict of interests

VD declares that he has had a research contract with Pfizer Inc unrelated to this work within the past year. B.L.D. serves on the advisory board of Latus Biosciences, Patch Bio, Spirovant Biosciences, Resilience, and Carbon Biosciences and has sponsored research unrelated to this work from Roche, Latus, and Spirovant. All other authors declare no competing interests.

References

1. The Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72**, 971–983 (1993).
2. Fu, Y. H. *et al.* An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* **255**, 1256–1258 (1992).
3. Tabrizi, S. J. *et al.* Potential disease-modifying therapies for Huntington's disease: lessons learned and future opportunities. *Lancet Neurol.* **21**, 645–658 (2022).
4. Pascual-Gilabert, M., Artero, R. & López-Castel, A. The myotonic dystrophy type 1 drug development pipeline: 2022 edition. *Drug Discov. Today* **28**, 103489 (2023).
5. McAllister, B. *et al.* Timing and Impact of Psychiatric, Cognitive, and Motor Abnormalities in Huntington Disease. *Neurology* **96**, e2395–e2406 (2021).
6. Morales, F. *et al.* Individual-specific levels of CTG•CAG somatic instability are shared across multiple tissues in myotonic dystrophy type 1. *Hum. Mol. Genet.* **32**, 621–631 (2023).
7. Kennedy, L. *et al.* Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. *Hum. Mol. Genet.* **12**, 3359–3367 (2003).
8. Mätlik, K. *et al.* Cell-type-specific CAG repeat expansions and toxicity of mutant Huntingtin in human striatum and cerebellum. *Nat. Genet.* 1–12 (2024) doi:10.1038/s41588-024-01653-6.
9. Bettencourt, C. *et al.* DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. *Ann. Neurol.* **79**, 983–990 (2016).

10. Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium. Identification of Genetic Factors that Modify Clinical Onset of Huntington's Disease. *Cell* **162**, 516–526 (2015).
11. Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium. CAG Repeat Not Polyglutamine Length Determines Timing of Huntington's Disease Onset. *Cell* **178**, 887-900.e14 (2019).
12. Wheeler, V. C. & Dion, V. Modifiers of CAG/CTG Repeat Instability: Insights from Mammalian Models. *J. Huntingt. Dis.* **10**, 123–148 (2021).
13. Yang, S. *et al.* CRISPR/Cas9-mediated gene editing ameliorates neurotoxicity in mouse model of Huntington's disease. *J. Clin. Invest.* **127**, 2719–2724 (2017).
14. Merienne, N. *et al.* The Self-Inactivating KamiCas9 System for the Editing of CNS Disease Genes. *Cell Rep.* **20**, 2980–2991 (2017).
15. Duarte, F. *et al.* Limitations of Dual-Single Guide RNA CRISPR Strategies for the Treatment of Central Nervous System Genetic Disorders. *Hum. Gene Ther.* **34**, 958–974 (2023).
16. Duarte, F. *et al.* Semi-automated workflows to quantify AAV transduction in various brain areas and predict gene editing outcome for neurological disorders. *Mol. Ther. - Methods Clin. Dev.* **29**, 254–270 (2023).
17. Lo Scrudato, M. *et al.* Genome Editing of Expanded CTG Repeats within the Human DMPK Gene Reduces Nuclear RNA Foci in the Muscle of DM1 Mice. *Mol. Ther.* **27**, 1372–1388 (2019).
18. Cardinali, B. *et al.* Time-controlled and muscle-specific CRISPR/Cas9-mediated deletion of CTG-repeat expansion in the DMPK gene. *Mol. Ther. - Nucleic Acids* **27**, 184–199 (2022).
19. Oura, S. *et al.* Precise CAG repeat contraction in a Huntington's Disease mouse

model is enabled by gene editing with SpCas9-NG. *Commun. Biol.* **4**, 1–13 (2021).

20. Monteys, A. M., Ebanks, S. A., Keiser, M. S. & Davidson, B. L. CRISPR/Cas9 Editing of the Mutant Huntingtin Allele In Vitro and In Vivo. *Mol. Ther.* **25**, 12–23 (2017).
21. Fang, L. *et al.* Haplotyping SNPs for allele-specific gene editing of the expanded huntingtin allele using long-read sequencing. *Hum. Genet. Genomics Adv.* **4**, 100146 (2023).
22. Xue, C. & Greene, E. C. DNA Repair Pathway Choices in CRISPR-Cas9-Mediated Genome Editing. *Trends Genet.* **37**, 639–656 (2021).
23. Fichter, K. M., Setayesh, T. & Malik, P. Strategies for precise gene edits in mammalian cells. *Mol. Ther. Nucleic Acids* **32**, 536–552 (2023).
24. Yan, S. *et al.* Cas9-mediated replacement of expanded CAG repeats in a pig model of Huntington's disease. *Nat. Biomed. Eng.* **7**, 629–646 (2023).
25. van Agtmaal, E. L. *et al.* CRISPR/Cas9-Induced (CTG·CAG)n Repeat Instability in the Myotonic Dystrophy Type 1 Locus: Implications for Therapeutic Genome Editing. *Mol. Ther. J. Am. Soc. Gene Ther.* **25**, 24–43 (2017).
26. Wang, Y. *et al.* Therapeutic Genome Editing for Myotonic Dystrophy Type 1 Using CRISPR/Cas9. *Mol. Ther.* **26**, 2617–2630 (2018).
27. Mosbach, V. *et al.* Resection and repair of a Cas9 double-strand break at CTG trinucleotide repeats induces local and extensive chromosomal deletions. *PLOS Genet.* **16**, e1008924 (2020).
28. Zhang, X.-H., Tee, L. Y., Wang, X.-G., Huang, Q.-S. & Yang, S.-H. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Mol. Ther. - Nucleic Acids* **4**, e264 (2015).

29. Han, H. A., Pang, J. K. S. & Soh, B.-S. Mitigating off-target effects in CRISPR/Cas9-mediated in vivo gene editing. *J. Mol. Med.* **98**, 615–632 (2020).
30. Shen, B. *et al.* Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat. Methods* **11**, 399–402 (2014).
31. Choi, D. E. *et al.* Base editing strategies to convert CAG to CAA diminish the disease-causing mutation in Huntington's disease. 2023.04.28.538700 Preprint at https://doi.org/10.1101/2023.04.28.538700 (2023).
32. Bečanović, K. *et al.* A SNP in the HTT promoter alters NF-κB binding and is a bidirectional genetic modifier of Huntington disease. *Nat. Neurosci.* **18**, 807–816 (2015).
33. Van Raamsdonk, J. M. *et al.* Loss of wild-type huntingtin influences motor dysfunction and survival in the YAC128 mouse model of Huntington disease. *Hum. Mol. Genet.* **14**, 1379–1392 (2005).
34. Cinesi, C., Aeschbach, L., Yang, B. & Dion, V. Contracting CAG/CTG repeats using the CRISPR-Cas9 nickase. *Nat. Commun.* **7**, 13272 (2016).
35. Aeschbach, L. & Dion, V. Minimizing carry-over PCR contamination in expanded CAG/CTG repeat instability applications. *Sci. Rep.* **7**, 18026 (2017).
36. HD iPSC Consortium, T. Induced Pluripotent Stem Cells from Patients with Huntington's Disease Show CAG-Repeat-Expansion-Associated Phenotypes. *Cell Stem Cell* **11**, 264–278 (2012).
37. Jy, S. *et al.* Expression of mutant huntingtin in glial cells contributes to neuronal excitotoxicity. *J. Cell Biol.* **171**, (2005).
38. Garcia, V. J. *et al.* Huntington's Disease Patient-Derived Astrocytes Display Electrophysiological Impairments and Reduced Neuronal Support. *Front. Neurosci.* **13**, (2019).

39. Ah, J. *et al.* Frequency of nuclear mutant huntingtin inclusion formation in neurons and glia is cell-type-specific. *Glia* **65**, (2017).
40. Estrada-Sánchez, A. M. & Rebec, G. V. Corticostriatal dysfunction and glutamate transporter 1 (GLT1) in Huntington's disease: Interactions between neurons and astrocytes. *Basal Ganglia* **2**, 57–66 (2012).
41. Faideau, M. *et al.* In vivo expression of polyglutamine-expanded huntingtin by mouse striatal astrocytes impairs glutamate transport: a correlation with Huntington's disease subjects. *Hum. Mol. Genet.* **19**, 3053–3067 (2010).
42. 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).
43. Bates, G. P. *et al.* Huntington disease. *Nat. Rev. Dis. Primer* **1**, 15005 (2015).
44. De Serres-Bérard, T., Pierre, M., Chahine, M. & Puymirat, J. Deciphering the mechanisms underlying brain alterations and cognitive impairment in congenital myotonic dystrophy. *Neurobiol. Dis.* **160**, 105532 (2021).
45. Wilm, A. *et al.* LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res.* **40**, 11189–11201 (2012).
46. Bae, S., Park, J. & Kim, J.-S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014).
47. Horlbeck, M. A. *et al.* Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *eLife* **5**, e12677 (2016).
48. Taylor, A. S. *et al.* Repeat Detector: versatile sizing of expanded tandem repeats

and identification of interrupted alleles from targeted DNA sequencing. *NAR Genomics Bioinforma.* **4**, lqac089 (2022).

49. Smolka, M. *et al.* Detection of mosaic and population-level structural variants with Sniffles2. *Nat. Biotechnol.* 1–10 (2024) doi:10.1038/s41587-023-02024-y.
50. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
51. Mangiarini, L. *et al.* Exon 1 of the HD Gene with an Expanded CAG Repeat Is Sufficient to Cause a Progressive Neurological Phenotype in Transgenic Mice. *Cell* **87**, 493–506 (1996).
52. Gomes-pereira, M., Foiry, L. & Gourdon, G. CHAPTER 36 - Transgenic Mouse Models of Unstable Trinucleotide Repeats: Toward an Understanding of Disease-Associated Repeat Size Mutation. in *Genetic Instabilities and Neurological Diseases (Second Edition)* (eds. Wells, R. D. & Ashizawa, T.) 563–583 (Academic Press, Burlington, 2006). doi:10.1016/B978-012369462-1/50037-5.
53. Msh2 Acts in Medium-Spiny Striatal Neurons as an Enhancer of CAG Instability and Mutant Huntingtin Phenotypes in Huntington's Disease Knock-In Mice | PLOS ONE. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0044273.
54. Tsai, Y.-C. *et al.* Identification of a CCG-Enriched Expanded Allele in Patients with Myotonic Dystrophy Type 1 Using Amplification-Free Long-Read Sequencing. *J. Mol. Diagn. JMD* **24**, 1143–1154 (2022).
55. Suzuki, K. *et al.* In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* **540**, 144–149 (2016).
56. Bubeck, F. *et al.* Engineered anti-CRISPR proteins for optogenetic control of

CRISPR–Cas9. *Nat. Methods* **15**, 924–927 (2018).

57. Lee, J.-M. *et al.* A novel approach to investigate tissue-specific trinucleotide repeat instability. *BMC Syst. Biol.* **4**, 29 (2010).
58. Wang, D., Zhang, F. & Gao, G. CRISPR-Based Therapeutic Genome Editing: Strategies and In Vivo Delivery by AAV Vectors. *Cell* **181**, 136–150 (2020).
59. Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186–191 (2015).
60. Estevez-Fraga, C., Tabrizi, S. J. & Wild, E. J. Huntington's Disease Clinical Trials Corner: August 2023. *J. Huntingt. Dis.* **12**, 169–185 (2023).
61. Hasuike, Y. *et al.* CAG repeat-binding small molecule improves motor coordination impairment in a mouse model of Dentatorubral-pallidoluysian atrophy. *Neurobiol. Dis.* **163**, 105604 (2022).
62. Nakamori, M. *et al.* A slipped-CAG DNA-binding small molecule induces trinucleotide-repeat contractions in vivo. *Nat. Genet.* **52**, 146–159 (2020).
63. Serio, A. *et al.* Astrocyte pathology and the absence of non-cell autonomy in an induced pluripotent stem cell model of TDP-43 proteinopathy. *Proc. Natl. Acad. Sci.* **110**, 4697–4702 (2013).
64. Shi, Y., Kirwan, P. & Livesey, F. J. Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks. *Nat. Protoc.* **7**, 1836–1846 (2012).
65. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
66. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
67. Pedersen, B. S. & Quinlan, A. R. Mosdepth: quick coverage calculation for

genomes and exomes. *Bioinformatics* **34**, 867–868 (2018).

68. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* **339**, 823–826 (2013).
69. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
70. Ruiz Buendía, G. A. *et al.* Three-dimensional chromatin interactions remain stable upon CAG/CTG repeat expansion. *Sci. Adv.* **6**, eaaz4012 (2020).
71. Dion, V., Lin, Y., Hubert, L., Waterland, R. A. & Wilson, J. H. Dnmt1 deficiency promotes CAG repeat expansion in the mouse germline. *Hum. Mol. Genet.* **17**, 1306–1317 (2008).

Supplementary Material

Cas9 nickase-mediated contraction of CAG/CTG repeats at multiple disease loci

Alvaro Murillo[1], Melanie Alpaugh[2], Meghan Larin[1,3], Emma L. Randall[1], Laura Heraty[1], Ruban Rex Durairaj[1], Alys N. Aston[1], Alysha S. Taylor[1], Alex Mas Monteys[4], Nina Stöberl[5], Aeverie E. R. Heuchan[1], Pascale Aeschlimann[1], Soumyasree Bhattacharyya[1], Nicholas D. Allen[5], Jack Puymirat[6], Beverly L. Davidson[4], Francesca Cicchetti[7], Mariah Lelos[5], and Vincent Dion[1,8*]

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

2: Department of Molecular and Cellular Biology, University of Guelph, Guelph, Ontario, Canada N1G 2W1

3: Current address: AstraZeneca R&D 1 Francis Crick Avenue Cambridge, UK CB2 0AA.

4: Raymond G. Perelman Center for Cellular and Molecular Therapeutics, Children's Hospital of Philadelphia, Philadelphia, 19104, USA.

5: School of Biosciences, Cardiff University, The Sir Martin Evans Building, Museum Avenue, Cardiff, CF10 3AX

6: LOEX, CHU de Québec-Université Laval Research Center, Quebec City, QC, Canada.

7: Centre de Recherche du CHU de Québec, Axe Neurosciences; Département de Psychiatrie et Neurosciences, Université Laval, Québec, QC, Canada.

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

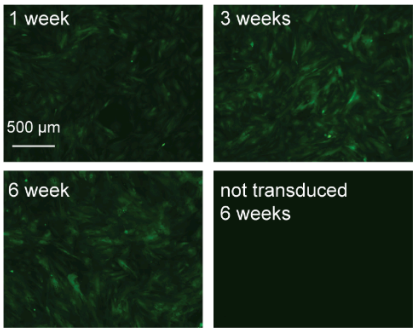
*: corresponding author: dionv@cardiff.ac.uk

A.

HD iPSC-derived astrocytes

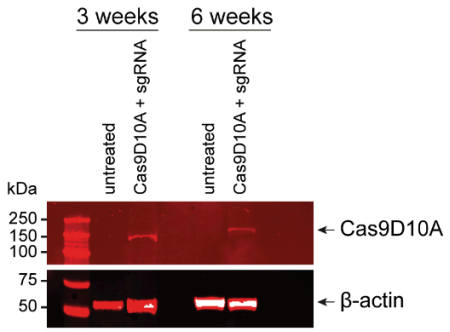
Time post transduction	no. nuclei	Ki67⁺ cells	% of Ki67⁺ cells
3 weeks	423	1	0.24
6 weeks	868	2	0.23

B.



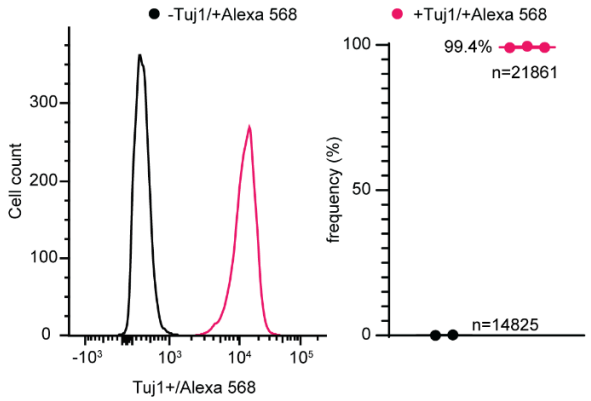
GFP

C.

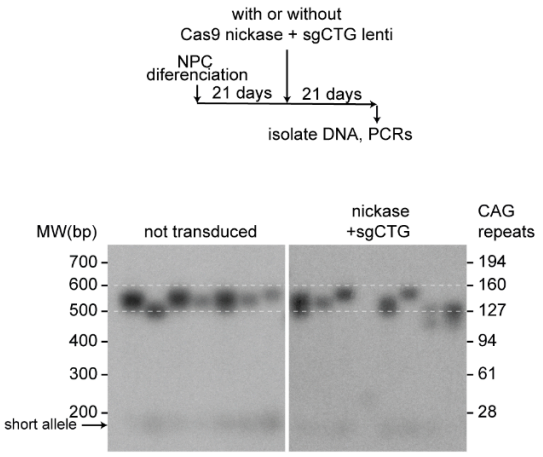


D.

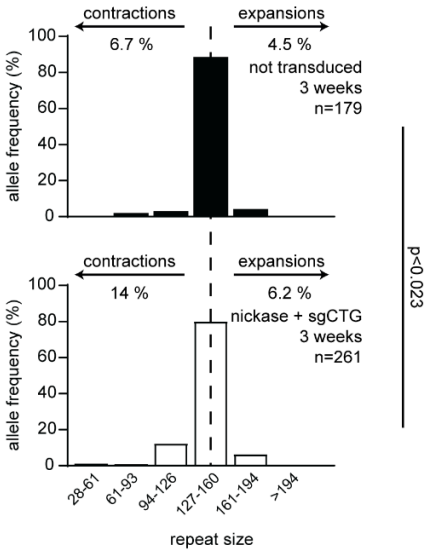
HD iPSC-derived cortical neurons



E.



F.



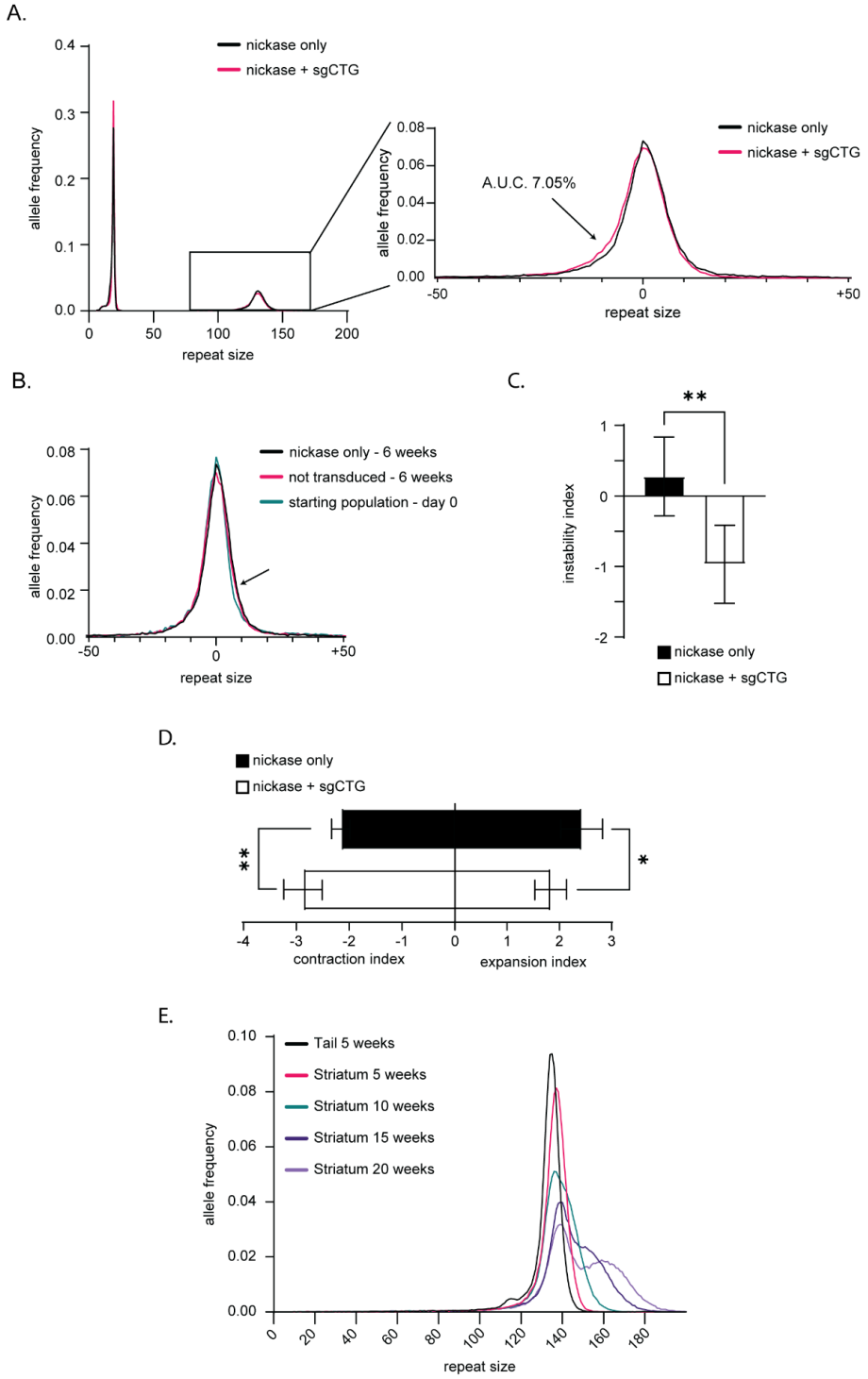
G.

DM1 iPSC-derived cortical neurons

no. nuclei	Tuj1⁺	Ki67⁺ cells	% of Tuj1⁺ cells	% of Ki67⁺ cells
1148	1128	19	98 %	1.7 %

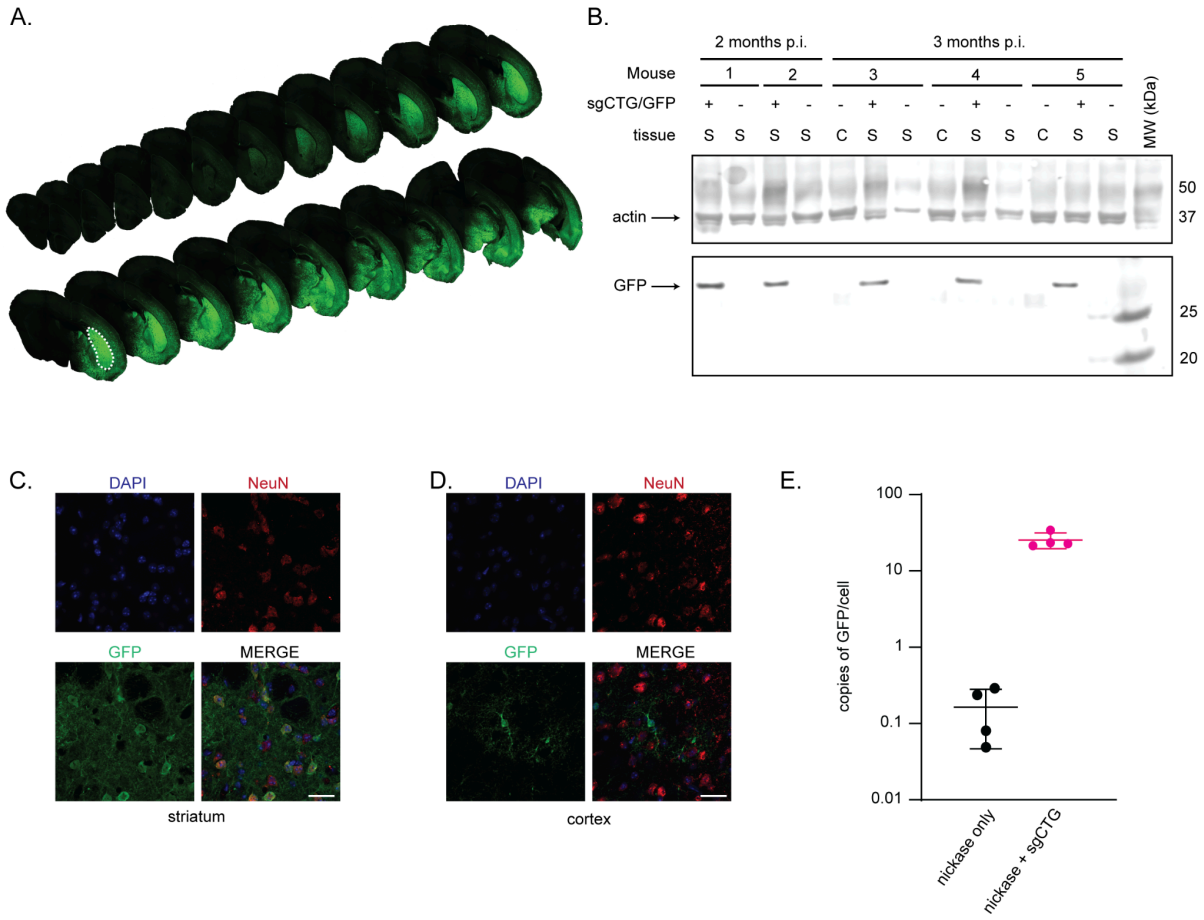
Supplementary Figure 1: CAG/CTG repeat contractions in HD iPSC-derived astrocytes and neurons. A) HD iPSC-derived astrocyte cultures stained for Ki67 as a measure of proliferating potential 3 and 6 weeks after transduction with Cas9D10A and sgCTG. B) Lentivirus-encoded GFP expression on the sgCTG vector over a 6 week time course in HD iPCS-derived S100□⁺ astrocytes (pLV[gRNA]-CMV-EGFP-U6>{sgCTG}). C) Cas9D10A expression in HD iPSC-derived S100□⁺

astrocytes over 3 and 6 weeks from (pLenti-EF1alpha-Cas9D10A nickase-Blast). D) Characterisation of the HD iPSC-derived cortical neurons. Left: Representative flow cytometry histogram of cultures stained with □-tubulin III (Tuj1; magenta) and only with the secondary antibody (black). Right: quantification of 2 to 3 independent experiments showing that over 99% of the cells expressed this neuronal marker. N = number of cells. E) Timeline and representative small pool PCR blot of HD iPSC-derived neuronal cultures grown in the presence of both Cas9D10A and sgCTG or untreated for 3 weeks. F) Quantification of the small pool PCRs. P-value determined using a Mann-Whitney U test comparing non-transduced and Cas9D10A + sgCTG. G) DM1 iPSC-derived neuronal cultures stained for Ki67 and Tuj1, showing that over 98% of the cells expressed a neuronal marker and few cells had the potential to proliferate.

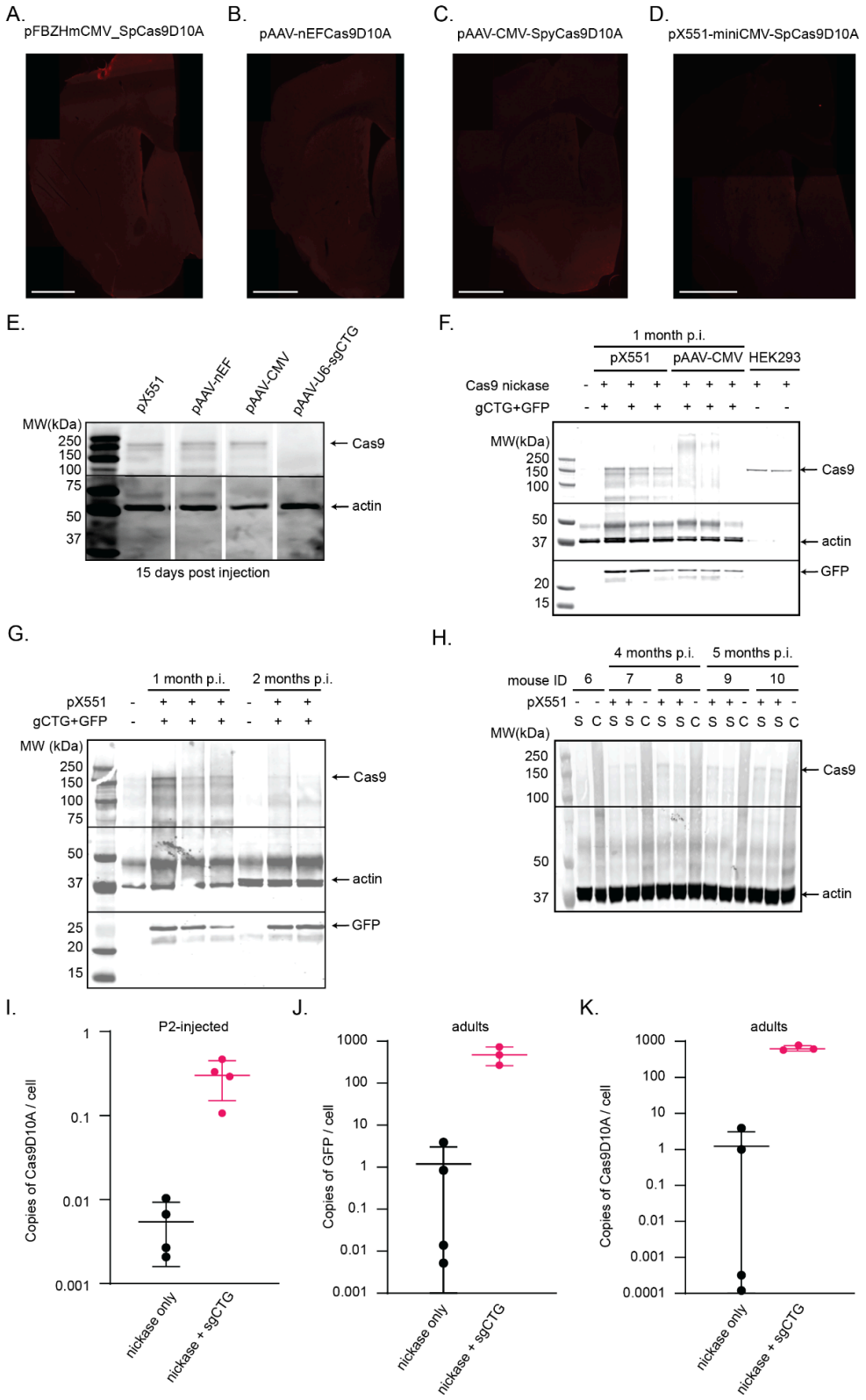


Supplementary figure 2: SMRT sequencing reproduces expected patterns of repeat size distribution. A) SMRT sequencing of DNA from HD iPSC-derived cortical neurons showing both the non-pathogenic and the expanded allele (left) and the change in repeat size between the cells treated with Cas9D10A only versus those treated with both Cas9D10A and sgCTG for 6 weeks (n=4 per condition). Analysis of the difference in the area under the curve (AUC) shows that 7% of the cells transduced with both Cas9D10A and sgCTG had shorter alleles above what is seen in cells treated with Cas9D10A treated only. This is similar to what we found with small-pool PCR. B) Changes in

repeat size between cultures of HD-derived cortical neurons that have not been transduced (n=4), transduced with Cas9D10A only (n=4), compared to the starting repeat size distribution (n=2). Arrow indicates the accumulation of expansions over the 6 week period. The calculated AUC difference for expansions between day 0 and non-transduced cells at 6 weeks was 5.6 % whereas between D0 Cas9D10A only the AUC was 7.8%. C&D) Indices of the HD iPSC-derived cortical neurons treated with Cas9D10A only (black boxes) and Cas9D10A + sgCTG (white boxes) (n=4 per condition). *: $P<0.05$, **:$P<0.01$ calculated using a Student's t-test test. E) Representative example time course analysis using SMRT sequencing showing the expected increase in the frequency of longer alleles over several weeks in the R6/1 striatum. These data show that the approach can measure somatic expansions readily.

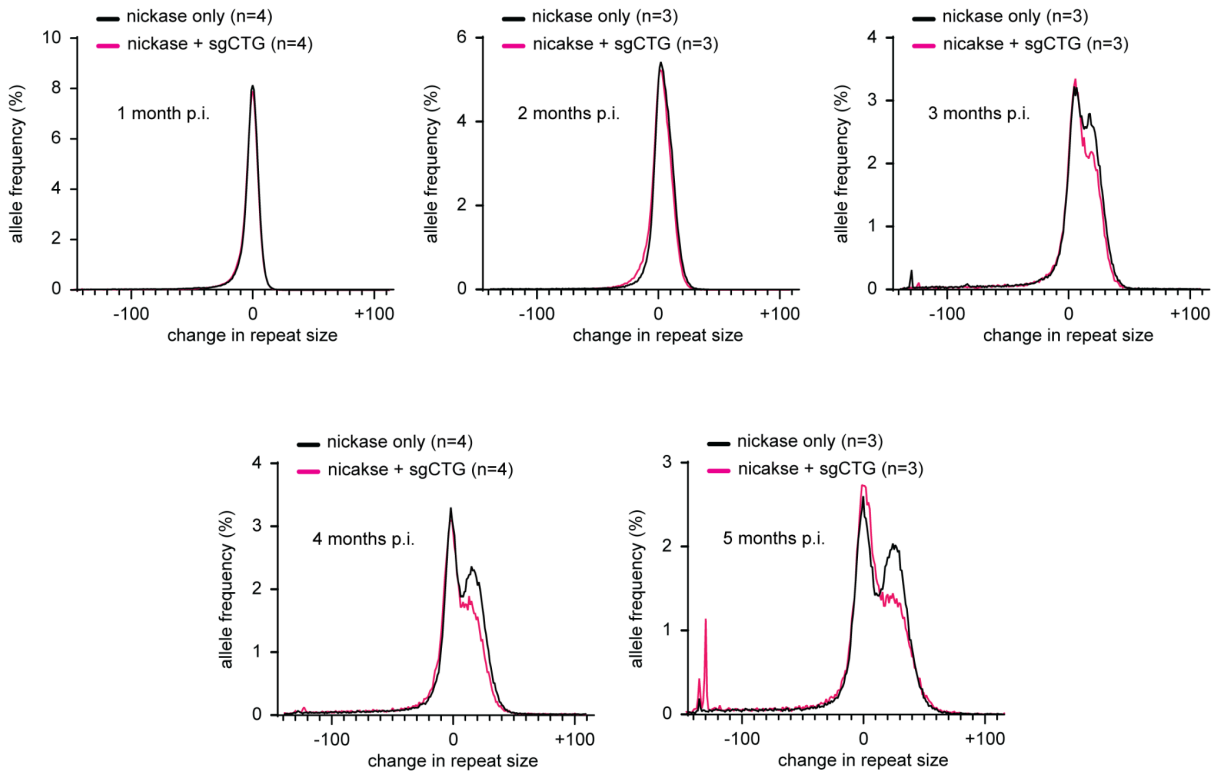


Supplementary Figure 3: GFP expression in P2-injected mice. A) Coronal sections through the mouse brain showing GFP expression in the injected hemisphere and throughout the striatum (dashed lines) and into much of the cortex. B) GFP and actin expression 2 and 3 months post injection (p.i.) in striatum (S) and cerebellum (C). C) Representative example of GFP colocalising with NeuN+ neurons in the striatum. D) GFP expression in the cortex did not colocalise with NeuN, but rather the GFP positive cells had an astrocytic morphology. Scale bar = 25µm. E) qPCR results for the number of GFP DNA found in nickase-only or nickase + sgCTG-injected striatum, showing that there were, on average, 30 copies of the AAV per cell 1 month after injection n=4 per condition.

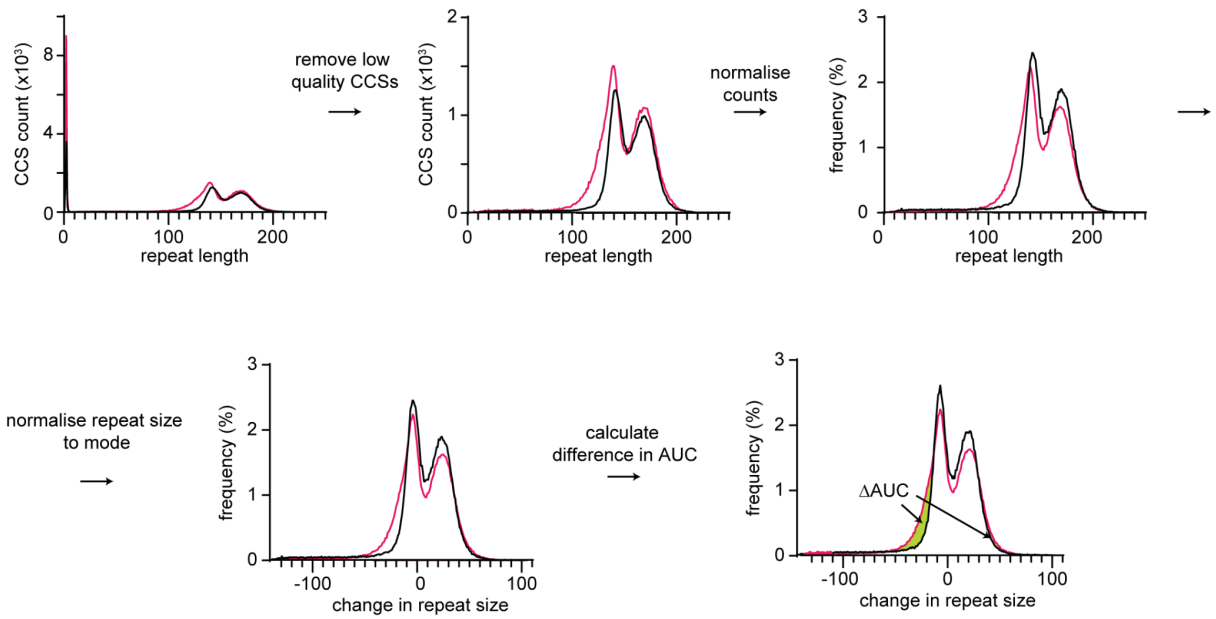


Supplementary Figure 4: Cas9D10A expression *in vivo*. A) Immunofluorescence of Cas9D10A expressed from pFBZHmCMV_SpCas9D10A AAV9 injected into wild type striatum. B) Same as A, but using pAAV-nEFCas9D10A. C) Same as A but using pAAV-CMV-SpyCas9D10A. D) Same as A but using pX551-miniCMV-SpCas9D10A. Scale bars = 1000 µm E) Western blot of Cas9D10A expression from the indicated AAVs in the striatum of adult wild type mice 15 days post injection (p.i.). This gel

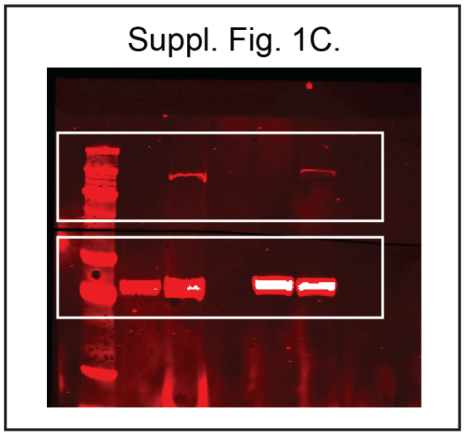
was spliced for ease of comparison. Full blots are found in Supplementary Figure 7. F) Western blot of Cas9D10A expression in the striatum of adult R6/1 mice injected with the indicated AAVs 1 month post injection. G) Western blot of Cas9D10A expression in the striatum of adult R6/1 mice from an AAV9 packaged using pX551-miniCMV-SpCas9D10A, 1 and 2 months after injection. H) Cas9D10A expression in the striatum of P2-injected R6/1 mice using an AAV9 packaged from pX551-miniCMV-SpCas9D10A, 4 and 5 months post injection. I) qPCR of Cas9D10A DNA in the striatum of R6/1 P2-injected mice 1 month after injection n=4 per condition. J) qPCR of GFP DNA in the striatum of R6/1 mice injected at 3.5 months and analysed after 1 month. K) Same as J, but for Cas9D10A DNA (Number of animals: Cas9D10A nickase only n=4; Cas9D10A nickase + sgCTG n=3).

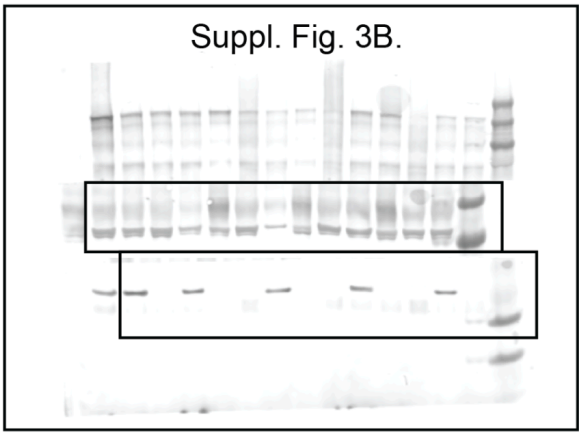


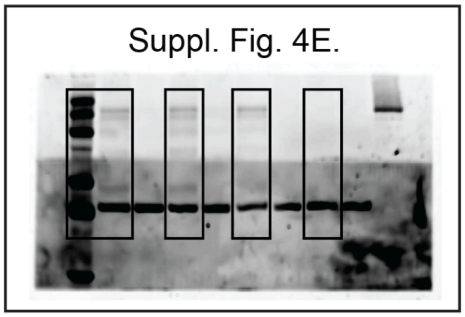
Supplementary Figure 5: Instability of CAG repeats in P2-injected mice over time. Same experiments quantified in Figure 2. The data at 1 month post-infection and 5 months post-infection is the same as in Figure 2. P.i. = post-injection.

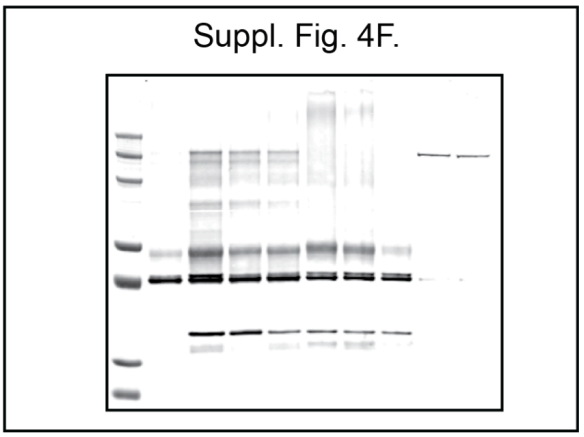


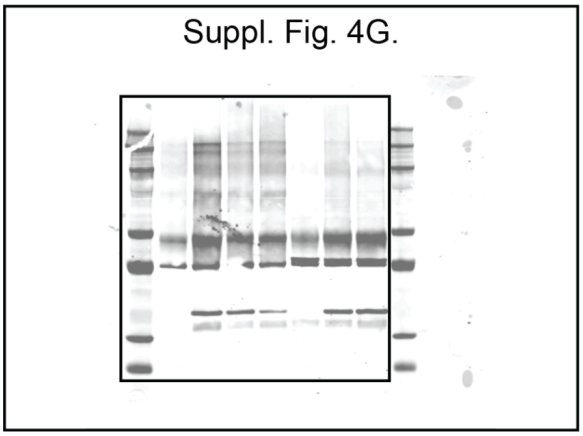
Supplementary Figure 6: Explanation of how the difference in the area under the curve was calculated. This is an example dataset. Repeat Detector is used to determine repeat sizes. Then repeat size between 0 and 5 CAGs are removed as they are low quality reads that do not align to the *HTT* locus. Then the graphs are normalised from CCS counts and for the modal repeat size of each sample. This prevents taking small expansions in the untreated samples to inflate the number of contractions. The difference in the area under the curve is calculated for both shorter and longer alleles.

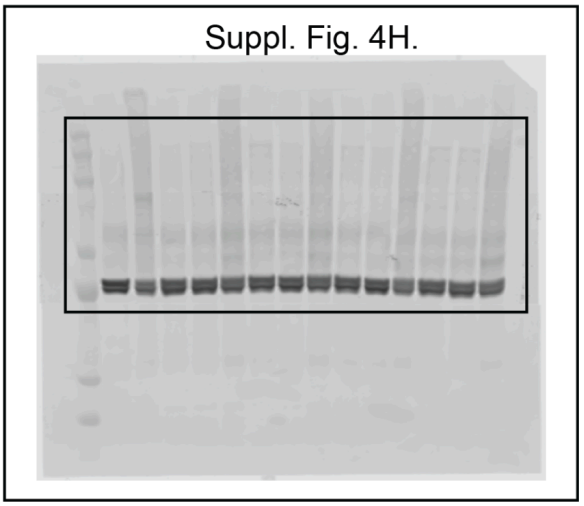




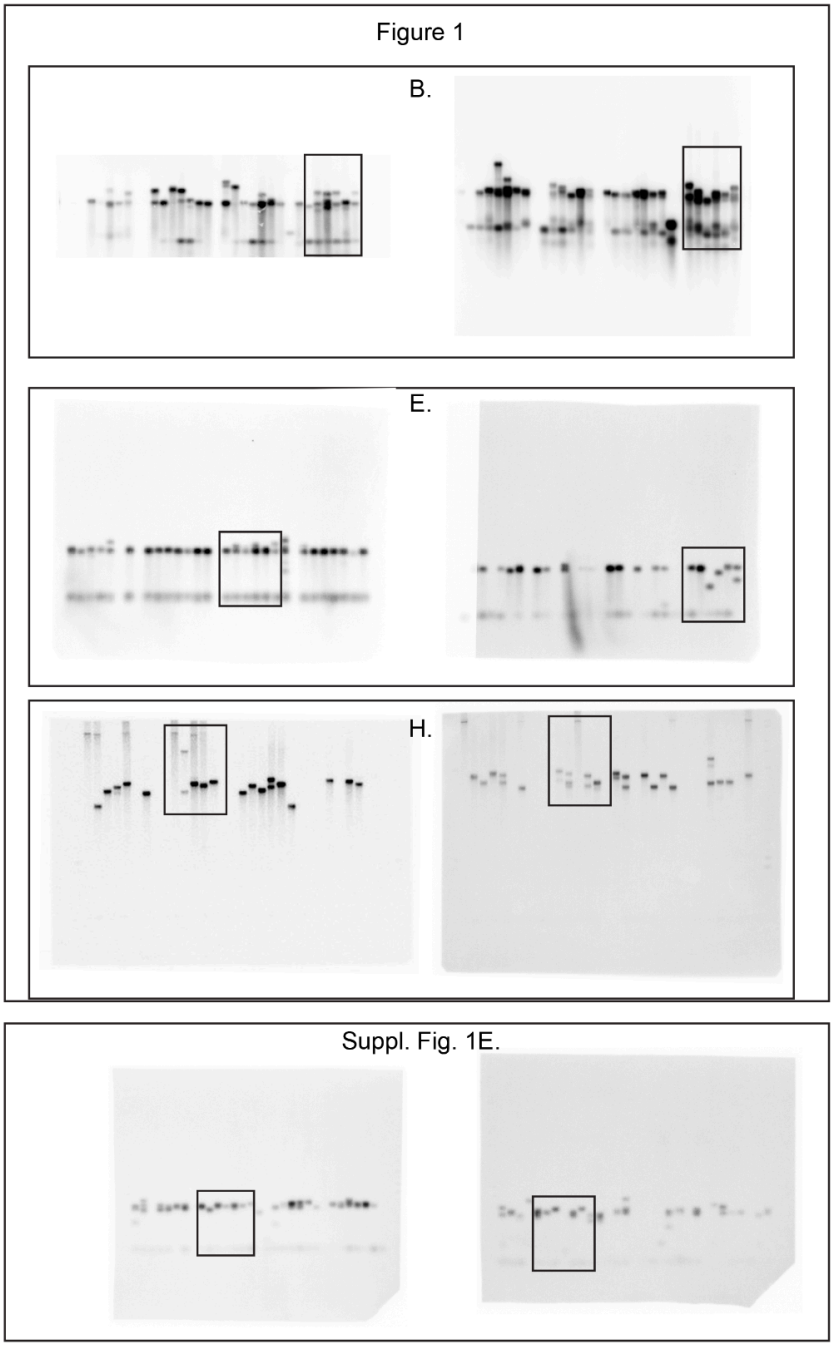








<u>Supplementary Figure 7:</u> Unaltered full western blots membranes from Supplementary Figure 1a, Supplementary Figure 3b and Supplementary Figure 4efgh. Black boxes indicate where the blots were cropped.



Supplementary Figure 8: Unaltered full small-pool PCR membranes from Figure 1beh and Supplementary Figure 1e. Black boxes indicate where the membranes were cropped.

Supplementary Table 1: Plasmids used herein.

Name	Description	Addgene number	Reference
pcDNA3.3-TOPO - Cas9D10A nickase	Cas9_D10A expression plasmid. Also harbours a G418 resistance gene. Used as a positive control for Cas9D10A detection in HEK293 cells using western blotting.	41816	[68]
pPN10 – sgCTG	Expresses the sgRNA against $(CTG)_6$ from a U6 promoter. Also contains a Puromycin resistance gene.	114385	[34]
pLenti-EF1alpha-Cas9D10A nickase -Blast	Expresses Cas9D10A together with a blasticidin resistance gene. Suitable for lentivirus packaging.	63593	[69]
pLV(gRNA)-CMV-eGFP-U6(sgCTG)	Contains a eGFP gene along with a U6 promoter driving the sgCTG (target sequence: $(CUG)_6$). Used with the HD iPSC-derived astrocytes.	216732	This study
pLenti-U6-sgCTG-CAG-Cas9D10A nickase -Blast	Expresses Cas9D10A under a CAG promoter together with a blasticidin resistance gene, along with a U6 promoter driving the sgCTG (target sequence: $(CUG)_6$). Suitable for lentivirus packaging.	216733	This study
pAAV-U6-sgCTG-CMV-GFP	AAV plasmid containing sgCTG (sgCTG (target sequence: $(CUG)_6$) driven by a U6 promoter and eGFP under the control of the CMV promoter. Used for the *in vivo* experiments.	216734	This study
pX551-miniCMV-SpCas9D10A nickase	Expresses SpCas9-D10A nickase from a miniCMV promoter. For AAV packaging. Modified from pX551-miniCMV-SpCas9, which was a gift from Alex Hewitt (Addgene #107031)	216735	This study
pAAV-nEF-SpCas9D10A nickase	Expresses Cas9-D10A nickase from a nEF promoter. For AAV packaging. Derived from pAAV-nEF-SpCas9 (Addgene #87115)[55].	216736	This study
pAAV-CMV-SpCas9D10A nickase	Expresses Cas9-D10A nickase from a CMVd1 promoter. For AAV packaging. Derived from pAAV-CMV-SpCas9 (Addgene #113034)[56]	216737	This study

Supplementary Table 2: *List of potential off-target genes*.*

PAM	$(CAG)_x$	Chromosome	Direction	Gene
AGG	7	1	-	*CAMTA1* (calmodulin binding transcription activator 1)
TGG	7	1	-	ENSG00000215908 (CROCCP2 pseudo gene)
CGG	7	1	-	*ARID1A* (AT-rich interaction domain 1A)
GAG	11	1	-	*LRP8* (LDL receptor related protein 8)
CGG	7	1	-	*ST6GALNAC*5 (ST6 N-acetylgalactosaminide alpha-2,6-sialyltransferase 5)
AAG	9	1	-	*ADGRL2* (adhesion G protein-coupled receptor L2)
CAG	6	1	-	*LINC01709* (101,639,560..101,785,723)
AAG	14	1	+	*KCNN3* (potassium calcium-activated channel subfamily N member 3)
GAG	6	1	-	*NOS1AP* (nitric oxide synthase 1 adaptor protein)
TGG	6	1	+	*QSOX1* (quiescin sulfhydryl oxidase 1)
CGG	6	1	+	*TMEM63A* (transmembrane protein 63A)
GGG	6	1	-	*TTC13* (tetratricopeptide repeat domain 13)
CAG	6	1	-	intergenic
CAG	7	1	-	*FMN2* (formin 2)
CGG	7	2	+	*FBXO11* (F-box protein 11)
CAG	9	2	+	ENSG00000233723 (LINC01122)
GGG	8	2	-	*EIF2KA3* (eukaryotic translation initiation factor 2 alpha kinase 3)
GAG	7	2	+	intergenic
GGG	7	2	-	*PLA2R1* (phospholipase A2 receptor 1)
CAG	12	2	-	intergenic
CAG	6	2	+	intergenic
AGG	7	2	-	*RAPH1* (Ras association (RalGDS/AF-6) and pleckstrin homology domains 1)

PAM	$(CAG)_x$	Chromosome	Direction	Gene
GGG	7	2	+	*RS1I* (insulin receptor substrate 1)
CAG	8	2	+	*DNER* (delta/notch like EGF repeat containing)
TGG	6	2	-	*GIGYF2* (GRB10 interacting GYF protein 2)
GAG	6	3	-	*TPRXL* (tetrapeptide repeat homeobox like (pseudogene))*
GAG	6	3	-	*TPRXL* (tetrapeptide repeat homeobox like (pseudogene))
GAG	6	3	-	*LIMD1* (LIM domains containing 1)
TGG	6	3	+	*RNF123* (ring finger protein 123)
TAG	6	3	+	*FLNB* (filamin B)
AGG	8	3	-	*MGLL* (monoglyceride lipase)
CAG	6	3	+	*TMCC1* (transmembrane and coiled-coil domain family 1)
CAG	6	3	+	ENSG00000175110 (MRPS22)
CAG	6	4	+	intergenic
AAG	6	4	-	*SORCS2* (sortilin related VPS10 domain containing receptor 2)
CAG	8	4	-	intergenic
CAG	7	4	-	*F11-AS1* (F11 antisense RNA 1)
CAG	6	4	+	intergenic
CAG	6	5	-	*UBE2QL1* (ubiquitin conjugating enzyme E2 Q family like 1)
TAG	6	5	+	LOC107986418 (complement(63,343,882..63,528,654))
CAG	8	5	+	ENSG00000271904 (AC091826.2)
CAG	17	5	+	intergenic
CAG	10	5	+	*PPP2R2B* (protein phosphatase 2 regulatory subunit Bbeta)
GGG	9	6	+	*FARS2* (phenylalanyl-tRNA synthetase 2, mitochondrial)

PAM	$(CAG)_x$	Chromosome	Direction	Gene
TGG	6	6	-	*TREML2* (triggering receptor expressed on myeloid cells like 2)
AGG	8	6	+	*BACH2* (BTB domain and CNC homolog 2)
CAG	6	6	-	*KLHL32* (kelch like family member 32)
CAG	6	6	+	*FOXO3* (forkhead box O3)
GAG	10	7	-	*POU6F2* (POU class 6 homeobox 2)
GGG	7	7	-	*CAMK2B* (calcium/calmodulin dependent protein kinase II beta)
AGG	11	7	+	intergenic
TGG	6	7	+	*GIGYF1* (GRB10 interacting GYF protein 1)
CAG	6	7	-	*CUX1* (cut like homeobox 1)
GGG	7	7	+	*SMO* (smoothened, frizzled class receptor)
CGG	6	7	+	*AGAP3* (ArfGAP with GTPase domain, ankyrin repeat and PH domain 3)
GAG	7	7	+	*PAXIP1* (PAX interacting protein 1)
CAG	6	8	+	*STC1* (stanniocalcin 1)
CGG	6	8	-	ENSG00000254777 (AC022182.1)
AAG	6	9	-	LOC101927042 (uncharacterized LOC101927042, transcript variant X1)
AAG	6	9	-	intergenic
AAG	6	9	+	LOC101927602 (complement(67,365,072..67,414,285))
CAG	6	9	-	*SLC25A25-AS1* (SLC25A25 antisense RNA 1)
CAG	8	9	+	intergenic
TGG	6	9	+	*COL5A1* (collagen type V alpha 1 chain)
GGG	8	11	-	*DCHS1* (dachsous cadherin-related 1)
GAG	10	11	-	*CSRP3-AS1* (19,196,233..19,312,856)
CGG	6	11	-	*PACS1* (phosphofurin acidic cluster sorting protein 1)

PAM	$(CAG)_x$	Chromosome	Direction	Gene
GGG	6	11	-	intergenic
CAG	6	11	+	*BACE1* (beta-secretase 1)
AGG	6	11	-	*CEP164* (centrosomal protein 164)*
CAG	8	11	+	*PRDM10* (PR/SET domain 10)
AAG	7	12	+	*KMT2D* (lysine methyltransferase 2D)
CAG	6	12	+	*KMT2D* (lysine methyltransferase 2D)
TGG	10	12	-	ENSG00000287014 (AC087883.2)
CAG	9	12	-	*CFAP54* (cilia and flagella associated protein 54)
GGG	13	12	+	*ATXN2*
CAG	6	12	-	*TPCN1* (two pore segment channel 1)
GGG	8	12	-	*LRCOL1* (leucine rich colipase like 1)
TGG	6	13	-	intergenic
GGG	6	14	-	*ADCY4* (adenylate cyclase 4)*
CAG	6	14	-	*NPAS3* (neuronal PAS domain protein 3)
GGG	7	14	+	LOC105378178 (complement(48,427,255..49,288,023))
TGG	8	14	+	*MIDEAS* (mitotic deacetylase associated SANT domain protein)
AAG	6	15	-	*ZNF609* (zinc finger protein 609)
AAG	8	15	-	*ARID3B* (AT-rich interaction domain 3B)
CGG	7	15	+	*ADAMTS17* (ADAM metallopeptidase with thrombospondin type 1 motif 17)
CAG	8	16	+	intergenic
GGG	7	16	+	*ZFHX3* (zinc finger homeobox 3)
AGG	10	16	+	*CDH13* (cadherin 13)
GAG	10	17	-	*BCL6B* (BCL6B transcription repressor)*
AGG	6	17	+	*NUFIP2* (nuclear FMR1 interacting protein 2)

PAM	$(CAG)_x$	Chromosome	Direction	Gene
AAG	6	17	+	*KRT14* (keratin 14)
CAG	11	17	+	intergenic
CAG	6	17	+	*ERN1* (endoplasmic reticulum to nucleus signaling 1)
GGG	13	19	+	*CACNA1A*
GGG	6	19	-	*YIF1B* (Yip1 interacting factor homolog B, membrane trafficking protein)
CAG	6	19	+	*NKPD1* (NTPase KAP family P-loop domain containing 1)
GGG	20	19	-	DMPK
CGG	8	19	+	*FCGRT* (Fc fragment of IgG receptor and transporter)
CAG	6	20	+	intergenic
TGG	6	20	+	intergenic
TGG	7	21	+	*SCAF4* (SR-related CTD associated factor 4)
CAG	6	21	-	*PTTG1IP* (PTTG1 interacting protein)
TAG	7	22	-	*MED15* (mediator complex subunit 15)
GAG	8	22	-	*MED15* (mediator complex subunit 15)
GGG	6	22	+	*MN1* (MN1 proto-oncogene, transcriptional regulator)
CAG	8	22	-	*TNRC6B* (trinucleotide repeat containing adaptor 6B)
TAG	6	X	+	intergenic
CAG	12	X	+	LOC107985695 (48,617,240..48,639,305)
CAG	22	X	-	*AR* (androgen receptor)
CGG	7	X	-	intergenic

List obtained from Cas-OFFinder[46]

*: not analysed as part of the HD iPSC-derived cortical neurons treated for 3 weeks dataset.

Supplementary table 3: Primers used herein.

Primer	Sequence (5′ to 3’)	Application	Reference
oVIN-1333	CCGCTCAGGTTCTGCTTTTA	Amplification of the *HTT* locus in astrocytes for small-pool PCR.	[70]
oVIN-1334	CAGGCTGCAGGGTTACCG	Amplification of the *HTT* locus in astrocytes for small-pool PCR.	[70]
oVIN-1345	ATGAAGGCCTTCGAGTCCCTCAAGTCCTTC	Amplification of the *HTT* locus in neurons for small-pool PCR.	This study
oVIN-1347	CGGCTGAGGCAGCAGCGGCTGT	Amplification of the *HTT* locus in neurons for small-pool PCR.	This study
oVIN-1251	GAGCGTGGGTCTCCGCCCAG	Amplification of the *DMPK* locus for small pool PCR.	[70]
oVIN-1252	CACTTTGCGAACCAACGATA	Amplification of the *DMPK* locus for small pool PCR.	[70]
oVIN-100	AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC	SP-PCR probe for the repeats	[71]
oVIN-1321	GCCTCCCTTACCATGCAGT	Amplifies human *HTT* locus for PacBio amplicon sequencing.	This study
oVIN-3490	ACAAGGGAAGACCCAAGTGA	Amplifies human *HTT* locus for PacBio amplicon sequencing.	This study
oVIN2612	AAGGGCATCGACTTCAAGGA	Amplification of GFP by qPCR	This study
oVIN2613	GGCGGATCTTGAAGTTCACC	Amplification of GFP by qPCR	This study
oVIN2622	CGATAAGAACCTGCCCAACG	Amplification of SpCas9 by qPCR	This study
oVIN2623	GGTTGGTCTTGAACAGCAGG	Amplification of SpCas9 by qPCR	This study
oVIN3321	GGAGACGGGGTCACCCACAC	Amplification of mouse β-actin for qPCR	This study
oVIN3322	AGCCTCAGGGCATCGGAACC	Amplification of mouse β-actin for qPCR	This study

Supplementary Table 4: Antibodies used in this study.

Antibody	Reference	Species	Clonality	Method and concentration
Anti-S100 beta antibody [EP1576Y] - Astrocyte Marker	ab52642	Rabbit	Monoclonal	Immunofluorescence (1:100)
Anti-Beta III Tubulin, Cy3 Conjugate Antibody	AB15708C3	Rabbit	Polyclonal	Immunofluorescence (1:1000)
Anti-CRISPR/Cas9 monoclonal antibody 4G10	C15200216-100	Mouse	Monoclonal	Western blotting (1:1000)
Anti-β-Actin	A5441	Mouse	Monoclonal	Western blotting (1:2000)
Anti-Green Fluorescent Protein	MAB3580	Mouse	Monoclonal	Western blotting (1:1000)
Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 488	A11008	Goat	Polyclonal	Immunofluorescence (1:1000)
Goat anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 680	A21058	Goat	Polyclonal	Western blotting (1:1000)
Anti-β-Tubulin III antibody clone SDL.3D10	T8660	Mouse	Monoclonal	FACS (1:500)
Anti-human CD44 Antibody, FITC Conjugate Antibody	130-113-896	Mouse	Monoclonal	FACS (1:500)
Goat anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 568	A-11031	Goat	Polyclonal	FACS (1:1000)