ARTICLES
https://doi.org/10.1038/s41588-019-0575-8
nature genetics

# A slipped-CAG DNA-binding small molecule induces trinucleotide-repeat contractions in vivo

Masayuki Nakamori[1,12], Gagan B. Panigrahi[2,12], Stella Lanni[2,12], Terence Gall-Duncan[2,3], Hideki Hayakawa[1], Hana Tanaka[1], Jennifer Luo[2,3], Takahiro Otabe[4], Jinxing Li[4], Akihiro Sakata[4], Marie-Christine Caron[5,6], Niraj Joshi[5,6], Tanya Prasolava[2], Karen Chiang[2,3], Jean-Yves Masson[5,6], Marc S. Wold[7], Xiaoxiao Wang[8], Marietta Y. W. T. Lee[8], John Huddleston[9,10], Katherine M. Munson[9], Scott Davidson[2], Mehdi Layeghifard[2], Lisa-Monique Edward[2], Richard Gallon[11], Mauro Santibanez-Koref[11], Asako Murata[4], Masanori P. Takahashi[1], Evan E. Eichler[9,10], Adam Shlien[2], Kazuhiko Nakatani[4], Hideki Mochizuki[1] and Christopher E. Pearson[2,3*]

**In many repeat diseases, such as Huntington's disease (HD), ongoing repeat expansions in affected tissues contribute to disease onset, progression and severity. Inducing contractions of expanded repeats by exogenous agents is not yet possible. Traditional approaches would target proteins driving repeat mutations. Here we report a compound, naphthyridine-azaquinolone (NA), that specifically binds slipped-CAG DNA intermediates of expansion mutations, a previously unsuspected target. NA efficiently induces repeat contractions in HD patient cells as well as en masse contractions in medium spiny neurons of HD mouse striatum. Contractions are specific for the expanded allele, independently of DNA replication, require transcription across the coding CTG strand and arise by blocking repair of CAG slip-outs. NA-induced contractions depend on active expansions driven by MutSβ. NA injections in HD mouse striatum reduce mutant HTT protein aggregates, a biomarker of HD pathogenesis and severity. Repeat-structure-specific DNA ligands are a novel avenue to contract expanded repeats.**

Expansion of unstable trinucleotide repeats causes over 40 neurodegenerative diseases, including HD, and 17 of these are caused by gene-specific CAG/CTG trinucleotide-repeat expansions[1–5]. Ongoing repeat expansions in affected tissues correlate with disease age-of-onset, progression and severity[6,7], and dramatic repeat length variations exist between tissues of the same individual[1]. For at least six CAG diseases (HD, spinocerebellar ataxia type 1 (SCA1), SCA2, SCA3, SCA7 and SCA17), DNA repair proteins are major modifiers of age-of-onset[8–10] and disease progression[11], suggesting that ongoing somatic expansions contribute to age-of-onset and progression[2]. Therefore, arresting or reversing somatic CAG/CTG-repeat expansions may arrest or reverse disease progression[2]. Inheritance of larger HD CAG expansions also leads to earlier age-of-onset and accelerated disease progression[12,13]. Thus, inducing contractions of the expanded repeat to lengths shorter than the inherited tract length should have beneficial effects.

While the exact mechanism(s) of repeat expansions are unclear, all models involve the formation and aberrant repair of unusual slipped-DNA structures formed by the repeats[4,14–23]. Slipped-DNAs form by out-of-register annealing of complementary repeat strands during DNA repair, replication or transcription, and have slip-outs with intra-strand hairpins of CAG or CTG repeats[17,22]. Slipped-DNAs were detected at the mutant myotonic dystrophy (DM1) locus in patient tissues, including the central nervous system, and were not present at the non-expanded allele[14]. The levels of slipped-DNAs were greatest in tissues showing the largest somatic expansions, supporting the involvement of slipped-DNAs in the expansion processes.

A compound with specificity for slipped-DNAs may alter their formation and/or processing—attributes that could modulate repeat instability. We previously designed a CAG-specific DNA-binding compound, NA[24–26], which binds to a distorted intra-strand CAG hairpin, with the naphthyridine and azaquinolone moieties forming hydrogen bonds to guanine and adenine, respectively, causing two cytosines to flip out from the CAG hairpin (Fig. 1a,b)[26]. NA bound with high affinity and increased the melting temperature of $(CAG)_{10}$ hairpins by >30 °C (ref. [25]), and preferentially bound to longer (>30) CAG hairpins[26]. The high selectivity of NA for longer CAG hairpins coupled with the presence of slipped-DNAs at a mutant repeat locus[14] suggested that it may specifically target the expanded allele.

## Results

**NA binds long CAG slip-out structures.** To determine whether NA can bind disease-relevant (CAG)•(CTG) duplexes with and

[1]Department of Neurology, Osaka University Graduate School of Medicine, Osaka, Japan. [2]Program of Genetics & Genome Biology, The Hospital for Sick Children, The Peter Gilgan Centre for Research and Learning, Toronto, Ontario, Canada. [3]Program of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. [4]Department of Regulatory Bioorganic Chemistry, The Institute of Scientific and Industrial Research, Osaka University, Osaka, Japan. [5]Genome Stability Laboratory, CHU de Québec Research Center, HDQ Pavilion, Oncology Division, Quebec, Quebec, Canada. [6]Department of Molecular Biology, Medical Biochemistry and Pathology, Laval University Cancer Research Center, Quebec, Quebec, Canada. [7]Department of Biochemistry, Carver College of Medicine, University of Iowa, Iowa City, IA, USA. [8]Department of Biochemistry and Molecular Biology, New York Medical College, Valhalla, NY, USA. [9]Department of Genome Sciences, University of Washington, Seattle, WA, USA. [10]Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA. [11]Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK. [12]These authors contributed equally: Masayuki Nakamori, Gagan B. Panigrahi, Stella Lanni. *e-mail: cepearson.sickkids@gmail.com

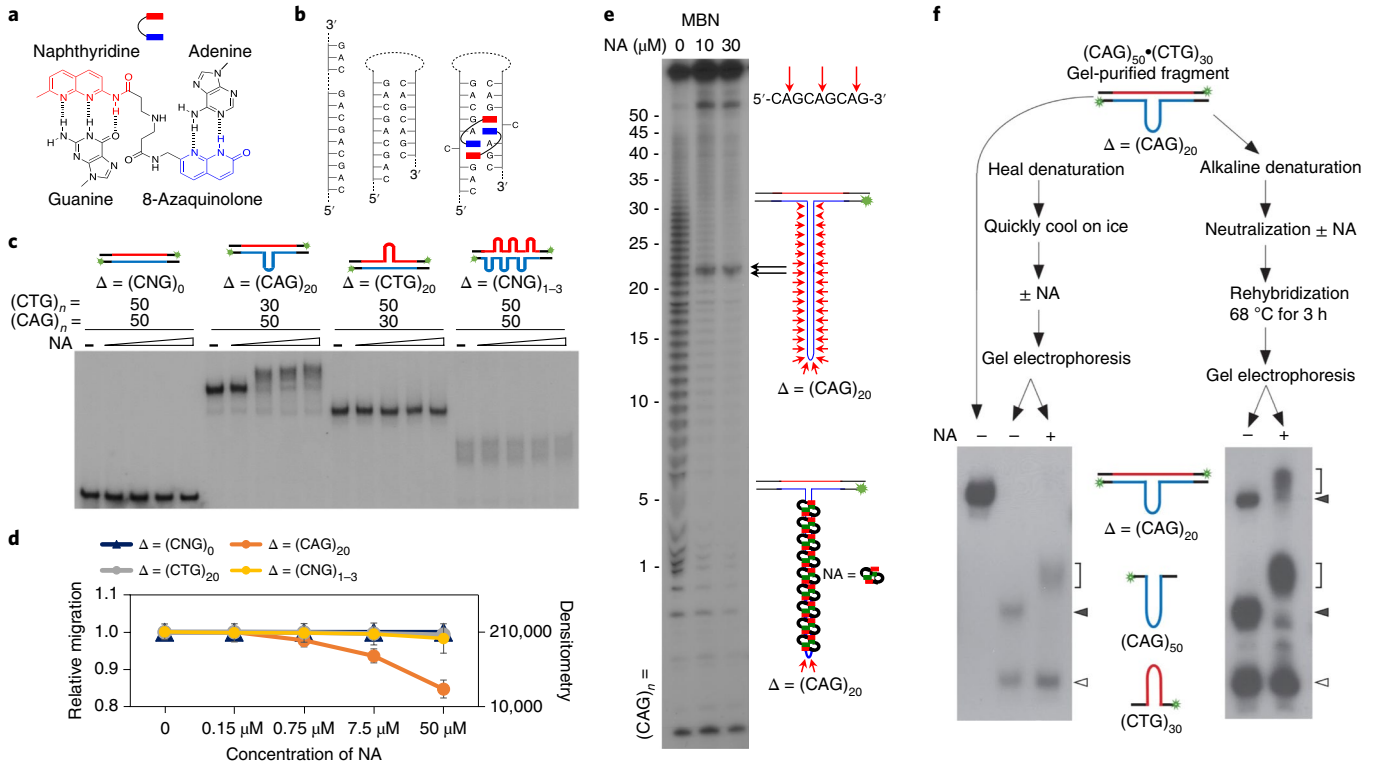



**Fig. 1 | NA binds to long CAG slip-outs. a**, The structure of NA comprising two heterocycles, a naphthyridine (red) and an 8-azaquinolone (blue) moiety[26]. **b**, A schematic of the NA–(CAG)•(CAG) triad complex revealed by NMR spectroscopy[26]. The $(CAG)_n$ DNA sequence (left) can fold into hairpins involving mismatched A–A pairs flanked by C–G and G–C pairs (middle). NA molecules intercalate into the DNA helix, with the 2-amino-1,8-naphthyridine moiety hydrogen-bonding to guanine (in red) and the 8-azaquinolone moiety hydrogen-bonding to adenine (in blue), forcing the flipped-out cytosine bases. **c**, Binding of NA to gel-purified DNA fragments with (CAG)•(CAG) repeats in both strands flanked by 59 base pairs (bp) and 54 bp of non-repetitive DNA labeled with $^{32}P$ on both strands (green star). **d**, Quantification of NA binding. The relative migration was measured as the ratio of the migration distance of each NA–DNA complex to the migration distance of free DNA. Densitometric analysis was performed for the $(CNG)_{1-3}$ DNA substrate. The graphs indicate the mean of three independent experiments ± the standard deviation (s.d.). **e**, Footprinting on $(CAG)_{50}$•$(CTG)_{30}$ uniquely $^{32}P$-labeled on the $(CAG)_{50}$ strand, cleaved throughout the repeat using MBN[17]. In the presence of NA, all scissile sites (red arrows) with the exception of the hairpin tip are protected, revealing binding specificity for the slip-out stem. Two independent experiments were performed with similar results. **f**, Gel-purified DNAs with a long $(CAG)_{20}$ slip-out from $(CAG)_{50}$•$(CTG)_{30}$, $^{32}P$-labeled on both strands, were treated as indicated and resolved on 4% polyacrylamide. NA–DNA complexes formed with both the $(CAG)_{50}$ strand (both panels) and the heteroduplexed $(CAG)_{50}$•$(CTG)_{30}$ (right panel) are shown by brackets; free DNA is indicated by arrowheads. NA did not bind the $(CTG)_{30}$ hairpin fragment in either experiment (white arrowheads), and did not inhibit re-hybridization of complementary strands. Two independent experiments were performed with similar results. Uncropped gels are available as source data.

without slip-outs, we performed band-shifts using slipped-DNAs[17] that mimic those at the mutant DM1 locus in patient tissues[14]. NA did not bind to fully duplexed DNAs containing $(CAG)_{50}$•$(CTG)_{50}$ repeats, and high concentrations of NA did not induce structural alterations of the DNA (Fig. 1c,d). NA bound molecules of $(CAG)_{50}$•$(CTG)_{50}$ with clustered short slip-outs[23] to a limited degree, but their short slip-outs and electrophoretic heterogeneity prohibited detection of a single shifted species (Fig. 1c,d). Slipped heteroduplexes with a single long slip-out of 20 CAG excess repeats in $(CAG)_{50}$•$(CTG)_{30}$ were extensively bound by NA (Fig. 1c,d). NA binding caused electrophoretic band-broadening, previously observed for other DNA-binding ligands[27]. The progressively slower migration of the slipped-DNA with added NA is consistent with NA binding to additional (CAG)•(CAG) pairs in the hairpin[26]. In contrast, DNAs with long slip-outs of CTG repeats, $(CAG)_{30}$•$(CTG)_{50}$, were not bound by NA (Fig. 1c,d). Thus, NA specifically bound CAG but not CTG slip-outs. We confirmed this with footprinting experiments on $(CAG)_{50}$•$(CTG)_{30}$ by using the single-strand DNA-specific mung bean nuclease (MBN), which cleaves only the slip-out regions between the AG of every CAG unit[17] (Fig. 1e). Following NA binding to the CAG slip-out, MBN digestion is inhibited, with the exception of the hairpin tip (Fig. 1e), not bound by NA. Denaturation of purified slipped $(CAG)_{50}$•$(CTG)_{30}$ to individual $(CAG)_{50}$ and $(CTG)_{30}$ single strands showed NA binding specifically to the single-stranded $(CAG)_{50}$ strand (Fig. 1f). Subsequent renaturation of the individual single strands revealed that NA did not block complementary strand hybridization, but bound specifically to the slipped-out CAG strand (Fig. 1f). The greater degree of electrophoretic shift of the isolated single-stranded $(CAG)_{50}$ tract compared to $(CAG)_{50}$•$(CTG)_{30}$ is caused by it being forced into a long hairpin of all 50 repeats that can bind more NA molecules. Note that the amount of NA bound to $(CAG)_{50}$•$(CTG)_{30}$ or $(CAG)_{50}$ following re-hybridization (Fig. 1f, 7.5 μM NA) is considerably greater than in the absence of re-hybridization (Fig. 1c, 50 μM NA), even though the amount of NA is more than fivefold higher in the latter. These results suggest that, in biological situations where the CAG repeat is denatured from its complementary CTG strand, the full length of the isolated CAG tract would become fully bound by NA in a hairpin-like conformation. Furthermore, NA binding was significantly greater for longer CAG slip-outs (Supplementary Fig. 1). We previously estimated a ratio of two NA molecules to one CAG–CAG, wherein a slip-out of $(CAG)_{20}$ would bind maximally 20 NA molecules[26]. The molar ratio of 2:1 NA to slipped-out repeats was evident for the longer slip-outs. The affinity of each

NA molecule/CAG–CAG was estimated as $1.8\times10^{6}\,M^{-1}$ (association constant, $K_a$) or $0.56\times10^{-6}$ M (dissociation constant, $K_d$)[26]. NA can be removed from the DNA using phenol/chloroform extraction (Extended Data Fig. 4c). Together, these findings support the structure-specificity of NA for long CAG-repeat slip-outs of disease-relevant tract lengths, and indicate that NA does not induce slip-out extrusion from an expanded fully duplexed molecule, but induces formation of longer slip-outs from a single-stranded CAG tract.

**NA inhibits repair of long CAG slip-outs.** We tested whether NA could block processing of slipped-DNAs, which could forecast its ability to modify repeat instability. NA specifically inhibited repair of slipped-DNAs with long $(CAG)_{20}$ slip-outs, but not $(CTG)_{20}$ slip-outs (Fig. 2a–c). This specificity is consistent with the inability of DNA polymerases to extend primers along NA-bound $(CAG)_{20}$ templates[24,25]. We assessed the effects of NA on repair of a single extra CAG-repeat slip-out, too small to be bound by NA. Repair of this substrate was unaffected by NA (Fig. 2d and Supplementary Fig. 1), consistent with its binding specificity for long slip-outs. NA did not bind to slip-outs of 3 excess repeats, but binding could be detected for 5, 11, 15 or 20 excess CAG repeats, with increased binding for slip-outs >10 (Supplementary Fig. 1, upper panel). The repair of DNA substrates with a similar range of slip-out sizes from 6, 10, 13 or 20 excess CAG repeats was inhibited by NA (Supplementary Fig. 1, lower panel), indicating that, once the DNA is bound by NA, repair is inhibited and not further altered by increased NA-bound lengths. These data suggest that there is a less-than-additive effect for both NA binding and NA's effect on slip-out repair. NA had no effect on the repair of a G–T mismatch, the most frequent base–base mismatch, which depends on mismatch repair (MMR) proteins[23,28] (Fig. 2e). Thus, NA is unlikely to cause genome-wide mutations that occur in the absence of MMR.

**NA modulates instability of expanded mutant repeats in HD cells, with limited off-target effects.** At 50 µM, NA is cell-permeable and enters nuclei without causing acute or long-term cytotoxicity, slowing proliferation, or altering DNA replication or transcription across CTG tracts (Fig. 3a–c and Extended Data Fig. 1). Tract lengths of >200 repeats are frequent in brain cells that have experienced somatic expansions in HD individuals that inherited $(CAG)_{40-50}$ (refs. [1,7,29]). NA induced a significant shift in repeat population towards contraction of a $(CAG)_{180}\bullet(CTG)_{180}$ tract at the huntingtin (*HTT*) locus in HD patient-derived primary fibroblasts[5]. NA enhanced the number of contractions of the expanded HD repeat (Fig. 3d; $P=7.25\times10^{-6}$, Supplementary Table 1), causing significant repeat loss (Fig. 3f; $P=0.0003$). NA also induced a significant reduction in the number of expansions of the HD repeat (Fig. 3d; $P=4.34\times10^{-5}$, Supplementary Table 1). We also treated HD patient fibroblasts with $(CAG)_{43}$, a mutation length common in the majority of patients with HD. A significant number of NA-treated cells showed contractions of the expanded repeat, with contractions to as low as 20 repeats, below the HD disease threshold of 35 units (Fig. 3e,g; $P=3.28\times10^{-5}$, Supplementary Table 1). NA also reduced the number of expansions (Fig. 3e; $P=8.32\times10^{-5}$, Supplementary Table 1). Thus, NA can induce contractions of expanded tract lengths common in inherited and somatically expanded HD alleles. In contrast, NA does not affect either the non-expanded HD repeat or other microsatellite repeats known to be prone to instability under stressed conditions (Fig. 3h and Extended Data Fig. 2a), suggesting that NA is specific for structures formed by expanded CAG/CTG repeats and will not deleteriously affect other repeats.

For additional controls, we first assessed the effect of NA at the very long but genetically stable (that is, not forming slipped-DNA structures) CAG tract of the *TBP* gene in HD patient cells. Stretches of >49 CAG repeats at the *TBP* locus cause fully penetrant SCA17 disease[30]. Our HD cell lines have 43 unstable repeats at the *HTT* gene and a similar length of 38 and 34 stable repeats of the non-mutant *TBP* gene. NA did not change the length distribution of *TBP* repeats[31,32] (Extended Data Fig. 2d and Supplementary Table 1), further supporting its specificity for actively unstable tracts that form slipped-DNAs.

To assess possible mutagenic activity of NA, we assessed its effect on a cancer diagnostic MSI panel of >20 slippage-prone mononucleotide repeats, using a sequencing-based assay capable of detecting low-level microsatellite instability (MSI) in non-neoplastic tissues[33]. There was no evidence of increased MSI from NA treatment in either HD cells with $(CAG)_{43}$ or $(CAG)_{180}$ (Extended Data Fig. 2f). As a control for non-repetitive regions, we assessed the *HPRT1* gene (exons 2 to 3), which is an established surrogate for evaluating mutagen-induced mutation spectra[34–37]. Since *HPRT1* is on the X chromosome, each of multiple single long-read sequences of *HPRT1* can be used as a proxy for a single cell in male HD cells (GM09197, $(CAG)_{180/21}$), permitting assessment of the effect of NA on individual cells without overlapping alleles. We sequenced up to 2,402 individual *HPRT1* alleles in 3 independent replicates and found no sequence differences between NA- and mock-treated cells (Extended Data Fig. 3a–c). We also assessed whether NA could induce an unsuspected mutation signature in HD cells ($(CAG)_{43}$ or $(CAG)_{180}$) using whole-genome sequencing. Environmental and anticancer agents can lead to specific mutation signatures[38–40]. We screened the NA-treated and untreated HD cells for mutational signatures shown in COSMIC using an established pipeline (SigProfiler version 0.0.5.75 (ref. [41])). We were unable to detect a statistically significant mutation signature. The genome of NA-treated HD cells showed very low numbers of single- or double-base substitutions (SBS or DBS) or insertions/deletions (indels), and no rearrangements (Fig. 3i). The numbers of SBS/indels/DBS/rearrangements for the $(CAG)_{43}$ and $(CAG)_{180}$ cells were 172/29/7/0 and 187/42/5/0, respectively. Similar background mutations attributed to DNA damage incurred during cell culture (~245 SBS, ~35 indels, ~6 DBS and ~0 rearrangements per genome) have been detected[42–44]. Moreover, cells treated for 24 h with the food additive potassium bromate (875 µM) or the anticancer drug cisplatin (12.5 µM) showed ~576/~31 and ~926/~65 SBS/indels, respectively[42], both substantially 'noisier' than our experiment treating HD cells with NA chronically for 40 d (50 µM). The 'silence' of NA off-target mutagenesis is contrasted by the 'noisy' mutational burden of the hyper- and ultrahyper-mutated genomes of brain cancers of MMR-deficient individuals assessed by the same pipeline[45–49]—a portent of non-specific targeting of the heterodimer MSH2–MSH3, MutSβ (Fig. 3i). Together, these results argue against NA as a general mutagen.

**Effect of NA depends on transcription but not replication through the repeat tract.** NA could affect repeat instability during replication or transcription. NA treatment induced a significant shift in repeat population towards contraction of a $(CAG)_{850}\bullet(CTG)_{850}$ tract in human cells, expressing r$(CAG)_{850}$ (ref. [21]) (Fig. 4a,b; $P=4.78\times10^{-5}$, Supplementary Table 1), and significantly shortened the average size of the repeat tract, with losses of up to 790 repeats (Fig. 4b; $P=6.44\times10^{-3}$). NA's effects were independent of cell proliferation and DNA replication, as near-complete arrest of proliferation using palbociclib, contact inhibition or serum starvation had no effect on NA-induced CAG contractions (Fig. 4d,e and not shown). This is consistent with NA's lack of effect on cell proliferation (Fig. 3c) or replication fork progression (Extended Data Fig. 1). To address the requirement for transcription, we used an established cell model, HT1080 non-transcribing $(CAG)_{850}$, having a single $(CAG)_{850}$ transgene floxed by transcription-terminator elements, so that the expanded repeat is not transcribed in either direction[21,50]. We confirmed the absence of transcription (Fig. 4c) and treated cells with NA in transcribing or non-transcribing conditions. NA induced contractions of the $(CAG)_{850}$ tract only when

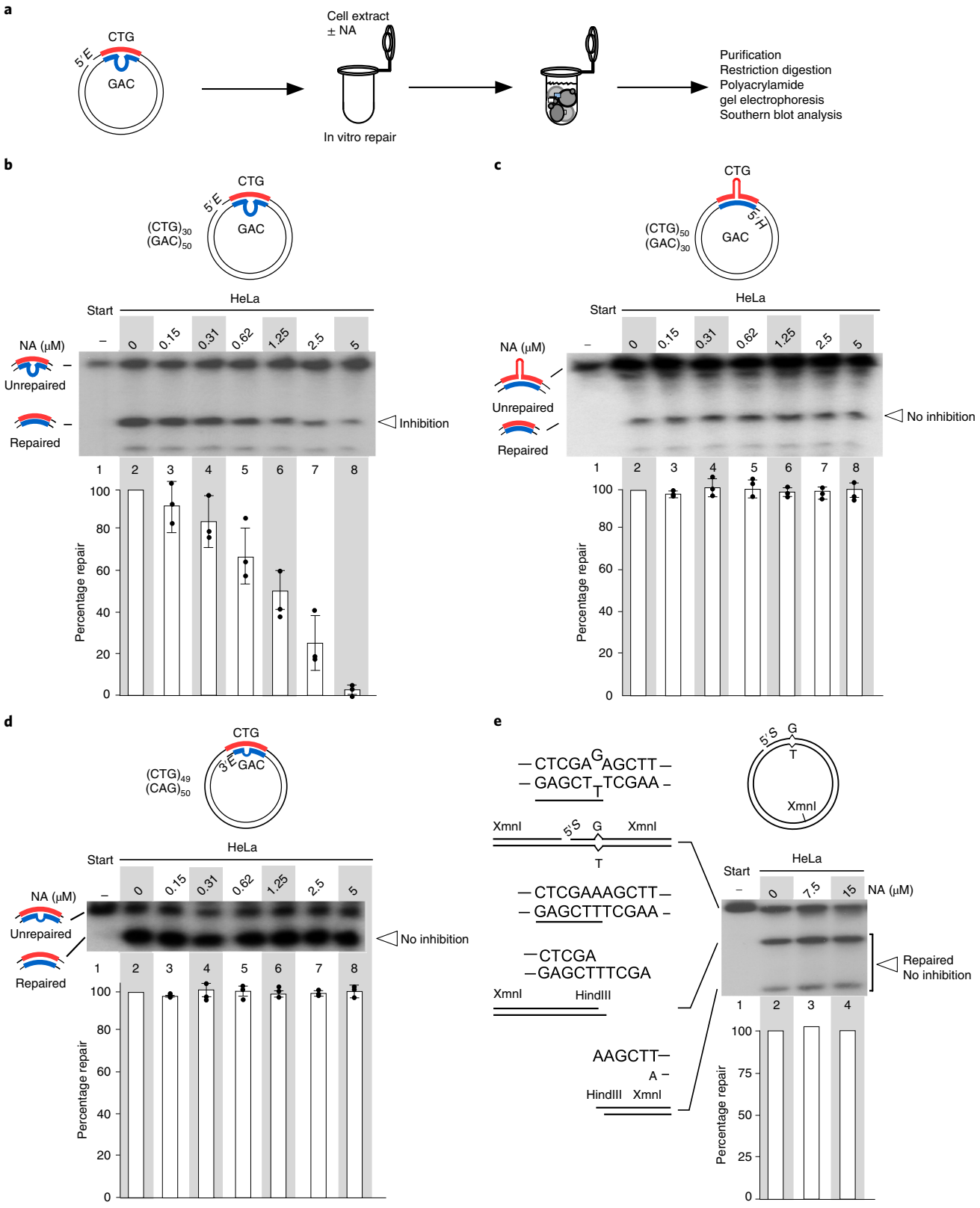



**Fig. 2 | NA specifically inhibits repair of long CAG slip-outs by human (HeLa) cell extracts. a**, A schematic of the in vitro repair assay. **b**–**e**, Starting DNAs and repair products (schematics) have the repeat-containing fragment released, resolved on PAGE and assessed on a molar level by Southern blotting and densitometry. The repair of DNA substrates containing a long $(CAG)_{20}$ slip-out (**b**), a long $(CTG)_{20}$ slip-out (**c**), a single CAG slip-out (**d**) or a G–T base–base mismatch (**e**) in the absence or presence of NA. Slipped-DNAs were hybrids of $(CAG)_{50}$•$(CTG)_{30}$, $(CAG)_{30}$•$(CTG)_{50}$ or $(CAG)_{50}$•$(CTG)_{49}$ (Methods)[23,57]. The repair of the G-T mismatch reconstitutes a HindIII restriction site. The graphs show percentage repair efficiencies to repaired product relative to all repeat-containing fragments in the lane; the values are normalized to the NA-free efficiency. The values represent the mean of three independent experiments ± s.d. Uncropped gels are available as source data. *E*, *H* and *S* represent EcoRI, HindIII and SapI, respectively.

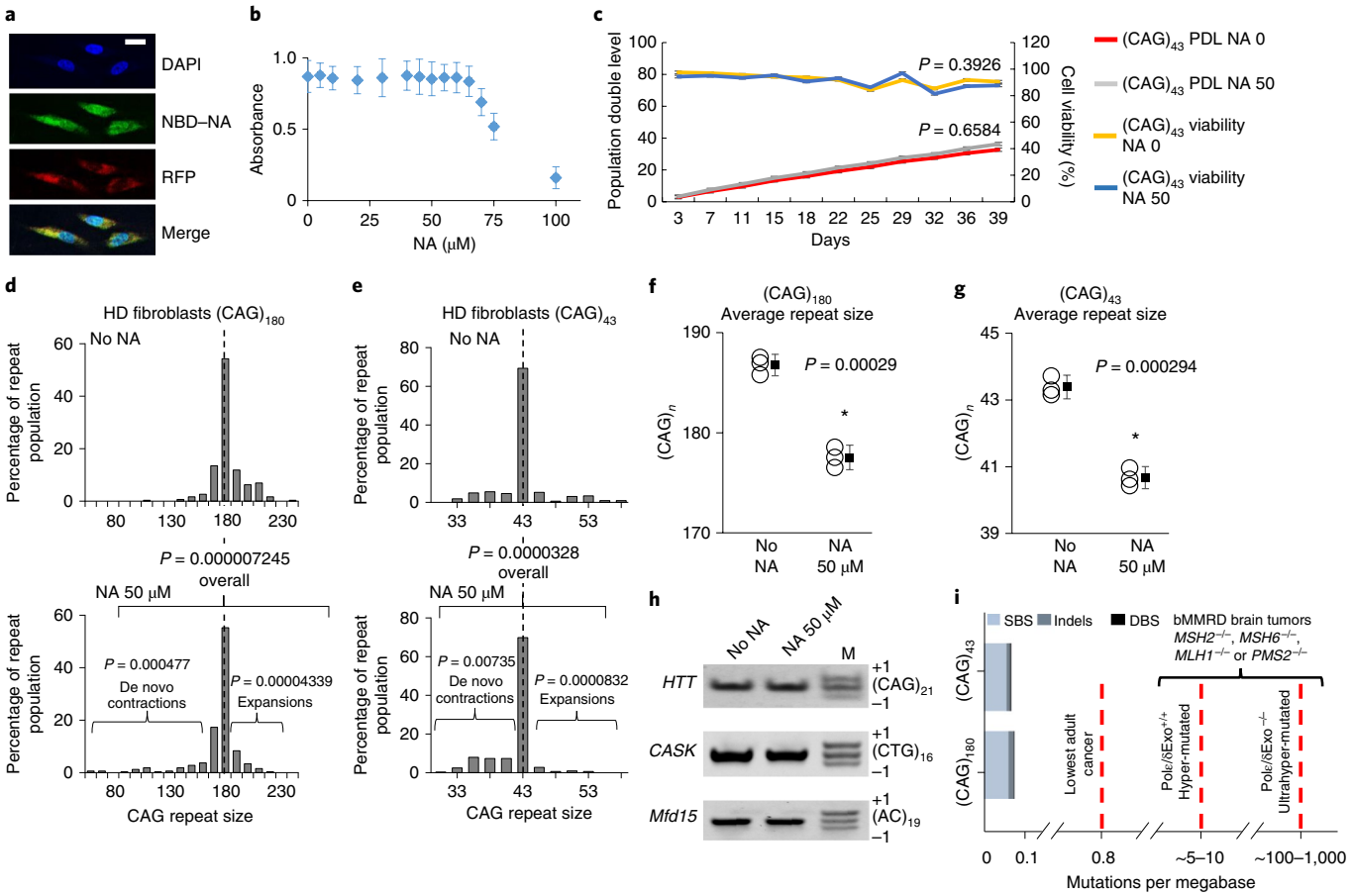



**Fig. 3 | NA cellular distribution, non-toxicity and effects on repeat instability in HD patient cells. a**, NBD-labeled NA (green) was distributed throughout nuclei and cytoplasm of human HD primary fibroblasts with $(CAG)_{43}$ (GM02191). Nuclei and cell membranes were stained with DAPI (blue) and CellLight Plasma Membrane-RFP (red), respectively. Scale bar, 20 μm. Three independent experiments were performed. **b**, Cell toxicity of $(CAG)_{43}$ cells treated with NA for 72 h. The viability was estimated using WST-1 assays. The values represent the means of three independent measures ± s.d. **c**, The population doubling levels (PDL) and percentage of cell viability of $(CAG)_{43}$ cells treated with or without 50 μM NA. The error bars indicate the s.d. of three independent experiments. An unpaired two-tailed *t*-test was used to calculate the *P* values. **d**,**e**, The repeat instability was analyzed by spPCR across the HD repeat tract (Methods, see Supplementary Table 1). The histograms show repeat length distributions in human HD primary fibroblasts with $(CAG)_{180}$ (ref. [5]) (GM09197) or with $(CAG)_{43}$ (GM02191), after 40 d of growth ± NA. The frequency distribution of repeat alleles is indicated as gray bars. The dashed line indicates the peak CAG size. The allele lengths are grouped in bins spanning 10 repeats; >230 alleles were sized per group. The percentage of repeat population was calculated by dividing the number of alleles grouped in bins spanning ten repeats by the number of total alleles. Shown is a summary of three independent experiments. The *P* values were calculated using a $\chi^2$ test comparing the frequencies of expanded, unchanged and contracted alleles in each set of experiments (Supplementary Table 1). **f**,**g**, The average (mean) repeat size in HD fibroblasts after a 40-d incubation with or without NA. *$P < 0.001$, two-sided Student's *t*-test. Error bars: 99% confidence limits. **h**, The repeat lengths of the *HTT* (normal allele), *CASK* and *Mfd15* loci in HD fibroblast cells (after a 40-d incubation ± NA). Three independent experiments were performed. **i**, The rate of SBS, indels and DBS in NA-treated HD cells with $(CAG)_{43}$ or $(CAG)_{180}$ compared to mock-treated cells. The numbers of SBS/indels/DBS/rearrangements detected in the NA-treated $(CAG)_{43}$ or $(CAG)_{180}$ cells were 172/29/7/0 and 187/42/5/0, respectively. The dashed lines indicate the 0.8 mutations per megabase level, which is the lowest burden observed in a typically 'quiet' adult cancer (bone marrow myelodysplastic syndrome[46]), and the 5–10 and 100–1,000 mutations per megabase levels in biallelic MMR-deficient (bMMRD) cancers, with and without polExonuclease activity[45–49]. Uncropped gels are available as source data.

transcription was permissible (Fig. 4b,f). Since NA did not alter transcription across the expanded repeat, either in these cells or HD patient fibroblasts (Fig. 4g and Extended Data Fig. 4a), this shows that NA induces CAG contractions during transcription across the expanded repeat, consistent with transcription having a driving role in inducing CAG instability and NA-induced contractions[16,21,51,52]. The ability of NA to induce contractions required transcription across the CTG repeat, but did not involve arrest of transcription. Again, NA did not affect non-expanded CAG/CTG tracts or other microsatellite repeats in the HT1080 cell model (Extended Data Fig. 2b,c). Thus, NA depends on transcription across the expanded repeat to induce repeat contractions, and is effective independently of cell proliferation.

**NA does not block in vitro R-loop formation, R-loop stability or RNase digestion, but enhances formation of repeat contractions.** Transcription-induced R-loops can lead to CAG/CTG instability[18,21] and may cause the transcription dependency of NA on repeat instability. NA did not alter the transcript levels of the expanded repeat or the translation of the mutant (m)HTT protein in cells (Extended Data Fig. 4a,b), nor did it affect transcription in vitro (Supplementary Fig. 2a–d). NA (120 μM) did not affect R-loop formation, or RNaseA or RNaseH processing of pre-formed R-loops (Supplementary Fig. 2c,d). At the NA concentrations used on cells (50 μM), R-loops remained detectable, suggesting that NA does not affect R-loop biophysical stability (Supplementary Fig. 2d). However, NA altered the processing of CAG/CTG R-loops in

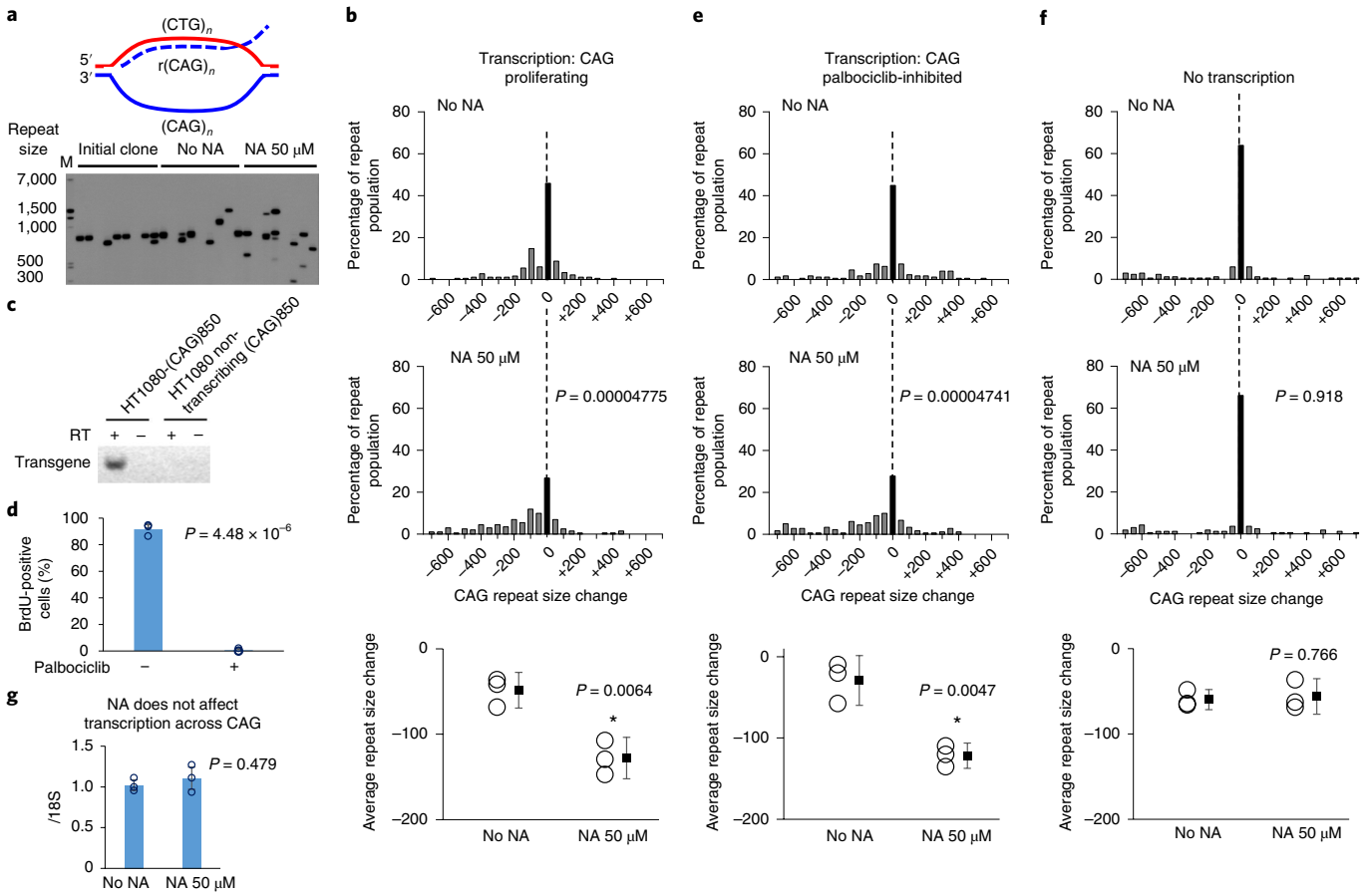



**Fig. 4 | NA induces CAG contractions independent of proliferation, dependent on rCAG transcription. a**, A schematic transcription bubble, and representative data showing spPCR repeat length analysis of HT1080-$(CAG)_{850}$ cells (initial cell clone and cells after a 30-d incubation ± NA) (Methods)[21,50]. The scale on the left shows molecular weight markers (M) converted into repeat number for CAG-repeat fragments of equivalent size. **b**, Histograms showing repeat length distributions, after a 30-d incubation ± NA, in proliferating HT1080-$(CAG)_{850}$ cells, in which transcription was permissible. The mean repeat size change is shown below. The frequency distribution of unstable and stable alleles is shown by gray bars and black bars, respectively. The dashed lines indicate the unchanged CAG size. The allele lengths are grouped in bins spanning 50 repeats. The percentage of repeat population was calculated by dividing the number of alleles grouped in bins by the number of total alleles; >50 alleles were sized for each group (Supplementary Table 1). The *P* values were calculated using a $\chi^2$ test. Error bars: 99% confidence limits. Shown is a summary of three independent experiments. **c**, PCR with reverse transcription (RT) analysis of the CAG repeat in HT1080 non-transcribing cells, showing that the transgene is integrated but not transcribed. Three independent experiments were performed. **d**, BrdU-positive cells after a 24-h incubation with BrdU in palbociclib-treated HT1080-$(CAG)_{850}$ cells. The *P* value was calculated using a *t*-test. The data are the mean ± s.d. of triplicates. The s.d. is smaller than the size of the data points. **e**, Histograms showing repeat length distributions after a 30-d incubation with or without NA, in non-proliferating (palbociclib-arrested) HT1080-$(CAG)_{850}$ cells. The mean repeat size change is shown below. The *P* values were calculated using a $\chi^2$ test (Supplementary Table 1). Shown is a summary of three independent experiments. **f**, Histograms showing repeat length distributions after a 30-d incubation with or without NA, in non-transcribing HT1080-$(CAG)_{850}$ cells. The mean repeat size change is shown below. The *P* values were calculated using a $\chi^2$ test (Supplementary Table 1). Error bars: 99% confidence limits. Shown is a summary of three independent experiments. **g**, The RNA transcript levels of the transgene (transcript in the CAG direction) in HT1080-$(CAG)_{850}$ cells treated with or without NA (50 μM)[50]. The data are the mean ± s.d. of triplicates; a *t*-test (two-sided) was used for independent biological triplicate experiments. Uncropped gels are available as source data.

extracts of neuron-like human SH-SY5Y cells, significantly increasing the number of contraction products with a milder effect on expansion products (Extended Data Fig. 5a–c, $P = 0.002$ versus $P = 0.04$). These results suggest that NA causes preferential R-loop processing to repeat contractions.

**NA alters activity of several DNA repair proteins on long CAG slip-outs.** NA could block the interaction of DNA repair proteins with slipped-DNAs, similarly to its ability to block MBN activity on $(CAG)_{20}$ slip-outs (Fig. 1e). As the proteins involved in large $(CAG)_{20}$ slip-out repair are unknown, we assessed four candidate proteins, MutSβ, replication protein A (RPA), polδ and polβ. In mice, the MMR MutSβ complex with a functioning ATPase drives CAG/CTG expansions[2,3,20,53,54]. MSH3 is a modifier of age-of-onset and disease progression in patients with HD, and a modifier of repeat instability in patients with HD and DM1 (refs. [55,56]). MutSβ is not required to repair long $(CAG)_{10–25}$ slip-outs[23,28,57,58], but may be involved in the formation of slipped-DNAs following resolution of transcriptionally induced R-loops[22,57,59]. This process, expected to involve MutSβ binding to DNA and ATP-mediated dissociation of the MutSβ–DNA complex, may be affected by NA. NA did not block the binding of MutSβ to long CAG slip-outs, nor did NA affect the ATP-mediated dissociation of this complex (Fig. 5a) that was previously reported[60]. Given also that NA does not block the formation of slip-outs from denatured repeat-containing DNAs (Fig. 1f), our data do not support a role of NA in blocking the formation of

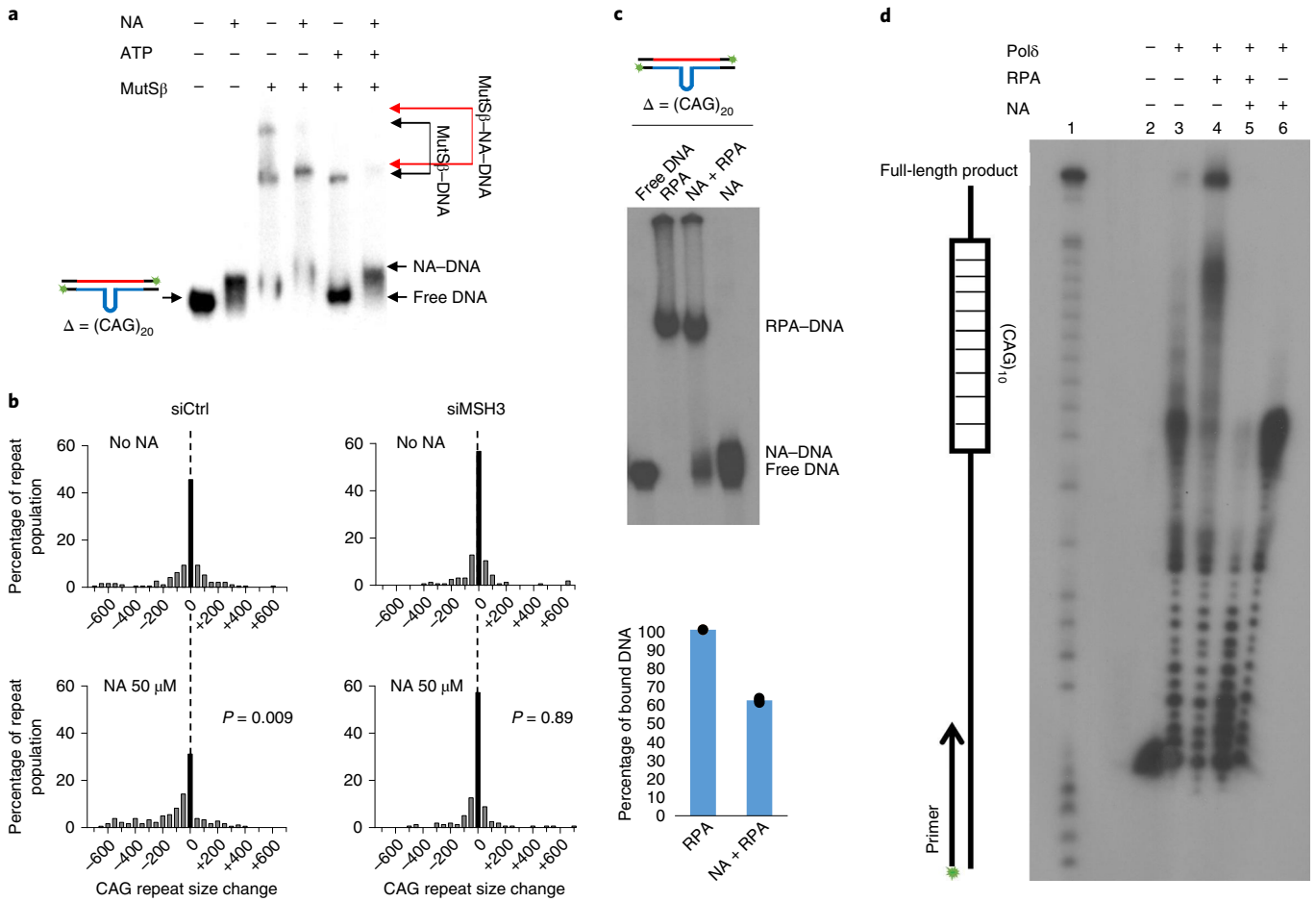



**Fig. 5 | NA affects interaction of DNA repair proteins on long CAG slip-outs. a**, MutSβ binding to radiolabeled slipped-DNA $(CAG)_{50}\bullet(CTG)_{30}$ containing a single $(CAG)_{20}$ slip-out at room temperature for 30 min. The addition of ATP (± $Mg^{2+}$) disrupts the binding of MutSβ to the DNA, as described previously[60]. The addition of NA has no effect on the binding or dissociation of MutSβ with this substrate. **b**, Histograms showing repeat length distributions in HT1080-$(CAG)_{850}$ cells treated with control short interfering RNA (siRNA; left) and siRNA against MSH3 (right), ± NA. The frequency distribution of unstable and stable alleles is shown by gray and black bars, respectively. The dashed lines indicate unchanged CAG size. The percentage of repeat population was calculated by dividing the number of alleles grouped in bins spanning 50 repeats by the number of total alleles; >50 alleles were sized for each group (Supplementary Table 1). The *P* values were calculated using a $\chi^2$ test. Shown is a summary of three independent experiments. **c**, RPA (250 nM) binding to slipped-DNA $(CAG)_{50}\bullet(CTG)_{30}$ containing a single $(CAG)_{20}$ slip-out (lane 2), $^{32}$P-labeled on both strands (schematic at the top). DNA was incubated with NA (50 μM) for 10 min at room temperature, before the addition of RPA. The percentage of RPA-bound DNA was quantified by densitometry. The histograms indicate the mean of three independent experiments ± s.d. The s.d. is smaller than the size of the data points. **d**, Polymerase extension assay performed as described previously[24]. The $(CAG)_{10}$ template oligonucleotide was annealed with a $^{32}$P-labeled primer and incubated ± NA (50 μM) for 30 min at room temperature. RPA (250 nM) and/or 20 nM Polδ was added and incubated for 15 min at 37 °C. The products were separated on a 6% sequencing gel together with Maxam–Gilbert sequencing reactions (lane 1). Lane 2 contains primer alone. Three independent experiments were performed. Uncropped gels are available as source data.

slipped-DNAs either with or without MutSβ. We further tested possible overlap of MutSβ and NA pathways. MSH3 knockdown suppressed CAG-repeat expansions in cells, confirming the essential role of MutSβ in active repeat instability[2,3,53,54,61]. MSH3 knockdown also blocked the effect of NA on repeat instability (Fig. 5b). Thus, NA's ability to induce CAG contractions depends upon MutSβ, as NA depends on active CAG instability.

RPA inhibits the formation of unusual DNA structures, such as hairpins, by binding and stabilizing single-stranded regions[62,63], and enhances DNA polymerase progression through structured DNA templates[64,65]. Polymerase delta (polδ) is implicated in CAG-repeat instability[66,67] and active in brains in a repair capacity[68]. Polδ was unable to synthesize across a CAG tract, and this was rescued by RPA (Fig. 5c). NA competitively blocked the interaction of RPA with slipped-CAG repeats (Fig. 5c) and blocked RPA-enhanced progression of polδ along the CAG template (Fig. 5d). NA was also able to block progression of polβ (data not shown). These results support a mechanism whereby NA induces CAG contractions by blocking RPA from binding slip-outs, which would otherwise facilitate polymerases to synthesize through CAG templates (Fig. 5c). NA may disrupt the interaction or activity of other repair proteins on CAG slip-outs.

**NA induces CAG contractions in the striatum of HD mice.** In HD mice and patients with HD, the largest somatic CAG expansions and most neurodegeneration occur in the striatum[1,2,15,69]. R6/2 mice harbor a single-copy transgene of *HTT* exon 1 (refs. [70,71]). R6/2 is one of the best characterized models of repeat instability, with instability patterns and tissue specificity comparable to HD knock-in models[1,29,72–75] and human HD patient brains[1,7,29,74–76]. CAG expansions are evident in the striatum of these mice as early as a few weeks, and continue as the mice age[15]. All HD CAG mouse models assessed show a similar pattern of ongoing spontaneous CAG expansions in striatum[75], which

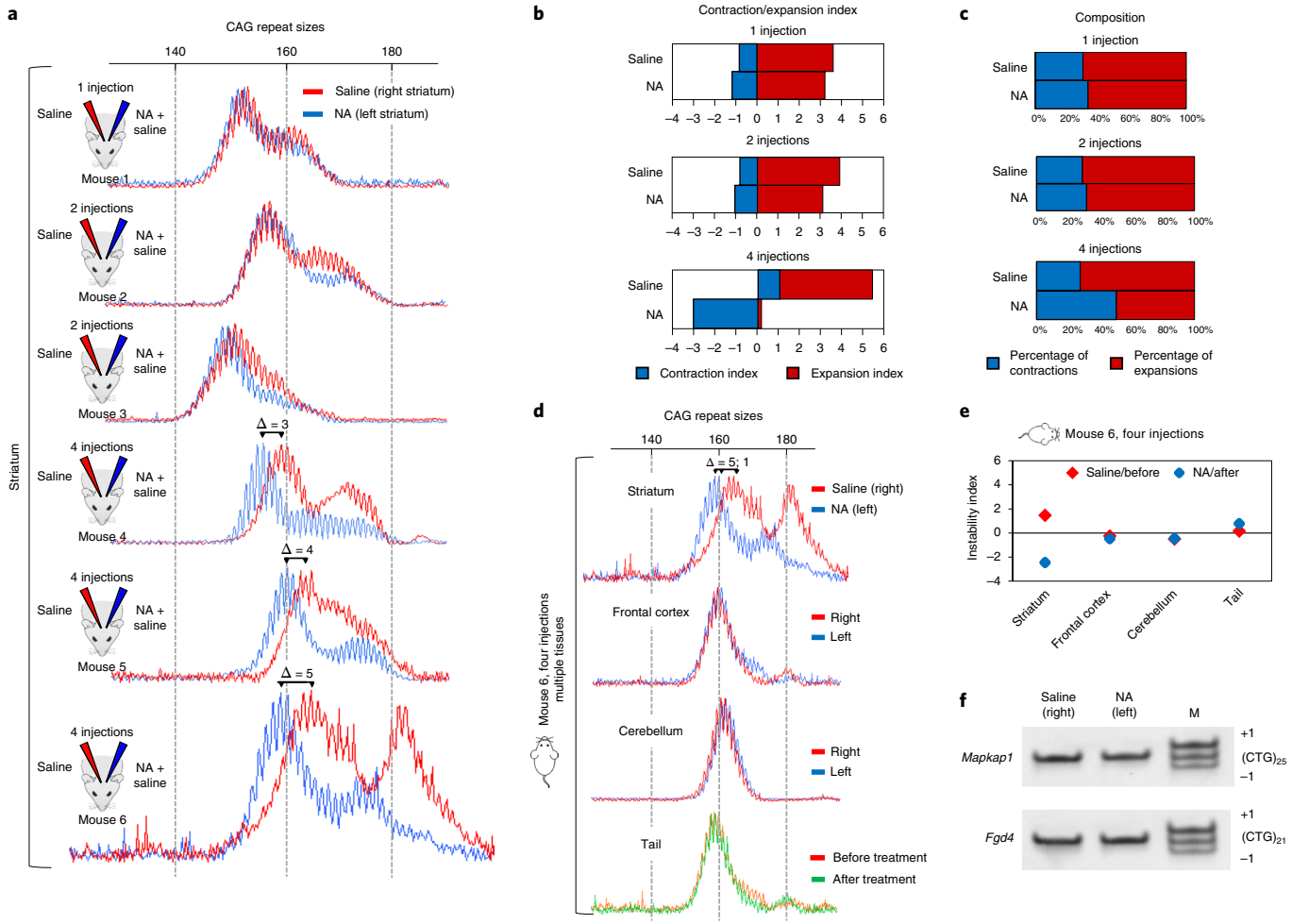



**Fig. 6 | NA induces CAG contractions in R6/2 mouse striatum. a**, GeneMapper traces showing the distribution of CAG-repeat lengths in striatum from six representative 10-week-old R6/2 mice that received one (mouse 1), two (mice 2 and 3) or four (mice 4–6) injections of NA, over a 4-week period. NA, dissolved in saline, was injected into the left striatum (blue) and saline alone was injected into the right striatum (red). All mice treated four times are shown in Supplementary Figs. 3–6. **b**, The effect of one, two or four NA injections reflected by contraction and expansion instability indices (Methods)[77,78]. **c**, The effect of one, two or four NA injections reflected by the relative composition of contractions and expansions (Methods)[77,78]. **d**, The CAG length distributions by GeneMapper traces in the striatum, frontal cortex, cerebellum and tail from one representative R6/2 mouse (mouse 6) following four injections of NA into the left striatum, over a 4-week period. DNAs were isolated from the left (NA, blue) and right (saline, red) sides of the striatum, frontal cortex and cerebellum, and from the tail before (red) and after (green) NA treatment. The repeat size change is in brackets, with the first number representing the NA-induced contractions of the major peak relative to the somatic expansions without NA, and the second number representing the contractions relative to the inherited (tail) allele. The brackets do not account for the size changes in the second mode of the bimodal distribution in the striatum. **e**, Instability indices in various tissues shown in **d**, where the red and blue diamonds represent values of the saline-treated/right and NA-treated/left sides of the striatum, respectively. **f**, The repeat tract lengths of the *Mapkap1* and *Fgd4* loci in both sides of the striatum from an R6/2 mouse with four injections. Three independent experiments were performed. Uncropped gels are available as source data.

was quantified in one model to show broadly distributed sizes of additional repeats gained at a rate of ~3.5 CAG units per month per cell[77,78].

Our dosing protocol involved one, two or four stereotactic injections into the striatum of 6-week-old R6/2 HD mice, each spanning a total of 4 weeks, when DNAs were collected (Extended Data Fig. 6). Since pups of the same litter inherit different CAG lengths, each inheriting around 150–160 CAG repeats, this complicates a direct comparison of the effect of NA between mice. Instead, we injected the left striatum with NA (in saline) and the right striatum with saline alone to serve as an internal control. NA intra-striatal injections led to contractions of the expanded repeat relative to the saline-only injections in the same mice (Fig. 6). In the absence of treatment, the CAG length distributions between the left and right halves of the striatum for a given mouse were identical within the limits of experimental resolution. Thus, NA specifically induces CAG length differences in mice (Fig. 6a).

Continued repeat contractions resulted from additional NA administrations over a 4-week period, with the most striking effect after the fourth administration. This was highly reproducible for a total of 13 mice (1 mouse for 1 injection, 2 mice for 2 injections, and 10 mice for 4 injections; Fig. 6a, Supplementary Figs. 3–6 and Extended Data Fig. 8). NA injection did not alter brain morphology or induce cell death or cell proliferation (Extended Data Fig. 9a–c). The effect of NA was similar between mice, regardless of inherited repeat length. A bimodal distribution of repeat sizes is present in HD patient brains[7,29,76]. The second peak of larger expansions is evident in the striatum of most CAG mouse models and was found to consist of the most vulnerable HD cells, medium spiny neurons (MSNs)[1,4,69], and to have greater levels of expanded CAG transcripts[75]. NA had a greater contracting effect on the larger expansions of the bimodal repeat distribution (Fig. 6d). This portion of larger-sized expansions in the striatum may have arisen by mutation

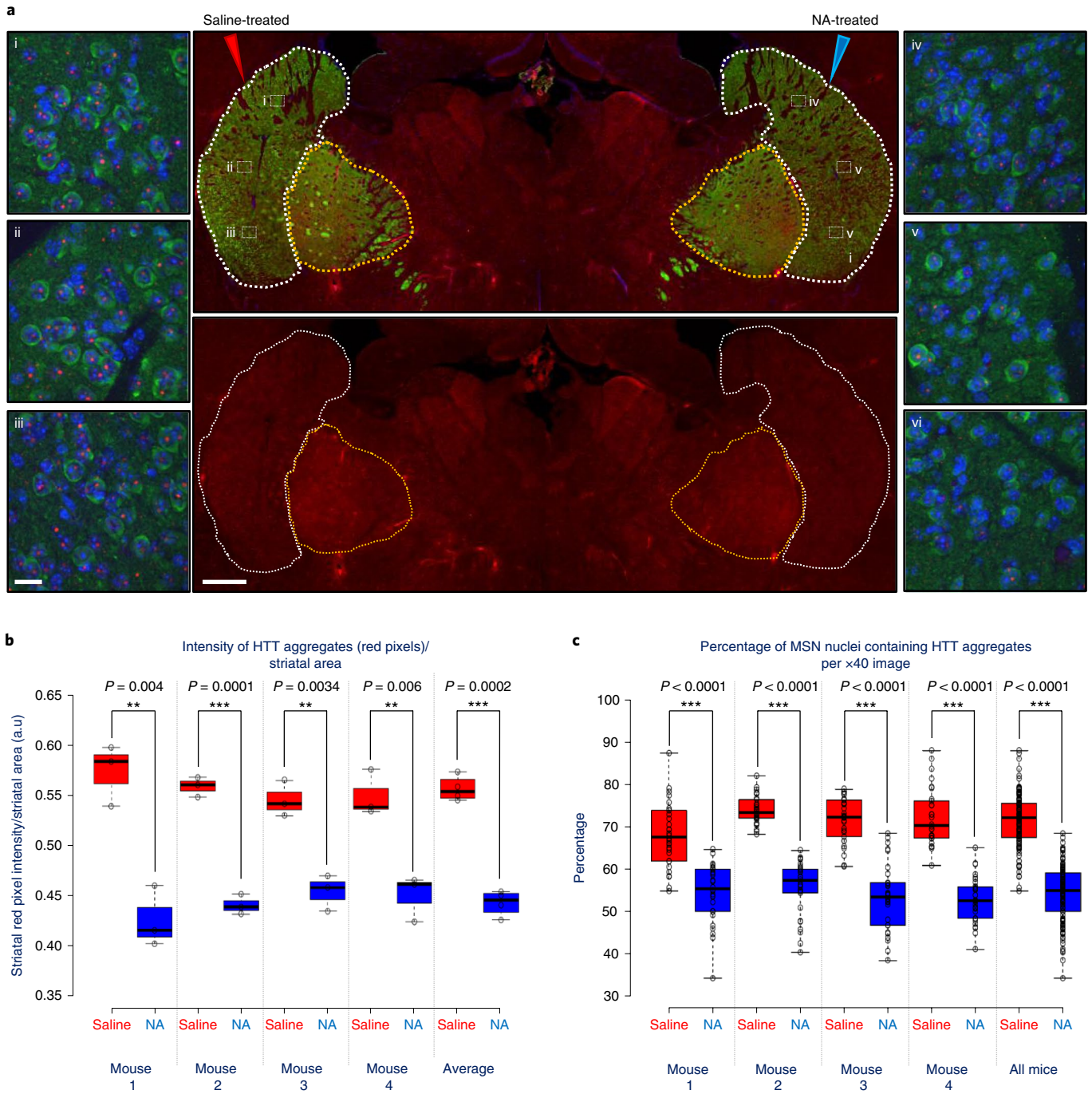



**Fig. 7 | NA induces a reduction in mHTT aggregates in R6/2 mice.** The effect of NA on mHTT aggregates in striatal MSNs of R6/2 mice treated with saline (right striata) and 50 µM NA (left striata), the slide indicates saline- and NA-treated halves with red and blue arrowheads, respectively. **a**, Top middle panel: a representative ×20 magnification epifluorescent image of a whole-brain slice stained for DARPP-32 (green, staining MSNs), mHTT aggregates (red) and counterstained with DAPI (blue). The numerals and associated dotted white squares mark the location of the corresponding representative ×40 magnification confocal images for saline-treated (i, ii, iii) and NA-treated (iv, v, vi) striata. The dotted lines demarcate borders of striatum (DARPP-32-positive cells) with the white dotted lines demarcating borders of striatal cell bodies (used for quantification) and the yellow dotted lines demarcating axons that form striatonigral bundles (not used for quantification). Bottom middle panel: the mHTT (red) channel alone. Scale bars, 19 µm (magnifications) and 500 µm (middle panels). **b**, Quantification by box plots of the red pixel intensity per striatal area in R6/2 mouse striatum treated with saline (right striata) and 50 µM NA (left striata) obtained via ImageJ quantification of ×20 mHTT/red channel epifluorescent images (4 mice total, 3 slides quantified per mouse, 30 images per mouse, 60–75 cells per image). An unpaired two-tailed *t*-test was used to compare the intensity of the red pixels in treated versus untreated striata. **c**, Quantification of the percentage of MSN nuclei containing mHTT aggregates in R6/2 mouse striatum treated with saline (right striata) and 50 µM NA (left striata) obtained via counting of mHTT-positive MSN nuclei in ×40 confocal images (4 mice total, 3 slides quantified per mouse, 10 images quantified per striata, ~50–100 cells per image). An unpaired two-tailed *t*-test was used to compare the percentage of MSN nuclei containing mHTT aggregates in treated versus untreated striata.

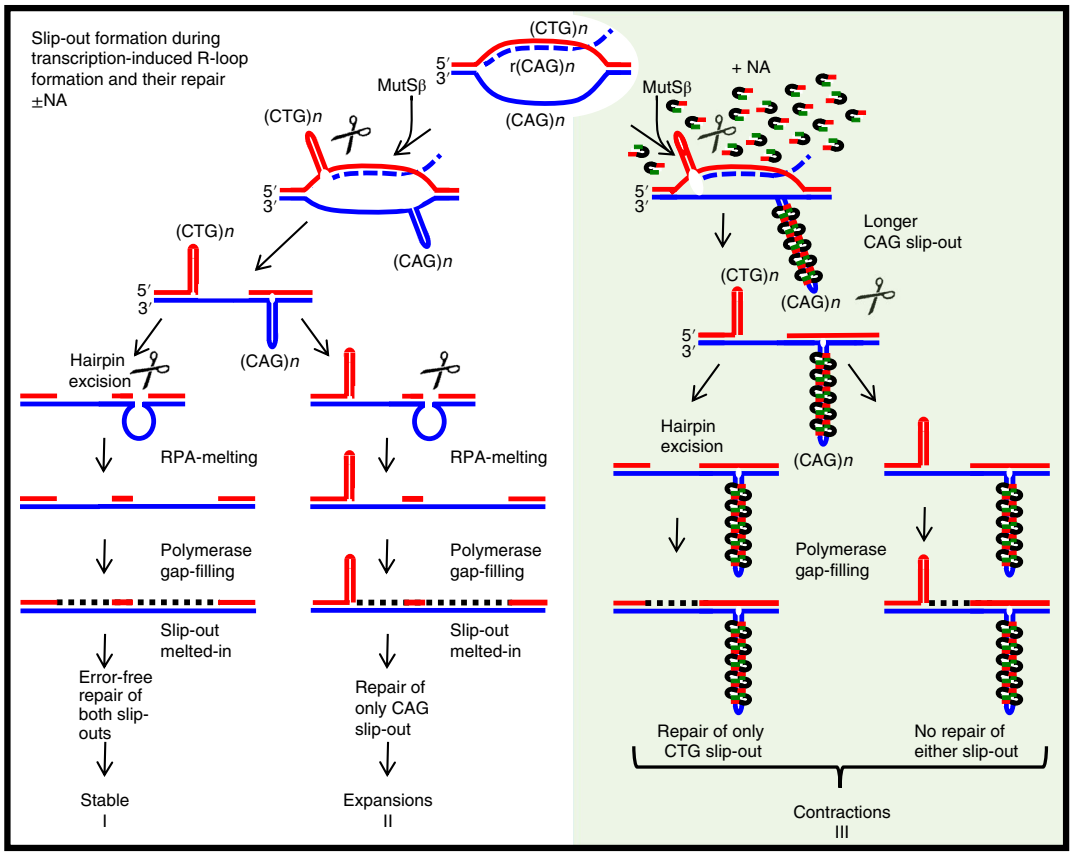



**Fig. 8 | Schematic of the plausible mechanisms through which NA may induce contractions of expanded CAG tracts.** A schematic of transcription-dependent slip-out formation and repair in the presence (green panel) or absence (white panel) of NA. See the text for details.

events involving large slip-outs, large enough to be bound by NA. Instability analyses in HD individuals suggest that the number of CAG units gained or lost at each mutation event are predominantly changes of one repeat unit, but may include changes of 5–15 repeat units, sizes that could be bound by NA[79–81]. Slipped-DNAs at the DM1 locus, isolated from various tissues of patients with DM1, including the brain, also presented a bimodal distribution of slip-out sizes, with peaks of ~30 and <10 repeats[14], where the former could be bound by NA.

Most alleles in the NA-treated striatum showed contractions of CAG repeats, indicating that NA affected most cells (Fig. 6a, Supplementary Figs. 3–6 and Extended Data Fig. 8). Notably, contractions were detected by standard PCR and did not require more sensitive small-pool (spPCR). Hence, they were robust, seeming to occur en masse in most cells. We quantified the somatic 'instability index'[77,78]: higher indices reflect greater expansions, whereas lower indices reflect lower expansions or greater contractions (Extended Data Fig. 7). The number of contracted versus expanded peaks (Fig. 6b) or the relative composition (%) of contracted versus expanded peaks (Fig. 6c) was greater with subsequent NA treatments. The repeat size distributions in striatum treated four times with NA were significantly different from those of mock saline-treated striatum (Mann–Whitney, $P = 0.00034$). The effect of NA was localized to the site of injection, as the CAG tract in cerebral cortex and cerebellum from the same mice showed identical patterns of CAG length heterogeneity between the right and left sides (Fig. 6d,e and Supplementary Figs. 3–6).

NA induced contractions rather than only arresting expansions because the peak repeat length in the NA-treated striatum was shorter than the inherited tract length in the tail of the same mouse, which remains constant throughout the mouse's lifetime (Fig. 6d,e and Supplementary Figs. 3–6). For a given mouse, the repeat size distributions in the NA-treated half of the striatum shifted towards contraction, relative to the inherited length in the tail, while the saline-treated half continued to incur expansions (Fig. 6b–e and Supplementary Figs. 3–6; see Δ of the main peaks where saline- versus NA-treated versus tail CAG lengths are reduced by 3–7 CAG units). This suggests that injections of NA into the striatum induced contractions of the expanded CAG tract, and thereby countered the expansion bias in that tissue (showing a spontaneous expansion rate of ~3.5 CAG units per month per cell, calculated as described previously[77,78]).

The effect of NA in vivo was specific for the expanded repeat, having no effect on non-expanded repeats (Fig. 6f and Extended Data Fig. 2). Furthermore, analysis of the mouse *Hprt1* sequence in treated male mice did not show significant sequence changes in NA-treated striatum (Extended Data Fig. 3d), indicating that NA is not a general mutagen in vivo, but specific for CAG tracts that are actively unstable. We also performed TUNEL analyses on HD mouse striatum to determine whether NA can induce DNA fragmentation and apoptosis[82,83]. Studies with R6/2 HD mice did not show obvious TUNEL signals[84–86]. Neither NA nor saline treatment (four administrations spanning 4 weeks) induced detectable TUNEL staining (Extended Data Fig. 9d).

To identify molecular phenotypes caused by NA-induced CAG contractions, we assessed the aggregation of mHTT protein in the MSNs of R6/2 mice. mHTT aggregates are a key biomarker of HD: the aggregation level and number of MSNs expressing aggregates are directly linked to CAG-repeat length and disease progression[87–92]. Furthermore, reductions in mHTT aggregate levels in the striatum of HD mice can increase cell viability, decrease neurodegeneration and improve behavioral phenotypes[93–98]. The rapid rate of mHTT

aggregate turnover (~39 h (refs. [93,96,98])) coupled with the considerable CAG contractions induced by NA over 4 weeks (Fig. 6a) may permit detection of NA-mediated alteration of mHTT aggregate levels. To test this, we assessed the intensity of fluorescent mHTT aggregates and the number of MSN nuclei expressing aggregates in NA-treated and saline-treated striatum of R6/2 mice. The intensity of mHTT aggregates in NA-treated striatum (four administrations), relative to that of the contralateral saline-treated striatum, was significantly reduced ($P=0.0002$) (Fig. 7a,b). Furthermore, NA-treated striata also showed significant reductions (10–20%) in the number of MSNs expressing mHTT aggregates ($P<0.0001$), with aggregates being smaller and less intense (Fig. 7a,c). Since NA does not affect transcription (Extended Data Figs. 4a and 9e) or translation of mHTT (Extended Data Fig. 4b), the reduced mHTT aggregation is most likely due to NA-induced CAG contractions and hence polyglutamine contractions arising from the contracted mRNA templates. This interpretation is consistent with the CAG/polyQ length sensitivity of mHTT aggregate formation[87–91,99–101]. Together, these results suggest that NA can disrupt the accumulation of mHTT into aggregates. Thus, NA can improve a pathological molecular phenotype by inhibiting the formation of a biomarker associated with neurodegeneration and disease progression.

## Discussion

Targeting DNA repeats to eliminate or correct the expansion in disease cells is of interest[2,50,102–108], with a substantial amount of mechanistic and human data validating somatic instability as a driver of disease age-of-onset, progression and severity[8,9,11,20,55,56,109]. Here we targeted slipped-DNA structures using a small molecule, NA, which induced contractions in vivo in patient cells and in the striatum of HD mice. Targeting this DNA intermediate permitted high specificity for the actively unstable mutant allele and limited off-target effects. There are many paths by which repeat instability can arise, and various ways through which NA may act, all of which involve slipped-DNAs[3,4,14]. Our data show that the effect of NA on CAG instability is independent of proliferation/replication, yet dependent on transcription through the repeat (Fig. 4d,e), consistent with the reported enhanced CAG instability following transcription[21]. Involvement of slipped-DNAs in NA's mechanism of action is supported by their existence in patient cells at disease loci[14], the ability to form slipped-DNAs during transcription across repeats and through R-loop processing[18,22], and NA's ability to block repair of slipped-DNAs (Fig. 2b).

We propose a mechanism by which NA enhances contractions of expanded repeats, while an expansion bias spontaneously arises in the absence of NA (Fig. 8). When *mHTT* is transcribed along the coding CTG DNA strand, the noncoding CAG DNA strand is displaced as single-stranded, which is extended and stabilized through the formation of an R-loop. Following removal of RNA from the R-loop, which can occur in the presence of NA (Supplementary Fig. 2), slipped-DNAs can arise on both strands[18,22]. Evidence suggests that MutSβ facilitates the formation of slipped-DNAs by binding to repeats[22,57,59,110], even in the presence of NA (Fig. 5a). To resolve these slipped-DNAs, transcription-coupled repair will be directed to the coding strand[111], which in *HTT* is the CTG strand, using the displaced noncoding CAG strand as a template for repair. Correct repair of both slip-outs would leave the tract length stable, with removal of the CTG slip-out and retention of the CAG hairpin. Repair of the CTG slip-outs in vitro is less efficient than repair of CAG slip-outs[23,28,57,58,112]; therefore, CTG slip-outs will often be retained, leading to an expansion bias. Repair would be facilitated by RPA-melting of slip-outs[63,66] and RPA-enhanced polymerase gap-filling across the repeat[62,63] (Fig. 5d). Thus, in the absence of NA, spontaneous mutations would lead to expansions. In contrast, NA binding will drive most of the transcriptionally displaced CAG repeats into long hairpins, essentially increasing the size and number of repeat units forming the NA-bound CAG slip-outs, which may be enhanced by MutSβ binding. The poor ability of RPA to melt the NA-bound CAG slip-outs would block the ability of polymerases to gap-fill[113] (Fig. 5c), and ultimately block the repair of NA-bound CAG slip-outs, leading to a contraction bias. Since NA bound more CAG repeats than were displaced from the CTG strand, this leads to fewer repeats incorporated in the repair product and hence contractions. That NA targets active repeat expansions is consistent with its increased efficacy in striatum over cultured fibroblasts, where the former displays higher levels of spontaneous expansions. That NA binds to and inhibits the repair of long, but not short, CAG slip-outs is consistent with NA preferentially affecting repeats in striatum, which undergo large saltatory expansions (Fig. 6).

NA efficiently induced contractions of expanded CAG repeats in most striatal cells, a tissue in HD individuals that shows selective neuronal vulnerability. Our findings over a 4-week period of NA treatment in mice with ~150 CAG repeats shows that neurons en masse incurred contractions of approximately ~0.5 repeats lost per week (on the basis of inherted tail lengths). Repeated administration led to continued contractions. Extrapolating this to an HD-affected human, we speculate that applying a drug such as NA before rapid onset of somatic CAG expansions could effectively block expansions and induce contractions of the inherited expanded allele to shorter lengths, where treatments spanning ~1 yr could be estimated to contract the repeat by 5–25 repeats. For an HD allele of 36–70 repeats, such changes could be clinically relevant.

mHTT-polyQ inclusions are biomarkers of HD, where inclusion size and numbers of neurons expressing inclusions are directly linked to CAG-repeat length and disease progression[87–92]. Reducing the levels of mHTT-polyQ aggregates in the striatum of HD mice can increase cell viability, decrease neurodegeneration and improve behavioral phenotypes[93–98]. Complete ablation of somatic CAG expansions (but no induced contractions) caused by genetic knockout of *Msh2*, *Msh3* or *Mlh1* delayed the detectable formation of mHTT-polyQ aggregates[114,115]. This effect was detectable after repair-deficient mice with $(CAG)_{110–117}$ had lived 5–10 months without any somatic CAG expansions. Our treatment of mice with NA for only 4 weeks was able to induce CAG contractions, and reduced both the size and frequency and intensity of mHTT aggregates in MSNs (Fig. 7), the cells in which NA is most effective in reducing CAG expansions (Fig. 6a).

Other approaches have been used to target expanded repeats in vivo. Targeting the HD locus has a history that predates the discovery of the *HTT* gene or the disease-causing mutation[116]. CRISPR–Cas9 targeting of the mutant *HTT* allele in HD cells deleted ~44 kilobases of DNA spanning the promoter region, transcription start site and CAG expansion, resulting in haploinsufficiency with a functional non-mutant allele[110]. A study using the CRISPR–Cas9 nickase that produced single-strand nicks within a transgenic repeat led to contractions of the repeat in an artificial selectable cell model[106]. Extension to an endogenous disease locus in cells or relevant tissues has yet to be demonstrated. CRISPR–Cas9 targeting only the expanded repeats at a disease locus proved not possible due to the absence of suitable protospacer-adjacent motif sequences and a likelihood of targeting other CAG/CTG repeats elsewhere in the genome. The inactivation of mutant *HTT* resulting in haploinsufficiency is possible[103–105]. Preferential targeting of fully duplexed expanded repeats by zinc-finger nucleases[117] or zinc-finger repressors[118] has been demonstrated. A small-molecule approach may overcome some of the in vivo hurdles of enzyme-mediated paths (delivery, editing efficiency, persistent nuclease/off-targets, immune response, haploinsufficiency and so on)[108,119–121]. Previous studies using cell models of CAG/CTG instability demonstrated that exogenously added compounds, including DNA-damaging agents, can modulate repeat instability[122]. However, these compounds lack specificity for expanded repeats, and would induce

mutations in non-mutant alleles and throughout the genome[123]. In a separate strategy, a $(CAG)_6$ antisense oligonucleotide was able to reduce CAG expansions, but did not induce contractions of an expanded repeat[50]. A mitochondria-directed compound, XJB-5-131, by unknown processes, mildly suppressed CAG expansions rather than inducing contractions[124]. Similarly, an inhibitor of histone deacetylase 3 suppressed somatic CAG expansions through unknown processes[107]. A sequence-specific polyamide directed to the fully paired duplex GAA/TTC repeat that is expanded in patients with Friedreich's ataxia prevented triple-stranded structure formation and suppressed GAA-repeat expansions in FRDA cells, but did not induce contractions[108]. The earliest non-enzymatic chemical approaches to target the HD locus aimed to modulate gene regulation using ligands targeted to fully paired duplex DNAs[116,125].

In contrast, NA, through rational design, targets the unusual DNA structures formed by CAG repeats that are actively unstable, and can induce en masse repeat contractions rather than only suppressing expansions. Moreover, NA is specific for the mutant HD allele in a disease-vulnerable brain region. These attributes of NA—its specificity for CAG-repeat slipped-DNA structures, its preferential effect on disease-causing expanded repeats and its ability to induce contractions in vivo in an affected brain region—make it a first-in-class example showing the potential of small-molecule DNA-binding compounds to impact somatic CAG-repeat instability, which can now be applied to any unstable repeat. Administration of such small molecules, once optimized for therapy, to human brains might effectively target the root cause of repeat expansion diseases and address downstream deleterious effects.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-019-0575-8.

Received: 20 October 2018; Accepted: 19 December 2019;
Published online: 14 February 2020

## References

1. 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).
2. Lopez Castel, A., Cleary, J. D. & Pearson, C. E. Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat. Rev. Mol. Cell Biol.* **11**, 165–170 (2010).
3. Mirkin, S. M. Expandable DNA repeats and human disease. *Nature* **447**, 932–940 (2007).
4. Pearson, C. E., Nichol Edamura, K. & Cleary, J. D. Repeat instability: mechanisms of dynamic mutations. *Nat. Rev. Genet.* **6**, 729–742 (2005).
5. Sathasivam, K., Amaechi, I., Mangiarini, L. & Bates, G. Identification of an HD patient with a (CAG)180 repeat expansion and the propagation of highly expanded CAG repeats in lambda phage. *Hum. Genet.* **99**, 692–695 (1997).
6. Morales, F. et al. Somatic instability of the expanded CTG triplet repeat in myotonic dystrophy type 1 is a heritable quantitative trait and modifier of disease severity. *Hum. Mol. Genet.* **21**, 3558–3567 (2012).
7. Swami, M. et al. Somatic expansion of the Huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset. *Hum. Mol. Genet.* **18**, 3039–3047 (2009).
8. Bettencourt, C. et al. DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. *Ann. Neurol.* **79**, 983–990 (2016).
9. Genetic Modifiers of Huntington's Disease Consortium Identification of genetic factors that modify clinical onset of Huntington's disease. *Cell* **162**, 516–526 (2015).
10. 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).
11. Hensman Moss, D. J. et al. Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study. *Lancet Neurol.* **16**, 701–711 (2017).
12. Gusella, J. F. & MacDonald, M. E. Molecular genetics: unmasking polyglutamine triggers in neurodegenerative disease. *Nat. Rev. Neurosci.* **1**, 109–115 (2000).
13. Rosenblatt, A. et al. Age, CAG repeat length, and clinical progression in Huntington's disease. *Mov. Disord.* **27**, 272–276 (2012).
14. Axford, M. M. et al. Detection of slipped-DNAs at the trinucleotide repeats of the myotonic dystrophy type I disease locus in patient tissues. *PLoS Genet.* **9**, e1003866 (2013).
15. Goula, A. V. et al. Stoichiometry of base excision repair proteins correlates with increased somatic CAG instability in striatum over cerebellum in Huntington's disease transgenic mice. *PLoS Genet.* **5**, e1000749 (2009).
16. Lin, Y., Dent, S. Y., Wilson, J. H., Wells, R. D. & Napierala, M. R loops stimulate genetic instability of CTG·CAG repeats. *Proc. Natl Acad. Sci. USA* **107**, 692–697 (2010).
17. Pearson, C. E. et al. Slipped-strand DNAs formed by long (CAG)·(CTG) repeats: slipped-out repeats and slip-out junctions. *Nucleic Acids Res.* **30**, 4534–4547 (2002).
18. Reddy, K. et al. Processing of double-R-loops in (CAG)·(CTG) and *C9orf72* (GGGGCC)·(GGCCCC) repeats causes instability. *Nucleic Acids Res.* **42**, 10473–10487 (2014).
19. Schmidt, M. H. & Pearson, C. E. Disease-associated repeat instability and mismatch repair. *DNA Repair* **38**, 117–126 (2016).
20. Tome, S. et al. *Msh3* polymorphisms and protein levels affect CAG repeat instability in Huntington's disease mice. *PLoS Genet.* **9**, e1003280 (2013).
21. Nakamori, M., Pearson, C. E. & Thornton, C. A. Bidirectional transcription stimulates expansion and contraction of expanded (CTG)·(CAG) repeats. *Hum. Mol. Genet.* **20**, 580–588 (2011).
22. Lin, Y. & Wilson, J. H. Nucleotide excision repair, mismatch repair, and R-loops modulate convergent transcription-induced cell death and repeat instability. *PLoS ONE* **7**, e46807 (2012).
23. Panigrahi, G. B., Slean, M. M., Simard, J. P., Gileadi, O. & Pearson, C. E. Isolated short CTG/CAG DNA slip-outs are repaired efficiently by hMutSβ, but clustered slip-outs are poorly repaired. *Proc. Natl Acad. Sci. USA* **107**, 12593–12598 (2010).
24. Hagihara, M. & Nakatani, K. Inhibition of DNA replication by a d(CAG) repeat binding ligand. *Nucleic Acids Symp. Ser.* **50**, 147–148 (2006).
25. Hagihara, M., He, H. & Nakatani, K. Small molecule modulates hairpin structures in CAG trinucleotide repeats. *ChemBioChem* **12**, 1686–1689 (2011).
26. Nakatani, K. et al. Small-molecule ligand induces nucleotide flipping in (CAG)*n* trinucleotide repeats. *Nat. Chem. Biol.* **1**, 39–43 (2005).
27. Nielsen, P. E., Zhen, W. P., Henriksen, U. & Buchardt, O. Sequence-influenced interactions of oligoacridines with DNA detected by retarded gel electrophoretic migrations. *Biochemistry* **27**, 67–73 (1988).
28. Pluciennik, A. et al. Extrahelical (CAG)/(CTG) triplet repeat elements support proliferating cell nuclear antigen loading and MutLα endonuclease activation. *Proc. Natl Acad. Sci. USA* **110**, 12277–12282 (2013).
29. Shelbourne, P. F. et al. Triplet repeat mutation length gains correlate with cell-type specific vulnerability in Huntington disease brain. *Hum. Mol. Genet.* **16**, 1133–1142 (2007).
30. Silveira, I. et al. Trinucleotide repeats in 202 families with ataxia: a small expanded (CAG)*n* allele at the SCA17 locus. *Arch. Neurol.* **59**, 623–629 (2002).
31. Sanchez-Contreras, M. & Cardozo-Pelaez, F. Age-related length variability of polymorphic CAG repeats. *DNA Repair* **49**, 26–32 (2017).
32. Gao, R. et al. Instability of expanded CAG/CAA repeats in spinocerebellar ataxia type 17. *Eur. J. Hum. Genet.* **16**, 215–222 (2008).
33. Gallon, R. et al. A sensitive and scalable microsatellite instability assay to diagnose constitutional mismatch repair deficiency by sequencing of peripheral blood leukocytes. *Hum. Mutat.* **40**, 649–655 (2019).
34. Keohavong, P., Xi, L. & Grant, S. G. Molecular analysis of mutations in the human *HPRT* gene. *Methods Mol. Biol.* **291**, 161–170 (2005).
35. Keohavong, P., Xi, L. & Grant, S. G. Molecular analysis of mutations in the human *HPRT* gene. *Methods Mol. Biol.* **1105**, 291–301 (2014).
36. Albertini, R. J. et al. Mutagenicity monitoring following battlefield exposures: longitudinal study of *HPRT* mutations in Gulf War I veterans exposed to depleted uranium. *Environ. Mol. Mutagen.* **56**, 581–593 (2015).
37. Nicklas, J. A. et al. Mutagenicity monitoring following battlefield exposures: molecular analysis of *HPRT* mutations in Gulf War I veterans exposed to depleted uranium. *Environ. Mol. Mutagen.* **56**, 594–608 (2015).
38. Poon, S. L., McPherson, J. R., Tan, P., Teh, B. T. & Rozen, S. G. Mutation signatures of carcinogen exposure: genome-wide detection and new opportunities for cancer prevention. *Genome Med.* **6**, 24 (2014).
39. Behjati, S. et al. Mutational signatures of ionizing radiation in second malignancies. *Nat. Commun.* **7**, 12605 (2016).

40. Phillips, D. H. Mutational spectra and mutational signatures: insights into cancer aetiology and mechanisms of DNA damage and repair. *DNA Repair* **71**, 6–11 (2018).
41. Alexandrov, L. B. et al. The repertoire of mutational signatures in human cancer. Preprint at *bioRxiv* https://doi.org/10.1101/322859 (2019).
42. Kucab, J. E. et al. A compendium of mutational signatures of environmental agents. *Cell* **177**, 821–836.e16 (2019).
43. Behjati, S. et al. Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature* **513**, 422–425 (2014).
44. Rouhani, F. J. et al. Mutational history of a human cell lineage from somatic to induced pluripotent stem cells. *PLoS Genet.* **12**, e1005932 (2016).
45. Shlien, A. et al. Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermutated cancers. *Nat. Genet.* **47**, 257–262 (2015).
46. Chalmers, Z. R. et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med.* **9**, 34 (2017).
47. Campbell, B. B. et al. Comprehensive analysis of hypermutation in human cancer. *Cell* **171**, 1042–1056.e10 (2017).
48. Hodel, K. P. et al. Explosive mutation accumulation triggered by heterozygous human Pol epsilon proofreading-deficiency is driven by suppression of mismatch repair. *eLife* **7**, e32692 (2018).
49. Rayner, E. et al. A panoply of errors: polymerase proofreading domain mutations in cancer. *Nat. Rev. Cancer* **16**, 71–81 (2016).
50. Nakamori, M., Gourdon, G. & Thornton, C. A. Stabilization of expanded (CTG)·(CAG) repeats by antisense oligonucleotides. *Mol. Ther.* **19**, 2222–2227 (2011).
51. Su, X. A. & Freudenreich, C. H. Cytosine deamination and base excision repair cause R-loop-induced CAG repeat fragility and instability in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **114**, E8392–E8401 (2017).
52. Lin, Y., Hubert, L. Jr. & Wilson, J. H. Transcription destabilizes triplet repeats. *Mol. Carcinog.* **48**, 350–361 (2009).
53. Tomé, S. et al. MSH2 ATPase domain mutation affects CTG·CAG repeat instability in transgenic mice. *PLoS Genet.* **5**, e1000482 (2009).
54. McMurray, C. T. Hijacking of the mismatch repair system to cause CAG expansion and cell death in neurodegenerative disease. *DNA Repair* **7**, 1121–1134 (2008).
55. Morales, F. et al. A polymorphism in the MSH3 mismatch repair gene is associated with the levels of somatic instability of the expanded CTG repeat in the blood DNA of myotonic dystrophy type 1 patients. *DNA Repair* **40**, 57–66 (2016).
56. Flower, M. et al. *MSH3* modifies somatic instability and disease severity in Huntington's and myotonic dystrophy type 1. *Brain* **142**, 1876–1886 (2019).
57. Panigrahi, G. B., Lau, R., Montgomery, S. E., Leonard, M. R. & Pearson, C. E. Slipped (CTG)*(CAG) repeats can be correctly repaired, escape repair or undergo error-prone repair. *Nat. Struct. Mol. Biol.* **12**, 654–662 (2005).
58. Zhang, T., Huang, J., Gu, L. & Li, G. M. In vitro repair of DNA hairpins containing various numbers of CAG/CTG trinucleotide repeats. *DNA Repair* **11**, 201–209 (2012).
59. Lai, Y. et al. Crosstalk between MSH2–MSH3 and polβ promotes trinucleotide repeat expansion during base excision repair. *Nat. Commun.* **7**, 12465 (2016).
60. Tian, L. et al. Mismatch recognition protein MutSβ does not hijack (CAG)*n* hairpin repair in vitro. *J. Biol. Chem.* **284**, 20452–20456 (2009).
61. Nakatani, R., Nakamori, M., Fujimura, H., Mochizuki, H. & Takahashi, M. P. Large expansion of CTG·CAG repeats is exacerbated by MutSβ in human cells. *Sci. Rep.* **5**, 11020 (2015).
62. Chen, H., Lisby, M. & Symington, L. S. RPA coordinates DNA end resection and prevents formation of DNA hairpins. *Mol. Cell* **50**, 589–600 (2013).
63. Nguyen, B. et al. Diffusion of human replication protein A along single-stranded DNA. *J. Mol. Biol.* **426**, 3246–3261 (2014).
64. Tsurimoto, T. & Stillman, B. Multiple replication factors augment DNA synthesis by the two eukaryotic DNA polymerases, alpha and delta. *EMBO J.* **8**, 3883–3889 (1989).
65. Tsurimoto, T. & Stillman, B. Replication factors required for SV40 DNA replication in vitro. I. DNA structure-specific recognition of a primer–template junction by eukaryotic DNA polymerases and their accessory proteins. *J. Biol. Chem.* **266**, 1950–1960 (1991).
66. Chan, N. L. et al. The Werner syndrome protein promotes CAG/CTG repeat stability by resolving large $(CAG)_n$/$(CTG)_n$ hairpins. *J. Biol. Chem.* **287**, 30151–30156 (2012).
67. Callahan, J. L., Andrews, K. J., Zakian, V. A. & Freudenreich, C. H. Mutations in yeast replication proteins that increase CAG/CTG expansions also increase repeat fragility. *Mol. Cell. Biol.* **23**, 7849–7860 (2003).
68. Raji, N. S., Krishna, T. H. & Rao, K. S. DNA-polymerase α, β, Δ and ε activities in isolated neuronal and astroglial cell fractions from developing and aging rat cerebral cortex. *Int. J. Dev. Neurosci.* **20**, 491–496 (2002).
69. Kovalenko, M. et al. 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* **7**, e44273 (2012).
70. Mangiarini, L. et al. Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. *Nat. Genet.* **15**, 197–200 (1997).
71. Chiang, C. et al. Complex reorganization and predominant non-homologous repair following chromosomal breakage in karyotypically balanced germline rearrangements and transgenic integration. *Nat. Genet.* **44**, 390–397 (2012).
72. Larson, E., Fyfe, I., Morton, A. J. & Monckton, D. G. Age-, tissue- and length-dependent bidirectional somatic CAG·CTG repeat instability in an allelic series of R6/2 Huntington disease mice. *Neurobiol. Dis.* **76**, 98–111 (2015).
73. Kennedy, L. & Shelbourne, P. F. Dramatic mutation instability in HD mouse striatum: does polyglutamine load contribute to cell-specific vulnerability in Huntington's disease? *Hum. Mol. Genet.* **9**, 2539–2544 (2000).
74. Ishiguro, H. et al. Age-dependent and tissue-specific CAG repeat instability occurs in mouse knock-in for a mutant Huntington's disease gene. *J. Neurosci. Res.* **65**, 289–297 (2001).
75. Gonitel, R. et al. DNA instability in postmitotic neurons. *Proc. Natl Acad. Sci. USA* **105**, 3467–3472 (2008).
76. De Rooij, K. E., De Koning Gans, P. A., Roos, R. A., Van Ommen, G. J. & Den Dunnen, J. T. Somatic expansion of the $(CAG)_n$ repeat in Huntington disease brains. *Hum. Genet.* **95**, 270–274 (1995).
77. Lee, J. M., Pinto, R. M., Gillis, T., St Claire, J. C. & Wheeler, V. C. Quantification of age-dependent somatic CAG repeat instability in Hdh CAG knock-in mice reveals different expansion dynamics in striatum and liver. *PLoS ONE* **6**, e23647 (2011).
78. Lee, J. M. et al. A novel approach to investigate tissue-specific trinucleotide repeat instability. *BMC Syst. Biol.* **4**, 29 (2010).
79. Wheeler, V. C. et al. Factors associated with HD CAG repeat instability in Huntington disease. *J. Med. Genet.* **44**, 695–701 (2007).
80. Higham, C. F., Morales, F., Cobbold, C. A., Haydon, D. T. & Monckton, D. G. High levels of somatic DNA diversity at the myotonic dystrophy type 1 locus are driven by ultra-frequent expansion and contraction mutations. *Hum. Mol. Genet.* **21**, 2450–2463 (2012).
81. Veitch, N. J. et al. Inherited CAG.CTG allele length is a major modifier of somatic mutation length variability in Huntington disease. *DNA Repair* **6**, 789–796 (2007).
82. Hornsby, P. J. & Didenko, V. V. In situ ligation: a decade and a half of experience. *Methods Mol. Biol.* **682**, 49–63 (2011).
83. Majtnerova, P. & Rousar, T. An overview of apoptosis assays detecting DNA fragmentation. *Mol. Biol. Rep.* **45**, 1469–1478 (2018).
84. Iannicola, C. et al. Early alterations in gene expression and cell morphology in a mouse model of Huntington's disease. *J. Neurochem.* **75**, 830–839 (2000).
85. Turmaine, M. et al. Nonapoptotic neurodegeneration in a transgenic mouse model of Huntington's disease. *Proc. Natl Acad. Sci. USA* **97**, 8093–8097 (2000).
86. Yu, Z. X. et al. Mutant huntingtin causes context-dependent neurodegeneration in mice with Huntington's disease. *J. Neurosci.* **23**, 2193–2202 (2003).
87. DiFiglia, M. et al. Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. *Science* **277**, 1990–1993 (1997).
88. Li, S. H. & Li, X. J. Aggregation of N-terminal huntingtin is dependent on the length of its glutamine repeats. *Hum. Mol. Genet.* **7**, 777–782 (1998).
89. Becher, M. W. et al. Intranuclear neuronal inclusions in Huntington's disease and dentatorubral and pallidoluysian atrophy: correlation between the density of inclusions and *IT15* CAG triplet repeat length. *Neurobiol. Dis.* **4**, 387–397 (1998).
90. Li, H. et al. Ultrastructural localization and progressive formation of neuropil aggregates in Huntington's disease transgenic mice. *Hum. Mol. Genet.* **8**, 1227–1236 (1999).
91. Li, H., Li, S. H., Johnston, H., Shelbourne, P. F. & Li, X. J. Amino-terminal fragments of mutant huntingtin show selective accumulation in striatal neurons and synaptic toxicity. *Nat. Genet.* **25**, 385–389 (2000).
92. Carty, N. et al. Characterization of HTT inclusion size, location, and timing in the zQ175 mouse model of Huntington's disease: an in vivo high-content imaging study. *PLoS ONE* **10**, e0123527 (2015).
93. Kaytor, M. D., Wilkinson, K. D. & Warren, S. T. Modulating huntingtin half-life alters polyglutamine-dependent aggregate formation and cell toxicity. *J. Neurochem.* **89**, 962–973 (2004).
94. Coufal, M. et al. Discovery of a novel small-molecule targeting selective clearance of mutant huntingtin fragments. *J. Biomol. Screen.* **12**, 351–360 (2007).
95. Chopra, V. et al. A small-molecule therapeutic lead for Huntington's disease: preclinical pharmacology and efficacy of C2-8 in the R6/2 transgenic mouse. *Proc. Natl Acad. Sci. USA* **104**, 16685–16689 (2007).
96. Butler, D. C. & Messer, A. Bifunctional anti-huntingtin proteasome-directed intrabodies mediate efficient degradation of mutant huntingtin exon 1 protein fragments. *PLoS ONE* **6**, e29199 (2011).
97. Perucho, J. et al. Striatal infusion of glial conditioned medium diminishes huntingtin pathology in r6/1 mice. *PLoS ONE* **8**, e73120 (2013).

98. Tsvetkov, A. S. et al. Proteostasis of polyglutamine varies among neurons and predicts neurodegeneration. *Nat. Chem. Biol.* **9**, 586–592 (2013).
99. Penney, J. B. Jr, Vonsattel, J. P., MacDonald, M. E., Gusella, J. F. & Myers, R. H. CAG repeat number governs the development rate of pathology in Huntington's disease. *Ann. Neurol.* **41**, 689–692 (1997).
100. Wheeler, V. C. et al. Long glutamine tracts cause nuclear localization of a novel form of huntingtin in medium spiny striatal neurons in HdhQ92 and HdhQ111 knock-in mice. *Hum. Mol. Genet.* **9**, 503–513 (2000).
101. Rosenblatt, A. et al. Does CAG repeat number predict the rate of pathological changes in Huntington's disease? *Ann. Neurol.* **44**, 708–709 (1998).
102. Wild, E. J. & Tabrizi, S. J. Therapies targeting DNA and RNA in Huntington's disease. *Lancet Neurol.* **16**, 837–847 (2017).
103. Dabrowska, M., Juzwa, W., Krzyzosiak, W. J. & Olejniczak, M. Precise excision of the CAG tract from the Huntingtin gene by Cas9 nickases. *Front. Neurosci.* https://doi.org/10.3389/fnins.2018.00075 (2018).
104. Shin, J. W. et al. Permanent inactivation of Huntington's disease mutation by personalized allele-specific CRISPR/Cas9. *Hum. Mol. Genet.* **25**, 4566–4576 (2016).
105. 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).
106. Cinesi, C., Aeschbach, L., Yang, B. & Dion, V. Contracting CAG/CTG repeats using the CRISPR–Cas9 nickase. *Nat. Commun.* **7**, 13272 (2016).
107. Suelves, N., Kirkham-McCarthy, L., Lahue, R. S. & Gines, S. A selective inhibitor of histone deacetylase 3 prevents cognitive deficits and suppresses striatal CAG repeat expansions in Huntington's disease mice. *Sci. Rep.* **7**, 6082 (2017).
108. Eisenstein, M. CRISPR takes on Huntington's disease. *Nature* **557**, S42–S43 (2018).
109. Martins, S. et al. Modifiers of $(CAG)_n$ instability in Machado–Joseph disease (MJD/SCA3) transmissions: an association study with DNA replication, repair and recombination genes. *Hum. Genet.* **133**, 1311–1318 (2014).
110. Guo, J., Gu, L., Leffak, M. & Li, G. M. MutSβ promotes trinucleotide repeat expansion by recruiting DNA polymerase β to nascent $(CAG)_n$ or $(CTG)_n$ hairpins for error-prone DNA synthesis. *Cell Res.* **26**, 775–786 (2016).
111. Hanawalt, P. C. & Spivak, G. Transcription-coupled DNA repair: two decades of progress and surprises. *Nat. Rev. Mol. Cell Biol.* **9**, 958–970 (2008).
112. Hou, C., Chan, N. L., Gu, L. & Li, G. M. Incision-dependent and error-free repair of $(CAG)_n/(CTG)_n$ hairpins in human cell extracts. *Nat. Struct. Mol. Biol.* **16**, 869–875 (2009).
113. Chan, N. L. et al. Coordinated processing of 3′ slipped $(CAG)_n/(CTG)_n$ hairpins by DNA polymerases β and Δ preferentially induces repeat expansions. *J. Biol. Chem.* **288**, 15015–15022 (2013).
114. Pinto, R. M. et al. Mismatch repair genes *Mlh1* and *Mlh3* modify CAG instability in Huntington's disease mice: genome-wide and candidate approaches. *PLoS Genet.* **9**, e1003930 (2013).
115. Wheeler, V. C. et al. Mismatch repair gene *Msh2* modifies the timing of early disease in $Hdh^{Q111}$ striatum. *Hum. Mol. Genet.* **12**, 273–281 (2003).
116. Strobel, S. A., Doucette-Stamm, L. A., Riba, L., Housman, D. E. & Dervan, P. B. Site-specific cleavage of human chromosome 4 mediated by triple-helix formation. *Science* **254**, 1639–1642 (1991).
117. Mittelman, D. et al. Zinc-finger directed double-strand breaks within CAG repeat tracts promote repeat instability in human cells. *Proc. Natl Acad. Sci. USA* **106**, 9607–9612 (2009).
118. Zeitler, B. et al. Allele-selective transcriptional repression of mutant HTT for the treatment of Huntington's disease. *Nat. Med.* **25**, 1131–1142 (2019).
119. Mosbach, V., Poggi, L. & Richard, G. F. Trinucleotide repeat instability during double-strand break repair: from mechanisms to gene therapy. *Curr. Genet.* **65**, 17–28 (2019).
120. Malankhanova, T. B., Malakhova, A. A., Medvedev, S. P. & Zakian, S. M. Modern genome editing technologies in Huntington's disease research. *J. Huntington's Dis.* **6**, 19–31 (2017).
121. Babacic, H., Mehta, A., Merkel, O. & Schoser, B. CRISPR-cas gene-editing as plausible treatment of neuromuscular and nucleotide-repeat-expansion diseases: a systematic review. *PLoS ONE* **14**, e0212198 (2019).
122. Gomes-Pereira, M. & Monckton, D. G. Chemical modifiers of unstable expanded simple sequence repeats: what goes up, could come down. *Mutat. Res.* **598**, 15–34 (2006).
123. Pineiro, E. et al. Mutagenic stress modulates the dynamics of CTG repeat instability associated with myotonic dystrophy type 1. *Nucleic Acids Res.* **31**, 6733–6740 (2003).
124. Budworth, H. et al. Suppression of somatic expansion delays the onset of pathophysiology in a mouse model of Huntington's disease. *PLoS Genet.* **11**, e1005267 (2015).
125. Gottesfeld, J. M., Neely, L., Trauger, J. W., Baird, E. E. & Dervan, P. B. Regulation of gene expression by small molecules. *Nature* **387**, 202–205 (1997).

## Methods

**NA synthesis.** The synthesis of NA is described in the Supplementary Note.

**Cell culture.** Construction of the HT1080-$(CAG)_{850}$ cell model was described previously[21]. The HT1080 model cells, the HD primary fibroblasts, GM09197 (Coriell Biorepository) with $(CAG)_{180/21}$ (ref. [5]), and the HD primary fibroblasts, GM02191 (Coriell Biorepository) with $(CAG)_{43/19}$, were cultured at 37 °C with 5% $CO_2$ in DMEM supplemented with 10% fetal bovine serum. Cells were treated with or without continuous exposure to 50 μM NA for 30 d (HT1080 cells) or 40 d (primary fibroblasts). The WST-1 assay was performed according to the manufacturer's instructions (Roche). All cell lines were tested to be mycoplasma free. We performed three independent experiments for each cell line.

Proliferation of HT1080-$(CAG)_{850}$ cells was inhibited by treatment with 0.5 μM palbociclib, as described previously[126]. The degree of proliferation arrest was assessed in living cells by counting BrdU-positive cells after a 24-h incubation with BrdU in proliferating or palbociclib-treated HT1080-$(CAG)_{850}$ cells. Sustained knockdown of MSH3 by siRNA was as described previously[70].

**Microscopy.** HT1080-$(CAG)_{850}$ cells were incubated with 50 μM NBD-labeled NA for 48 h and Cell Light Plasma Membrane-RFP, BacMam 2.0 (Life Technologies, catalogue no. C10608), then fixed for 15 min at room temperature with 4% paraformaldehyde and washed twice for 10 min in PBS. Cells were mounted with Vectashield hard-set mounting medium that contains DAPI (Vector Laboratories, catalogue no. H-1500). Fluorescence images were obtained using an Olympus FV1000D confocal laser scanning microscope (Olympus).

**Repeat length analysis.** The (CTG)•(CAG) repeats were sized by spPCR with the input of 1.4–1.7 genome equivalents, followed by Southern blot[21]. Repeat size differences in models are at most 3,000 bp; therefore, being strictly conservative, a bias to amplifying shorter alleles can be possible even under optimized PCR conditions in our study. For HD primary fibroblasts, spPCR and Southern blots were performed as described previously[21] where DNA was diluted to 1.0–1.6 genome equivalents. Amplified products (PCR primers, Supplementary Table 2) were detected by Southern blot using a digoxigenin-labeled $(CAG)_7$ locked nucleic acid probe[127]. We analyzed at least 50 alleles for each of the three experiments (more than 150 alleles per experiment). The repeat analyses are summarized in Supplementary Table 1. The percentage of repeat population was calculated by determining the proportion of >50 individual spPCR reactions across the CAG-repeat tract that harbored a certain size of repeat product[19]. A $\chi^2$ test was performed to compare the frequencies of expanded, unchanged and contracted alleles in each set of experiments[21,50]. The trinucleotide and dinucleotide tract lengths of the *HTT*, *CASK*, *ATXN8*, *Mfd15*, *TBP* and mouse *Mapkap1*, *Fgd4* and *Tbp* loci were PCR-amplified (primers listed in Supplementary Table 2) following amplification conditions described elsewhere[31,32,128–132]. Human *TBP* CAG length variability was assessed by spPCR.

**Messenger RNA quantification.** RNA was collected using the RNeasy Plus Micro Kit (Qiagen). Total RNA was primed with random hexamers and reverse transcribed with Superscript III (Life Technologies), followed by RNAseH treatment. Quantitative PCR with reverse transcription was performed using TaqMan Gene Expression assays or PrimeTime qPCR assays on an ABI PRISM 7900HT (Life Technologies). The level of transgene-derived mRNA was normalized to *18S* rRNA. Results were statistically analyzed using a paired *t*-test. Primer sequences are listed in Supplementary Table 2.

**Western blot analysis.** Lysates were prepared from HD patient fibroblasts in RIPA buffer. A 20 μg quantity of lysate was electrophoresed at 100 V for 4 h at room temperature in 1× MOPS buffer (Thermo Fisher Scientific, catalogue no. NP0001) in 3–8% Tris-acetate pre-cast protein gel (Invitrogen, catalogue no. EA0375BOX), and transferred to PVDF western blotting membranes (Sigma-Aldrich, catalogue no. 3010040001) overnight at 4 °C at a constant voltage of 20 V. Membranes were blocked in 10% milk in TBS + 0.1% Tween-20, incubated with primary antibody at room temperature for 2 h, and with secondary antibody at room temperature for 1 h, and then detected with ECL (GE Healthcare, catalogue no. RPN2232). Primary antibodies: anti-HTT protein antibody amino acids 181–810 clone 1HU-4C8 (1:1,000, mouse) (Millipore Sigma, catalogue no. MAB2166), anti-actin protein antibody (C4 clone, 1:30,000, mouse) (BD Transduction, catalogue no. 612657). Secondary antibodies: peroxidase-AffiniPure sheep anti-mouse IgG H+L (1:2,000) (Cedarlane Labs, catalogue no. 515-035-062).

**NA–DNA binding.** Homoduplex slipped structures of 50 repeats and heteroduplexes with long $(CAG)_{20}$ or $(CTG)_{20}$ slip-outs were formed by alkaline denaturation/renaturation, as described previously[17]. Briefly, plasmids containing human DM1 genomic $(CTG)_n$•$(CAG)_n$ repeats ($n = 30$ or 50) were linearized by HindIII digestion. DNAs of $(CAG)_{50}$ and $(CTG)_{30}$ repeats, or DNAs of $(CAG)_{30}$ and $(CTG)_{50}$ repeats, were mixed in equimolar amounts and then heteroduplexed by alkaline denaturation/renaturation. Repeat-containing fragments were released by EcoRI digestion and electrophoretically resolved on a 4% polyacrylamide gel. Gel-purified fragments were radiolabeled with [α-$^{32}$P]deoxynucleoside triphosphates on both strands by a fill-in reaction, the radioactivity of each structure was determined using Cerenkov counting, and an equivalent radioactive concentration of each structure was incubated with an increasing concentration of NA for 30 min at room temperature with 1× hypotonic buffer. Binding products were resolved by electrophoresis on a 4% (w/v) polyacrylamide gel in 1× TBE buffer at a constant 200 V for 2.5 h.

**Replication assay.** In vitro replication templates containing $(CTG)_{79}$•$(CAG)_{79}$ were previously described[133,134]. The SV40 origin of bidirectional replication was placed 103 or 98 bp 5′ or 3′ of the CAG repeat for pDM79EF and pDM79HF, respectively. These templates and an SV40-ori template with no repeats (pKN16) were replicated in vitro by HeLa cell extract, adding [α-$^{32}$P]dCTP and T antigen, as described previously[133,134]. Replication was performed in the presence or absence of NA (7.5 μM or 15 μM). Radioactive replication products were purified, linearized (BamHI) and treated with DpnI. Equal amounts of unreplicated pKN16 DNA were treated with DpnI to show the complete digestion of unreplicated plasmid DNA. Equal quantities of reaction products were resolved by electrophoresis on a 15-cm 1% agarose gel run for 16 h at 4 V cm$^{-1}$ in TBE buffer, dried and exposed to Kodak film.

**Repair assay.** To determine the effect of NA binding on slipped-DNA repair, a series of circular slipped-DNA substrates were prepared with an excess of repeats with a nick located 5′ or 3′ of the slip-out[23,57]. G–T mismatched substrate was prepared as described previously[23,57]. All DNA substrates were processed in vitro by HeLa extracts. Analysis of repair products was performed by Southern blotting, comparing with starting material. Each repair assay has been performed in three independent experiments.

**R-loop formation and processing.** Plasmids bearing a $(CAG)_{79}$•$(CTG)_{79}$ tract with convergent T3 and T7 RNA polymerase promoters were described previously[18,135]. Transcription reactions were performed as described previously[18]. Briefly, 500 ng of template DNA in 1× transcription buffer (Roche) and 1× bovine serum albumin (New England Biolabs) were mixed for 1 h with 20 U of the appropriate RNA polymerase: T7, T3 or T7 + T3 (Roche), with or without NA (120 μM). Products were purified and treated with either 1 g of RNaseA (Roche) alone or mixed with 1 U of *Escherichia coli* RNaseH (Roche), at room temperature for 30 min, in the presence or absence of NA (120 μM). All in vitro transcription reaction products were analyzed on 1% agarose gels run in 1× TBE buffer at 80 V for 3 h. Gel-borne products were visualized with ethidium bromide (0.5 mg ml$^{-1}$) under ultraviolet light.

R-loop templates prepared from in vitro transcription and RNaseA treatments were incubated with NA and then processed by extracts of HeLa or SH-SY5Y neuroblastoma cells, where the latter were terminally differentiated by retinoic acid[57]. Nucleic acids were purified and transformed into bacteria for STRIP analysis as described previously[18]. Briefly, products of human cell extract processing were transformed into *E. coli* XL1-MutS (Agilent). Individual bacterial colonies (each representing one processed template) were picked and cultured for 6 h. The magnitudes of the repeat length changes were determined by electrophoretic sizing of the repeat-containing fragments on 4% polyacrylamide gels relative to the starting length material and a size marker. Data are derived from three independent in vitro transcription and human cell extract processing reactions with ~150 colonies (~50 colonies per replicate) representing 150 individual products of cell extract treatment analyzed for each R-loop configuration. Individual experiments were compared using the $\chi^2$ test.

**MutSβ binding assay.** MutSβ was purified from baculovirus-infected Sf9 cells expressing His-tagged hMSH2 and hMSH3 as described previously[23]. Binding reactions were performed at room temperature. Slipped-DNA with long $(CAG)_{20}$ was prepared and end-labeled as described above. Protein was incubated with DNA for 30 min in a buffer containing 10 mM HEPES-KOH pH 7.5, 110 mM KCl, 1 mM EDTA and 1 mM dithiothreitol with or without ATP in the buffer, as indicated. Reactions were loaded onto a 4% native polyacrylamide gel with non-denaturing loading dye (20 mM Tris-HCl pH 7.4, 4% glycerol, bromophenol blue). Gel was run in 1× TBE buffer for 2 h.

**MBN footprinting.** The experimental conditions used were single-hit kinetics, as described previously[17].

**RPA DNA-binding.** Recombinant RPA protein was expressed in BL21(DE3) cells and binding was performed as described previously[136].

**Polymerase extension assay.** Recombinant human polδ was prepared in insect cells using a baculovirus vector and purified by immunoaffinity column chromatography, as described previously[137]. The polδ extension assay was performed as described previously[138], using an oligonucleotide containing $(CAG)_{10}$ repeats, as described earlier[25]. Briefly, 0.1 μM primer and 0.1 μM oligonucleotide were denatured at 95 °C for 3 min, annealed for 30 min at room temperature, and incubated with NA for 30 min at room temperature. RPA and/or polδ were added and the reaction was started by adding 0.1 mM deoxynucleoside triphosphates in

10 µl reaction volume, and incubated at 37 °C for 15 min. The reaction was stopped by adding 20 mM EDTA and purified by extraction with phenol/chloroform/isoamyl alcohol (25:24:1, v/v/v) followed by ethanol precipitation. Pellets were resuspended in formamide buffer, denatured at 95 °C for 10 min, and run on a 6% sequencing gel at 2,000 V and 90 W for 40 min.

**Stereotaxic injections into R6/2 mice.** Mouse handling and experimental procedures were conducted in accordance with the Osaka University guidelines for the welfare of animals. A single NA application involved six separate stereotactic injections (three injections of NA or saline into three different striatal regions of either the left or right striatum, respectively). Under sterile conditions, 13 male R6/2 mice (B6CBA-Tg(HDexon1)62Gpb/1J, Jackson Laboratory, catalogue no. 002810), 6-week-old, were anesthetized, and stereotaxically injected with 5 µl of saline (PBS solution) (right side) or 500 µM NA dissolved in saline (left side). The control for NA treatment is therefore the contralateral side of the striatum of each mouse. Mice received injections once or twice bi-weekly, or weekly for four consecutive weeks. Stereotaxic injections were delivered to three sites within the striatum with the following coordinates: (anterior–posterior (AP) = 0.0 mm, medial–lateral (ML) = 1.5 mm from the bregma, dorsal–ventral (DV) = 2.5 mm below the dural surface; AP = 1.0 mm, ML = 1.5 mm, DV = 2.5 mm; and AP = 0.5 mm, ML = 1.5 mm, DV = 2.5 mm), using a 10-µl Hamilton microsyringe at a rate of 0.5 µl $min^{-1}$. Both the right and left striata were assessed for repeat length at the HD CAG transgene and at endogenous CAG tracts. For the mice treated four times, DNAs from the tail before and following NA administration, as well as left and right frontal cortex and cerebellum, were collected. Coronal sections of the striatum were cut serially at 20 µm thickness using a cryostat (CM1850UV; Leica Microsystems). Striatum sections of 0.62, 1.10 and 1.58 mm posterior from the bregma were washed in PBS containing 0.05% Triton X-100 (PBS-T). They were then immersed in a solution of 3% $H_2O_2$, 10% methanol in PBS for 10 min. After washing three times with PBS-T, the sections were incubated with primary antibody (mouse anti-NeuN, 1:100, Millipore; rabbit anti-doublecortin, 1:200, Abcam) diluted in 10% Block-Ace blocking solution (Yukijirushi Co.) overnight at 4 °C and then with biotinylated secondary antibody (1:500, VECTOR Laboratories) in PBS-T for 1 h at room temperature followed by VECTASTAIN ABC reagents (1:100, VECTOR Laboratories) in PBS for 1 h at room temperature. The sections were washed three times with PBS-T in between the steps. The bound complex was visualized with 3,3′-diaminobenzidine (Sigma).

**Genescan analysis.** At 4 weeks after the first injection, DNA was isolated from mouse brain tissue as described previously[127]. PCR was performed as described previously[139], and PCR products were sized on an ABI310 Gene Analyser using GENESCAN 3.1 software (Life Technologies).

**Instability index calculation.** The instability index was calculated as described previously[77,78] with modifications as outlined in Extended Data Fig. 7.

**Immunofluorescence.** MSN-specific mHTT aggregates were assessed using a multiplexed immunofluorescence approach that stains for HTT protein (EM48 antibody, catalogue no. MAB5374, Millipore-Sigma) and anti-DARPP-32 (19A3 antibody, catalogue no. 2306S, New England Biolabs—expressed almost exclusively by striatal MSNs), as described previously[140]. After rehydration, slides were subjected to antigen retrieval in 10 mM sodium citrate for 20 min in a steamer and cooled to room temperature for 1 h. Slides were washed twice in 1× PBS + 0.05% Tween-20 (1× PBST) (2 min, 2 times). Slides were blocked in 10% normal goat serum for 1 h at room temperature and incubated with primary antibody overnight at 4 °C. Secondary antibody incubation was performed for 1 h at room temperature. Slides were mounted with Hardset VectaShield with DAPI. Aggregates were then blindly assessed in two ways, as follows.

The first method was a whole striatal assessment of the red pixel intensity (mHTT aggregates) via ImageJ. The intensity was measured from ×20 images obtained via a 3DHistech Pannoramic 250 Flash II slide scanner. Each individual striatal intensity per slice was normalized to the total striatal intensity of the slice and then plotted. Three separate brain slices were assessed per mouse, assessing a total of four mice (each treated four times over 4 weeks).

The second method was quantification of MSN nuclei containing mHTT aggregates from confocal images obtained from a Quorum spinning-disk confocal (Olympus IX81) microscope. Ten ×40 (water) confocal images were taken at 10 different locations per striatum with ~100 cells per image (~1,000 cells per striatum, or ~2,000 cells total per brain slice) being assessed. Three striatal slices were assessed per mouse (4 mice total, ~3,000 cells assessed per striatum per mouse). The percentages of cells containing mHTT aggregates were taken relative to the total number of cells in the image.

**Whole-genome sequencing, alignment and variant calling.** Whole-genome sequencing was performed using established protocols on Illumina instruments and paired-end FASTQ files were aligned to the human genome (hg19/GRCh37) using BWA-MEM (v.0.7.8), with Picard MarkDuplicates (v.1.108) being used to mark PCR duplicates. Indel realignment and base quality scores were recalibrated using the Genome Analysis Toolkit (v.2.8.1). Somatic mutations were detected using MuTect2 (part of GATK v.3.5)[141]. Mutations present only in treated cells (as opposed to non-treated cells) were retained. For substitutions, we also removed common single-nucleotide polymorphisms as described previously[45]. To remove common germline variants, we used an in-house panel of controls and removed any putative substitutions present in two or more controls. Putative substitutions were also removed if they were overlapped with a highly repetitive sequence (using DUST[142] with score >60) or were located in excessively high-depth alignments in difficult-to-align regions of the genome, as described previously[143]. We apply a maximum depth threshold of $d + 4\sqrt{d}$, where $d$ is the average normal mean read depth of the chromosome in the corresponding untreated cell.

Mutational signatures shown in COSMIC were screened for using a further elaborated version (v0.0.5.75 (ref. [41])) of an established non-negative matrix factorization-based method, SigProfiler[144–146]. This framework works in two steps: de novo extraction of signatures on the catalogue of substitutions using a non-negative matrix factorization approach; and comparing deciphered de novo signatures to a set of previously described mutational signatures (COSMIC Mutational Signatures v3) using cosine similarity to identify the etiology of the underlying mutagenic process. It is common practice to consider only mutational signatures with >100 mutations as reliable.

***HPRT1* sequence analysis by single-molecule real-time circular consensus sequencing.** DNAs were extracted from three independent NA or saline treatments of the HD primary fibroblasts (GM09197) and three independent NA-treated and control striata of three R6/2 mice. The details of the sequencing results are described in the Supplementary Note. Mutation rates were not significantly different between NA- and saline-treated sequences (two-sample Kolmogorov–Smirnov test).

**MSI assay.** MSI was assessed as per Gallon et al.[33]. Briefly, >20 mononucleotide repeats were amplified from 50 ng of template DNA using single-molecule molecular inversion probes, and sequenced to a median (Q1–Q3) depth of 1,007 (733–1,515) reads per marker per sample using a MiSeq (Illumina) sequencer. Sequencing reads were aligned to the human reference genome (build hg19) using BWA[147]. Frequencies of microsatellite length variants were extracted for each marker in each sample, and converted to a probability by comparison to a reference distribution generated from 40 microsatellite-stable, peripheral blood leukocyte controls. A sample score is calculated by combining probabilities from each marker using Fisher's method, and multiplying the decadic logarithm of this combined probability by −1. Scores >1.3 are therefore equivalent to a >95% probability that the observed frequency of microsatellite length variants for a sample is greater than for the reference set, and is used as a classification threshold. Residual samples from Gallon et al.[33], comprising peripheral blood leukocyte DNA of three constitutional MMR-deficiency and eight control patients, were used as controls.

**TUNEL assay.** The TUNEL assay was conducted using the Click-iT Plus TUNEL Assay for In Situ Apoptosis Detection, Alexa Fluor 488 dye (Invitrogen, catalogue no. C10617), multiplexed with anti-DARPP-32 (19A3 antibody, catalogue no. 2306S, New England Biolabs), according to the manufacturer's instructions.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Raw sequencing data have been deposited at the Sequence Read Archive (SRA) under accession numbers SRR10532698, SRR10532697, SRR10532700 and SRR10532699. Source data for Figs. 1–8 and Extended Data Figs. 1 and 4 are provided online.

## References

126. Leontieva, O. V. & Blagosklonny, M. V. CDK4/6-inhibiting drug substitutes for p21 and p16 in senescence: duration of cell cycle arrest and MTOR activity determine geroconversion. *Cell Cycle* **12**, 3063–3069 (2013).
127. Nakamori, M., Sobczak, K., Moxley, R. T. & Thornton, C. A. Scaled-down genetic analysis of myotonic dystrophy type 1 and type 2. *Neuromuscul. Disord.* **19**, 759–762 (2009).
128. Brook, J. D. et al. Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. *Cell* **69**, 385 (1992).
129. Dietmaier, W. et al. Diagnostic microsatellite instability: definition and correlation with mismatch repair protein expression. *Cancer Res.* **57**, 4749–4756 (1997).
130. Kabbarah, O. et al. A panel of repeat markers for detection of microsatellite instability in murine tumors. *Mol. Carcinog.* **38**, 155–159 (2003).
131. Koob, M. D. et al. An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). *Nat. Genet.* **21**, 379–384 (1999).
132. Kremer, B. et al. Sex-dependent mechanisms for expansions and contractions of the CAG repeat on affected Huntington disease chromosomes. *Am. J. Hum. Genet.* **57**, 343–350 (1995).

133. Cleary, J. D., Nichol, K., Wang, Y. H. & Pearson, C. E. Evidence of *cis*-acting factors in replication-mediated trinucleotide repeat instability in primate cells. *Nat. Genet.* **31**, 37–46 (2002).
134. Panigrahi, G. B., Cleary, J. D. & Pearson, C. E. In vitro (CTG)*(CAG) expansions and deletions by human cell extracts. *J. Biol. Chem.* **277**, 13926–13934 (2002).
135. Reddy, K. et al. Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. *Nucleic Acids Res.* **39**, 1749–1762 (2011).
136. Binz, S. K., Dickson, A. M., Haring, S. J. & Wold, M. S. Functional assays for replication protein A (RPA). *Methods Enzymol.* **409**, 11–38 (2006).
137. Zhou, Y., Meng, X., Zhang, S., Lee, E. Y. & Lee, M. Y. Characterization of human DNA polymerase delta and its subassemblies reconstituted by expression in the MultiBac system. *PLoS ONE* **7**, e39156 (2012).
138. Mason, A. C., Roy, R., Simmons, D. T. & Wold, M. S. Functions of alternative replication protein A in initiation and elongation. *Biochemistry* **49**, 5919–5928 (2010).
139. Tome, S. et al. Tissue-specific mismatch repair protein expression: MSH3 is higher than MSH6 in multiple mouse tissues. *DNA Repair* **12**, 46–52 (2013).
140. Jeon, I. et al. Human-to-mouse prion-like propagation of mutant huntingtin protein. *Acta Neuropathol.* **132**, 577–592 (2016).
141. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
142. Morgulis, A., Gertz, E. M., Schaffer, A. A. & Agarwala, R. A fast and symmetric DUST implementation to mask low-complexity DNA sequences. *J. Comput. Biol.* **13**, 1028–1040 (2006).
143. Li, H. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* **30**, 2843–2851 (2014).
144. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
145. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* **3**, 246–259 (2013).
146. Nik-Zainal, S. et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).
147. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).


## Acknowledgements

This work was partially supported by the Canadian Institutes of Health Research (FRN388879, J.-Y.M. and FRN148910, C.E.P.), Natural Sciences and Engineering Research Council (RGPIN-2016-08355, C.E.P.), Muscular Dystrophy Canada (C.E.P.), Tribute Communities (C.E.P.), The Petroff Family Fund (C.E.P.), The Kazman Family Fund (C.E.P.), The Marigold Foundation (C.E.P.), The National Center of Neurology and Psychiatry (29-4, M.N.), a JSPS KAKENHI Grant-in-Aid for Young Scientists (Start-up A, 24890110 and 25713034, M.N.), Scientific Research (B, 16H05321, M.N.) and Specially Promoted Research (26000007, K.N.), the Cancer Research UK Catalyst Award (C569/A24991, R.G.), the US National Institutes of Health (2 R01 ES014737, M.Y.W.T.L.) and a US National Institutes of Health (NIH) grant (HG010169 to E.E.E.). E.E.E. is an investigator of the Howard Hughes Medical Institute, A. Shlien holds the Canada Research Chair in Childhood Cancer Genomics, J.-Y.M. holds the Fonds de Recherche de Santé Québec Chair in Genome Stability, and C.E.P. holds the Canada Research Chair in Disease-Associated Genome Instability. We acknowledge the technical support of R. Manabe, K. Hayashi, P. Wang, L. Yu, M. Mirceta, N. Thakkar, I. Panigrahi, D. Ripsman, H. Adhikary, S. Bérubé, Y. Coulombe, A. Couturier, M. Scofield Sorensen and K. Hoekzema. We acknowledge TCAG (SickKids) for expedited sequencing.

## Author contributions

M.N. performed repeat length analysis, and cell and mouse treatments. G.B.P. performed replication, repair, footprinting and R-loop processing. S.L. performed NA binding, repeat instability analysis, protein interactions, polδ extension assay and instability index assessment. H.H., H.T., M.P.T. and H.M. performed mouse treatments. T.O., J.Li, A. Sakata, A.M. and K.N. synthesized, purified and characterized NA. J.Luo and T.P. performed cell treatments. T.G.-D. performed assessment of NA on mHTT aggregates, HTT translation and TUNEL assay. M.-C.C., N.J., J.-Y.M., M.S.W., X.W. and M.Y.W.T.L. performed protein purification. S.L., G.B.P. and K.C. performed in vitro repair/binding assays. J.H., K.M.M. and E.E.E. performed single-molecule sequencing and bioinformatic analysis. R.G. and M.S.-K. performed MSI assay. S.D., M.L., L.-M.E. and A. Shlien performed whole-genome sequencing and mutation signature analysis. C.E.P., M.N., K.N., G.B.P. and S.L. conceived experiments, analyzed data and wrote the manuscript. M.N., G.B.P. and S.L. contributed equally to the study. All authors discussed the results and commented on the manuscript.

## Competing interests

The authors declare no competing interests.

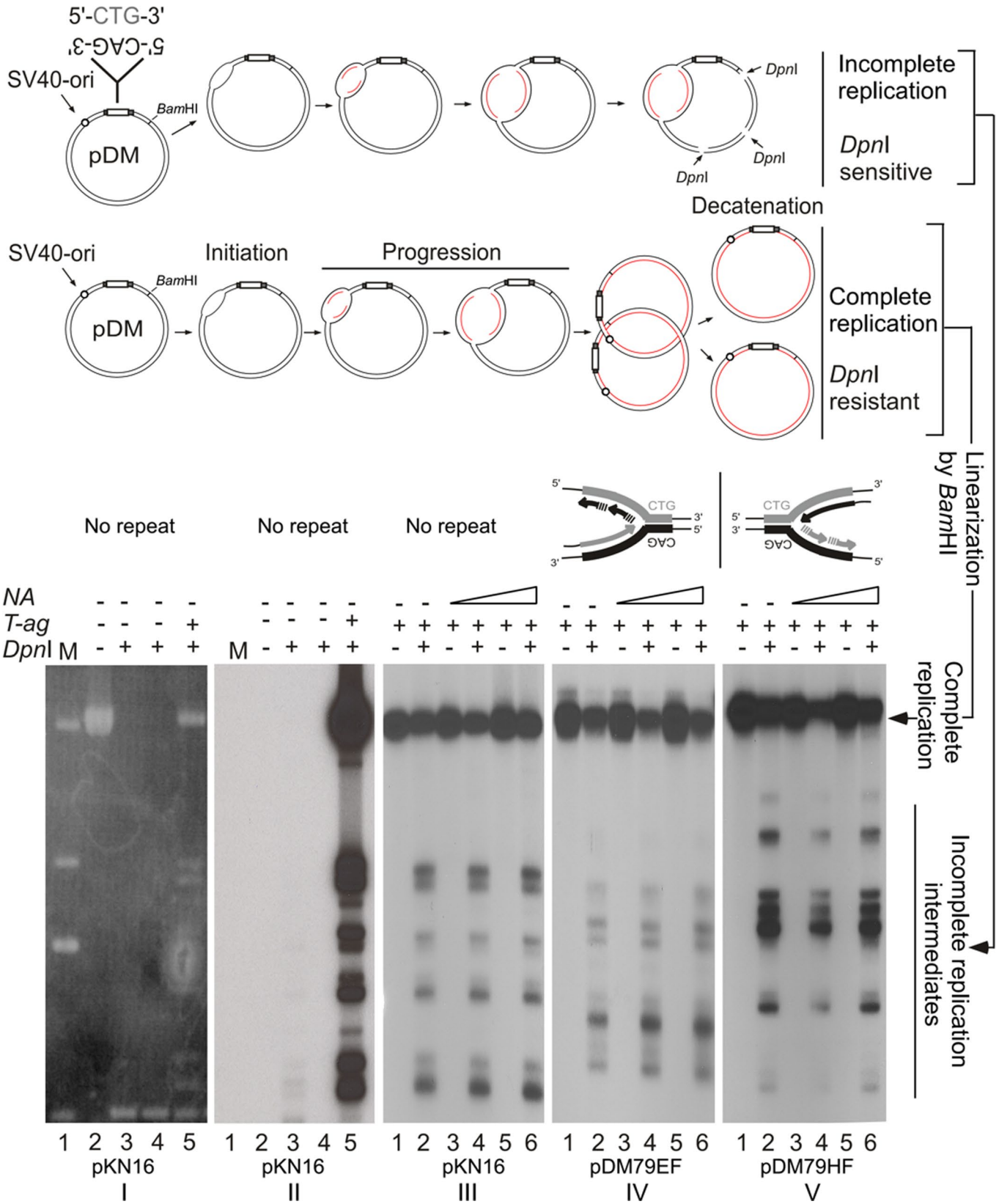
## Additional information

**Extended data** is available for this paper at https://doi.org/10.1038/s41588-019-0575-8.

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41588-019-0575-8.

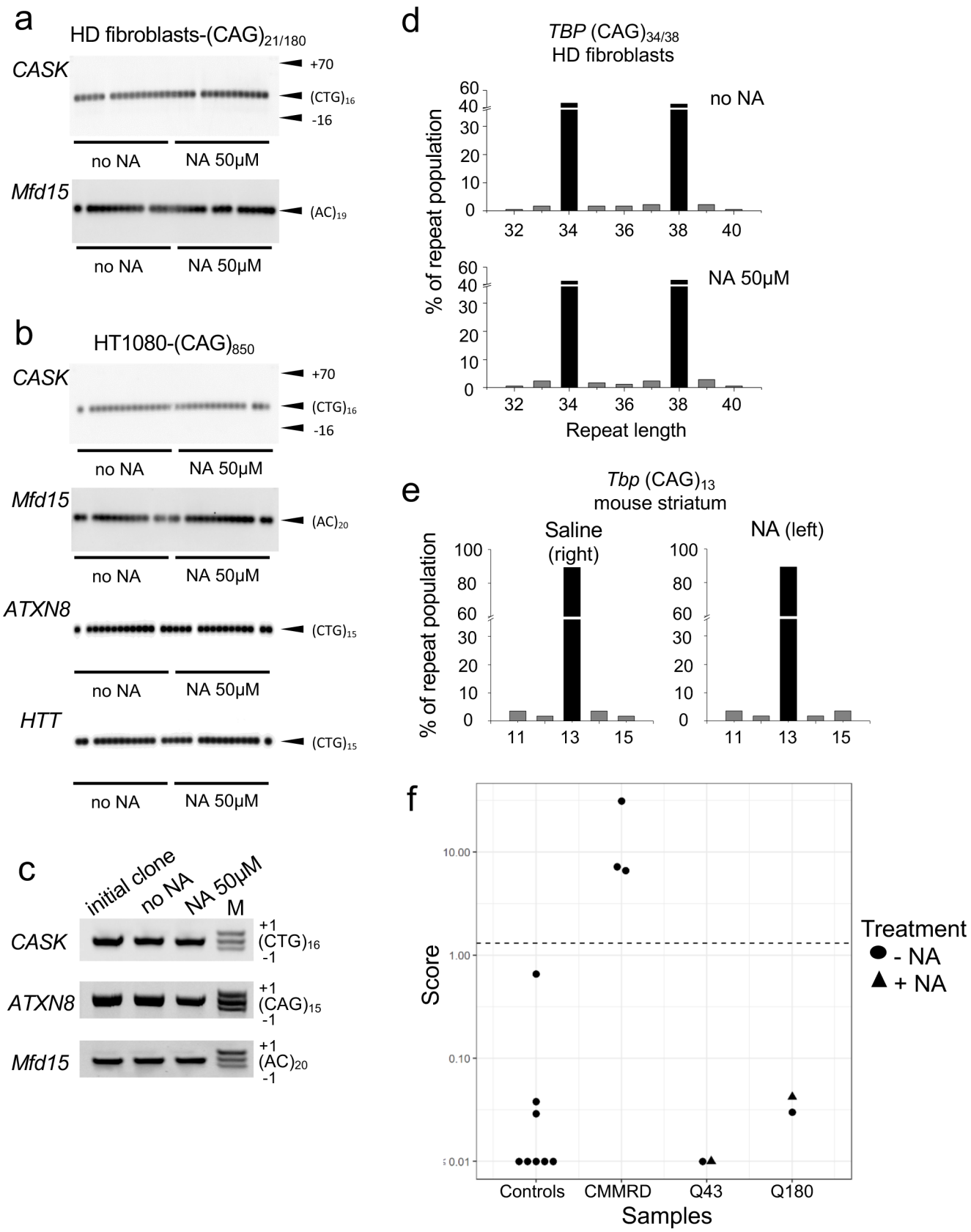
**Correspondence and requests for materials** should be addressed to C.E.P.

**Reprints and permissions information** is available at www.nature.com/reprints.



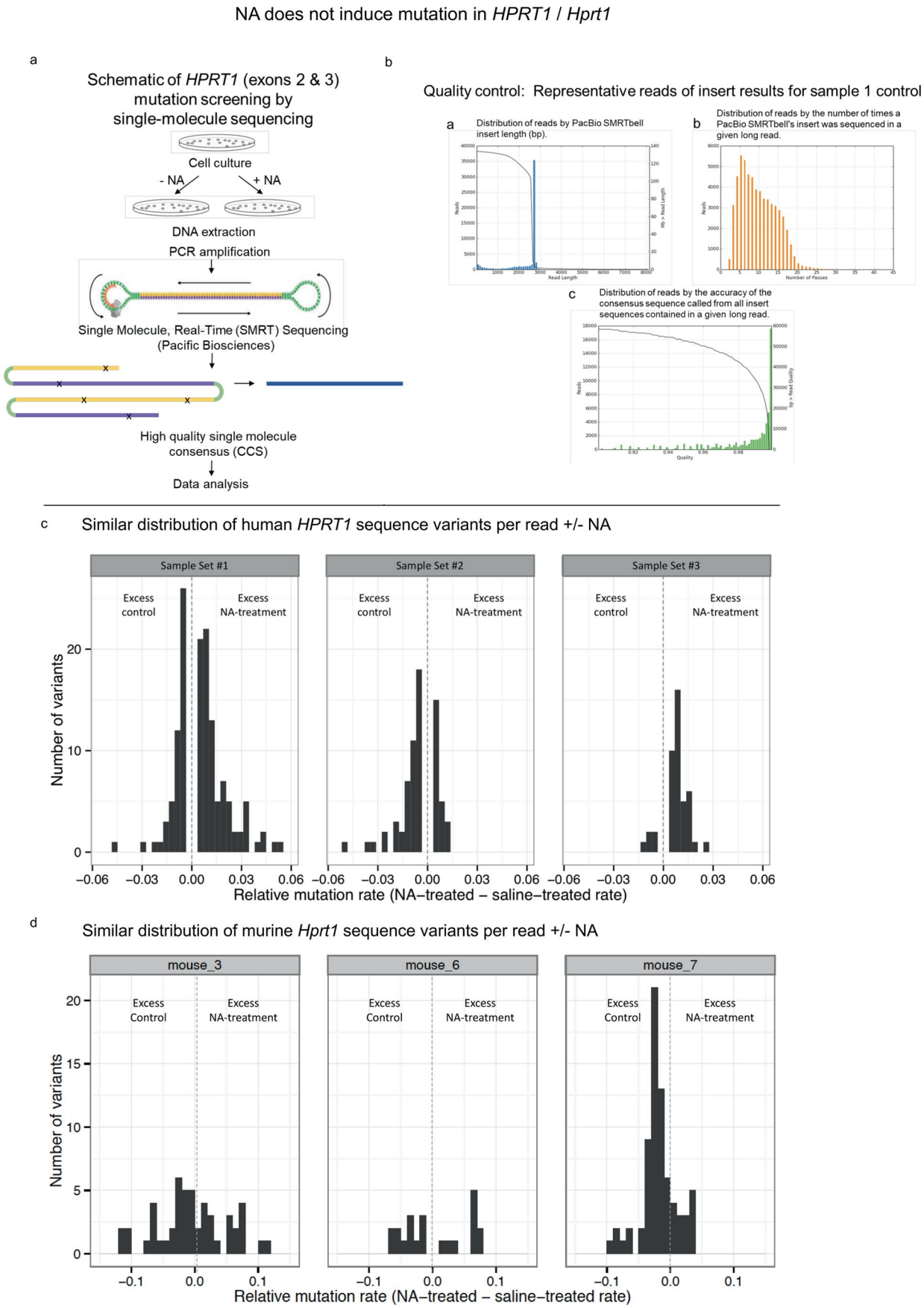

**Extended Data Fig. 1 |** See next page for caption.

**Extended Data Fig. 1 | NA does not affect replication efficiency or replication fork progression.** Three circular plasmids containing the SV40 origin of replication, and an expanded (CAG)79•(CTG)79 repeat tract (pDM79EF and pDM79HF) or no repeats (pKN16), were replicated *in vitro* by human (HeLa) cell extracts without or with NA (7.5 μM or 15 μM) treatment. The location of SV40-*ori* determines the replication direction and which strand will be used as the leading or the lagging strand template. pDM79HF uses the CAG strand as the lagging strand template, while pDM79EF uses the CTG strand as the lagging strand template (schematic on the top of the gel panel). Replication products were purified and linearized with *Bam*HI. An equal portion of the reaction material was also digested with *Bam*HI and *Dpn*I as *Dpn*I digests un-replicated and partially-replicated material, as shown in the schematic (top figure). The digestion products were electrophoresed on a 1% agarose gel to resolve completely replicated and un-replicated material (bottom figure). Equal amount of unreplicated plasmid DNA was digested with *Dpn*I and stained with Ethidium Bromide to show the complete digestion of unreplicated plasmid DNA (Bottom panel). Panel I, ethidium bromide stained, Panel II, autorad: marker (lane 1); *Dpn*I undigested plasmid DNA (lane 2); *Dpn*I digested unreplicated plasmid DNA (lane 3-4); replicated plasmid DNA, *Dpn*I resistant (lane 5). No difference in *Dpn*I resistant material is observed between replication in the presence or absence of NA, in all the three templates tested (panel III, IV, V). Blots have been cropped and the corresponding full blots are available in the Source Data files.



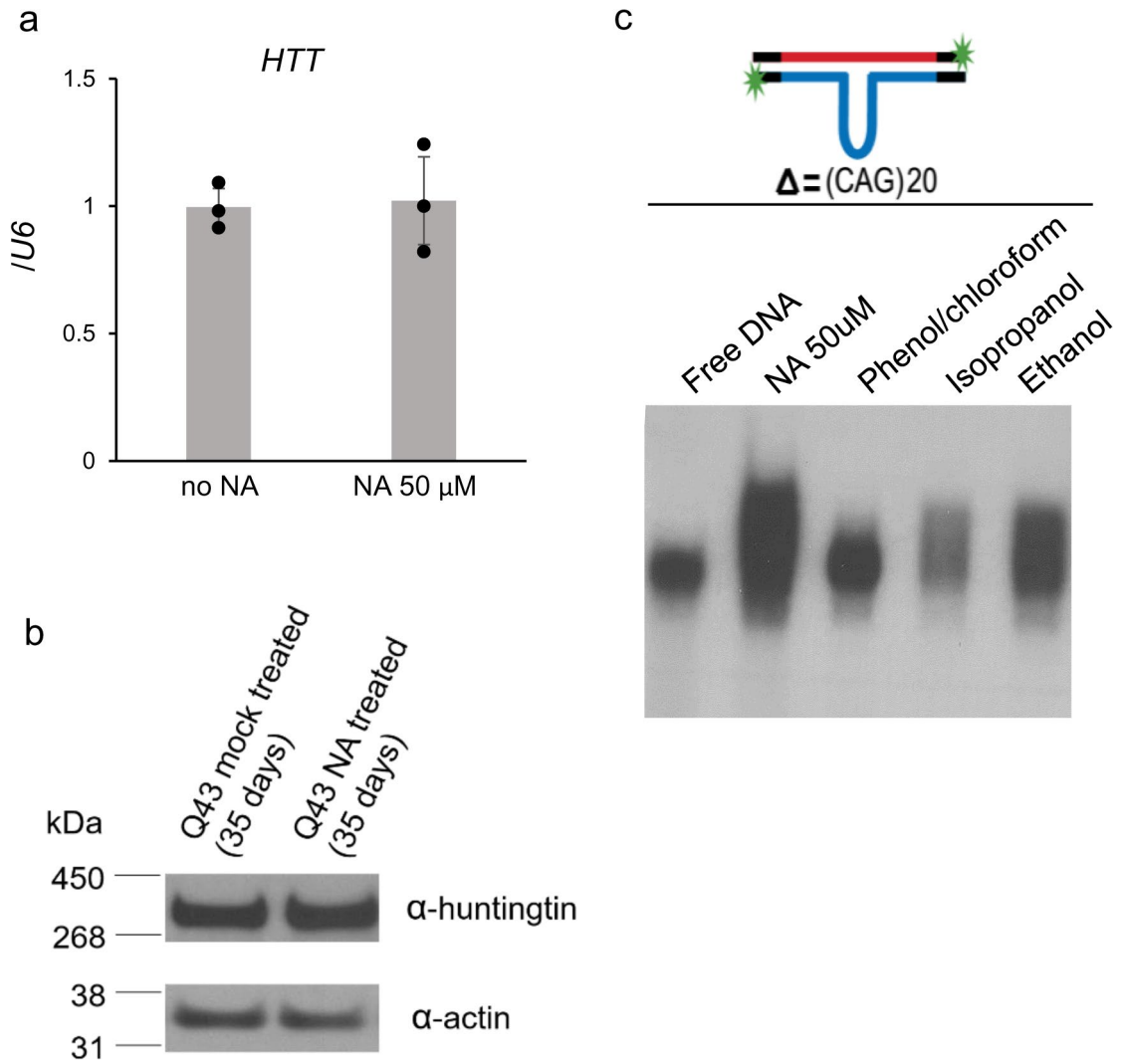

**Extended Data Fig. 2 |** See next page for caption.

**Extended Data Fig. 2 | NA does not affect non-mutant genetically stable repeats. a**, **b**, Representative data showing small-pool PCR (spPCR) for the non-expanded CAG/CTG repeat length of *CASK* and *Mdf15* in HD primary fibroblast cells (**a**) and spPCR for the non-expanded CAG/CTG repeat length of *CASK*, *Mdf15, ATXN8* and the non-expanded *HTT* allele genes in HT1080-(CAG)850 cells (**b**). Even under the sensitive mutation detection capacity of spPCR, length variation was not observed in either NA treated- and untreated-cells. Notably, some reactions did not show any product as is typical of the low genomic DNA template dilutions used in spPCR. **c**, The repeat-tract lengths of the *CASK, ATXN8*, and *Mfd15* loci in HT1080-(CAG)850 cells (initial clone and cells after 30 days incubation with or without NA). Length variation was not observed at any of these repeats of normal length loci in HT1080-(CAG)850 cells (after 30 days incubation with or without NA). Three independent experiments were performed. **d**, spPCR for the non-expanded CAG tracts in TBP alleles in HD patient fibroblasts treated with or without NA for 40 days. **e**, spPCR for the non-expanded CAG tracts in TBP alleles in HD R6/2 mouse striatum with four injections of NA or saline. **f**, Microsatellite instability assay. Assay scores >1.3 indicate increased MSI relative to a control sample set from peripheral blood leukocytes. Both NA positive and NA negative HD cells with (CAG)43 or (CAG)180 scored <1.3, indicating no effect of treatment on MSI. Eight known CMMRD-negative controls and 3 known CMMRD-positive controls were included in the assay.

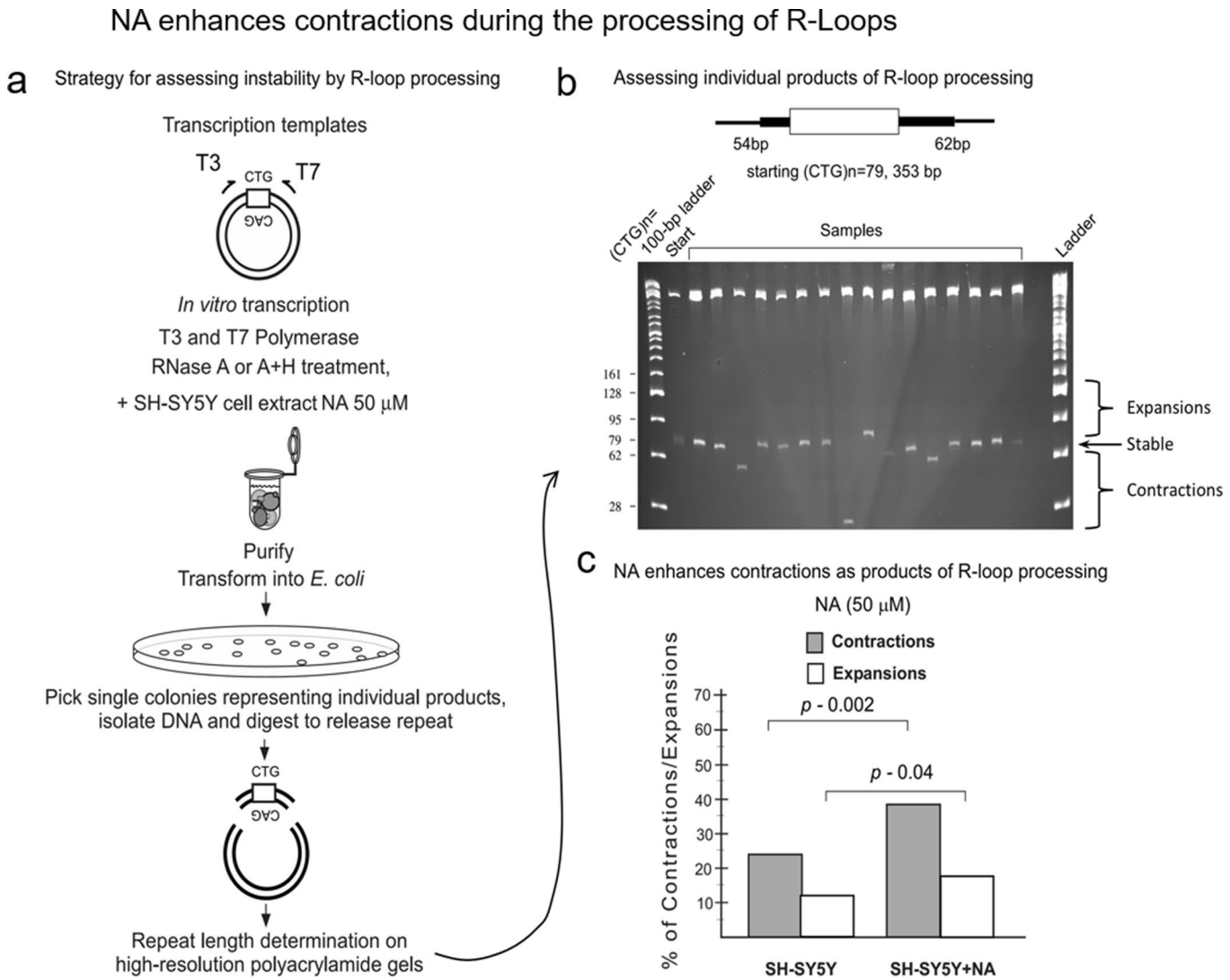



**Extended Data Fig. 3 | See next page for caption.**

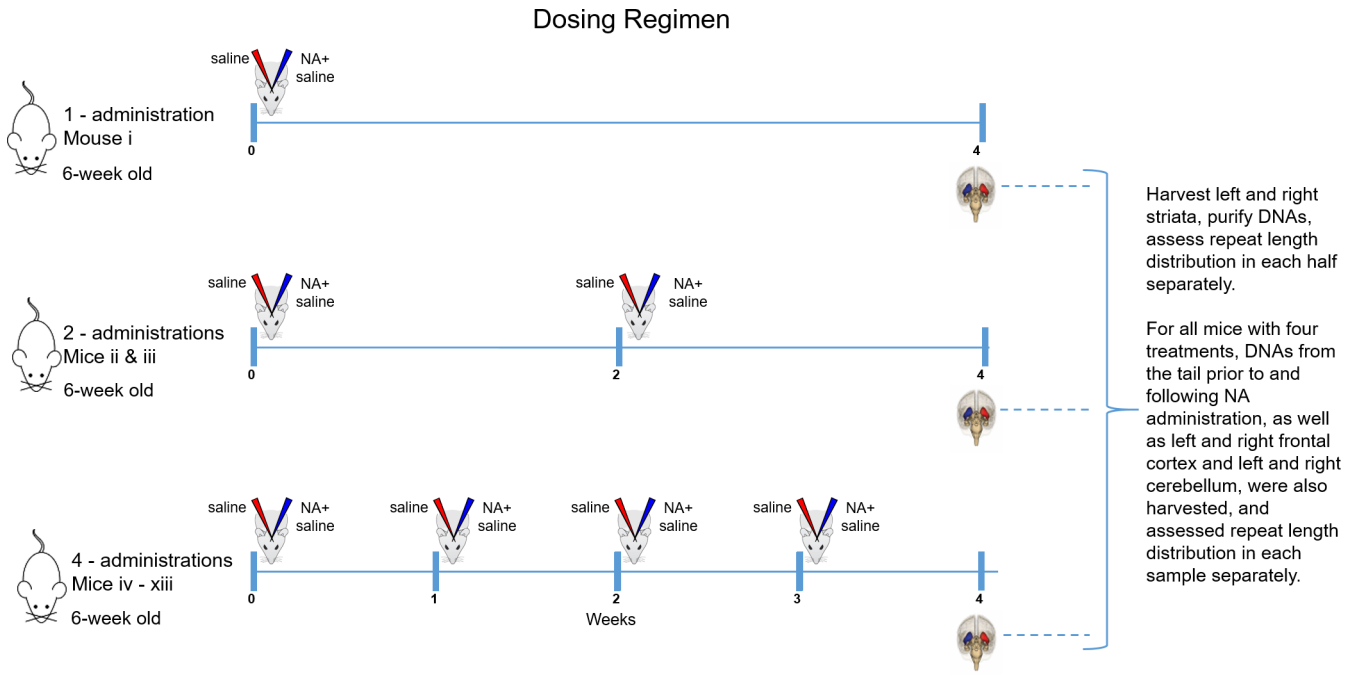
**Extended Data Fig. 3 | NA is not a general mutagen.** Towards assessing whether NA-treatment acted as a general mutagen to sequences other than CAG slip-outs, we harnessed the high read accuracy and depth of single molecule, real time, circular consensus sequencing (SMRT-CCS). Single-molecule sequencing was done on the *HPRT1* gene – widely used as a surrogate indicator of the global effect of induced genetic variation. For each replicate, we calculated the relative mutation rate between NA- and saline-treated cells as the mutation rate for NA-treated cells minus the rate for saline-treated cells and identified excess mutation rates based on an absolute relative rate >0.5%. **a**, Schematic of *HPRT1* sequencing for mutation detection. Briefly, cells were grown under identical conditions differing only by the addition of NA (50 μM) or saline, DNAs were isolated, *HPRT1* exons 2 and 3 PCR amplified and sequenced. **b**, Quality control for our analysis. **c,d**, Comparison of sequence variations between NA-treated and saline treated is presented. We chose to compare the single-molecule sequence reads of individual X chromosome-linked *HPRT1* alleles (exons 2 and 3) from our male HD patient-derived cells (**c**), and our male R6/2 mice (**d**), that had been NA- or saline-treated. Each read represents a single cell (Supplementary Note). Graphs show the distribution of sequence variants by relative mutation rate between three experimental replicates of NA-treated and saline-treated cells sequenced with PacBio single-molecule long reads.



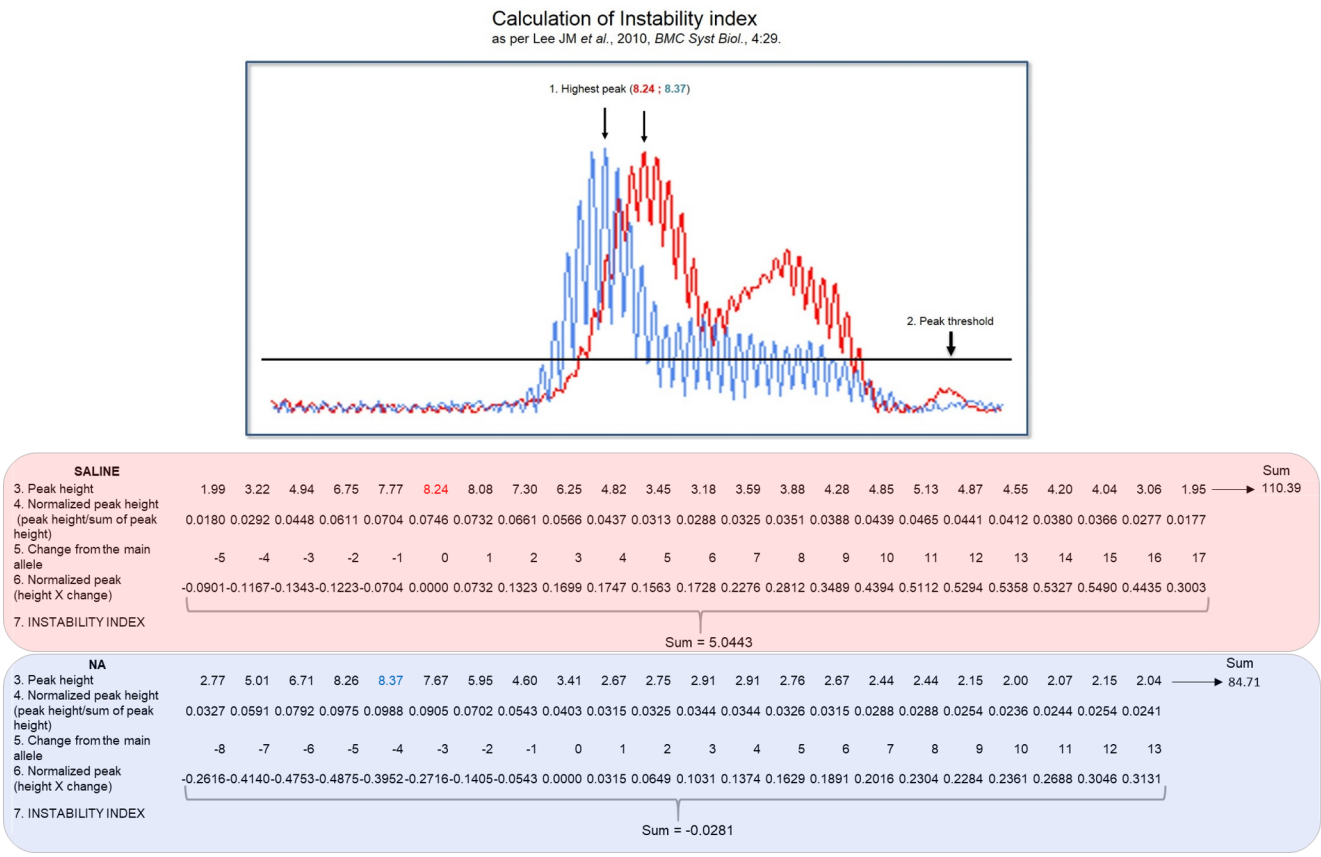

**Extended Data Fig. 4 | NA does not affect HTT transcription or translation. a**, NA does not affect transcription across expanded repeats in HTT in HD patient cells, determined by quantitative real-time reverse transcriptase (qRT)-PCR and normalized to U6 RNA. Data are indicated as the mean ± s.d. of independent triplicates. **b**, Western blot showing that NA does not affect HTT translation in HD patient cells with (CAG)43. Western blots were repeated 4 times with similar results. Blots have been cropped and the corresponding full blots are available in the Source Data files. **c**, Extraction of NA from DNA by solvents.Source data



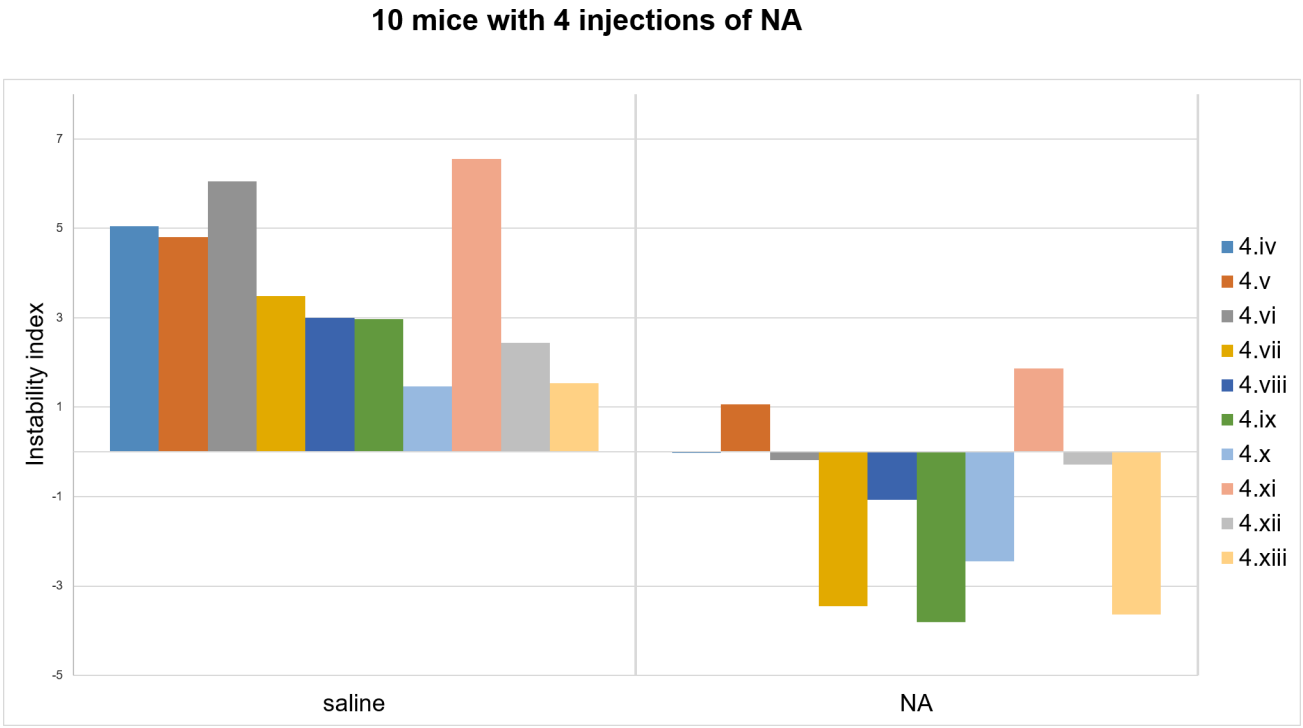

**Extended Data Fig. 5 | NA induces contractions during R-loops processing. a**, Schematic of R-loop formation, processing, and analysis. Pre-formed double-R-loops were processed by terminally differentiated (retinoic acid) human neuron-like cell extracts (SH-SY5Y) in the absence or presence of NA (50 μM), as described and DNA repeat lengths scored as expansions, contractions, or stable, by the STRIP assay (Methods). **b**, Representative example of STRIP analysis. Transcription products were isolated, processed and transformed in E. coli cells, previously shown to stably maintain the (CAG)79•(CTG)79 lengths (Methods). Plasmids isolated from individual bacterial colonies were digested with restriction enzymes to release the repeat containing fragment, resolved on 4% polyacrylamide gels and scored for instability. **c**, Graphical analysis of STRIP results. Two-sided $\chi 2$ test was performed to compare 191 untreated colonies vs. 100 NA-treated colonies.



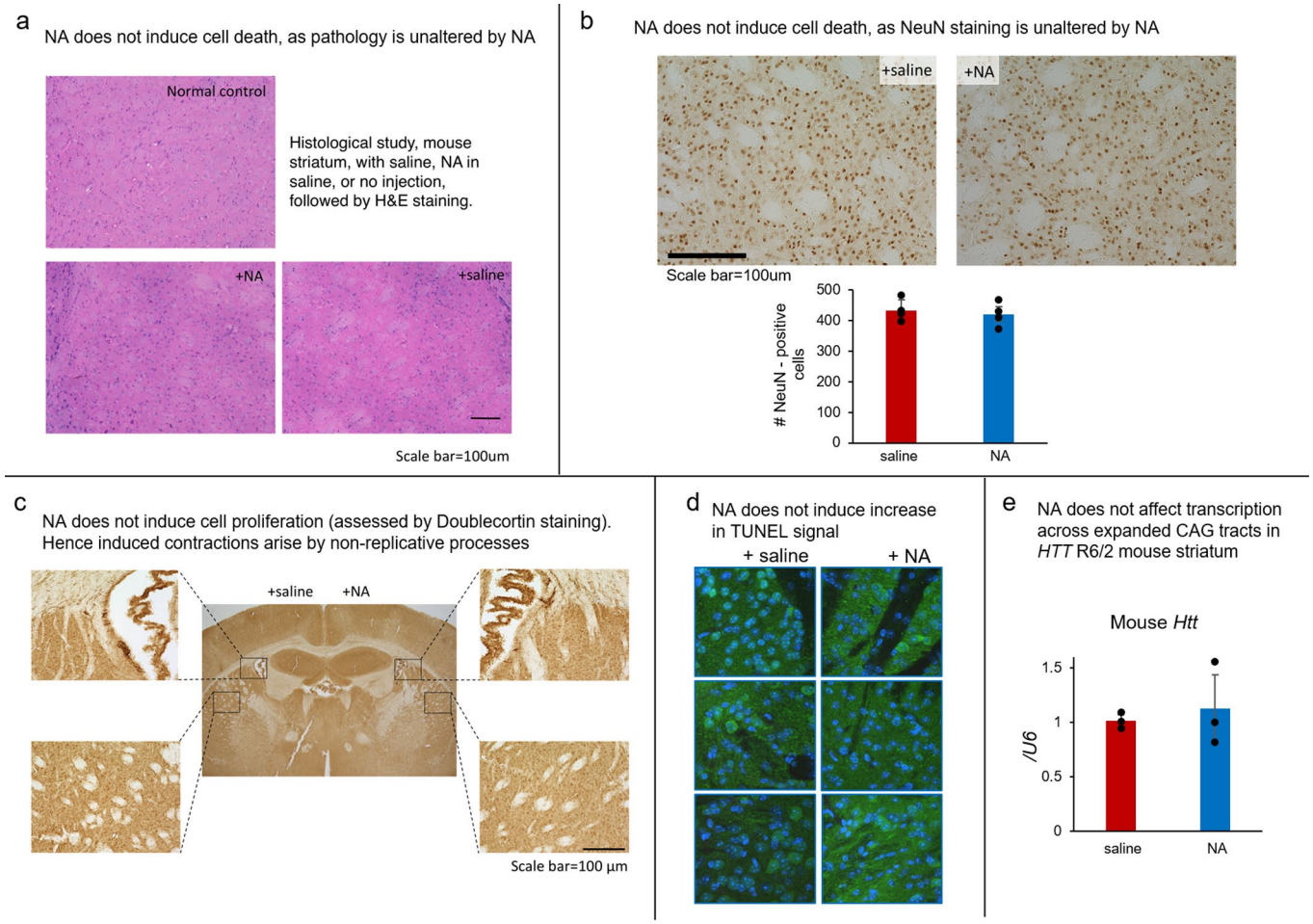

**Extended Data Fig. 6 | Dosing regimen.** A single drug administration involved six separate stereotactic injections (three injections of drug in saline or saline into three different striatal regions of either the left or right striatum, respectively). At the onset mice were 6-weeks old.




**Extended Data Fig. 7 | Instability Index calculation.** Instability index determination was as described[77,78], using a relative peak height threshold, with modifications. To quantify the levels of instability from GeneMapper traces peak height was used to determine a relative threshold of 20% based upon the main peak in the shorter mode of the control striatum (see points 1 & 2 in the figure). We used a conservative threshold factor (20%) as this detects peaks with good signal intensity and is more resistant to amplification variation than lower thresholds. Lower thresholds (10%, 5%) can provide more sensitive quantification. Peaks falling below his threshold were excluded from analysis. Peak heights were scored (see point 3) and normalized to the total of all peak heights in a given scan (see point 4). Since we are comparing the effect of NA versus saline upon instability in the striatum, the CAG length distribution in tail is not a factor in this comparison, but is for determining absolute instability, as in previous studies[62,66]. So as to facilitate comparison between NA and saline-treated striatum, these were normalized by multiplying the values by the change in CAG length of each peak relative to the highest peak in saline-treated striatum (see point 5), as opposed to the highest peak in the tail, as previously done[77,78]. These normalized values (see point 6) were summed to generate the instability index (see point 7). Striatum analysis for mouse vi is shown as an example R6/2, 6-weeks treated with four injections spanning 4 weeks of saline (red) or NA (blue). Peaks of the main allele in the saline-treated striatum, NA-treated striatum and tail of the same mouse, are indicated by triangle-brackets at the top (see point 1).




**Extended Data Fig. 8 | A total of ten HD mice revealed consistent NA-induced contractions of expanded CAG repeats.** Instability Indices in striatum of ten mice (iv-xiii) treated four times with saline in the right striatum and NA in the left striatum. Indices in NA-treated striatum were significantly different from the control saline-treated striatum (Mann-Whitney, $P = 0.00035$). Instability Indices for mouse v and xi are positive for both NA and saline as there are less data points to the left of the highest peak compared to the points to the right. Still, after NA treatment there is a reduction in the index.




**Extended Data Fig. 9 | NA does not induce cell death in the CNS and cell proliferation, and does not affect transcription across the *Htt* locus.** **a**, Histological study, mouse striatum with saline, NA in saline, or no injection, followed by H&E staining. Three independent experiments were performed. **b**, NeuN staining showing that NA does not induce cell death. Quantification of NeuN positive cells below. Data are indicated as mean ± s.d. of triplicates. **c**, Doublecortin staining showing that NA does not induce cell proliferation. Three independent experiments were performed. **d**, The effect of NA on TUNEL signal as assessed via fluorescent microscopy and immunohistochemistry. Representative 40x magnification confocal images of striatal medium spiny neurons (MSNs) of R6/2 mice treated with saline (right striata) and 50 µM NA (left striata) stained for TUNEL (red, staining apoptotic cells), and DARPP32 (green, staining MSNs). Panel locations (i-vi) correspond to the locations outlined in Fig. 7 (middle panels). **e**, NA does not affect transcription across expanded repeats in *HTT* in HD patient cells and mouse striatum, determined by quantitative real-time reverse transcriptase (qRT)-PCR and normalized to U6 RNA, expressed as the ratio of NA-treated vs. PBS-treated R6/2 striatum. Data are indicated as mean ± s.d. of independent triplicates.

nature research

Corresponding author(s): Christopher E Perason

Last updated by author(s): Nov 26, 2019

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☐ | ☒ | A description of all covariates tested |
| ☒ | ☐ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☒ | ☐ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | No software used for Data Collection |
| Data analysis | WGS was performed using established protocols on Illumina instruments and paired-end FASTQ files were aligned to the human genome (hg19/GRCh37) using BWA-MEM (v.0.7.8), with Picard MarkDuplicates (v.1.108) being used to mark PCR duplicates. Indel realignment and base quality scores were recalibrated using the Genome Analysis Toolkit (v.2.8.1). Somatic mutations were detected using MuTect2 (part of GATK v.3.5). |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw sequencing data has been deposited at the Sequence Read Archive (SRA) under accession number SUB6615083.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | Sample size was chosen based on preliminary experimental results and previous published analyses for similar experiments (Methods section). |
| Data exclusions | No data were excluded from the study |
| Replication | A minimum of 3 independent experiments was done for each experiment (indicated in the Methods section for each figure), and the required statistical test was performed. All the replicate were successful. For repeat length analysis >50 individual small-pool PCR reactions across the CAG repeat tract were performed. Results were analyzed using χ2-tests. For HPRT gene SMRT-sequencing, an average of 1,343 consensus sequences per replicate/treatment pair was analyzed. Mutation rate was evaluated using a two-sample Kolmogorov-Smirnov test |
| Randomization | NO randomization was necessary for the experiments described in the manuscript. All the experiments have internal controls. |
| Blinding | Blinding was applied for repeat length analysis (fragment analysis) and HTT foci analysis. When performing the experiment, the researcher did not know which side of the striatum (treated or untreated) the samples came from. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☐ | ☒ Antibodies |
| ☐ | ☒ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology |
| ☐ | ☒ Animals and other organisms |
| ☒ | ☐ Human research participants |
| ☒ | ☐ Clinical data |

## Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

## Antibodies

| | |
|---|---|
| Antibodies used | Anti-Huntingtin Protein Antibody a.a. 181-810 Clone 1HU-4C8 (Millipore Sigma, Cat #MAB2166), Anti-Actin Protein Antibody (BD Transduction Laboratory, Cat #612657). Secondary antibodies: Peroxidase-AffiniPure Sheep Anti-Mouse IgG H+L (Cedarlane Labs, Cat #515-035-062).<br>mouse anti-NeuN antibody (Millipore, Cat # MAB377), rabbit anti-doublecortin antibody (Abcam, Cat # ab18723).<br>For IF:<br>α-mutant HTT: clone mEM48, 1:250 dilution, mouse monoclonal, Millipore-Sigma, cat number: MAB5374<br>α-DARPP-32: clone 19A3, 1:250 dilution, rabbit monoclonal, New England Biolabs, cat number: 2306S |
| Validation | For validation details please refer to the manufacturer's website:<br>-Anti-Huntingtin: http://www.emdmillipore.com/CA/en/product/Anti-Huntingtin-Protein-Antibody-a.a.-181-810-clone-1HU-4C8, MM_NF-MAB2166#overview<br>- Anti-actin: http://www.bdbiosciences.com/ca/applications/research/stem-cell-research/mesoderm-markers/human/purifiedmouse-anti-actin-ab-5-c4actin/p/612657<br>-Anti NeuN: http://www.merckmillipore.com/JP/ja/product/Anti-NeuN-Antibody-clone-A60,MM_NF-MAB377<br>-Anti doublecortin: https://www.abcam.co.jp/doublecortin-antibody-ab18723.html<br>- α-mutant HTT: https://www.sigmaaldrich.com/catalog/product/mm/mab5374?lang=en&region=CA<br>- α-DARPP-32: https://www.neb.ca/detail.php?id=2306 |

## Eukaryotic cell lines

Policy information about cell lines

| | |
|---|---|
| Cell line source(s) | HT1080 and Sf9 cells were purchased from ATCC. HD primary fibroblast cell lines GM09197 and GM02191 were purchased from Coriell Biorepository. The human SH-SY5Y neuroblastoma cells were purchased from the American Type Culture Collections and cultured in recommended medium. HeLa cells were from National Cell Culture Center, National Centre for |

| | |
|---|---|
| | Resources, National Institute of Health. |
| Authentication | All the cell lines were authenticated by the respective Biobank |
| Mycoplasma contamination | All cell lines were tested for micoplasma contamination and were proved to be free of mycoplasma |
| Commonly misidentified lines (See ICLAC register) | The cell lines used are not in ICLAC database |

## Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

| | |
|---|---|
| Laboratory animals | In this study we used R6/2 mice (B6CBA-Tg(HDexon1)62Gpb/1J, Jackson Laboratory, cat# 002810), 6-week-old, males |
| Wild animals | No wild animals used in the study |
| Field-collected samples | No field-collected samples used in the study |
| Ethics oversight | Mouse handling and experimental procedures were conducted in accordance with the Osaka University guidelines for the welfare of animals. |

Note that full information on the approval of the study protocol must also be provided in the manuscript.