Somatic CAG expansion in Huntington's disease is dependent on the MLH3 endonuclease domain, which can be excluded via *MLH3* splice redirection to suppress expansion

Jennie C. L. Roy[1#], Antonia Vitalo[2,3#], Marissa A. Andrew[2], Eduarda Mota-Silva[2], Marina Kovalenko[2], Zoe Burch[2], Anh M. Nhu[2], Paula E. Cohen[4,5], Ed Grabczyk[1], Vanessa C. Wheeler[2,3*], Ricardo Mouro Pinto[2,3,6*]

[1] Department of Genetics, Louisiana State University Health Sciences Center, New Orleans, LA, USA

[2] Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA.

[3] Department of Neurology, Harvard Medical School, Boston, MA, USA.

[4] Department of Biomedical Sciences, Cornell University, Ithaca, NY, USA.

[5] Center for Reproductive Genomics, Cornell University, Ithaca, NY, USA.

[6] Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA

# equal contribution

* Corresponding authors:
Ricardo Mouro Pinto
Center for Genomic Medicine
Simches Research Building, 6.102
185 Cambridge St.
Boston, MA 02114
USA
Tel: 617-309-6683
Email: rmouropinto@mgh.harvard.edu

Vanessa C. Wheeler
Center for Genomic Medicine
Simches Research Building, 5.808
185 Cambridge St.
Boston, MA 02114
USA
Tel: 617-643-3103
Email: Wheeler@helix.mgh.harvard.edu

Abstract

Somatic expansion of the CAG repeat tract that causes Huntington's disease (HD) is thought to contribute to the rate of disease pathogenesis. Therefore, factors influencing repeat expansion are potential therapeutic targets. Genes in the DNA mismatch repair pathway are critical drivers of somatic expansion in HD mouse models. Here, we have tested, using genetic and pharmacological approaches, the role of the endonuclease domain of the mismatch repair protein MLH3 in somatic CAG expansion in HD mice and patient cells. A point mutation in the MLH3 endonuclease domain completely eliminated CAG expansion in the brain and peripheral tissues of a HD knock-in mouse model (Htt^{Q111}). To test whether the MLH3 endonuclease could be manipulated pharmacologically, we delivered splice switching oligonucleotides in mice to redirect *Mlh3* splicing to exclude the endonuclease domain. Splice redirection to an isoform lacking the endonuclease domain was associated with reduced CAG expansion. Finally, CAG expansion in HD patient-derived primary fibroblasts was also significantly reduced by redirecting *MLH3* splicing to the endogenous endonuclease domain-lacking isoform. These data indicate the potential of targeting the MLH3 endonuclease domain to slow somatic CAG repeat expansion in HD, a therapeutic strategy that may be applicable across multiple repeat expansion disorders.

Introduction

Huntington's disease (HD) is a dominantly inherited neurodegenerative disorder caused by the inheritance of an expanded CAG trinucleotide repeat in the huntingtin gene (*HTT*), encoding an extended polyglutamine tract in the huntingtin protein (1). The expanded CAG mutation ultimately results in neuronal dysfunction and death (2), via mechanisms that are as yet unclear (3). The mutant CAG repeat undergoes further expansion in somatic cells in a CAG length-dependent and tissue- or cell-type-specific manner (4–8). Somatic expansion is recapitulated in HD mouse models that also show progressive repeat tract lengthening over time (9–12), with the striatum and liver showing high levels of expansion, particularly the medium spiny neurons and hepatocytes, respectively (9, 10). Studies of postmortem HD brains revealed that longer somatic CAG expansions are associated with an earlier age of disease onset (7), with more recent studies showing correlations between age of onset or disease progression and the degree of somatic expansion in HD patient blood (13). Significantly, recent genome-wide association studies (GWAS) of over 9,000 HD patients have demonstrated that the rate of disease onset is determined by the length of the pure CAG repeat, rather than the length of the glutamine tract in huntingtin (14). This is supported by additional studies (13, 15) and provides compelling support for a model in which the rate of somatic CAG expansion drives the rate of disease onset. Thus, therapies targeting somatic repeat expansions have the potential to delay disease onset.

Genetic knockout studies in HD mouse models have shown that *Msh2*, *Msh3*, *Mlh1*, and *Mlh3* genes in the DNA mismatch repair (MMR) pathway are required for somatic expansion (16–19). Human GWAS have identified several genes in this pathway (*MSH3, MLH1, PMS1, PMS2*) to be associated with age at motor onset (14, 20), with *MSH3* also showing association with a measure of HD progression (21). The mouse ortholog of an additional HD onset-modifier gene, *FAN1* (14, 20), with no known function in the MMR pathway, suppressed somatic expansion in HD knock-in mice in a manner that was dependent on *Mlh1* (8). Genetic variation in some human onset or progression modifier genes (*MSH3, MLH1, MLH3, FAN1*) also modified CAG expansion in HD patient blood or blood-derived cells (13, 14, 22). Together, these data indicate that HD pathogenesis is modulated by naturally occurring variation in genes that act by altering the rate of somatic CAG expansion, and that genes driving this process are potential therapeutic targets for HD. MMR pathway genes also modify somatic expansion in mouse models and in

mouse or human cell-based models of a number of other DNA repeat expansion disorders, including myotonic dystrophy type I (DM1) (23–25), Friedreich ataxia (FRDA) (26–29), and the fragile X-related disorders (FXD) (30–34). In addition, polymorphisms in two DNA repair genes (*FAN1, PMS2*) associated with HD onset were also associated with disease onset in the spinocerebellar ataxias (35). Thus, therapeutic strategies targeting MMR genes may be applicable to multiple repeat expansion diseases.

Our previous genetic knockout studies in *Htt*^{Q111} knock-in mice demonstrated that MLH3 is an essential component of the CAG repeat expansion process (17). MLH3 is known to play an important role in meiosis, but a relatively minor role in canonical MMR (36–38). MLH3 contains an evolutionary conserved endonuclease domain (39) that we postulated may be important in its CAG expansion-promoting role in HD (17), with inactivation or elimination of this domain hypothesized to eliminate or reduce somatic expansions. Interestingly, human MLH3 exists primarily as two protein isoforms: isoform 1 (UniProt Q9UHC1-1) is specified by splice variant 1 (CCDS32123) containing exon 7, which encodes the endonuclease domain, while isoform 2 (UniProt Q9UHC1-2) is specified by splice variant 2 (CCDS9837), lacking exon 7, and therefore has no endonuclease domain (28, 40). *MLH3* splice redirection from predominantly splice variant 1 to predominantly splice variant 2 can be achieved with splice switching oligonucleotides (SSOs) that mask the acceptor and donor sites that surround exon 7 of the *MLH3* pre-mRNA (28). In a previous FRDA study, splice redirection of *MLH3* to exclude the endonuclease domain resulted in reduced rates of GAA expansion in a human cell-based model and in patient cells (28). This indicated that the MLH3 endonuclease domain was important for GAA repeat expansion and that splice redirection could be used to reduce its contribution to the expansion process.

Here we have tested the role of the MLH3 endonuclease domain in somatic *HTT* CAG expansion, using both genetic studies and a pharmacological approach based on splice redirection. We demonstrate that: 1) a single point mutation in the MLH3 endonuclease domain, predicted to abrogate endonuclease activity, eliminates somatic CAG expansion in *Htt*^{Q111} mice; 2) systemic delivery of SSOs to redirect *Mlh3* splicing to exclude the endonuclease domain suppresses somatic CAG expansion in *Htt*^{Q111} mice; and 3) *MLH3* splice redirection with SSOs

also effectively reduces CAG expansion in HD patient-derived primary fibroblast cells, in the absence of replication. These data highlight the therapeutic potential of targeting the MLH3 endonuclease domain to slow *HTT* CAG repeat expansion, a strategy that may be applicable across multiple repeat expansion disorders.

Materials and Methods

Mice

Htt^{Q111} HD knock-in mice (formerly *Hdh*^{Q111}) (41) were maintained on a C57BL/6J background (9) by breeding heterozygous males to C57BL/6J wild-type females from The Jackson Laboratory (Bar Harbor, ME). *Mlh3*^{DN} mice (42) were also maintained on C57BL/6J background. *Htt*$^{Q111/+}$ and *Mlh3*$^{WT/DN}$ mice were crossed together to generate *Htt*$^{Q111/+}$ *Mlh3*$^{WT/DN}$ and *Htt*$^{+/+}$ *Mlh3*$^{WT/DN}$ mice, which were subsequently intercrossed to generate heterozygous *Htt*$^{Q111/+}$ mice (hereafter referred to as *Htt*^{Q111}, for simplicity) that were either wild type (*Mlh3*$^{WT/WT}$), heterozygous (*Mlh3*$^{WT/DN}$), or homozygous (*Mlh3*$^{DN/DN}$) for the *Mlh3*^{DN} point mutation. Genotyping of the *Htt*^{Q111} and *Mlh3*^{DN} alleles in genomic DNA extracted from tail or ear biopsies at weaning, or in adult tissues for genotype confirmation, was carried out as previously described (16, 42). All animal procedures were carried out to minimize pain and discomfort, under approved IACUC protocols of the Massachusetts General Hospital. Animal husbandry was performed under controlled temperature and light/dark cycles.

HTT CAG repeat instability analysis

Genomic DNA was isolated from mouse tissues and patient fibroblast cells as previously described (17, 28). Analyses of *HTT* CAG repeat size in both *Htt*^{Q111} mice and in patient fibroblasts was performed by PCR using human-specific *HTT* primers and *Taq* PCR Core Kit with Q solution (Qiagen), as previously described (17, 43). The resulting FAM-labelled PCR products, encompassing the *HTT* CAG repeat, were resolved on the ABI3730xl automated DNA analyzer (Applied Biosystems), with either GeneScan 500-LIZ or 1200-LIZ internal size standards, and analyzed with GeneMapper v5 (Applied Biosystems). When necessary, to increase the PCR signal strength, up to three PCR replicates were combined and concentrated with MinElute PCR Purification Kit (Qiagen). CAG expansion indices were quantified in *Htt*^{Q111} mice from GeneMapper peak height data, taking into account only expansion peaks (*i.e*. larger

than main allele) and using a 5% peak height threshold, as described previously (44). In the HD patient fibroblasts, CAG expansion was reflected in a shift of the entire CAG length distribution to greater repeat lengths over time, without any discernible impact on the shape of the distribution. To quantify repeat expansion, we determined the average gain of CAG units by calculating the difference in average CAG length before (t = 0, n = 4) and after the 6-week culture period (t = 6 weeks, n = 10). To determine average CAG length in an individual sample, we first applied a 20% peak height threshold to the distribution (i.e. only peaks with height ≥ 20% of tallest peak were used), then multiplied the frequency of each peak (peak height divided by the sum of all peak heights) by the corresponding CAG size, and finally summed these values.

Treatment of *Htt*^{Q111} mice with *Mlh3* SSOs

Mouse-specific SSOs were designed to target the pre-mRNA of *Mlh3* exon 6 (Chr12:85,250,302-85,250,373; GRCm38/mm10): mMLH3ac5 (acceptor site) 5'-TCCACCTACAAAATAATCCAGGATT-3' and mMLH3dr7 (donor site) 5'-AACTACAGACAGATACTTACCAGTA-3'. Oligonucleotides were obtained as "Vivo-Morpholinos" from Gene Tools, LLC (45). *Htt*^{Q111} mice (CAG 104-109; two males and two females per group) were treated with either 25 mg/kg SSOs (mMLH3ac5 and mMLH3dr7; 1:1 ratio), or PBS, by single tail vein injection of 200 µl. Treatment was initiated at 4 weeks of age and injections were performed three times per week (every 48-72hr). Weight measurements were taken at the beginning of each week to determine weekly weight-adjusted doses. Treatment was stopped at 12 weeks of age, with animal euthanasia and immediate dissection performed 24 hours after the final injection. Tissues were rapidly frozen using liquid nitrogen to preserve RNA integrity.

***MLH3* splicing analysis**

Frozen mouse tissues were mechanically homogenized prior to RNA extraction. Total RNA was extracted from *Htt*^{Q111} mouse tissues and HD patient cultured primary fibroblasts with TRIzol (Thermo Fisher Scientific) following the manufacturer's protocol. Total RNA (250 ng) was used to generate cDNA with the High Capacity cDNA Reverse Transcription Kit following the manufacturer's protocol (Thermo Fisher Scientific). Human *MLH3* splice variants were characterized by Reverse Transcription PCR (RT-PCR) as previously described (28). Mouse

Mlh3 splice variants were characterized by RT-PCR with primers mMLH3x3-3328F (5'-CCAGTGTTTGCTCGATACCC-3') and mMLH3x11-4028R (5'-GGACAAGGACAGAGCTTCGA-3), at 59°C annealing temperature, for 33 cycles. PCR products were resolved by electrophoresis on a 1.4% agarose gel containing 1 µg/ml ethidium bromide. Real-time RT-PCR (RT-qPCR) characterization of *Mlh3* splicing in mouse tissues was performed with TaqMan Gene Expression Assays (Applied Biosystems) following the manufacturer's protocol on the CFX96 Real-Time PCR Detection System (Bio-Rad). The assay used for the *Mlh3* splice variant 1 detects the splice junction between exon 6 and exon 7 (Mm00520819_m1, FAM-labeled probe). The assay for *Mlh3* splice variant 2, binds at the splice junction between exon 5 and exon 7 (custom design, FAM-labeled probe) and only binds when exon 6 is spliced out. Relative *Mlh3* splice variant levels were normalized to the endogenous mouse β-actin (*Actb*) reference gene (Mm00607939_s1, primer-limited, VIC-labeled probe) in duplex reactions.

Treatment of HD patient primary fibroblasts with *MLH3* SSOs

Primary fibroblasts from a juvenile HD patient (GM09197; CAG ~180/18) (46) were obtained from the Coriell Institute for Medical Research. Cells were maintained in Dulbecco's modified Eagle's medium (DMEM) high glucose (Thermo Fisher Scientific) with 10% fetal bovine serum (FBS; Sigma) at 37°C in an atmosphere containing 5% CO_2. To determine the ability of the SSOs to induce *MLH3* splice redirection in HD primary fibroblasts, confluent cell cultures of unaltered primary GM09197 cells were treated with a single dose of either 500 nM of *MLH3* SSOs (MLH3acr6 and MLH3dnr8, 1:1 ratio, 250 nM each) (28) or 500 nM of scrambled control oligonucleotide (5'-CCTCTTACCTCAGTTACAATTTATA-3'), alongside cells that were left untreated. Oligonucleotides were obtained as "Vivo-Morpholinos" from Gene Tools, LLC (45). Cells were treated for six hours in reduced-serum Opti-MEM media (Thermo Fisher Scientific). The media was then replaced with 20% FBS high glucose DMEM and cells were harvested for RNA isolation after 48 hours to analyze changes to *MLH3* mRNA splice variants. To test the long-term effect of *MLH3* splice redirection on *HTT* CAG expansion, GM09197 primary HD fibroblasts were first transduced with MSH3- and EGFP-expressing lentiviral particles (PNL-MSH3-IRES2EGFP construct), as previously described (27, 28), to induce CAG expansion. Cells were then grown to confluent cultures and treated twice weekly with 500 nM of *MLH3*

SSOs, as descried above, or left untreated, for a total of six weeks. Cells were then harvested for DNA for *HTT* CAG instability analysis. A total of four wells were harvested immediately prior to treatment ($t = 0$) and a total of ten independent wells were harvested at the end of treatment ($t = 6$ weeks).

Statistical analysis

Statistical analyses were performed using GraphPad Prism v.8. Comparison of *HTT* somatic CAG expansion indices between *Htt*^{Q111} mice with different *Mlh3*^{DN} genotypes and across different tissues was performed using two-way ANOVA, with Tukey multiple comparisons *post hoc* correction. Comparison of *HTT* somatic CAG expansion indices in *Htt*^{Q111} mice treated with *Mlh3* SSOs was performed using two-tailed unpaired *t* tests comparing *Mlh3* SSO-treated and PBS-treated mice for each tissue. Comparison of average *HTT* CAG gain in HD patient fibroblasts treated with *MLH3* SSOs was performed using one-way ANOVA ("$t = 0$" vs "*MLH3* SSOs, $t = 6$wks" vs "untreated, $t = 6$wks"), with Tukey multiple comparisons *post hoc* correction. Not significant, $p > 0.05$; *, $p \leq 0.05$; **, $p \leq 0.01$; ***, $p \leq 0.001$; ****, $p \leq 0.0001$.

Results

Mutation in the MLH3 endonuclease domain prevents HD CAG expansion in *Htt*^{Q111} mice

We previously demonstrated that MLH3 is required for somatic CAG repeat expansion in *Htt*^{Q111} knock-in mice (17). To test the hypothesis that MLH3 endonuclease activity is involved in repeat expansion, we used an *Mlh3* mutant mouse line (*Mlh3*^{D1185N}, abbreviated to *Mlh3*^{DN}) harboring a point mutation in the endonuclease domain (42). In these mice, GAC is changed to AAC in the genomic sequence to replace the aspartic acid "D" at amino acid 1185 with an asparagine "N" within the conserved DQHA$(X)_2$E$(X)_4$E motif (DQHAAHERIRLE in mouse MLH3) (**Figure 1A**). An analogous mutation in *S. cerevisiae* or human MLH3 eliminates its endonuclease activity (47–50). Importantly, this D to N substitution in purified mouse, human, and *S. cerevisiae* proteins did not alter the stability of MLH3 or its interaction with MLH1 (42, 50, 51). In addition, in the mouse there was no impact on the number or localization of MLH3 or MLH1 foci in spermatocytes or on the recruitment of factors required to initiate crossovers in meiosis (42). Molecular modeling of mouse MLH3 was consistent with the lack of impact of the D1185N substitution on the structure or stability of MLH3 (30). Therefore, the *Mlh3*^{DN} mutation in the

mouse is predicted to disrupt MLH3 endonuclease function in the absence of significant effects on protein expression, stability, or interactions.

To determine the effect of *Mlh3*^{DN} allele on somatic *HTT* CAG expansion, we crossed *Mlh3*^{DN} mice (42) with *Htt*^{Q111} HD knock-in mice (41) to generate heterozygous *Htt*^{Q111} mice that were either wild type (*Mlh3*$^{WT/WT}$), heterozygous (*Mlh3*$^{WT/DN}$), or homozygous (*Mlh3*$^{DN/DN}$) for the *Mlh3* D1185N point mutation. Analysis of CAG expansions in striatum, liver, and kidney – tissues previously shown to exhibit high levels of instability in *Htt*^{Q111} mice (9, 17) – revealed the absence of somatic expansions in *Mlh3*$^{DN/DN}$ mice, both at three (Error! Reference source not found.) and six months of age (**Figure 1B, C)**. Notably, the effect of the homozygous *Mlh3*$^{DN/DN}$ mutation on CAG expansion was indistinguishable from that of a homozygous *Mlh3* null mutation (17). Quantification of repeat expansions (44) showed statistically significant reductions in the CAG expansion index in all tissues analyzed (striatum, liver, and kidney) of *Mlh3*$^{DN/DN}$ mice compared to *Mlh3*$^{WT/WT}$ mice (**Figure 1B** and Error! Reference source not found.). There was a slight reduction in expansion index in the *Mlh3*$^{WT/DN}$ mice, consistent at both ages and in all tissues analyzed, though only reaching statistical significance in 6-month striata (**Figure 1B** and Error! Reference source not found.). The dramatic effect of this mutation, phenocopying the *Mlh3* null (17), indicates that an intact endonuclease domain is required for somatic *HTT* CAG repeat expansion.

In vivo Mlh3 **splice redirection suppresses CAG expansion in** ***Htt*^{Q111} **mice**

Having demonstrated that the MLH3 endonuclease domain was critical for somatic *HTT* CAG expansion in *Htt*^{Q111} mice, we were interested in testing the potential of an *Mlh3* splice redirection approach, previously used in FRDA cell-based models and patient cells (28), to slow CAG expansion in the same HD mice. Mice express predominantly one endogenous *Mlh3* mRNA splice variant in which exon 6, the homologous exon to human *MLH3* exon 7, encodes the endonuclease domain (**Figure 2A**). For this study, mouse-specific vivo-morpholino-based SSOs were designed to bind both the acceptor and donor splice sites of *Mlh3* exon 6, as previously reported for human *MLH3* (28). Starting at 4 weeks of age, heterozygous *Htt*^{Q111} mice (n = 4) were treated with 25 mg/kg of *Mlh3* SSOs (or PBS) by tail vein injection every 48-72 hours, until they reached 12 weeks of age (**Figure 2B**). *Htt*^{Q111} tissues were harvested 24 hours

after the final injection and *Mlh3* splice variants were characterized by RT-PCR analysis in liver and kidney (**Figure 2C)**, two peripheral tissues in which CAG expansion was shown to be dependent on an intact MLH3 endonuclease domain (**Figure 1B, C**) and which can be effectively targeted with systemic delivery of vivo-morpholinos (45). We analyzed *Mlh3* splice variants using RT-PCR, generating a 720 bp product for the exon 6-containing splice variant 1 (Var1) and a 648 bp product for *Mlh3* splice variant 2 (Var2), which lacks exon 6 (**Error! Reference source not found.**) and is the mouse analog of human *MLH3* splice variant 2. As expected, all four PBS-treated mice showed Var1 only in both kidney and liver, while all four *Mlh3* SSO-treated mice showed successful splice redirection favoring Var2 in the kidney, with much lower, yet still detectable, Var2 levels in the livers of two treated mice (**Figure 2C**). We also established a TaqMan-based real-time RT-PCR assay for quantitative analysis of *Mlh3* splice variants and increased sensitivity (Error! Reference source not found.). Using this more sensitive assay, only Var1 was detected in PBS-treated mice, whereas variable expression of Var2 was detected in all SSO-treated samples, with the global pattern of Var1 and Var2 expression in the two tissues consistent with levels detected in the gel-based RT-PCR assay (Error! Reference source not found.). The higher levels of Var2 in the kidney relative to the liver may reflect better cellular SSO uptake in the kidney than in the liver, or may be the result of SSOs being actively removed from circulation by the kidney for excretion, limiting SSO availability to other peripheral tissues. In summary, we found that *Mlh3* SSO-mediated splice redirection to exclude the exon encoding the endonuclease domain of MLH3 can be successfully achieved *in vivo* in a mouse model of HD.

To test whether such *Mlh3* splice redirection impacted somatic CAG expansion, we performed qualitative and quantitative analyses of CAG repeat length profiles from the same Htt^{Q111} mice ($n = 4$) treated with either *Mlh3* SSOs or PBS (**Figure 2D, E**). When compared to the PBS-treated mice, the *Mlh3* SSO-treated mice displayed significantly less somatic CAG expansion in the kidney ($p \leq 0.0001$), and to a lower extent in the liver ($p \leq 0.05$). These effects are consistent with the different degrees of splice redirection achieved in the two tissues. As expected, *Mlh3* SSO treatment had no effect on CAG expansion in the striatum, consistent with the inability of these SSOs to efficiently cross the blood brain barrier (45). As *Mlh3* splice redirection from Var1 to Var2 increases the proportion of mRNAs encoding an endonuclease-lacking MLH3 protein,

these data support our prior results in *Mlh3*^{DN} mice showing that an intact endonuclease domain is required for somatic *HTT* CAG expansion. Further, these results demonstrate that a pharmacological splice redirection strategy can be used *in vivo* to impact a function of MLH3 that is critical for somatic CAG expansion.

***MLH3* splice redirection suppresses CAG expansion in HD patient-derived fibroblasts**

Alternative splicing of human *MLH3* pre-mRNA results in endogenous expression of two predominant splice variants: variant 1 (Var1), containing the exon7-encoded endonuclease domain, and variant 2 (Var2), lacking this domain (**Figure 3A** and Error! Reference source not found.). We previously showed that *MLH3* splice redirection from Var1 to Var2 can be achieved in FRDA patient primary fibroblasts using a combination of SSOs targeting the exon 7 splice acceptor (MLH3acr6) and donor (MLH3dnr8) sites (28). Prolonged treatment with these *MLH3* SSOs resulted in suppression of GAA repeat expansion in FRDA cell-based models and patient cells (28). With the goal of testing this strategy in HD, we used primary fibroblasts derived from a juvenile HD patient (GM09197; CAG ~180/18) (46) with a long repeat tract, predicted to allow repeat expansion in cell culture within a reasonable time frame (27, 52). We first tested our ability to induce *MLH3* splice redirection following 48 hour treatment with 500 nM of *MLH3* SSOs (MLH3acr6 and MLH3dnr8, 1:1 ratio, 250nM each) compared to treatment with 500 nM of a control vivo-morpholino oligonucleotide with a scrambled sequence or no treatment (**Figure 3B**). RT-PCR analysis of the *MLH3* splice variants generates a 434 bp product for the full-length Var1 and a 362 bp product for Var2. RT-PCR analysis detects *MLH3* Var1 as the primary isoform expressed in GM09197 fibroblasts, while treatment with *MLH3* SSOs resulted in full redirection of splicing towards Var2 (**Figure 3B**). No impact on *MLH3* splicing was detected following treatment with scrambled control oligonucleotides, where splice variants were identical to untreated cells. We then determined whether splice redirection of *MLH3* was capable of slowing CAG expansion in these cells over time. In our studies on FRDA primary fibroblast cells, we found that the GAA repeat tract was stable due to low levels of expression of MSH3, one of the key factors that contribute to both GAA and CAG repeat expansion (27, 53). Therefore, we induced CAG expansion in the HD fibroblasts via ectopic expression of MSH3 using lentiviral vector delivery as previously described for FRDA fibroblasts (27, 28). These MSH3-expressing fibroblasts were then grown to confluent cultures to prevent replication,

before initiating SSO treatment. Prior to SSO treatment, genomic DNA was isolated from untreated cells (t = 0; n = 4) to determine the starting average CAG repeat size. Cells were then treated with 500 nM *MLH3* SSOs twice-weekly, or left untreated, for a total of six weeks (t = 6wks) (**Figure 3C**). Quantitative GeneMapper analysis revealed an overt shift of the entire population of CAG alleles towards larger lengths in untreated HD fibroblasts (t = 6wks), relative to the starting population (t = 0), reflected by an increase in average CAG lengths of 3.0 repeats ($p \leq 0.0001$) (**Figure 3D, E**). Treatment with *MLH3* SSOs for 6 weeks significantly suppressed CAG expansion, with cells showing only a modest and non-statistically significant increase in average CAG length of 0.6 repeats during this period ($p \leq 0.0001$, when compared with untreated cells, t = 6 weeks; p = 0.3120 when compared with t = 0) (**Figure 3D, E**). These data further support our prior results in *Mlh3*^{DN} and *Mlh3* SSO-treated *Htt*^{Q111} mice, suggesting that the MLH3 endonuclease domain is essential for somatic *HTT* CAG expansion in HD patient cells, and that the use of *MLH3* SSOs constitutes a viable and effective strategy to pharmacologically target this pathway in HD patients.

Discussion

Somatic expansion of the *HTT* CAG repeat is thought to drive the rate of HD pathogenesis. Repeat expansion is dependent on proteins in the MMR pathway and, therefore, these present potential therapeutic targets. Insight into specific functions or activities of MMR proteins that modify repeat expansion will provide possible opportunities for therapeutic intervention. Here, we provide evidence, using two distinct, yet complementary approaches, that the endonuclease domain of MLH3 is required for somatic *HTT* CAG expansion. First, a point mutation in a conserved endonuclease domain motif abolishes CAG expansion in *Htt*^{Q111} mouse tissues, having an effect indistinguishable from a *Mlh3* null mutation. Second, CAG expansion both in *Htt*^{Q111} mice and in HD patient-derived cells is reduced by modulating *Mlh3*/*MLH3* splicing in a manner that excludes the exon encoding the endonuclease motif.

The endonuclease domain of MLH3 is encompassed in the C-terminal domain of the protein, which interacts with MLH1 (54). As described above, whereas the D>N substitution in the first amino acid of the DQHA(X)2E(X4)E motif eliminates MLH3 endonuclease activity (47–50), there is no evidence that this mutation alters the stability of MLH3 or its interaction with MLH1

(30, 42, 50, 51). We have been unable to directly determine MLH3 protein levels in our study due to the lack of availability of suitable antibodies to detect MLH3, which is present at very low levels in most tissues (28, 38). More specifically, we could not detect MLH3 in mouse tissues using the same antibody used to detect MLH3 in mouse spermatocytes of *Mlh3*^{DN} mice, where this protein is relatively highly expressed (42). Importantly, we find that heterozygous or homozygous *Mlh3*^{DN} alleles have impacts on CAG expansion comparable to heterozygous or homozygous *Mlh3* null alleles, respectively (17). This indicates that the *Mlh3*^{DN} mutation would need to dramatically reduce MLH3 protein levels in order to have an impact on CAG expansion that phenocopies the null allele. Given existing data in the literature (30, 42, 50, 51), this scenario appears highly improbable. Further, the comparable effects of the *Mlh3*^{DN} and *Mlh3* null alleles argue against a dominant negative effect of the *Mlh3*^{DN} mutation. It is therefore likely that the absence of *HTT* CAG expansion in *Htt*^{Q111} mice harboring the *Mlh3* D1185N mutation is due to loss of MLH3-dependent endonuclease activity.

The impact of *Mlh3/MLH3* splice redirection that excludes the DQHA(X)2E(X4)E motif on *HTT* CAG expansion is also consistent with a requirement for MLH3-dependent endonuclease activity, although we cannot rule out the possibility that deletion of the endonuclease domain-encoding exon in MLH3 due to splice redirection has additional effects on MLH3 stability or function. Published data in this regard provide evidence that the human MLH3Δ7 isoform can interact with MLH1 in yeast 2-hybrid (54) and GST pull down experiments (55), and is recruited to repair foci in response to DNA damage in the same way as the full length MLH3 (56), implying normal protein interactions. However, one study found that MLH3Δ7 was unable to interact with MLH1 in a mammalian 2-hybrid assay (57). Although the endonuclease domain is encompassed in MLH3 and in the homologous PMS2 protein, endonuclease activity is dependent on the interaction of these proteins with MLH1 (MLH1-MLH3 = MutLγ dimer; MLH1-PMS2 = MutLα dimer), and structural analysis has revealed that MLH1 contributes an amino acid residue to the endonuclease active site of *S. cerevisiae* MutLα (58, 59). Thus, destabilizing this interaction would also be predicted to abrogate endonuclease activity. Overall, our data from orthogonal genetic and pharmacological-based approaches together provide strong support for a critical role of the MLH3 endonuclease in driving somatic expansion of the *HTT* CAG repeat.

As described previously, exclusion of *MLH3* exon 7 slowed GAA repeat expansion in FRDA cell models and in FRDA patient cells (28), and more recently, the same *Mlh3* D1185N point mutation was found to eliminate expansion of the *FXN* CGG repeat in a FXD mouse embryonic stem cell model (30). Together, data indicate that the MLH3 endonuclease domain is critical for somatic expansion of different trinucleotide repeat sequences, with implications for cross-disease applicability of therapeutic approaches that target this activity. Recent biochemical studies in cell-free systems have suggested a plausible mechanism by which MLH3 endonuclease activity might lead to repeat expansion (50). These studies provided support that MutLγ cleaves a closed circular plasmid harboring a CAG- or CTG-containing loop on the loop-lacking strand, with repair leading to loop incorporation, effectively corresponding to an expansion event. This observation distinguished the MutLγ endonuclease from the MutLα endonuclease, where MutLα was able to cleave on both the loop-containing and loop-lacking strands, potentially leading to both deletion and expansion events (60). This specific strand-directed activity of MutLγ may, in part, explain the necessity for MLH3 in the somatic expansion of several disease-associated repeats (17, 28, 31). This is in contrast to PMS2, whose knockout can fully (32) or partially (61) suppress expansion, or can promote expansion (26, 28), depending on the repeat sequence and/or biological system. Understanding the role played by PMS2 endonuclease in somatic repeat instability *in vivo* would also be of significant interest.

The fact that *MLH3* splice variants that both include or exclude the endonuclease domain-containing exon 7 occur naturally in humans indicates that the relative levels of these variants might have biological relevance. As exon 7 exclusion resulted in reduced CAG expansion, this prompted us to examine whether the levels of *MLH3* Var2 might correlate with tissue-specific expansion propensities of the *HTT* CAG repeat (43). Examination of *MLH3* Var1 and Var2 mRNA expression in the GTEx database (**Error! Reference source not found.**) did not reveal any obvious correlation with levels of CAG expansion in tissues; where forebrain regions such as the cortex and basal ganglia tend to have high instability, cerebellum exhibits low instability, and in the periphery, liver exhibits high instability (43). While more refined cell type-based analyses would be needed to better understand relationships between CAG instability and *trans*-modifying factors, these initial observations do not indicate any obvious relationship between naturally occurring *MLH3* Var2 and tissue-specific CAG expansion.

How is the MutLγ endonuclease activated to promote repeat expansion? MutLγ endonuclease is also required in canonical MMR in *S. cerevisiae* and in meiosis (42, 51), however the mechanism(s) of endonuclease activation are not yet well understood. In reconstituted systems, MutLγ can bind directly to DNA oligonucleotides, including those containing loop or branched structures, but does not cleave these structures (47–49, 62). This, and the further observation that larger DNA molecules are better substrates for the MutLγ endonuclease suggested that MutLγ binds cooperatively to DNA and that higher order polymers are required for endonuclease activation (62). Interestingly, the nicking activity of a 50:50 mixture of wild-type Mlh1-Mlh3 and Mlh1-Mlh3D523N dimers (*S. cerevisiae* D523N is orthologous to the D1185N mouse mutation) was the same as that elicited by 100% wild-type Mlh1-Mlh3, consistent with the idea that a critical level of DNA binding is needed for endonuclease activity (62). It is unknown whether MutLγ polymers are required *in vivo*, but this observation suggests one possible explanation for the relatively minor impact on CAG instability in heterozygous *Mlh3*$^{WT/DN}$ mice. In contrast, it appears that when the endonuclease-competent MLH3 is reduced below 50%, as in the kidney of *Htt*^{Q111} mice treated with *Mlh3* SSOs (Error! Reference source not found.), the effect on expansion is more pronounced (**Figure 2C, E**).

In reconstituted systems, both human and *S. cerevisiae* MutLγ are stimulated by MutSβ (human MSH2-MSH3 or *S. cerevisiae* Msh2-Msh3 respectively), potentially via direct interaction of these two complexes (47, 50, 63), but human MutLγ was not stimulated by MSH2-MSH6 (MutSα) (63). As MSH3, but not MSH6, is required for trinucleotide repeat expansion (16), MutSβ-stimulation of MutLγ endonuclease may be integral to repeat expansion *in vivo.* Other protein-protein interactions have been implicated in MutLγ endonuclease activation (63–65) and, interestingly, *S. cerevisiae Mlh3* separation of function mutations provided evidence that distinct protein interactions may be required for MutLγ endonuclease activity in MMR and in meiosis (63, 64). Activity appears to be variably stimulated by proliferating cell nuclear antigen (PCNA) and replication factor C (RFC), depending on the assay system (47, 50, 63, 65). Studies in *S. cerevisiae* have also indicated that *Mlh3* endonuclease creates DNA breaks in the presence of RNA/DNA hybrids (R-loops) that can lead to increased instability, suggesting a link with transcription (66), though the relevance to repeat expansion *in vivo* is unclear. Future

biochemical and genetic studies will be needed to dissect the mechanism of MutLγ endonuclease activation in its repeat expansion-promoting function.

Is *MLH3* a good therapeutic target for HD? In contrast to other MMR pathway genes (*MSH3, MLH1 PMS2, PMS1*), *MLH3* is not currently validated as a disease modifier in human GWAS (14). This may relate to the effect size of a putative modifier SNP(s) and the power to detect a significant effect on a clinical endpoint with the current patient sample size. Nevertheless, the following provide support for *MLH3* as a viable therapeutic target in HD: (1) *MLH3* is part of a pathway involved in disease modification that likely acts via somatic repeat expansion, (2) *MLH3* is essential for somatic CAG repeat expansion in HD mice (17), (3) Functional MLH3 promotes *HTT* CAG expansion in HD patient cells (this study), and (4) *MLH3* genetic variation is associated with somatic *HTT* CAG expansion measured in HD patient blood DNA (13). From a safety standpoint, consistent with a minor role of MLH3 in MMR, only a handful of studies have reported associations of *MLH3* mutations with cancer, however the pathogenicity of these mutations and their clinical significance are still unclear (67–70). Further, the genome Aggregation Database (gnomAD) reports an extremely low probability for intolerance to loss of function mutations for *MLH3* (pLI = 0.00) (71). Therefore, MLH3 appears to be a promising candidate as a safe and effective therapeutic target.

Here, we demonstrate applicability to HD of one possible therapeutic approach targeting the critical endonuclease domain of MLH3 using SSOs to redirect *MLH3* splicing. Future studies will be needed to demonstrate the efficacy of *MLH3* splice redirection in the brain. The splice redirection approach has the advantage of specifically altering the MLH3 endonuclease whilst leaving intact the PMS2 endonuclease that is critical to the major MMR activity in cells. Further insights into distinguishing features of these endonucleases may provide additional opportunities for specific inactivation of the MLH3 endonuclease. An alternative oligonucleotide-based approach of simply reducing *MLH3* expression would also be indicated based on genetic data in mice. In contrast to splice redirection, this strategy would reduce or eliminate all MLH3 function(s). Indeed, by some measures of MLH3 activity, the *Mlh3*$^{DN/DN}$ is indistinguishable from *Mlh3*$^{WT/WT}$, suggesting that the splice redirected MLH3 protein would also retain such nuclease-independent activities (42). Therefore, additional understanding of the cellular roles of

the MLH3 endonuclease, and of the endonuclease-lacking variant, will help to inform on optimal therapeutic approaches applicable to both HD and other repeat expansion diseases.

Funding

This work was supported by the Huntington's Disease Society of America Berman/Topper Family HD Career Development Fellowship and HD Human Biology Project (R.M.P.); the National Institutes of Health (NS049206 to V.C.W); the Dake Family Foundation (R.M.P.); the Friedreich Ataxia Research Alliance (E.G.); and the Harrington Discovery Institute (E.G.).

Conflict of Interest Disclosure

V.C.W. is a scientific advisory board member of Triplet Therapeutics, a company developing new therapeutic approaches to address triplet repeat disorders such Huntington's disease and Myotonic Dystrophy, and of LoQus23 Therapeutics, and has provided paid consulting services to Alnylam and Acadia Pharmaceuticals. Her financial interests in Triplet Therapeutics were reviewed and are managed by Massachusetts General Hospital and Partners HealthCare in accordance with their conflict of interest policies. R.M.P. and V.C.W have received sponsored research funding from Pfizer unrelated to the content of this manuscript. E.G. is a named inventor on a patent related to this work (US10669542B2).

Acknowledgments

We would like to thank Tammy Gillis and Jacqi Siciliano of the MGH Mission Driven Service Core for DNA Fragment Analysis.

References

1. Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell*, **72**, 971–983.
2. Vonsattel,J.-P., Myers,R.H., Stevens,T.J., Ferrante,R.J., Bird,E.D. and Richardson Jr.,E.P. (1985) Neuropathological Classification of Huntington's Disease. *J. Neuropathol. Exp. Neurol.*, **44**, 559–577.
3. McColgan,P. and Tabrizi,S.J. (2018) Huntington's disease: a clinical review. *Eur. J. Neurol.*, **25**, 24–34.
4. Kennedy,L., Evans,E., Chen,C.-M.M., Craven,L., Detloff,P.J., Ennis,M. and Shelbourne,P.F. (2003) Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. *Hum. Mol. Genet.*, **12**, 3359–3367.
5. Shelbourne,P.F., Keller-McGandy,C., Bi,W.L., Yoon,S.R., Dubeau,L., Veitch,N.J., Vonsattel,J.P., Wexler,N.S., Norman,A. and Augood,S.J. (2007) Triplet repeat mutation length gains correlate with cell-type specific vulnerability in Huntington disease brain. *Hum. Mol. Genet.*, **16**, 1133–1142.
6. Gonitel,R., Moffitt,H., Sathasivam,K., Woodman,B., Detloff,P.J., Faull,R.L.M. and Bates,G.P. (2008) DNA instability in postmitotic neurons. *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 3467–3472.
7. Swami,M., Hendricks,A.E., Gillis,T., Massood,T., Mysore,J., Myers,R.H. and Wheeler,V.C. (2009) 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.
8. Loupe,J.M., Mouro Pinto,R., Kim,K.-H., Gillis,T., Mysore,J.S., Andrew,M.A., Kovalenko,M., Murtha,R., Seong,I., Gusella,J.F., *et al.* (2020) Promotion of somatic CAG repeat expansion by Fan1 knock-out in Huntington's disease knock-in mice is blocked by Mlh1 knock-out. *Hum. Mol. Genet.*, 10.1093/hmg/ddaa196.
9. Lee,J.M., Mouro Pinto,R., Gillis,T., St. Claire,J.C. and Wheeler,V.C. (2011) 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.
10. Kovalenko,M., Dragileva,E., St. Claire,J., Gillis,T., Guide,J.R., New,J., Dong,H., Kucherlapati,R., Kucherlapati,M.H., Ehrlich,M.E., *et al.* (2012) 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.

11. Larson,E., Fyfe,I., Morton,A.J. and Monckton,D.G. (2015) 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.
12. Ament,S.A., Pearl,J.R., Grindeland,A., Claire,J.S., Earls,J.C., Kovalenko,M., Gillis,T., Mysore,J., Gusella,J.F., Lee,J.M., *et al.* (2017) High resolution time-course mapping of early transcriptomic, molecular and cellular phenotypes in Huntington's disease CAG knock-in mice across multiple genetic backgrounds. *Hum. Mol. Genet.*, **26**, 913–922.
13. Ciosi,M., Maxwell,A., Cumming,S.A., Hensman Moss,D.J., Alshammari,A.M., Flower,M.D., Durr,A., Leavitt,B.R., Roos,R.A.C., Holmans,P., *et al.* (2019) A genetic association study of glutamine-encoding DNA sequence structures, somatic CAG expansion, and DNA repair gene variants, with Huntington disease clinical outcomes. *EBioMedicine*, **48**, 568–580.
14. GeM-HD (2019) CAG Repeat Not Polyglutamine Length Determines Timing of Huntington's Disease Onset. *Cell*, **178**, 887-900.e14.
15. Wright,G.E.B., Collins,J.A., Kay,C., McDonald,C., Dolzhenko,E., Xia,Q., Bečanović,K., Drögemöller,B.I., Semaka,A., Nguyen,C.M., *et al.* (2019) Length of Uninterrupted CAG, Independent of Polyglutamine Size, Results in Increased Somatic Instability, Hastening Onset of Huntington Disease. *Am. J. Hum. Genet.*, **104**, 1116–1126.
16. Dragileva,E., Hendricks,A., Teed,A., Gillis,T., Lopez,E.T., Friedberg,E.C., Kucherlapati,R., Edelmann,W., Lunetta,K.L., MacDonald,M.E., *et al.* (2009) Intergenerational and striatal CAG repeat instability in Huntington's disease knock-in mice involve different DNA repair genes. *Neurobiol. Dis.*, **33**, 37–47.
17. Mouro Pinto,R., Dragileva,E., Kirby,A., Lloret,A., Lopez,E., St. Claire,J., Panigrahi,G.B., Hou,C., Holloway,K., Gillis,T., *et al.* (2013) Mismatch Repair Genes Mlh1 and Mlh3 Modify CAG Instability in Huntington's Disease Mice: Genome-Wide and Candidate Approaches. *PLoS Genet.*, **9**, e1003930.
18. Wheeler,V.C., Lebel,L.-A., Vrbanac,V., Teed,A., te Riele,H. and MacDonald,M.E. (2003) Mismatch repair gene Msh2 modifies the timing of early disease in HdhQ111 striatum. *Hum. Mol. Genet.*, **12**, 273–281.

19. Owen,B.A.L., Yang,Z., Lai,M., Gajek,M., Badger,J.D., Hayes,J.J., Edelmann,W., Kucherlapati,R., Wilson,T.M. and McMurray,C.T. (2005) (CAG)n-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition. *Nat. Struct. Mol. Biol.*, **12**, 663–670.

20. GeM-HD (2015) Identification of Genetic Factors that Modify Clinical Onset of Huntington's Disease. *Cell*, **162**, 516–526.

21. Moss,D.J.H., Tabrizi,S.J., Mead,S., Lo,K., Pardiñas,A.F., Holmans,P., Jones,L., Langbehn,D., Coleman,A., Santos,R.D., *et al.* (2017) Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study. *Lancet Neurol.*, **16**, 701–711.

22. Flower,M., Lomeikaite,V., Ciosi,M., Cumming,S., Morales,F., Lo,K., Hensman Moss,D., Jones,L., Holmans,P., Investigators, the T.-H., *et al.* (2019) MSH3 modifies somatic instability and disease severity in Huntington's and myotonic dystrophy type 1. *Brain*, **142**, 1876–1886.

23. van den Broek,W.J.A.A., Nelen,M.R., Wansink,D.G., Coerwinkel,M.M., te Riele,H., Groenen,P.J.T.A. and Wieringa,B. (2002) Somatic expansion behaviour of the (CTG)n repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch–repair proteins. *Hum. Mol. Genet.*, **11**, 191–198.

24. Savouret,C., Brisson,E., Essers,J., Kanaar,R., Pastink,A., Te Riele,H., Junien,C. and Gourdon,G. (2003) CTG repeat instability and size variation timing in DNA repair-deficient mice. *EMBO J.*, **22**, 2264–2273.

25. Foiry,L., Dong,L., Savouret,C., Hubert,L., te Riele,H., Junien,C. and Gourdon,G. (2006) Msh3 is a limiting factor in the formation of intergenerational CTG expansions in DM1 transgenic mice. *Hum. Genet.*, **119**, 520–526.

26. Bourn,R.L., de Biase,I., Mouro Pinto,R., Sandi,C., Al-Mahdawi,S., Pook,M.A. and Bidichandani,S.I. (2012) Pms2 Suppresses Large Expansions of the (GAA·TTC)n Sequence in Neuronal Tissues. *PLoS One*, **7**, e47085.

27. Halabi,A., Ditch,S., Wang,J. and Grabczyk,E. (2012) DNA mismatch repair complex MutSb promotes GAA-TTC repeat expansion in human cells. *J. Biol. Chem.*, **287**, 29958–29967.

28. Halabi,A., Fuselier,K.T.B. and Grabczyk,E. (2018) GAA•TTC repeat expansion in human cells is mediated by mismatch repair complex MutLγ and depends upon the endonuclease

domain in MLH3 isoform one. *Nucleic Acids Res.*, **46**, 4022–4032.

29. Ezzatizadeh,V., Sandi,C., Sandi,M., Anjomani-Virmouni,S., Al-Mahdawi,S. and Pook,M.A. (2014) MutLα Heterodimers Modify the Molecular Phenotype of Friedreich Ataxia. *PLoS One*, **9**, e100523.
30. Hayward,B.E., Steinbach,P.J. and Usdin,K. (2020) A point mutation in the nuclease domain of MLH3 eliminates repeat expansions in a mouse stem cell model of the Fragile X-related disorders. *Nucleic Acids Res.*, **48**, 7856–7863.
31. Zhao,X., Zhang,Y., Wilkins,K., Edelmann,W. and Usdin,K. (2018) MutLγ promotes repeat expansion in a Fragile X mouse model while EXO1 is protective. *PLOS Genet.*, **14**, e1007719.
32. Miller,C.J., Kim,G.-Y., Zhao,X. and Usdin,K. (2020) All three mammalian MutL complexes are required for repeat expansion in a mouse cell model of the Fragile X-related disorders. *PLOS Genet.*, **16**, e1008902.
33. Lokanga,R.A., Zhao,X.-N. and Usdin,K. (2014) The Mismatch Repair Protein MSH2 is Rate Limiting for Repeat Expansion in a Fragile X Premutation Mouse Model. *Hum. Mutat.*, **35**, 129–136.
34. Zhao,X.-N., Lokanga,R., Allette,K., Gazy,I., Wu,D. and Usdin,K. (2016) A MutSβ-Dependent Contribution of MutSα to Repeat Expansions in Fragile X Premutation Mice? *PLOS Genet.*, **12**, e1006190.
35. Bettencourt,C., Hensman-Moss,D., Flower,M., Wiethoff,S., Brice,A., Goizet,C., Stevanin,G., Koutsis,G., Karadima,G., Panas,M., *et al.* (2016) DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. *Ann. Neurol.*, **79**, 983–990.
36. Lipkin,S.M., Moens,P.B., Wang,V., Lenzi,M., Shanmugarajah,D., Gilgeous,A., Thomas,J., Cheng,J., Touchman,J.W., Green,E.D., *et al.* (2002) Meiotic arrest and aneuploidy in MLH3-deficient mice. *Nat. Genet.*, **31**, 385–390.
37. Cannavo,E., Marra,G., Sabates-Bellver,J., Menigatti,M., Lipkin,S.M., Fischer,F., Cejka,P. and Jiricny,J. (2005) Expression of the MutL homologue hMLH3 in human cells and its role in DNA mismatch repair. *Cancer Res.*, **65**, 10759–10766.
38. Charbonneau,N., Amunugama,R., Schmutte,C., Yoder,K. and Fishel,R. (2009) Evidence that hMLH3 functions primarily in meiosis and in hMSH2-hMSH3 mismatch repair. *Cancer Biol. Ther.*, **8**, 1411–1420.

39. Kadyrov,F.A., Dzantiev,L., Constantin,N. and Modrich,P. (2006) Endonucleolytic Function of MutLα in Human Mismatch Repair. *Cell*, **126**, 297–308.

40. Yates,A.D., Achuthan,P., Akanni,W., Allen,J., Allen,J., Alvarez-Jarreta,J., Amode,M.R., Armean,I.M., Azov,A.G., Bennett,R., *et al.* (2020) Ensembl 2020. *Nucleic Acids Res.*, **48**, D682–D688.

41. Wheeler,V.C., Auerbach,W., White,J.K., Srinidhi,J., Auerbach,A., Ryan,A., Duyao,M.P., Vrbanac,V., Weaver,M., Gusella,J.F., *et al.* (1999) Length-dependent gametic CAG repeat instability in the Huntington's disease knock-in mouse. *Hum. Mol. Genet.*, **8**, 115–122.

42. Toledo,M., Sun,X., Brieño-Enríquez,M.A., Raghavan,V., Gray,S., Pea,J., Milano,C.R., Venkatesh,A., Patel,L., Borst,P.L., *et al.* (2019) A mutation in the endonuclease domain of mouse MLH3 reveals novel roles for MutLγ during crossover formation in meiotic prophase I. *PLoS Genet.*, **15**, e1008177.

43. Mouro Pinto,R., Arning,L., Giordano,J. V, Razghandi,P., Andrew,M.A., Gillis,T., Correia,K., Mysore,J.S., Grote Urtubey,D.M., Parwez,C.R., *et al.* (2020) Patterns of CAG repeat instability in the central nervous system and periphery in Huntington's disease and in spinocerebellar ataxia type 1. *Hum. Mol. Genet.*, **29**, 2551–2567.

44. Lee,J.-M., Zhang,J., Su,A.I., Walker,J.R., Wiltshire,T., Kang,K., Dragileva,E., Gillis,T., Lopez,E.T., Boily,M.-J., *et al.* (2010) A novel approach to investigate tissue-specific trinucleotide repeat instability. *BMC Syst. Biol.*, **4**, 29.

45. Morcos,P.A., Li,Y. and Jiang,S. (2008) Vivo-Morpholinos: A non-peptide transporter delivers Morpholinos into a wide array of mouse tissues. *Biotechniques*, **45**, 613–623.

46. Sathasivam,K., Amaechi,I., Mangiarini,L. and Bates,G. (1997) 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.

47. Rogacheva,M. V., Manhart,C.M., Chen,G., Guarne,A., Surtees,J. and Alani,E. (2014) Mlh1-Mlh3, a meiotic crossover and DNA mismatch repair factor,is a Msh2-Msh3-stimulated endonuclease. *J. Biol. Chem.*, **289**, 5664–5673.

48. Ranjha,L., Anand,R. and Cejka,P. (2014) The Saccharomyces cerevisiae Mlh1-Mlh3 heterodimer is an endonuclease that preferentially binds to holliday junctions. *J. Biol. Chem.*, **289**, 5674–5686.

49. Claeys Bouuaert,C. and Keeney,S. (2017) Distinct DNA-binding surfaces in the ATPase and

linker domains of MutLγ determine its substrate specificities and exert separable functions in meiotic recombination and mismatch repair. *PLOS Genet.*, **13**, e1006722.

50. Kadyrova,L.Y., Gujar,V., Burdett,V., Modrich,P.L. and Kadyrov,F.A. (2020) Human MutLγ, the MLH1–MLH3 heterodimer, is an endonuclease that promotes DNA expansion. *Proc. Natl. Acad. Sci. U. S. A.*, **117**, 3535–3542.
51. Nishant,K.T., Plys,A.J. and Alani,E. (2008) A mutation in the putative MLH3 endonuclease domain confers a defect in both mismatch repair and meiosis in Saccharomyces cerevisiae. *Genetics*, **179**, 747–755.
52. Goold,R., Flower,M., Moss,D.H., Medway,C., Wood-Kaczmar,A., Andre,R., Farshim,P., Bates,G.P., Holmans,P., Jones,L., *et al.* (2019) FAN1 modifies Huntington's disease progression by stabilizing the expanded *HTT* CAG repeat. *Hum. Mol. Genet.*, **28**, 650–661.
53. Nakatani,R., Nakamori,M., Fujimura,H., Mochizuki,H. and Takahashi,M.P. (2015) Large expansion of CTG·CAG repeats is exacerbated by MutSβ in human cells. *Sci. Rep.*, **5**, 11020.
54. Kondo,E., Horii,A. and Fukushige,S. (2001) The interacting domains of three MutL heterodimers in man: hMLH1 interacts with 36 homologous amino acid residues within hMLH3, hPMS1 and hPMS2. *Nucleic Acids Res.*, **29**, 1695–1702.
55. Santucci-Darmanin,S., Neyton,S., Lespinasse,F., Saunières,A., Gaudray,P. and Paquis-Flucklinger,V. (2002) The DNA mismatch-repair MLH3 protein interacts with MSH4 in meiotic cells, supporting a role for this MutL homolog in mammalian meiotic recombination. *Hum. Mol. Genet.*, **11**, 1697–1706.
56. Roesner,L.M., Mielke,C., Fähnrich,S., Merkhoffer,Y., Dittmar,K.E.J.J., Drexler,H.G. and Dirks,W.G. (2013) Stable expression of MutLgamma in human cells reveals no specific response to mismatched DNA, but distinct recruitment to damage sites. *J. Cell. Biochem.*, **114**, 2405–2414.
57. Lipkin,S.M., Wang,V., Jacoby,R., Banerjee-Basu,S., Baxevanis,A.D., Lynch,H.T., Elliott,R.M. and Collins,F.S. (2000) MLH3: A DNA mismatch repair gene associated with mammalian microsatellite instability. *Nat. Genet.*, **24**, 27–35.
58. Gueneau,E., Dherin,C., Legrand,P., Tellier-Lebegue,C., Gilquin,B., Bonnesoeur,P., Londino,F., Quemener,C., Le Du,M.H., Márquez,J.A., *et al.* (2013) Structure of the MutLα C-terminal domain reveals how Mlh1 contributes to Pms1 endonuclease site. *Nat. Struct.*

Mol. Biol., **20**, 461–468.

59. Guarné,A. (2012) The functions of MutL in mismatch repair: The power of multitasking. In *Progress in Molecular Biology and Translational Science*. Elsevier B.V., Vol. 110, pp. 41–70.
60. Pluciennik,A., Burdett,V., Baitinger,C., Iyer,R.R., Shi,K. and Modrich,P. (2013) Extrahelical (CAG)/(CTG) triplet repeat elements support proliferating cell nuclear antigen loading and MutLa endonuclease activation. *Proc. Natl. Acad. Sci. U. S. A.*, **110**, 12277–12282.
61. Gomes-Pereira,M., Fortune,M.T., Ingram,L., McAbney,J.P. and Monckton,D.G. (2004) Pms2 is a genetic enhancer of trinucleotide CAG-CTG repeat somatic mosaicism: Implications for the mechanism of triplet repeat expansion. *Hum. Mol. Genet.*, **13**, 1815–1825.
62. Manhart,C.M., Ni,X., White,M.A., Ortega,J., Surtees,J.A. and Alani,E. (2017) The mismatch repair and meiotic recombination endonuclease Mlh1-Mlh3 is activated by polymer formation and can cleave DNA substrates in trans. *PLOS Biol.*, **15**, e2001164.
63. Cannavo,E., Sanchez,A., Anand,R., Ranjha,L., Hugener,J., Adam,C., Acharya,A., Weyland,N., Aran-Guiu,X., Charbonnier,J.B., *et al.* (2020) Regulation of the MLH1–MLH3 endonuclease in meiosis. *Nature*, 10.1038/s41586-020-2592-2.
64. Al-Sweel,N., Raghavan,V., Dutta,A., Ajith,V.P., Di Vietro,L., Khondakar,N., Manhart,C.M., Surtees,J.A., Nishant,K.T. and Alani,E. (2017) mlh3 mutations in baker's yeast alter meiotic recombination outcomes by increasing noncrossover events genome-wide. *PLOS Genet.*, **13**, e1006974.
65. Kulkarni,D.S., Owens,S.N., Honda,M., Ito,M., Yang,Y., Corrigan,M.W., Chen,L., Quan,A.L. and Hunter,N. (2020) PCNA activates the MutLγ endonuclease to promote meiotic crossing over. *Nature*, 10.1038/s41586-020-2645-6.
66. Su,X.A. and Freudenreich,C.H. (2017) Cytosine deamination and base excision repair cause R-loop–induced CAG repeat fragility and instability in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.*, **114**, E8392–E8401.
67. Hansen,M.F., Johansen,J., Sylvander,A.E., Bjørnevoll,I., Talseth-Palmer,B.A., Lavik,L.A.S., Xavier,A., Engebretsen,L.F., Scott,R.J., Drabløs,F., *et al.* (2017) Use of multigene-panel identifies pathogenic variants in several CRC-predisposing genes in patients previously tested for Lynch Syndrome. *Clin. Genet.*, **92**, 405–414.

68. Xavier,A., Olsen,M.F., Lavik,L.A., Johansen,J., Singh,A.K., Sjursen,W., Scott,R.J. and Talseth□Palmer,B.A. (2019) Comprehensive mismatch repair gene panel identifies variants in patients with Lynch□like syndrome. *Mol. Genet. Genomic Med.*, **7**.

69. Olkinuora,A., Nieminen,T.T., Mårtensson,E., Rohlin,A., Ristimäki,A., Koskenvuo,L., Lepistö,A., Silander,G., Kuchinskaya,E., Aravidis,C., *et al.* (2019) Biallelic germline nonsense variant of MLH3 underlies polyposis predisposition. *Genet. Med.*, **21**, 1868–1873.

70. Yang,M., Zhu,L., Xu,D., Weng,S., Fang,X., Dong,C., Hu,H. and Yuan,Y. (2019) How should we correctly interpret biallelic germline truncating variant of MLH3 in hereditary colorectal cancer? *Genet. Med.*, **21**, 2650–2651.

71. Karczewski,K.J., Francioli,L.C., Tiao,G., Cummings,B.B., Alföldi,J., Wang,Q., Collins,R.L., Laricchia,K.M., Ganna,A., Birnbaum,D.P., *et al.* (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, **581**, 434–443.

Figure Legends

Figure 1. *MLH3* endonuclease domain mutation eliminates *HTT* CAG expansion in *Htt*^{Q111} mice.

(**A**) Schematic representation of MLH3 protein and its functional domains, as well as mutation of the endonucleolytic motif in *Mlh3*^{DN} mice (adapted from Toledo *et al*. 2019 (42)). (**B**) Quantification of *HTT* CAG expansions in striatum, liver, and kidney at 6 months of age in *Htt*^{Q111} knock-in mice that are either wild type (*Mlh3*$^{WT/WT}$, CAG 114-121), heterozygous (*Mlh3*$^{WT/DN}$, CAG 116-121), or homozygous (*Mlh3*$^{DN/DN}$, CAG 116-121) for the D1185N mutation shows that an intact MLH3 endonuclease domain is required for somatic CAG expansion. Striatum and liver, $n = 3$; kidney, $n = 2$-3; error bars represent standard deviation of the mean; two-way ANOVA (with genotype and tissue as variables) with Tukey multiple comparisons *post hoc* correction. Genotype effect: *ns*, not significant; ***, $p \leq 0.001$; ****, $p \leq 0.0001$. (**C**) Representative GeneMapper traces across different tissues at 6 months of age in *Htt*^{Q111} knock-in mice (CAG 121) with different *Mlh3* genotypes.

Figure 2. Successful *in vivo Mlh3* splice redirection and suppression of somatic *HTT* CAG expansion in *Htt*^{Q111} mice by systemic *Mlh3* SSO treatment.

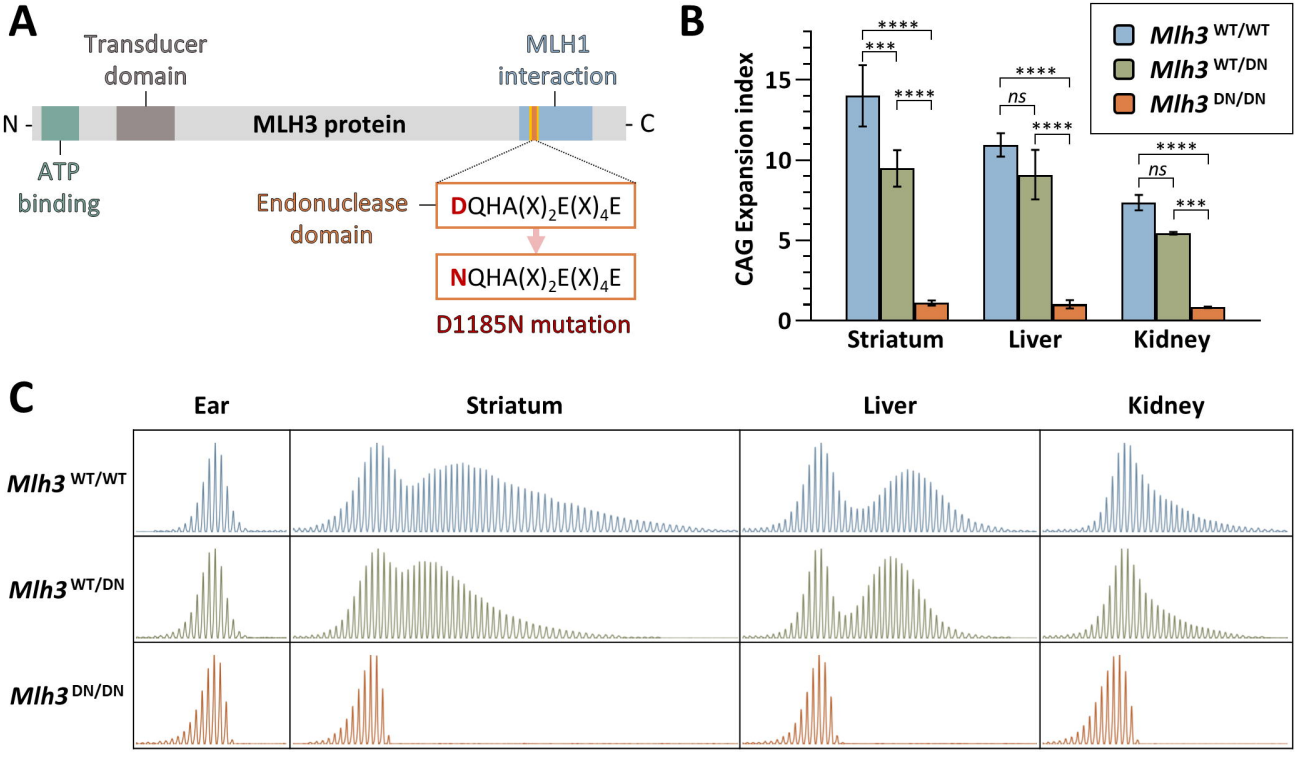
(**A**) Schematic representation of mouse *Mlh3* splice redirection with splice switching oligonucleotides (SSOs). SSOs were designed to bind the intron–exon junctions in mouse *Mlh3* pre-mRNA at either the splice acceptor (mMlh3ac5) or donor (mMlh3dr7) regions of the endonuclease-coding exon 6, which is analogous to the human MLH3 exon 7, inducing exon skipping and preferential production of *Mlh3* splice variant 2, which lacks the endonuclease domain. (**B**) Summary of *Mlh3* SSOs systemic treatment in HD mice. *Htt*^{Q111} mice (two males and two females per group) were treated with either 25 mg/kg *Mlh3* SSOs (mMLH3ac5 and mMLH3dr7, 1:1 ratio; CAG 106-109), or PBS (CAG 104-108), by tail vein injection. Treatment was initiated at 4 weeks of age and injections were performed three times per week (every 48-72hr). Weight measurements were taken at the beginning of each week to determine weekly weight-adjusted doses. Treatment was stopped at 12 weeks of age, with tissue collection performed 24 hours after the final injection. (**C**) RT-PCR analyses of *Mlh3* mRNA splice variants in 12-week-old *Htt*^{Q111} mice, following treatment with *Mlh3* SSOs, reveals successful splice redirection from splice variant 1 (Var1) to variant 2 (Var2) in the kidney (K), and to a

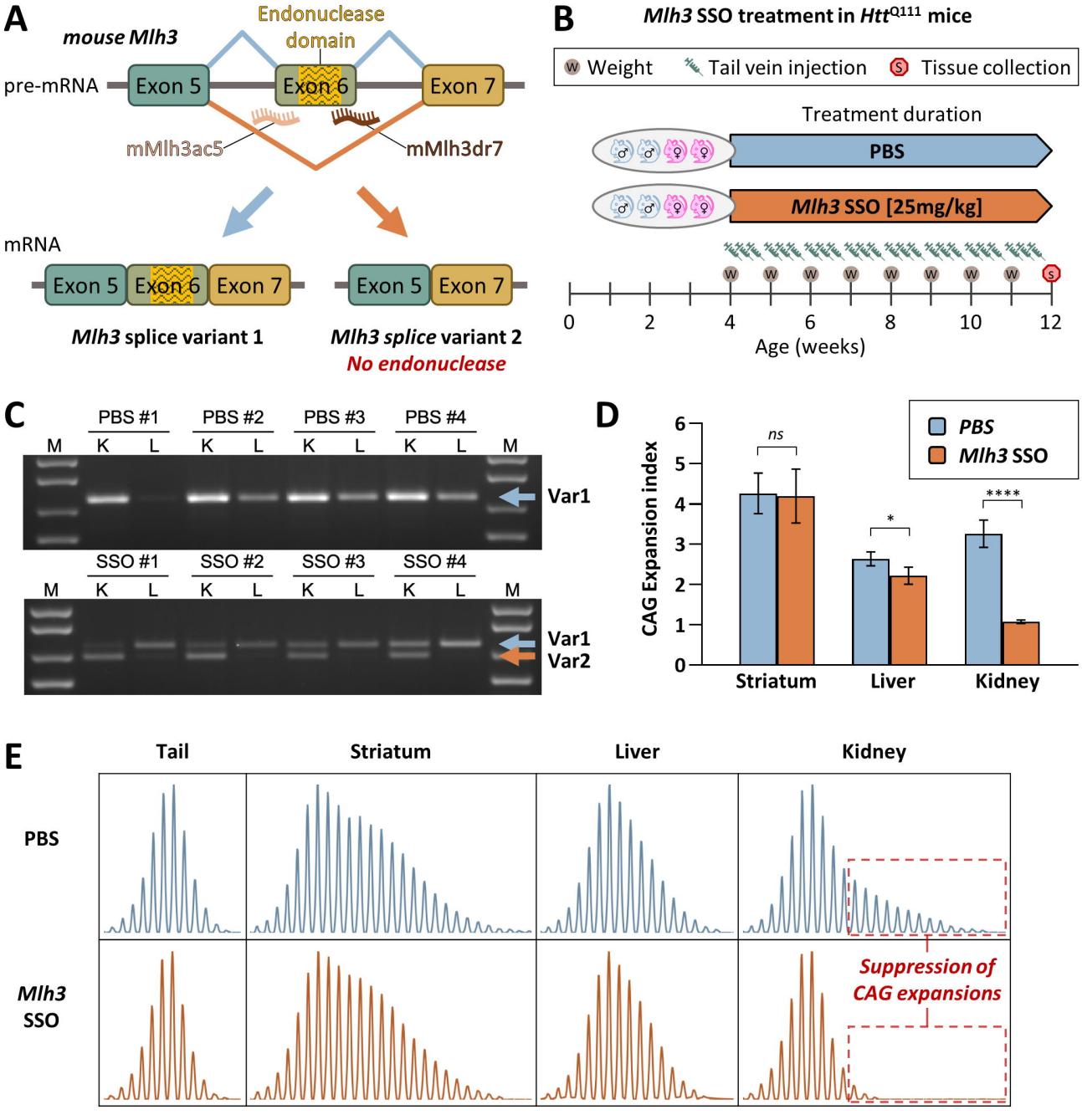
lesser extent in the liver (L). (**D**) Quantification of CAG expansion indices reveals statistically significant reductions in somatic *HTT* CAG expansion in kidney and, to a lesser extent, in liver of mice that received *Mlh3* SSO treatment when compared to PBS. $n = 4$; error bars represent standard deviation of the mean; 2-tailed unpaired *t* test: *ns*, not significant; *, $p \leq 0.05$; ****, $p \leq 0.0001$. (**E**) Representative GeneMapper traces of somatic *HTT* CAG repeat size distributions across different tissues at 12 weeks of age in Htt^{Q111} knock-in mice (CAG 108), following PBS or *Mlh3* SSO treatment.

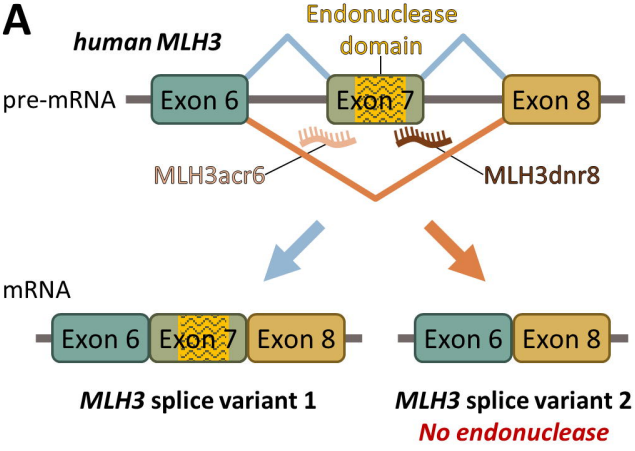
Figure 3. Successful *MLH3* splice redirection and suppression of *HTT* CAG expansion in HD patient-derived fibroblasts following treatment with *MLH3* SSOs.

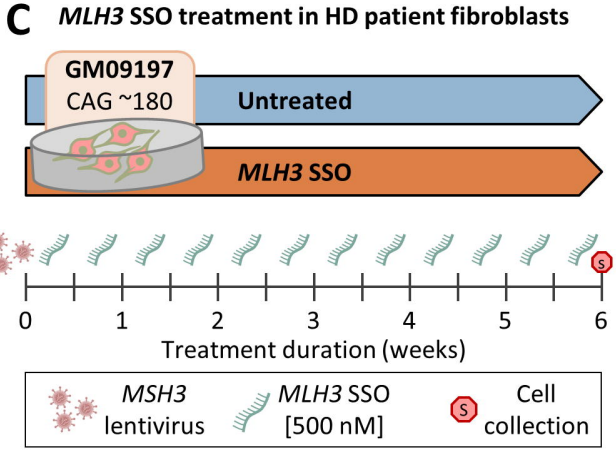
(**A**) Schematic representation of human *MLH3* splice redirection with splice switching oligonucleotides (SSOs) (adapted from Halabi *et al*. 2018 (28)). SSOs were designed to bind the intron–exon junctions in human *MLH3* pre-mRNA at either the splice acceptor (MLH3ac6) or donor (MLH3dnr8) regions of the endonuclease-coding exon 7, inducing exon skipping and preferential production of *MLH3* splice variant 2, which lacks the endonuclease domain. (**B**) RT-PCR analyses of *MLH3* mRNA splice variants in HD patient-derived primary fibroblasts (GM09197, CAG ~180/18) following treatment with 500 nM of either scrambled control vivo-morpholino oligonucleotide (Scr) or *MLH3* SSOs (SSO) for 48 hours, alongside untreated controls (Unt). Splice redirection from predominantly splice variant 1 (Var1) to predominantly splice variant 2 (Var2), which lacks exon 7 and the endonuclease domain, was efficiently achieved with *MLH3* SSOs treatment. No splice redirection was observed in scrambled oligo-treated cells, which were identical to untreated cells. (**C**) In order to induce CAG expansion in HD patient fibroblasts (GM09197, CAG ~180/18), cells were initially treated with lentiviral particles for stable ectopic expression of MSH3. Cells were then cultured to confluency, to inhibit replication, and treated twice weekly with 500 nM of *MLH3* SSOs (MLH3acr6 and MLH3dnr8, 1:1 ratio, 250 nM each), or left untreated, for a total of six weeks when cells were harvested for *HTT* CAG instability analysis. (**D**) Quantification of average *HTT* CAG gain during the 6 week-long treatment (relative to average CAG prior to treatment, "$t = 0$", $n = 4$) reveals a potent inhibition of CAG expansion by *MLH3* SSOs (average gain of 0.6 CAGs, $n = 10$), when compared to untreated cells (average gain of 3.0 CAGs, $n = 10$). Error bars represent standard deviation of the mean; one-way ANOVA with Tukey multiple comparisons *post hoc*

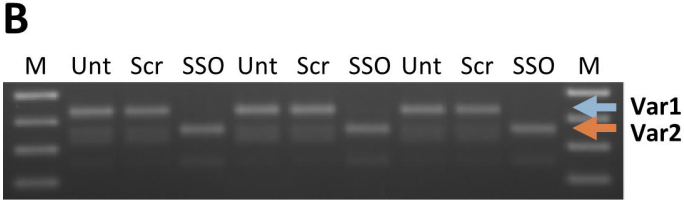
correction; ****, $p \leq 0.0001$. (**E**) Representative GeneMapper traces of *HTT* CAG repeat size distributions from GM09197 HD patient fibroblasts before ($t = 0$) and after ($t = 6$ weeks) treatment with *MLH3* SSOs (CAG gain = 0.7), or no treatment (CAG gain = 3.6).

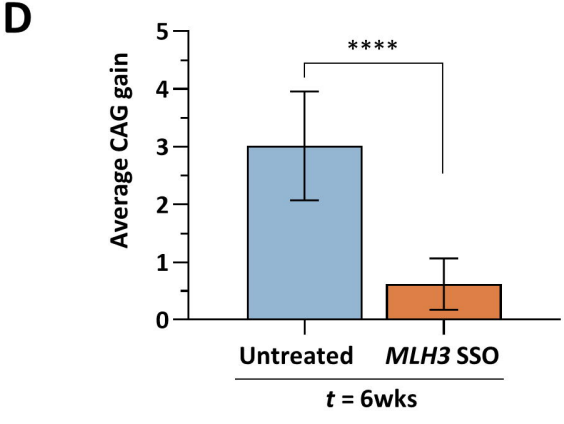

A
Transducer domain
MLH1 interaction
N -
MLH3 protein
- C
ATP binding
Endonuclease domain
DQHA(X)2E(X)4E
NQHA(X)2E(X)4E
D1185N mutation
B
CAG Expansion index
Mlh3 WT/WT
Mlh3 WT/DN
Mlh3 DN/DN
****
***
ns
Striatum
Liver
Kidney
C
Ear
Striatum
Liver
Kidney
Mlh3 WT/WT
Mlh3 WT/DN
Mlh3 DN/DN

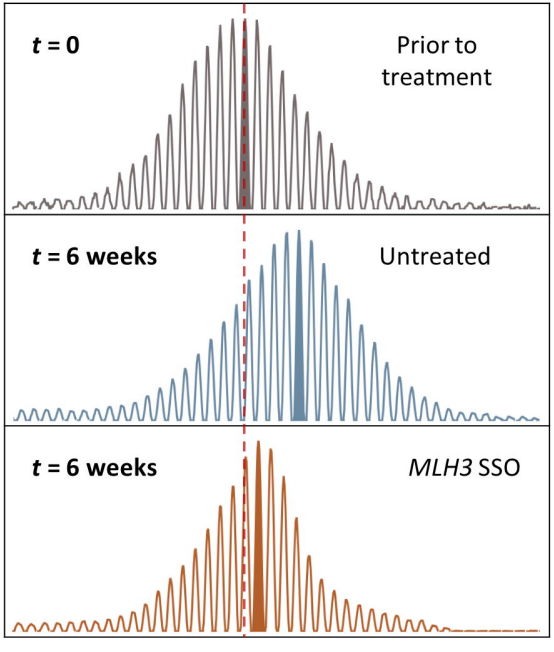

A
mouse Mlh3
Endonuclease domain
pre-mRNA
Exon 5
Exon 6
Exon 7
mMlh3ac5
mMlh3dr7
mRNA
Exon 5
Exon 6
Exon 7
Exon 5
Exon 7
Mlh3 splice variant 1
Mlh3 splice variant 2
No endonuclease
B
Mlh3 SSO treatment in Htt^Q111 mice
Weight
Tail vein injection
Tissue collection
Treatment duration
PBS
Mlh3 SSO [25mg/kg]
0
2
4
6
8
10
12
Age (weeks)
C
PBS #1
PBS #2
PBS #3
PBS #4
M
K
L
Var1
SSO #1
SSO #2
SSO #3
SSO #4
Var1
Var2
D
PBS
Mlh3 SSO
CAG Expansion index
ns
*
****
Striatum
Liver
Kidney
E
Tail
Striatum
Liver
Kidney
PBS
Mlh3 SSO
Suppression of CAG expansions


A
human MLH3
Endonuclease domain
pre-mRNA
Exon 6
Exon 7
Exon 8
MLH3acr6
MLH3dnr8
mRNA
Exon 6
Exon 7
Exon 8
Exon 6
Exon 8
MLH3 splice variant 1
MLH3 splice variant 2
No endonuclease


C
MLH3 SSO treatment in HD patient fibroblasts
GM09197
CAG ~180
Untreated
MLH3 SSO
0
1
2
3
4
5
6
Treatment duration (weeks)
MSH3 lentivirus
MLH3 SSO [500 nM]
Cell collection


B
M
Unt
Scr
SSO
Unt
Scr
SSO
Unt
Scr
SSO
M
Var1
Var2


D
Average CAG gain
0
1
2
3
4
5
****
Untreated
MLH3 SSO
t = 6wks


E
t = 0
Prior to treatment
t = 6 weeks
Untreated
t = 6 weeks
MLH3 SSO