

Somatic expansion of the Huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset

Meera Swami¹, Audrey E. Hendricks², Tammy Gillis¹, Tiffany Massood³, Jayalakshmi Mysore¹, Richard H. Myers³ and Vanessa C. Wheeler^{1,*}

¹Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA 02114, USA, ²Department of Biostatistics and ³Department of Neurology, Boston University School of Public Health, Boston, MA 02118, USA

Received March 26, 2009; Revised and Accepted May 19, 2009

The age of onset of Huntington's disease (HD) is determined primarily by the length of the *HD* CAG repeat mutation, but is also influenced by other modifying factors. Delineating these modifiers is a critical step towards developing validated therapeutic targets in HD patients. The *HD* CAG repeat is somatically unstable, undergoing progressive length increases over time, particularly in brain regions that are the targets of neurodegeneration. Here, we have explored the hypothesis that somatic instability of the *HD* CAG repeat is itself a modifier of disease. Using small-pool PCR, we quantified somatic instability in the cortex region of the brain from a cohort of HD individuals exhibiting phenotypic extremes of young and old disease onset as predicted by the length of their constitutive *HD* CAG repeat lengths. After accounting for constitutive repeat length, somatic instability was found to be a significant predictor of onset age, with larger repeat length gains associated with earlier disease onset. These data are consistent with the hypothesis that somatic *HD* CAG repeat length expansions in target tissues contribute to the HD pathogenic process, and support pursuing factors that modify somatic instability as viable therapeutic targets.

INTRODUCTION

Huntington's disease (HD) is a dominantly inherited, fatal neurodegenerative disorder characterized by chorea, cognitive and psychiatric decline (1). It is caused by the expansion over 35 repeats of a polymorphic CAG repeat tract within exon 1 of the *HD* gene that lengthens a glutamine tract in the huntingtin protein (2). The lengthened glutamine tract is thought to confer a novel toxic property on huntingtin that initiates the eventual demise of neurons particularly in the striatum and cortex (3).

Although the mechanism(s) by which mutant huntingtin elicits its toxic effects are not clearly understood, studies of genotype–phenotype relationships in patients have provided critical information regarding the underlying pathogenic process. Central to these studies is the demonstration that the age of disease onset is strongly inversely correlated with the length of the expanded *HD* CAG repeat (4–8), implying that the mechanism(s) that determine onset age are repeat length-dependent. However, although *HD* CAG repeat length

is the overriding factor that determines age at onset, repeat number only accounts for ~70% of the variability in age at onset, and no more than 50% of the variability for the vast majority of HD patients with repeats less than 60 (9,10). Further, there is evidence for strong heritability for that portion of onset age not explained by CAG repeat size (10,11), providing evidence for genetic modifiers of onset age, as demonstrated by several studies (10–22). As modifiers, by definition, alter onset age in patients, delineating these factors provides a direct route to therapies aimed at slowing the pathogenic process.

Interestingly, the expanded *HD* CAG repeat is somatically unstable, undergoing progressive length increases over time (23–27). Somatic instability is tissue-specific, with particularly high levels found in striatum and cortex (25–27) and occurs in post-mitotic neurons (28,29). Furthermore, somatically expanded *HD* CAG repeats are transcribed and translated (29–31). The fact that the *HD* CAG repeat somatically expands in tissues that are the targets of pathogenesis raises

*To whom correspondence should be addressed. Tel: +1 6176433103; Fax: +1 6176433203; Email: wheeler@helix.mgh.harvard.edu

the hypothesis that somatic instability itself contributes to the HD pathogenic process. This is supported by experiments in a genetically accurate Huntington’s disease homologue (*Hdh*) knock-in mouse model (*Hdh*^{Q111}), in which an early presymptomatic, *HD* CAG length-dependent phenotype was significantly delayed in mice that lacked somatic instability as a result of the deletion of mismatch repair genes *Msh2* and *Msh3* (31,32).

Therefore, due to the progressive increases in *HD* CAG repeat length in target tissues, somatic instability may itself be a modifier of the CAG repeat length-dependent pathogenic process in HD patients, beyond the contribution of the constitutional CAG repeat size. Specifically, given the relationship of *HD* CAG repeat length to onset age, and the clear evidence for the existence of additional modifiers of onset age, we have explored whether inter-individual differences in somatic instability in HD patient brain might explain some of the variation in onset age that is unaccounted for by the length of the constitutive *HD* CAG repeat.

RESULTS

HD individuals

The association of onset age with the length of the mutant *HD* CAG repeat predicts that somatic instability, which progressively lengthens the *HD* CAG repeat tract, may influence the onset of HD. We have performed a genetic test of this hypothesis by quantifying *HD* CAG somatic instability in the brains of HD individuals and determining whether somatic instability was associated with the age of neurological onset (motor symptoms) not explained by constitutive *HD* CAG length. Our study design utilized those HD individuals that exhibit phenotypic extremes of young and old onset. These are individuals whose onset ages deviate the most from those predicted by the lengths of their constitutive mutant *HD* CAG repeats, thereby affording the greatest power to detect modifiers of onset. Constitutive *HD* CAG repeat lengths were determined from cerebellar DNA, shown to be somatically stable (27). We classified ‘extreme young’ and ‘extreme old’ onset cases as those in which the residual of the repeat adjusted onset age (that not explained by constitutive repeat length) was <0.5 SD below the mean, or >0.5 SD above the mean, respectively (see Materials and Methods for determination of residuals). We thus identified 48 individuals, 24 with extreme young onset and 24 with extreme old onset. These were closely matched for both mutant and normal constitutive *HD* CAG repeat lengths (*t*-test extreme young versus extreme old: mutant repeat *P*=0.12, wild-type repeat *P*=0.87), but had mean onset ages differing by approximately 30 years (Table 1 and Supplementary Material, Table S1).

SP-PCR profiling of somatic instability in HD cortex

Profiling of *HD* CAG repeat lengths was carried out using small-pool PCR (SP-PCR) amplification of genomic DNA isolated from frontal cortex, dissected from brains obtained at autopsy, of the 48 individuals described above using *HD* CAG repeat-specific primers. SP-PCR is a highly sensitive technique for the quantification of the repeat length

Table 1. Summary of HD samples

	Extreme young	Extreme old
<i>n</i>	24	24
CAG _{mut} : mean (range)	44.5 (40–48)	43.4 (40–47)
CAG _{norm} : mean (range)	18.5 (9–26)	18.6 (10–27)
Age of onset: mean (range)	30 (20–41)	61 (40–81)
Residual: mean (range)	−1.41 (−2.66 to −0.55)	1.66 (0.58–2.33)

Constitutive mutant (CAG_{mut}), normal (CAG_{norm}) *HD* CAG repeat lengths (determined in cerebellar DNA), age of onset and residual onset age (in standard deviations from the mean) in the extreme young (*n*=24) and extreme old (*n*=24) onset phenotypic groups. See Materials and Methods for determination of residuals.

distribution of a population of molecules (33) that has previously revealed the presence of dramatic somatic instability in HD patient brain that was undetected or underestimated using standard PCR amplification of ‘bulk’ genomic DNA (27). Frontal cortex was chosen as this brain region is a target of the HD disease process that has previously been shown to display relatively high levels of somatic instability in end-stage brain, in contrast to striatum, which shows little somatic instability in end-stage brain, presumably due to extensive neuronal cell loss (27). For each cortex sample we determined the length of the *HD* CAG repeat of 100 or more individual mutant alleles, assayed from PCR products of single molecule input DNA. *HD* CAG repeat lengths of the normal alleles were also determined. There was no significant difference in the number of normal and mutant *HD* alleles amplified by SP-PCR (mean difference in number of mutant alleles and number of normal alleles=1.85; paired *t*-test, *P*=0.34). This indicated the absence of bias in the size of allele amplified, that the genomic targets were indeed single molecules, and therefore that the SP-PCR products were an accurate reflection of the *HD* CAG repeat sizes present in the genomic DNA. All SP-PCR data are available in Supplementary Material, Table S1.

Cortex from all 48 individuals exhibited various degrees of somatic instability of the mutant *HD* CAG repeat; most of the repeat length changes were expansions, although somatic contractions also occurred. Interestingly, the normal *HD* CAG repeat also exhibited some somatic instability, although these repeat length changes were smaller and less frequent than mutant *HD* CAG length changes, and were not biased towards expansions. We did not find a correlation between measures of variation of the normal and mutant repeats (data not shown), either due to insufficient variation in the normal repeat, or possibly suggesting that different factors influence normal and mutant repeat lengths.

Figure 1A shows examples of frequency distributions of cortical *HD* CAG repeat lengths for four individuals, two with extreme young onset HD (a, c) and two with extreme old onset HD (b, d), with pairs a and b and pairs c and d matched for constitutive mutant *HD* CAG repeat size.

Figure 1B displays the overall frequencies of contractions (blue bars), unchanged alleles (grey bars) and expansions (red bars; repeat size change ≥1) of the mutant *HD* CAG repeat for the same cortex samples shown in Figure 1A. We also broke down the overall expansion frequency into

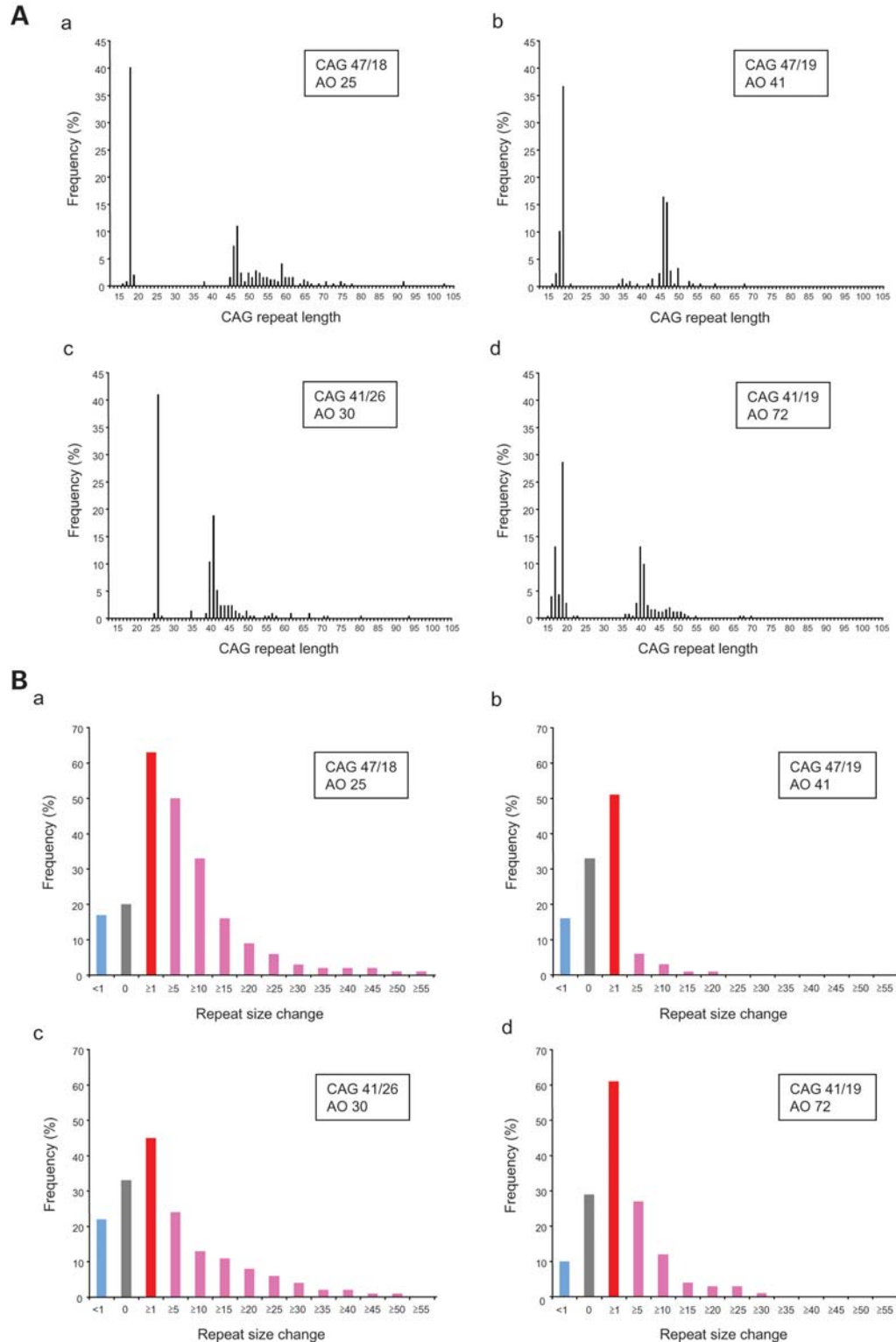


Figure 1. Small-pool PCR analyses of the *HD* CAG repeat amplified from human *HD* cortex. The graphs show examples from four individual cortex samples, two exhibiting extreme young onset (a and c) and two exhibiting extreme old onset (b and d) as predicted from the sizes of their constitutive mutant *HD* CAG repeats, determined in cerebellar DNA. (a) and (b) are matched for mutant *HD* CAG repeat length and (c) and (d) are matched for mutant *HD* CAG repeat length. The inset boxes indicate constitutive *HD* CAG repeat sizes (mutant and normal alleles) and ages of onset (AO). Number of amplified alleles from single molecule inputs: a = 136 mutant, 106 normal; b = 103 mutant, 104 normal; c = 122 mutant, 90 normal; d = 115 mutant, 136 normal. **(A)** Frequency distributions of the *HD* CAG repeat sizes of the mutant and normal alleles. **(B)** Frequency distributions of repeat size changes of the mutant *HD* CAG repeat relative to the modal mutant *HD* CAG repeat size determined in the small-pool PCR analysis. Blue bars: contractions (repeat size change <1); grey bars: unchanged alleles (repeat size change = 0); red bars: all expanded alleles (repeat size change ≥ 1); pink bars: alleles that are expanded beyond specific thresholds (repeat size change ≥ 5, ≥ 10 etc.).

Table 2. Frequencies of mutant *HD* CAG repeat length changes in HD cortex

Repeat length change	Frequency, %
<0	14.6
0	34.6
≥1	50.8
≥5	22.1
≥10	11.1
≥15	7.00
≥20	5.23
≥25	3.44
≥30	1.42
≥35	0.60
≥40	0.38
≥45	0.25
≥50	0.15
≥55	0.06
≥60	0.02

The mutant *HD* CAG repeat was amplified from HD cortex using SP-PCR, and repeat length changes relative to the modal *HD* CAG repeat length were determined. Shown are the mean frequencies, averaged over the 48 cortex samples, of contractions (repeat length change <0), unchanged alleles (repeat length change = 0), total expansions (repeat length change ≥1), and subsets of expansions over thresholds of increasing CAG length (repeat length changes ≥5, ≥10 etc.). The sum of repeat length changes <1, 0 and ≥1 is equal to 100%.

subsets, each representing the frequency of allele size changes that occurred over thresholds of progressively increasing magnitude (pink bars; repeat size change ≥5, ≥10, ≥15 etc.). Frequency data for all 48 individuals are shown in Table 2.

These data (Fig. 1 and Table 2) show that, on average, ~50% of mutant *HD* CAG alleles are expanded by at least one CAG repeat. The data also suggest that, on average, 22% of alleles have expanded by at least five repeats, and 11% of alleles have expanded by at least 10 repeats. Given that an increase in constitutive repeat size of 10 repeats (from ~45 to ~55 repeats) is predicted to precipitate onset by ~13 years (from ~40 to ~27 years) (8), one might expect somatic expansions of 10 or more repeats occurring in 11% of cells to have a considerable impact on disease onset. Larger repeat length increases of at least 35 CAGs were rare, occurring, on average, in <1% of mutant alleles. The largest repeat size change that we observed was 68 CAGs. The frequency and magnitude of repeat length changes that we have observed by profiling ~100 mutant molecules is consistent with previously published analyses of somatic instability in HD brain (27–29).

Somatic instability in HD cortex is predictive of onset age

We observed marked inter-individual differences in the degree of somatic instability present in cortex. Our general observation was that extreme young onset individuals tended to have larger somatic expansions than extreme old onset individuals, as exemplified in Figure 1A and B. Therefore, we determined whether inter-individual differences in somatic instability might explain some of the residual variation in age of onset. We first surveyed whether the magnitude and frequency of somatic repeat expansions differed between extreme young and extreme old onset groups. As shown in Table 3,

Table 3. Measures of *HD* CAG repeat expansion in cortex from extreme young onset and extreme old onset HD individuals

	Extreme young	Extreme old
<i>n</i>	24	24
Mean CAG expansion	7.5 (3.0–16.0)	6.5 (2.8–13.0)
Total expansion frequency, %	49.8 (37–85)	51.8 (24–74)
Frequency of expansions ≥10 CAGs, %	11.8 (4–33)	10.3 (3–33)
Maximum CAG expansion	42 (24–68)	29 (12–44)
Skewness	2.9 (1.9–4.4)	2.2 (0.3–3.4)

The mutant *HD* CAG repeat was amplified from HD cortex using SP-PCR, and repeat expansions relative to the modal *HD* CAG repeat length were determined. For each sample, the mean CAG expansion, the total expansion frequency, the frequency of expansions of at least 10 CAGs, the maximum CAG expansion (maximum repeat size-modal repeat size) and the skewness were determined. For each of these parameters, we determined the mean value, averaged over either the 24 extreme young samples or averaged over the 24 extreme old samples, and the range of values within each of these two groups (given in parentheses).

there were no obvious differences between the extreme young and extreme old groups in the mean expansion size, in the total expansion frequency, or the expansion frequency of 10 or more CAGs between the extreme young and extreme old groups. However, there was a marked difference in the magnitude of the average maximum expansion for each group (extreme young: mean 42 CAGs, extreme old: mean 29 CAGs). This suggested that repeat expansions might be biased towards much longer alleles in individuals with earlier disease onset.

The maximum expansion statistic, shown in Table 3, is dependent on a single repeat size measurement in each cortical sample, and is thus subject to error. In order to evaluate statistically whether repeat expansions are biased towards longer alleles in individuals with earlier disease onset, we also used skewness, a measurement of the degree of symmetry of a distribution, as a measure of somatic instability. Skewness uses all of the data from an individual sample rather than just one observation such as the maximum. As repeat length changes were biased towards expansions, their distributions were skewed to the right; we therefore reasoned that cortex samples showing greater expansions would exhibit greater right skewness than those exhibiting smaller repeat expansions.

On average, extreme young individuals exhibited a greater right skewness (positive value) than extreme old individuals (Table 3). As shown in Figure 2, there was a negative correlation ($r = -0.37$) between skewness and residual onset age, indicating an association between greater somatic expansion and earlier disease onset. A simple linear regression revealed that skewness was a significant predictor of residual onset age ($P = 0.009$), with increased right skewness being associated with a lower residual age of onset (beta estimate = -0.71).

Somatic instability has previously been shown to depend on CAG repeat length (25,26,34), prompting our choice of extreme young and extreme old onset individuals who were closely matched for constitutive mutant *HD* CAG repeat length (Table 1). Nevertheless, to ensure that effects of constitutive CAG length did not confound the association, we had

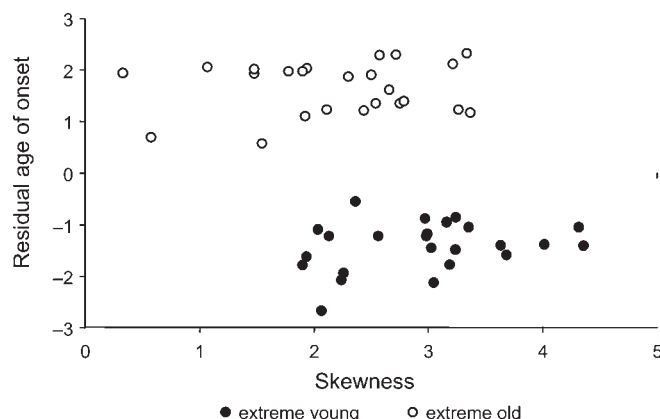


Figure 2. Residual age of onset versus skewness. Shown are the residual onset age and skewness for each of the 24 extreme young onset individuals (filled circles, residuals less than -0.5) and 24 extreme old onset individuals (open circles, residuals >0.5). Skewness values were calculated from the distribution of mutant *HD* CAG repeat lengths in cortex, determined using SP-PCR.

detected between somatic instability and residual onset age, we repeated the regression analysis above, including constitutive mutant *HD* CAG repeat length as a covariate. Controlling for effects of constitutive mutant *HD* CAG repeat length did not alter the association between skewness and residual onset age ($P=0.004$, beta estimate = -0.71).

We also investigated whether there was an association between residual age of onset and somatic instability of the normal allele. Statistical significance was not seen in a simple linear regression between the residual age of onset and skewness of the normal allele ($P=0.4551$), or in a regression model that controlled for the constitutive normal allele ($P=0.5123$).

Therefore, our results demonstrate that larger somatic expansions of the *HD* CAG repeat expansion in HD patient cortex are significantly associated with an earlier age of disease onset, independent of any effects of constitutive CAG repeat length on either somatic instability or onset age.

DISCUSSION

Understanding the factors that modify disease phenotypes in HD patients provides a critical route to therapeutic targets. The age of clinical motor onset, a well-defined phenotypic milestone, that reflects the rate of the HD pathogenic process, is strongly dependent on *HD* CAG repeat length, but is also modified by other environmental and genetic factors, some of which have been identified (10–22). Intriguingly, the *HD* CAG repeat is somatically unstable, progressively increasing in length over time, particularly the brain regions (striatum and cortex) that succumb earliest to disease pathogenesis (25–27). Given the dependence of onset age on *HD* CAG repeat length, this raises the hypothesis that somatic *HD* CAG repeat instability in these tissues accelerates the pathogenic process. Here, we present the first study to explore this hypothesis in HD individuals by rigorously quantifying, at the single molecule level, *HD* CAG repeat lengths in a large number of patient cortex samples, and

correlating the level of somatic instability in cortex with the age of disease onset. We find that, after controlling for effects of constitutive *HD* CAG repeat length on both somatic instability and onset age, somatic instability is significantly associated with onset age, with greater somatic expansions seen with earlier ages of onset. Our findings suggest that factors that contribute to differences in somatic instability between individuals may also be modifiers of disease.

The association between somatic instability and clinical onset does not directly demonstrate that increased somatic instability accelerates the pathogenic process. It is possible that somatic instability is consequence of pathogenesis, as suggested (35), with greater levels of somatic instability simply reflecting a more rapid pathogenic process in individuals with early onset disease. However, several lines of evidence would argue against this possibility. First, high levels of somatic instability in striatum and cortex are seen in other CAG/CTG repeat disorders, notably spinocerebellar ataxia type 1 (SCA1) and myotonic dystrophy type 1 (DM1) (36,37). As the major targets of neurodegeneration in these disorders lie in tissues and brain regions outside striatum and cortex, somatic instability is unlikely to be a consequence of a disease process, but rather to be due to normal tissue-specific factors that are unrelated to disease state. Similarly, greater *HD* CAG somatic expansion is not seen in transgenic mouse models that exhibit dramatic phenotypes in response to the C-terminal fragment of the *HD* gene compared with *Hdh* CAG knock-in mice exhibiting a slow disease course (29). Furthermore, data from our laboratory demonstrate that accelerating the disease process in *Hdh* CAG knock-in mice does not increase somatic instability (J.-M. Lee and V.C. Wheeler, unpublished data).

Rather than instability being a consequence of pathogenesis, data from *Hdh* knock-in mice suggest that instability is a modifier of the pathogenic process (31,32). When somatic instability in the striatum of *Hdh*^{Q111} CAG knock-in mice was eliminated by crossing these mice onto genetic backgrounds deficient in mismatch repair genes *Msh2* or *Msh3*, nuclear mutant huntingtin immunoreactivity in striatal nuclei, a *HD* CAG repeat length-dependent presymptomatic phenotype, was delayed (31,32). Conversely, deficiency of *Msh6* had no effect on somatic instability and did not alter nuclear mutant huntingtin immunoreactivity (32). These data imply that somatic instability contributes to the *HD* CAG pathogenic process in *Hdh* knock-in mice. While additional experiments are needed in the mouse to unambiguously demonstrate the role of somatic instability in the disease process, our data showing an inverse correlation between somatic instability in HD patient cortex and age of disease onset are consistent with, and support the hypothesis, that somatic instability contributes to the pathogenic process in HD.

How significant a role might somatic instability play in modulating the HD pathogenic process? It has recently been proposed that somatic expansion of the *HD* CAG repeat beyond a certain 'pathological threshold' of ~ 115 CAG repeats is required before overt HD symptoms ensue (38), i.e. that somatic expansion is *necessary* for disease onset. This would predict that, starting at the same constitutive repeat length, individuals with more somatic expansion

would reach this threshold earlier than individuals with less somatic expansion, and therefore exhibit earlier disease onset. While our data are consistent with this hypothesis, they cannot distinguish somatic expansion over a threshold as the instigator of disease, as proposed by Kaplan *et al.* (38), from somatic instability as modifier of a disease process that would proceed even in its absence. These possibilities may best be addressed by determining the effect of somatic instability on phenotypes in accurate genetic mouse models of HD containing repeat lengths that are below the predicted 'pathological threshold'.

It is difficult to determine from our study the possible contribution of somatic instability to onset age. As our HD sample included only those individuals displaying phenotypic extremes of young and old onset, the proportion of the variation of onset age that is accounted for by somatic instability is not representative of the HD population as a whole. In addition, as neurons with the longest repeat expansions may be preferentially lost during the disease course (27), the HD CAG repeat length distribution in end-stage brain may not be an accurate reflection of that present at disease onset. Therefore, we believe that the association between somatic instability and onset age in our study is likely an underestimate of the true association if somatic instability could be measured in the brain at the time disease onset.

While analyses of HD CAG repeat lengths in brains of rare presymptomatic HD gene carriers (27) provide insight into the extent of somatic instability preceding the onset of symptoms, they do not allow an estimate of the correlation with disease onset. It may, however, be possible to correlate measures of somatic instability in other tissues, such as blood/lymphoblasts, fibroblasts or buccal cells with onset age (34,39). Interestingly, as in the present study, analyses of the variation in buccal cell somatic instability between individuals provided evidence for modifiers of instability other than constitutive HD CAG repeat length (34). Thus, if an individual's propensity for somatic instability in the brain is reflected in peripheral tissues, it is possible that somatic instability in these tissues could be used as a surrogate for CAG repeat instability in the brain.

It would also be interesting to determine whether somatic instability is a predictor of additional clinical markers of disease, particularly early phenotypes that precede the overt onset of motor symptoms, e.g. cognitive and psychiatric symptoms and cortical changes (40–42). The expectation is that any phenotypes that are CAG repeat length-dependent would be modifiable by somatic instability, and would therefore show an association with somatic instability.

In summary, we suggest that somatic instability is a modifier of the HD pathogenic process. This predicts that factors that determine somatic instability in HD patients will also modify disease pathogenesis, and conversely, that disease modifiers may also influence somatic instability. Somatic instability is also predicted to alter disease phenotypes in other trinucleotide repeat disorders in which somatic instability is prevalent in tissues affected by the disorder, notably muscle in DM1 (43). Profiling somatic repeat instability over a large number of HD patient samples, as we have begun in this HD study, may provide a resource to identify genetic factors that determine somatic instability in humans.

Interestingly, we observed inter-individual differences in the instability of the normal HD CAG repeat, as well as in the mutant HD CAG repeat (Fig. 1A). Factors such as *Msh2*, *Msh3* and *Ogg1*, found to be important in generating somatic instability of an expanded HD CAG repeat in mouse models of HD (31,32,44,45,46), provide sources of candidate genes predicted to alter both somatic instability and age of onset in HD patients. Thus, with the aim of slowing the pathogenic process that leads to this destructive disease, our data support pursuing somatic instability as a novel and viable therapeutic target.

MATERIALS AND METHODS

HD postmortem brain samples

All work was conducted under an approved IRB protocol. Autopsy tissue from the frontal cortex and cerebellum of HD individuals with known age of neurological onset (motor symptoms) was obtained from the Harvard Brain Tissue Resource Center (Belmont, MA, USA) and the National Neurological Research Bank (Los Angeles, CA, USA).

Determination of extreme young and extreme old onset individuals

We identified HD individuals with 'extreme young' and 'extreme old' ages of motor onset as predicted from their constitutive mutant HD CAG repeat lengths. Constitutive HD CAG repeat sizes were determined from cerebellar DNA using a PCR assay that does not amplify the adjacent proline (CCG) tract (47). Regression analysis was performed using a natural log transformation of age at onset as the dependent variable and the size of the expanded HD CAG repeat as the independent variable [$\log(\text{onset}) = \alpha + \beta(\text{HD})\text{CAG}$], as previously described (19). Random effect models (*Proc MIXED* in SAS) were used to account for familial clustering. The regression model was used to determine the expected age at onset for a given expanded CAG repeat, and the residual was computed as the difference between the observed and expected age at onset. Residuals were standardized to a mean of zero and an SD of 1. We classified cases as extreme young when the standardized residual of the repeat adjusted onset age was >0.5 SD below the mean. Similarly, we classified individuals as extreme old when the standardized residual of the repeat adjusted onset age was >0.5 SD above the mean. We then identified 24 extreme young individuals and 24 extreme old individuals that exhibited the largest residuals while having the most closely matched mutant HD CAG repeat lengths (Table 1).

SP-PCR profiling of cortical somatic instability

DNA was extracted from 100 to 200 mg of cortical tissue using the 5' Genomic DNA extraction kit (Qiagen) according to the manufacturer's protocols. A sensitive nested SP-PCR assay was optimized that allowed resolution of HD CAG repeat-containing alleles derived from the genomic equivalent of a single cell. The SP-PCR assay did not amplify the proline (CCG) tract adjacent to the CAG repeat. Genomic DNA was

initially digested with *Hind*III (Roche) at a final DNA concentration of 20 ng/μl. The DNA was then serially diluted to a theoretical average of 1 molecule per amplification reaction using dilution buffer (1 × TE containing 1 μM LKH-1 primer as a carrier). A Poisson distribution was assumed, and if necessary, the DNA dilution was adjusted for each sample in order to generate PCR products from single molecules.

A first round PCR was carried out in a reaction mix containing the primers LKH-1 (5'-CCCATTTCATGCCCCGGTGCTG-3', 0.5 μM) and LKH-5 (5'-TGGGTTGCTGGGTCACCTCTGTC-3', 0.5 μM), Custom 1 × Mastermix [45 mM Tris-HCl, pH 8.8, 11 mM (NH₄)₂SO₄, 4.5 mM MgCl₂, 4.4 μM EDTA, 1 mM dNTPs, 113 μg/ml BSA] (ABgene), 10% DMSO (Sigma-Aldrich) and 1 U *Taq* polymerase (Fisher) in a final volume of 10 μl. The cycling conditions consisted of an initial denaturation step of 5 min at 94°C, followed by 30 cycles of 30 s at 94°C, 30 s at 65°C, 90 s at 70°C and a final extension of 10 min at 70°C. The PCR products from the first round reaction were diluted 1/100 in water and 1 μl of the dilution was used as a template for the second round reaction. The reaction mix contained the primers CAG1 (5'-ATGAAGGCCTTCGAGTCCCTCAAGTCCTTC-3', 0.125 μM) and CAG2 (5'-GGCGGTGGCGGCTGTTGCTGCTGCTGCTGC-3', 0.125 μM), 1 × AmpliTaq Gold Buffer II (Applied Biosystems), 1.25 mM MgCl₂ (Applied Biosystems), 0.25 mM dNTPs (Invitrogen), 12% DMSO (Sigma-Aldrich) and 1 U AmpliTaq Gold polymerase (Applied Biosystems). The cycling conditions were an initial denaturation of 5 min at 94°C, followed by 28 cycles of 30 s at 94°C, 30 s at 60°C, 90 s at 70°C and a final extension of 10 min at 70°C. The forward primer in the second round amplification was fluorescently labeled with 6-FAM (Perkin Elmer), allowing resolution of the PCR products and accurate quantification of CAG repeat length using an automated ABI3730XL sequencer and GeneMapper v.3.7 software with GS 500 LIZ internal size standard (Applied Biosystems). Control samples of known *HD* CAG repeat length were included in every run. In order to control for contamination, PCR reactions were set up in a laminar flow hood and 20% of the SP-PCRs set up per run were negative controls without any DNA as a template. The PCR products obtained from single molecule inputs comprised a cluster of peaks, each differing by one CAG repeat, due to PCR stutter. Allele size was assigned to be the highest peak in the cluster.

We found that there was a small degree of variability in the allele size called by the ABI software (allele size calling using the highest peak), such that the repeat size of the same sample, run at different times, sometimes differed by one CAG. For this reason, for our analyses of repeat size change of alleles amplified by SP-PCR, we compared repeat lengths to the modal repeat length determined in the SP-PCR, to ensure internal consistency. The modal repeat length in cortex and the constitutive repeat size, determined in cerebellum, did not differ by more than a single CAG (see Supplementary Material, Table S1).

Statistical analyses

To estimate the association between age of onset and somatic instability we ran a regression with residual age of onset (see

Materials and Methods section) as the outcome and skewness as the predictor.

Skewness is a measure of a distribution's symmetry. In SAS v9.1, the sample skewness of a distribution is found through the following formula:

$$\text{skewness} = \left(\frac{n}{n-2} \right) \cdot \left(\frac{1}{n-1} \right) \sum_{i=1}^n \left[\frac{y_i - \bar{y}}{s} \right]^3$$

where *s* is the sample standard deviation of *y*. A perfectly symmetrical distribution will result in a skewness of zero while a distribution with a long left tail will have a negative skewness and a distribution with a long right tail will have a positive skewness. For example, Figure 1Aa shows a brain with a long right tail and thus a positive skewness. We used skewness to measure somatic instability towards increasing CAG lengths because (i) skewness uses all of the observations in its calculation rather than just a one measure such as the maximum and (ii) skewness is more sensitive to the lengthened tail of a distribution than a central measurement such as mean or median.

As both age of onset and somatic instability are dependent on constitutive CAG length, we controlled for CAG length by (i) matching early and late onset brains by CAG repeat length, (ii) using the residual age of onset from a model based on CAG length and (iii) running an additional regression model with CAG repeat length as a predictor. All regression models were run using the generalized linear model procedure in SAS v9.1. Significance of all tests of association was assessed at the 0.05 level of significance.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We would like to thank Dr Marcy MacDonald for critical reading of the manuscript.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the National Institutes of Health [NS049206 to V.C.W., P50 NS16367-28 'Huntington Disease Center Without Walls' to R.H.M.], and the Jerry MacDonald HD Research Fund.

REFERENCES

- Harper, P.S. (1999) Huntington's disease: a clinical, genetic and molecular model for polyglutamine repeat disorders. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.*, **354**, 957–961.
- 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.
- Vonsattel, J.P., Myers, R.H., Stevens, T.J., Ferrante, R.J., Bird, E.D. and Richardson, E.P. (1985) Neuropathological classification of Huntington's disease. *J. Neuropathol. Exp. Neurol.*, **44**, 559–577.

4. Duyao, M., Ambrose, C., Myers, R., Novelletto, A., Persichetti, F., Frontali, M., Folstein, S., Ross, C., Franz, M., Abbott, M. *et al.* (1993) Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nat. Genet.*, **4**, 387–392.
5. Andrew, S.E., Goldberg, Y.P., Kremer, B., Telenius, H., Theilmann, J., Adam, S., Starr, E., Squitieri, F., Lin, B., Kalchman, M.A. *et al.* (1993) The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. *Nat. Genet.*, **4**, 398–403.
6. Snell, R.G., MacMillan, J.C., Cheadle, J.P., Fenton, I., Lazarou, L.P., Davies, P., MacDonald, M.E., Gusella, J.F., Harper, P.S. *et al.* (1993) Relationship between trinucleotide repeat expansion and phenotypic variation in Huntington's disease. *Nat. Genet.*, **4**, 393–397.
7. Stine, O.C., Pleasant, N., Franz, M.L., Abbott, M.H., Folstein, S.E. and Ross, C.A. (1993) Correlation between the onset age of Huntington's disease and length of the trinucleotide repeat in IT-15. *Hum. Mol. Genet.*, **2**, 1547–1549.
8. Gusella, J.F., McNeil, S., Persichetti, F., Srinidhi, J., Novelletto, A., Bird, E., Faber, P., Vonsattel, J.P., Myers, R.H. and MacDonald, M.E. (1996) Huntington's disease. *Cold Spring Harb. Symp. Quant. Biol.*, **61**, 615–626.
9. Myers, R., Marans, K.S. and MacDonald, M. (1998) Huntington's disease. In Wells, R.D. and Warren, S.T. (eds) *Genetic Instabilities and Hereditary Neurological Diseases*. Academic Press, San Diego, CA, USA, pp. 301–323.
10. Li, J.L., Hayden, M.R., Almqvist, E.W., Brinkman, R.R., Durr, A., Dode, C., Morrison, P.J., Suchowersky, O., Ross, C.A., Margolis, R.L. *et al.* (2003) A genome scan for modifiers of age at onset in Huntington disease: the HD MAPS Study. *Am. J. Hum. Genet.*, **73**, 682–687.
11. Wexler, N.S., Lorimer, J., Porter, J., Gomez, F., Moskowitz, C., Shackell, E., Marder, K., Penchaszadeh, G., Roberts, S.A., Gayan, J. *et al.* (2004) Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington's disease age of onset. *Proc. Natl. Acad. Sci. USA*, **101**, 3498–3503.
12. Rubinsztein, D.C., Leggo, J., Chiano, M., Dodge, A., Norbury, G., Rosser, E. and Craufurd, D. (1997) Genotypes at the GluR6 kainate receptor locus are associated with variation in the age of onset of Huntington disease. *Proc. Natl. Acad. Sci. USA*, **94**, 3872–3876.
13. MacDonald, M.E., Vonsattel, J.P., Shrinidhi, J., Couropmitree, N.N., Cupples, L.A., Bird, E.D., Gusella, J.F. and Myers, R.H. (1999) Evidence for the GluR6 gene associated with younger onset age of Huntington's disease. *Neurology*, **53**, 1330–1332.
14. Holbert, S., Denghien, I., Kiechle, T., Rosenblatt, A., Wellington, C., Hayden, M.R., Margolis, R.L., Ross, C.A., Dausset, J., Ferrante, R.J. *et al.* (2001) The Gln-Ala repeat transcriptional activator CA150 interacts with huntingtin: neuropathologic and genetic evidence for a role in Huntington's disease pathogenesis. *Proc. Natl. Acad. Sci. USA*, **98**, 1811–1816.
15. Naze, P., Vuillaume, I., Destee, A., Pasquier, F. and Sablonniere, B. (2002) Mutation analysis and association studies of the ubiquitin carboxy-terminal hydrolase L1 gene in Huntington's disease. *Neurosci. Lett.*, **328**, 1–4.
16. Chattopadhyay, B., Ghosh, S., Gangopadhyay, P.K., Das, S.K., Roy, T., Sinha, K.K., Jha, D.K., Mukherjee, S.C., Chakraborty, A., Singhal, B.S. *et al.* (2003) Modulation of age at onset in Huntington's disease and spinocerebellar ataxia type 2 patients originated from eastern India. *Neurosci. Lett.*, **345**, 93–96.
17. Djousse, L., Knowlton, B., Hayden, M.R., Almqvist, E.W., Brinkman, R.R., Ross, C.A., Margolis, R.L., Rosenblatt, A., Durr, A., Dode, C. *et al.* (2004) Evidence for a modifier of onset age in Huntington disease linked to the HD gene in 4p16. *Neurogenetics*, **5**, 109–114.
18. Metzger, S., Bauer, P., Tomiuk, J., Laccone, F., Didonato, S., Gellera, C., Soliveri, P., Lange, H.W., Weirich-Schwaiger, H., Wenning, G.K. *et al.* (2006) The S18Y polymorphism in the UCHL1 gene is a genetic modifier in Huntington's disease. *Neurogenetics*, **7**, 27–30.
19. Li, J.L., Hayden, M.R., Warby, S.C., Durr, A., Morrison, P.J., Nance, M., Ross, C.A., Margolis, R.L., Rosenblatt, A., Squitieri, F. *et al.* (2006) Genome-wide significance for a modifier of age at neurological onset in Huntington's disease at 6q23-24: the HD MAPS study. *BMC Med. Genet.*, **7**, 71.
20. Andresen, J.M., Gayan, J., Cherny, S.S., Brocklebank, D., Alkorta-Aranburu, G., Addis, E.A., Cardon, L.R., Housman, D.E. and Wexler, N.S. (2007) Replication of twelve association studies for Huntington's disease residual age of onset in large Venezuelan kindreds. *J. Med. Genet.*, **44**, 44–50.
21. Metzger, S., Rong, J., Nguyen, H.P., Cape, A., Tomiuk, J., Soehn, A.S., Propping, P., Freudenberg-Hua, Y., Freudenberg, J., Tong, L. *et al.* (2008) Huntingtin-associated protein-1 is a modifier of the age-at-onset of Huntington's disease. *Hum. Mol. Genet.*, **17**, 1137–1146.
22. Gayan, J., Brocklebank, D., Andresen, J.M., Alkorta-Aranburu, G., Zameel Cader, M., Roberts, S.A., Cherny, S.S., Wexler, N.S., Cardon, L.R. and Housman, D.E. (2008) Genomewide linkage scan reveals novel loci modifying age of onset of Huntington's disease in the Venezuelan HD kindreds. *Genet. Epidemiol.*, **32**, 445–453.
23. Telenius, H., Kremer, B., Goldberg, Y.P., Theilmann, J., Andrew, S.E., Zeisler, J., Adam, S., Greenberg, C., Ives, E.J. and Clarke, L.A. (1994) Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. *Nat. Genet.*, **6**, 409–414.
24. De Rooij, K.E., De Koning Gans, P.A., Roos, R.A., Van Ommen, G.J. and Den Dunnen, J.T. (1995) Somatic expansion of the (CAG)_n repeat in Huntington disease brains. *Hum. Genet.*, **95**, 270–274.
25. Wheeler, V.C., Auerbach, W., White, J.K., Srinidhi, J., Auerbach, A., Ryan, A., Duyao, M.P., Vrbanc, 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.
26. Kennedy, L. and Shelbourne, P.F. (2000) 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.
27. Kennedy, L., Evans, E., Chen, C., Craven, L., Detloff, P., Ennis, M. and Shelbourne, P. (2003) Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. *Hum. Mol. Genet.*, **12**, 3359–3367.
28. Shelbourne, P.F., Keller-McGandy, C., Bi, W.L., Yoon, S.R., Dubeau, L., Veitch, N.J., Vonsattel, J.P., Wexler, N.S., Arnheim, N. 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.
29. Gonitell, R., Moffitt, H., Sathasivam, K., Woodman, B., Detloff, P.J., Faull, R.L. and Bates, G.P. (2008) DNA instability in postmitotic neurons. *Proc. Natl. Acad. Sci. USA*, **105**, 3467–3472.
30. Aronin, N., Chase, K., Young, C., Sapp, E., Schwarz, C., Matta, N., Kornreich, R., Landwehrmeyer, B., Bird, E. and Beal, M.F. (1995) CAG expansion affects the expression of mutant Huntingtin in the Huntington's disease brain. *Neuron*, **15**, 1193–1201.
31. Wheeler, V.C., Lebel, L.A., Vrbanc, V., Teed, A., Te Riele, H. and MacDonald, M.E. (2003) Mismatch repair gene Msh2 modifies the timing of early disease in Hdh(Q111) striatum. *Hum. Mol. Genet.*, **12**, 273–281.
32. Dragileva, E., Hendricks, A., Teed, A., Gillis, T., Lopez, E.T., Friedberg, E.C., Kucherlapati, R., Edelmans, 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.
33. Gomes-Pereira, M., Bidichandani, S.I. and Monckton, D.G. (2004) Analysis of unstable triplet repeats using small-pool polymerase chain reaction. *Methods Mol. Biol.*, **277**, 61–76.
34. Veitch, N.J., Ennis, M., McAbney, J.P., Shelbourne, P.F. and Monckton, D.G. (2007) Inherited CAG-CTG allele length is a major modifier of somatic mutation length variability in Huntington disease. *DNA Repair*, **6**, 789–796.
35. Jung, J. and Bonini, N. (2007) CREB-binding protein modulates repeat instability in a Drosophila model for polyQ disease. *Science*, **315**, 1857–1859.
36. Fortune, M.T., Vassilopoulos, C., Coolbaugh, M.I., Siciliano, M.J. and Monckton, D.G. (2000) Dramatic, expansion-biased, age-dependent, tissue-specific somatic mosaicism in a transgenic mouse model of triplet repeat instability. *Hum. Mol. Genet.*, **9**, 439–445.
37. Watase, K., Venken, K.J., Sun, Y., Orr, H.T. and Zoghbi, H.Y. (2003) Regional differences of somatic CAG repeat instability do not account for selective neuronal vulnerability in a knock-in mouse model of SCA1. *Hum. Mol. Genet.*, **12**, 2789–2795.
38. Kaplan, S., Itzkovitz, S. and Shapiro, E. (2007) A universal mechanism ties genotype to phenotype in trinucleotide diseases. *PLoS Comput. Biol.*, **3**, e235.
39. Cannella, M., Maglione, V., Martino, T., Ragona, G., Frati, L., Li, G. and Squitieri, F. (2009) DNA instability in replicating Huntington's disease lymphoblasts. *BMC Med. Genet.*, **10**, 11.
40. Verry, C., Allain, P., Prudean, A., Malinge, M.C., Gohier, B., Scherer, C., Bonneau, D., Dubas, F. and Le Gall, D. (2007) Cognitive changes in

- asymptomatic carriers of the Huntington disease mutation gene. *Eur. J. Neurol.*, **14**, 1344–1350.
41. Paulsen, J.S., Langbehn, D.R., Stout, J.C., Aylward, E., Ross, C.A., Nance, M., Guttman, M., Johnson, S., MacDonald, M., Beglinger, L.J. *et al.* (2008) Detection of Huntington's disease decades before diagnosis: the Predict-HD study. *J. Neurol. Neurosurg. Psychiatr.*, **79**, 874–880.
 42. Rosas, H.D., Salat, D.H., Lee, S.Y., Zaleta, A.K., Pappu, V., Fischl, B., Greve, D., Hevelone, N. and Hersch, S.M. (2008) Cerebral cortex and the clinical expression of Huntington's disease: complexity and heterogeneity. *Brain*, **131**, 1057–1068.
 43. Ashizawa, T., Dubel, J.R. and Harati, Y. (1993) Somatic instability of CTG repeat in myotonic dystrophy. *Neurology*, **43**, 2674–2678.
 44. Owen, B.A., Yang, Z., Lai, M., Gajek, M., Badger, J.D. 2nd, 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.
 45. Kovtun, I.V., Liu, Y., Bjoras, M., Klungland, A., Wilson, S.H. and McMurray, C.T. (2007) OGG1 initiates age-dependent CAG trinucleotide expansion in somatic cells. *Nature*, **447**, 447–452.
 46. Manley, K., Shirley, T.L., Flaherty, L. and Messer, A. (1999) Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nat. Genet.*, **23**, 471–473.
 47. Warner, J.P., Barron, L.H. and Brock, D.J. (1993) A new polymerase chain reaction (PCR) assay for the trinucleotide repeat that is unstable and expanded on Huntington's disease chromosomes. *Mol. Cell Probes*, **7**, 235–239.