

ARTICLE

Instability of expanded CAG/CAA repeats in spinocerebellar ataxia type 17

Rui Gao^{1,7}, Tohru Matsuura^{2,7}, Mary Coolbaugh^{3,7}, Christine Zühlke⁴, Koichiro Nakamura⁵, Astrid Rasmussen⁶, Michael J Siciliano³, Tetsuo Ashizawa¹ and Xi Lin^{*1}

¹Department of Neurology, The University of Texas Medical Branch, Galveston, TX, USA; ²Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; ³Department of Molecular Genetics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; ⁴Institute of Human Genetics, University of Lubeck, Lubeck, Germany; ⁵Department of Neurology, The University of Tokyo School of Medicine, Tokyo, Japan; ⁶Department of Neurogenetics and Molecular Biology, Instituto Nacional de Neurología y Neurocirugía, México City, D.F. México

Trinucleotide repeat expansions are dynamic mutations causing many neurological disorders, and their instability is influenced by multiple factors. Repeat configuration seems particularly important, and pure repeats are thought to be more unstable than interrupted repeats. But direct evidence is still lacking. Here, we presented strong support for this hypothesis from our studies on spinocerebellar ataxia type 17 (SCA17). SCA17 is a typical polyglutamine disease caused by CAG repeat expansion in *TBP* (TATA binding protein), and is unique in that the pure expanded polyglutamine tract is coded by either a simple configuration with long stretches of pure CAGs or a complex configuration containing CAA interruptions. By small pool PCR (SP-PCR) analysis of blood DNA from SCA17 patients of distinct racial backgrounds, we quantitatively assessed the instability of these two types of expanded alleles coding similar length of polyglutamine expansion. Mutation frequency in patients harboring pure CAG repeats is 2–3 folds of those with CAA interruptions. Interestingly, the pure CAG repeats showed both expansion and deletion while the interrupted repeats exhibited mostly deletion at a significantly lower frequency. These data strongly suggest that repeat configuration is a critical determinant for instability, and CAA interruptions might serve as a limiting element for further expansion of CAG repeats in SCA17 locus, suggesting a molecular basis for lack of anticipation in SCA17 families with interrupted CAG expansion.

European Journal of Human Genetics (2008) 16, 215–222; doi:10.1038/sj.ejhg.5201954; published online 28 November 2007

Keywords: spinocerebellar ataxia [298]; trinucleotide repeat diseases [98]; cerebellum [312]

Introduction

Trinucleotide repeat expansion was uncovered as the genetic mutation for a number of hereditary neurological disorders in recent years.^{1–3} These disorders include

various neurodevelopmental, neurodegenerative, and neuromuscular diseases, such as Fragile X syndrome, Friedreich's ataxia, Huntington's disease, several spinocerebellar ataxias, and myotonic dystrophy. What is unique about these repeat expansions is their dynamic nature, that is changes of repeat number in both germline and somatic tissues. Repeat instability in germline is thought to be the basis for anticipation, a phenomenon characterized by younger age of onset and worse progression of disease in successive generations; whereas repeat instability in somatic tissues is believed to be important for

*Correspondence: Dr X Lin, Department of Neurology, The University of Texas Medical Branch, 301 University Blvd., Galveston, TX 77555-0653, USA. Tel: +1 409 772 2466; Fax: +1 409 747 4562; E-mail: xilin@utmb.edu

⁷Both the authors have contributed equally.

Received 7 August 2007; revised 9 October 2007; accepted 10 October 2007; published online 28 November 2007

variable expression of disease phenotypes in tissue- and age-dependent manners.^{4,5}

Multiple factors are implicated in trinucleotide repeat instability, which are largely divided into two categories. One is characterized as the *trans*-acting factors, which are often involved in DNA replication and repair, such as FEN1, Msh2, Msh3 and Msh6.^{6–9} Deficiency of these *trans*-acting factors results in genomic instability and cancer formation.¹⁰ The other category reflects the local properties of the loci, such as the proximity of CpG islands,¹¹ the orientation of adjacent replication origin,¹² and the configuration and copy number of repeats.¹³ It is generally accepted that the higher the number of trinucleotide repeat is, the more likely the repeat expansion is unstable.¹⁴ As for the configuration of repeat expansion, it has been suggested that pure trinucleotide repeats are more prone to expand and/or contract than repeats interrupted by distinct trinucleotides. However, to date, this notion is solely based on observations of repeat size changes during a limited number of parent-to-child transmissions, and there are few direct demonstrations of allele instability at the tissue level.^{15–17}

Spinocerebellar ataxia type 17 (SCA17) is caused by expansion of CAG trinucleotide repeats in exon 3 of the TATA binding protein (*TBP*) gene at chromosome 6q27.¹⁸ Patients with SCA17 exhibit cerebellar ataxia, pyramidal and extrapyramidal signs, cognitive impairments, psychosis, and seizures.^{18–20} It is of note that anticipation is not well documented for SCA17. The CAG expansion locus in SCA17 is polymorphic with a configuration of (CAG)₃ (CAA)₃ (CAG)_{n1} CAACAGCAA (CAG)_{n2} CAACAG, where *n1* ranges from 7 to 11 and *n2* from 9 to 21. This element has been dissected into five domains, that is, domain I = (CAG)₃ (CAA)₃; domain II = (CAG)_{n1}; domain III = CAACAGCAA; domain IV = (CAG)_{n2}, and domain V = CAACAG. Since both CAG and CAA in this repeat tract code for glutamine, the complex repeat array still encodes pure polyglutamine tract. Normal alleles code for 25–44 tandemly repeated glutamines, whereas mutant alleles code for expanded glutamine tracts in the range of 47–63, and reduced penetrance is seen in the middle range.^{19,21–23} Like other polyglutamine diseases,²⁴ SCA17 is thought to be caused by a gain-of-function mechanism due to the expanded polyglutamine tract within *TBP*, a general transcription factor that binds the basic promoter element, TATA box.

CAG repeats in SCA17 locus have two distinct configurations, which are differentiated by the absence or presence of domain III. Type I configurations are more complex with a domain III, and more prevalent. They have been reported in Japanese, German, and French patients. CAG repeat expansion usually occurs in domain IV ranging from 26 to 31, whereas domain II remains normal.^{18–20} Other more complicated patterns have also been observed, such as (CAG)₃ (CAA)₃ (CAG)₉ CAACAGCAA (CAG)₁₆ CAACAGCAA (CAG)₁₃ CAACAG and (CAG)₃(CAA)₃

(CAG)₉ CAACAGCAA (CAG)₉ (CAA)₃ (CAG)₉ CAACAGCAA (CAG)₁₉ CAACAG.^{18,25} This is presumably caused by unequal crossover during recombination. Type II configurations are not as frequent as type I, and have no domain III. As a result, domains II and VI are combined as longer pure CAG repeats. They have been reported in German and Italian families.^{19,26} It is of note that intergenerational instability and anticipation were documented in these families.

Small pool PCR (SP-PCR) is a highly sensitive assay in detecting low levels of somatic mosaicism. Through dilution of genomic DNA down to single genome equivalent, SP-PCR amplifies the alleles from single- or oligocopies of DNA templates. Thus, a large number of alleles can be analyzed individually, and the variant repeats can be detected by ABI 377 and GeneScan analysis. Here, we used SP-PCR to compare the instability of expanded CAG repeats of seven SCA17 families with either type I or type II configurations. Our results showed that instability of expanded CAG repeat is highly dependent on repeat configuration, and CAA interruption is a limiting factor for further CAG repeat expansion.

Materials and methods

Genotyping

Blood samples were obtained from SCA17 patients and age- and gender-matched controls. All these individuals have signed the consent form approved by the Local Institutional Review Boards. DNA was extracted from peripheral blood leukocytes using a standard chloroform/phenol extraction protocol. To determine the progenitor allele size of the SCA17 CAG/CAA repeat, the DNA from each individual was genotyped by PCR as described previously.¹⁸

DNA sample preparation for small pool PCR

The amount of input DNA in SP-PCR is critical for quantitative analysis of repeat size mutants. Thus, we have developed a systematic protocol to determine the number of amplifiable alleles in the input DNA. First, DNA concentration was estimated by fluorometer, SP-PCR (see below) was then performed using the samples that contain 6000, 600, 300, and 60 pg of genomic DNA. By Poisson analysis of amplified alleles, the genome equivalent (g.e.) of the amplifiable DNA amount was calculated. Using 1–10 g.e. of input DNA, multiple identical SP-PCRs were performed in 96-well plates to screen at least 180 alleles (Table 2). At least 8 Zero DNA controls (reaction mixture containing all reagents except for the template DNA) were included in each 96-well plate to monitor DNA contamination.

Small pool PCR

SP-PCR utilizes diluted DNA samples as the template. We developed a hemi-nested PCR protocol to detect

fluorescent signals from amplified product. The first PCR was performed with the following primer set: forward primer, HTFIID-A: 5'-ATGCCTTATGGCACTGGACTGACC-3' and reverse primer, HTFIID-B: 5'-CTGCTGGGACGTTGACTGCTGAAC-3'. Genomic DNA was first amplified in a 10 μ l reaction mixture containing 250 μ M dNTPs, 10 pmol of each primer, 1 μ l of 10 \times PCR buffer, and 1 U of Taq DNA polymerase. Initial DNA denaturation was carried out at 95°C for 6 min 45 s, which was followed by 30 PCR cycles of 1 min 15 s at 94°C, 1 min, 15 s at 58°C, 1 min, 35 s at 70°C, and final extension at 70°C for 25 min. Expected size of the PCR product was 120 + 3*n* bp, where *n* was the number of CAG or CAA repeat units.

For the primers in secondary PCR, we used HTFIID-B and a nested forward primer labeled with the fluorescent moiety, 6FAM. The sequence of the nested primer is FAM-5'-GTCTATTTTGGGAAGCAACAAAGG-3'. This fluorescent tag allows detection by the ABI 377 Sequencer. The primary PCR product was first diluted 10-folds, then adding 2 μ l of diluted initial PCR product to the secondary PCR mixture, for a final dilution of initial PCR of 50-fold. The same PCR conditions as described above were used for the secondary PCR. Expected PCR product is (67 + 3*n*) bp, where *n* was the number of CAG or CAA repeat units.

SP-PCR involves extensive amplification reactions, and consists of a primary PCR followed by hemi-nested secondary PCR across the repeat region. We rigorously controlled for contamination by including 'zero DNA controls' in every gel. In SP-PCR, as in regular PCR analysis of most trinucleotide repeat alleles, an amplification of a single stable normal allele results in an artifactual stuttering pattern, which consists of a large band of high intensity trailed by 2–4 smaller bands with decreasing intensity (Figure 1a). In a previous study on SP-PCR, we have demonstrated that the tallest peak corresponds to the actual allele and the smaller peaks are due to PCR artifacts.²⁷ When two DNA samples with different repeat sizes were mixed and analyzed by small pool PCR, the stuttering pattern was altered. If the size of allele differs by one repeat, the smaller allele is detected consistently with the higher intensity, as seen in Figures 1b and c, where the expanded progenitor allele is accompanied by mutant alleles with +1 and -1 repeat, respectively (Figure 1). When the size of alleles differed by two or more repeat units, two distinct peaks were observed. Figure 1d showed a mutant allele with two more repeats simultaneously amplified with the expanded progenitor allele. Thus, SP-PCR is able to differentiate repeat size variations from the stuttering artifacts.

TOPO cloning

We performed regular PCR using primers described above on each patient's blood DNA. High-fidelity Pfu polymerase was used, and the PCR product cloned into TOPO cloning vector as per the manufacturer's instruction manual

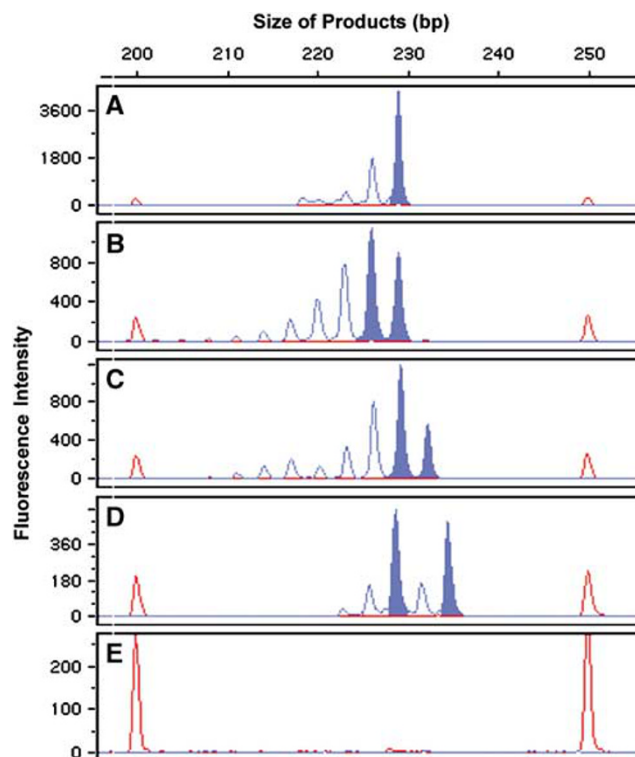


Figure 1 Small Pool PCR (SP-PCR) analysis of CAG/CAA repeat expansion in *sca17* locus. (a) An expanded progenitor allele of 229 bp (54 CAG/CAA repeats, filled with blue), is trailed by a series of progressively smaller peaks. (b) The expanded allele is coamplified with a mutant allele, which has a deletion of one repeat unit. (c) The expanded allele is coamplified with a mutant allele, which has a further expansion of one repeat unit. (d) The expanded allele is coamplified with a mutant allele, which has a further expansion of two repeat units. (e) Zero-DNA control. The red peaks are DNA standards of 200 and 250 bp. The unfilled blue peaks are PCR stutters.

(Invitrogen). Ten clones were picked for sequencing analysis for each patient.

Statistical analysis

Each gel contains SP-PCR product from SCA17 patients and controls amplified and analyzed simultaneously. The GeneScan analysis was performed to detect the progenitor alleles and mutant alleles. The allele and mutant frequencies for each patient are scored and computed in the following programs for statistical analysis:²⁸

- (1) Calibration estimation for Small-Pool PCR: The 'Calibrate' program (by BW Brown, UT-MDACC, Department of Biomathematics) uses SP-PCR results to quantify the DNA. The program uses the Poisson distribution of number of alleles observed in each reaction (0, 1, or 2+) combined with the maximum likelihood determination that any one observed allele has more than one copy of that allele present in the reaction. The program estimates the observed amount

of DNA present in that set of reactions (relative to the amount of expected DNA) and the 95% confidence interval for that set of data.

- (2) Mutation frequency estimations for SP-PCR: The 'FitFreq' program (by BW Brown, UT-MDACC, Department of Biomathematics) estimates the mutant frequency of a sample. It uses the quantitation data from 'Calibrate', the number of replicates, and the number of mutants observed for that sample to calculate mutant frequency, 95% confidence interval, and negative log likelihood.
- (3) A log-likelihood ratio test combined the negative log likelihoods of the matched normal control and the patient sample, $2((-\log a) + [-\log b]) - [-\log \text{set } a \text{ combined with set } b])$, yielding a χ^2 value that was then analyzed using 'STATTAB' (UT-MDACC Department of Biomathematics) to determine significant differences between the control and patient data. For these analyses, $P < 0.01$ was considered significant. STATTAB is available at <http://odin.mdacc.tmc.edu/anonftp/>.

Results

Somatic instability of the CAG/CAA repeat in SCA17 families

We analyzed DNA samples isolated from peripheral blood leukocytes of nine SCA17 patients. These patients are from different racial background and consist of four index patients of the Japanese families (Families J1, J2, J3, and J4),¹⁸ three patients from two German families (Families G1 and G2, including one parent–daughter pair from Family G1),¹⁹ and one aunt–niece pair from one Mexican family (Family M1). The allele sizes have been reported previously in the Japanese and German patients, and confirmed here by conventional, or genotyping, PCR analysis of the CAG/CAA repeat alleles (Table 1). In the two Mexican patients, we determined that their genotypes are 35/50 in the aunt and 36/55 in the niece.

In this study, we used SP-PCR to study somatic instability of expanded CAG/CAA repeats at the *TBP* locus in SCA17

patients. We analyzed, on average, 271 ± 68 alleles for each of the nine patients in this study. In each patient, the two most frequent alleles represent the wildtype and expanded progenitor alleles, which matches our genotyping results (Figure 2 and Table 2). It is remarkable that the distribution of mutant alleles around the expanded progenitor alleles showed two distinct patterns: patients 1–4 and 7 (group I) have one tailed distribution with the mutant alleles trailing the expanded alleles, but the distributions in patients 5, 6, 8, and 9 (group II) are two tailed with both deletion and expansion mutant alleles flanking the expanded alleles (Figure 2). Thus, mutation in group I patients is strongly biased toward contraction, and the expanded alleles in group II patients are more prone to continuing expansion.

Since the mutation load due to repeat instability might be cumulative and dependent on the age of the subjects, we selected normal control subjects who are age-matched and have equivalent repeat size to the wild-type allele of each SCA17 patient. Our data showed that alleles of the normal subjects were relatively stable; regardless of the age, the mean mutation frequency was $5.5 \pm 3.4\%$ ($n = 9$). Interestingly, similar to the mutant allele distributions described above, expanded alleles of SCA17 patients showed variable level of stability: in group I patients, the expanded alleles were relatively stable with a mean mutation frequency of $9.6 \pm 4.0\%$ ($n = 5$), which is not significantly different from controls ($P > 0.05$, Table 2); but in group II patients, the expanded alleles were remarkably unstable with a mean mutation frequency of $30.6 \pm 4.6\%$ ($n = 4$). Therefore, the mutation frequency in group II patients is significantly higher than that in group I patients ($P < 0.01$).

Correlation between somatic instability and sequence motif of expanded repeats

Repeat number is known as an important factor for instability. When we compare the mutation frequency with the number of total CAG/CAA repeats, we found a positive correlation (coefficient = 0.76, Figure 3). It is noteworthy that the average number of total CAG/CAA repeats for group I patients is three less than that of group

Table 1 SCA 17 patients from Japanese, German, and Mexican families

Patient ID	Family ID	Age	Sex	Genotype	Intergenerational instability	Reference
1	Japanese-1	37	Male	36/55	No	A-III-1 (Nakamura <i>et al</i> , 2001)
2	Japanese-2	37	Female	32/47	No	D-III-8 (Nakamura <i>et al</i> , 2001)
3	Japanese-3	53	Male	35/48	No	C-III-6 (Nakamura <i>et al</i> , 2001)
4	Japanese-4	50	Female	35/47	No	B-II-2 (Nakamura <i>et al</i> , 2001)
5	German-1	47	Female	37/52	Yes	Case I ₁ (Zühlke <i>et al</i> , 2001)
6	German-1	21	Female	35/54	Yes	Daughter of I ₁ (Zühlke <i>et al</i> , 2001)
7	German-2	44	Male	36/48	Unknown	Case II ₄ (Zühlke <i>et al</i> , 2002)
8	Mexican-1	55	Female	35/50	Yes	This study
9	Mexican-1	22	Female	36/55	Yes	This study

Patients 5 is the mother of patient 6 in the German-1 family, and patient 8 is an aunt of patient 9 in the Mexican-1 family. Other patients are not related to each other.

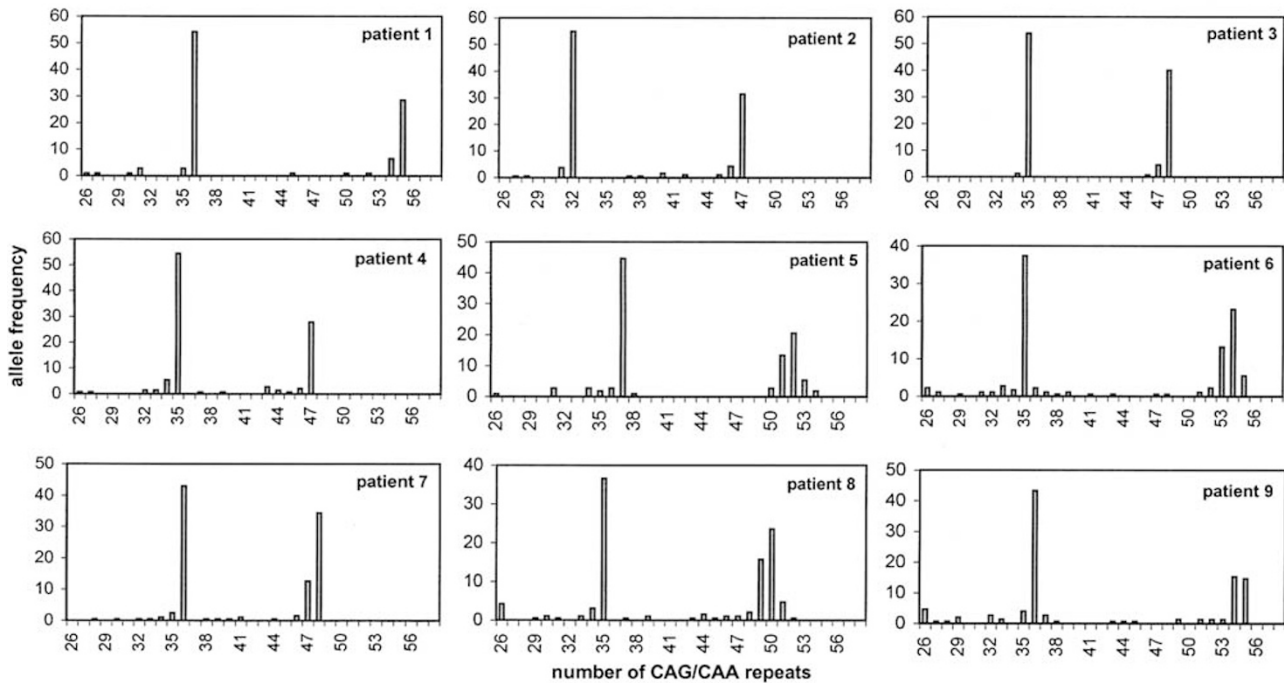


Figure 2 Two distinct distribution patterns of mutant alleles in SCA17 patients detected by SP-PCR. In each patient, the normal and expanded progenitor alleles were the most frequent. In patients 1–4 and 7 (group I), there is a one-tailed distribution of the mutant alleles trailing the expanded progenitor allele. In patients 5, 6, 8 and 9 (group II), the mutant allele distribution is two-tailed flanking the expanded progenitor allele.

Table 2 Mutation frequencies of the expanded CAG/CAA repeats in SCA17 patients

Patient	No. of CAG/CAA repeats	No. of alleles screened	No. of variants detected	Mutation frequency (95% CI)
1	55	184	21	12.8 (0.08–0.19)
2	47	354	26	7.9 (0.05–0.11)
3	48	288	10	3.6 (0.02–0.60)
4	47	250	26	10.5 (0.07–0.15)
5	52	196	39	24.0 (0.17–0.33)
6	54	271	68	31.7 (0.25–0.40)
7	48	377	45	13.4 (0.10–0.18)
8	50	320	76	31.8 (0.25–0.40)
9	55	231	62	34.9 (0.27–0.45)

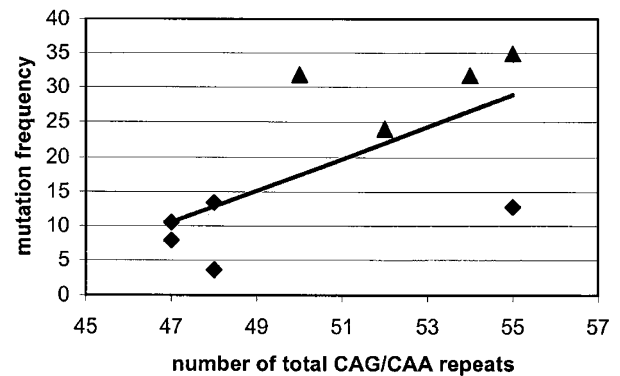


Figure 3 Correlation of mutation frequency and the number of total CAG/CAA repeats in SCA17. Diamond, group I patients; triangle, group II patients.

II patients (Table 2; 51 ± 3.4 for type I and 54 ± 2.2 for type II; $P=0.17$). However, this difference is within the range of variability in each group (Figure 3), and no dramatic changes of mutation frequency were found within either group. Therefore, the difference in total numbers of CAG/CAA repeats is unlikely to be the main determinant for the dramatic increase of mutation rate in group II patients. We thus examined the configuration of CAG/CAA repeats. We performed regular PCR on each patient's blood DNA using the high-fidelity Pfu polymerase and cloned the PCR product into TOPO cloning vector for sequencing analysis.

It is quite remarkable that the sequence of the expanded alleles in all group II patients showed the configuration of $(CAG)_3 (CAA)_3 (CAG)_{n1} CAA CAG$ (Table 3), and all group I patients had complex configurations of either $(CAG)_3 (CAA)_3 (CAG)_{n1} CAACAGCAA (CAG)_{n2} CAACAGCAA (CAG)_{n3} CAACAG^{18,19}$ or $(CAG)_3 (CAA)_3 (CAG)_9 CAACAG CAA (CAG)_9 (CAA)_3 (CAG)_9 CAACAGCAA (CAG)_{19} CAA CAG^{25}$ (Table 3). Thus, there is a strong correlation between CAG/CAA repeat configuration and instability: group I and II patients carry type I and II configurations, respectively.

Table 3 Configuration of expanded CAG/CAA repeat expansion in SCA17 patients

Patient	No. of CAG/CAA repeat expansion	Intergenerational instability	Somatic instability	Sequence of expanded alleles
1	55	No	No	(CAG) ₃ (CAA) ₃ (CAG) ₉ CAACAGCAA(CAG) ₁₆ CAACAGCAA(CAG) ₁₆ CAACAG
2	47	No	No	(CAG) ₃ (CAA) ₃ (CAG) ₆ CAACAGCAA(CAG) ₃₀ CAACAG
3	48	No	No	(CAG) ₃ (CAA) ₃ (CAG) ₆ CAACAGCAA(CAG) ₃₁ CAACAG
4	47	No	No	(CAG) ₃ (CAA) ₃ (CAG) ₈ CAACAGCAA(CAG) ₂₈ CAACAG
5	52	Yes	Yes	(CAG) ₃ (CAA) ₃ (CAG) ₄₄ CAACAG
6	54	Yes	Yes	(CAG) ₃ (CAA) ₃ (CAG) ₄₆ CAACAG
7	48	Unknown	No	(CAG) ₃ (CAA) ₃ (CAG) ₉ CAACAGCAA(CAG) ₂₈ CAACAG
8	50	Yes	Yes	(CAG) ₃ (CAA) ₃ (CAG) ₄₂ CAACAG
9	55	Yes	Yes	(CAG) ₃ (CAA) ₃ (CAG) ₄₇ CAACAG

Type II configuration with no CAA interruptions shows increased instability.

Correlation between somatic instability and intergenerational changes in the repeat number in SCA17 families

In group I patients, patients 1–4 are the index patients from four Japanese families, in which there were no changes in repeat size during transmission of the expanded alleles from one generation to the next;¹⁸ patient 7 is from the German family, in which the expanded allele was also stably transmitted across generations.¹⁹ However, in group II patients, patients 5 and 6 are from two German families, where the size of the expanded alleles changed during the transmission.¹⁹ Patients 8 and 9, reported in this study, are from one Mexican family; and their expanded alleles were unstably inherited. These data suggest that expanded alleles with the simple motif of (CAG)₃(CAA)₃(CAG)_n1 CAACAG are remarkably unstable while those of the complex configuration with more CAA interruptions are relatively stable during intergenerational transmission, and this coincides with the degree of instability shown in the expanded alleles of somatic cells, suggesting that a similar mechanism operates in these two situations.

Inverse correlation between the age of onset and the size of expanded CAG/CAA repeat alleles in SCA17

In spite of the dramatic difference in stability between interrupted and uninterrupted CAG expansions, they seem to be equally pathogenic at equivalent numbers. Here, the age of onset and the number of CAG/CAA repeats in the expanded allele were reviewed in 19 patients reported in the literature and two patients from our Mexican family (Figure 4). As previously reported,¹⁸ we found a significant inverse correlation between the age of onset and the number of CAG/CAA repeat in this extended series. Interesting outliers are a 48 CAG/CAA repeat allele reported in three asymptomatic members of a German SCA17 family¹⁹ and a 43-repeat allele in a patient reported by Silveira *et al.*²¹ Although the former asymptomatic patients may develop the disease later in their life, the latter may lower the pathogenic range of CAG expansion

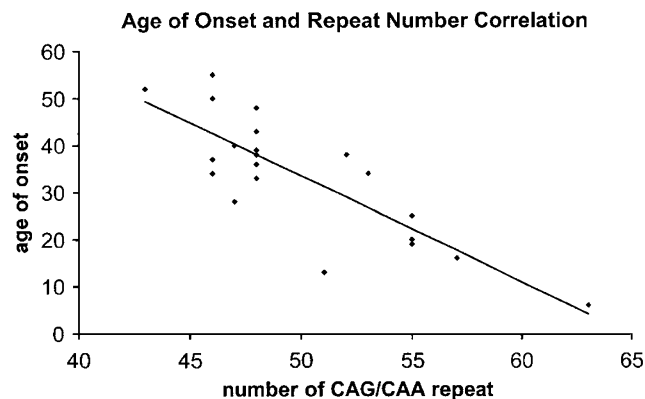


Figure 4 Correlation of age of onset and number of CAG/CAA repeats in SCA17.

down to 43 if potential phenocopy can be excluded. Thus, 43 CAGs might have been seen in both normal and SCA17 subjects, giving rise to a question of reduced penetrance associated with the lower end of the disease allele in SCA17.

Discussion

In this study, we used SP-PCR to study somatic instability of expanded CAG repeats at the *TBP* locus in SCA17 patients. SP-PCR is a highly sensitive assay in detecting low levels of somatic mosaicism. Our analysis on *TBP* locus showed that normal alleles of the CAG repeat have a low level of apparent mosaicism in peripheral blood leukocytes. Regular PCR using greater amount of input DNA from normal subjects cannot detect these rare mutant alleles that were detected by our SP-PCR analysis. Our evidence argues against the possibility that this apparent mosaicism represents artifacts: all zero-DNA controls were negative, and the distribution pattern of mutant alleles differs from sample to sample. Trinucleotide repeats within normal range have also been shown to have a low level of instability at another locus, the Huntington disease (*HD*) gene. At the upper end of the normal range of CAG repeats, *HD* alleles show repeat size mosaicism with occasional expansion into the full mutation range; when this event occurs in germline, it may be transmitted to the next

generation. This upper end of normal range of CAG repeat in HD alleles is called 'mutable normal alleles' (ACMG/ASHG Statement 1998). Single molecule PCR analysis of CAG repeat length at the *HD* locus has demonstrated that a 30-CAG-repeat allele has an 11% mutation frequency in the sperm of a normal individual.²⁹ A similar study has also been done for myotonic dystrophy type 1 (DM1); SP-PCR analysis of CTG repeat length in sperm showed low level instability of the normal alleles.³⁰

Interestingly, we observed two distinct patterns of mutant allele distribution around the expanded progenitor alleles: patients 1–4 and 7 (group I) have one tailed distribution with the mutant alleles trailing the expanded alleles, but the distributions in patients 5, 6, 8, and 9 (group II) are two tailed with both deletion and expansion mutant alleles flanking the expanded alleles (Figure 2). Remarkably, the mutation profiles match the configurations of CAG/CAA expansion. Group I patients had type I configuration of (CAG)₃ (CAA)₃ (CAG)_{n1} CAA CAG CAA (CAG)_{n2} CAA CAG with expanded (CAG)_{n2}. In contrast, group II patients had type II configuration of (CAG)₃(CAA)₃ (CAG)_{n1}CAACAG with expanded (CAG)_{n1}. Expanded alleles in the complex type I configuration have been shown to transmit from parent to child without changes in repeat size, while those in the simple type II configuration have intergenerational instability.^{19,31} Thus, pure expanded CAG repeats are associated with intergenerational instability, while CAG repeats with CAA interruption are stably transmitted from generation to generation.

A change of the allele size during the parent-to-child transmission is usually detected by comparing DNA isolated from blood. These alleles in somatic cells of the parent and the offspring, however, are separated by many cell divisions. Thus, the intergenerational changes of allele size are the combination of repeat size instability in the parental germline and somatic tissues, and the somatic instability in the offspring. Direct comparison of the progenitor allele across generations would be most appropriate, but this is not easy to do. To study the mechanisms of intergenerational instability, an alternative and parallel approach is to perform SP-PCR on DNA isolated from somatic tissues. Our data clearly demonstrated somatic instability in SCA17, which could contribute to repeat size variations between generations. Furthermore, somatic instability also raises the possibility of different expansion sizes in different tissues, which may contribute to the tissue specificity of the disease and genotype–phenotype correlation.

The relationship between the length and number of expanded repeats and their stability has been well characterized. It is widely accepted that the higher the number of repeats, the more unstable the expanded repeats are. However, for repeat configuration there have been no rigorous studies in the literature. The situation where the expanded repeats have similar numbers but different

configurations is unique to SCA17 among CAG expansion diseases. In SCA1, some intermediately sized alleles have CAT interruptions that disrupt the polyglutamine tract with histidine, making the allele non-pathogenic and apparently stable during parent-to-offspring transmissions.³² However, these alleles are not fully expanded, and the number of transmissions observed to date is limited. In SCA2, a 34-CAG allele with CAA interruptions has been associated with SCA2-like phenotype.³³ However, the stability of the repeat size during transmissions has not been extensively studied. In SCA8, expanded non-coding CTG/CAG repeats are associated with the disease. While the expanded repeats are intricately interrupted by various cryptic triplets and unstably transmitted from generation to generation, the effect of interruptions on the instability has not been systematically studied in SCA8.^{34,35} Thus, identification of two distinct repeat configurations in the expanded range in our study provided the opportunity to directly examine the effect of repeat interruptions on repeat instability. Our data clearly demonstrated that expanded CAG repeat alleles in SCA17 patients have variable somatic mosaicism, and the somatic mosaicism is greater with pure repeats. Thus, expanded alleles with pure repeat configuration are less stable than those with interrupted and more complex repeat configuration, at least in blood cells. It remains to be determined whether this observation holds true in other tissues and germ line. Studies in other tissues will address the issue of tissue-dependent repeat instability and provide insight into the mechanism of pathogenesis, and studies in germ line might provide molecular explanation for the lack of anticipation in most of the SCA17 families.

Our data support the theory that CAA interruption, serve as a limiting or stabilizing factor for CAG repeat expansion. Zuhlke *et al*²² also reported increased intergenerational instability in families carrying pure repeat expansions without domain III (CAACAGCAA). It is of interest that, evolutionarily, the *TBP* CAG/CAA microsatellite seems to be expanding from nonhuman primates,³⁶ with the most frequent human *SCA17* allele carrying 37 repeats, very close to the pathogenic range. Two major mutational processes have been proposed for microsatellite repeat instability, which are DNA strand slippage during replication and unequal chromosomal crossing-over during recombination.^{37,38} DNA slippage is considered the predominant mechanism underlying microsatellite repeat instability, particularly in somatic tissues. In germ lines, both processes might occur. Given the proximity of common repeat number and the critical pathogenic value in human *TBP* locus, mutations by either mechanism will lead to pathogenic repeat expansion. However, SCA17 is a rare disease, and majority of human alleles have complex configuration with all five domains. Therefore, CAA interruption is very like a stabilizing factor for the CAG/CAA microsatellite repeat in *TBP* locus.

Acknowledgements

This study was supported by grant from NIH (NS41547) to TA and CA112508 to MS.

References

- Pearson CE, Sinden RR: Trinucleotide repeat DNA structures: dynamic mutations from dynamic DNA. *Curr Opin Struct Biol* 1998; **8**: 321–330.
- Richards RI: Dynamic mutations: a decade of unstable expanded repeats in human genetic disease. *Hum Mol Genet* 2001; **10**: 2187–2194.
- Manto MU: The wide spectrum of spinocerebellar ataxias (SCAs). *Cerebellum (London, England)* 2005; **4**: 2–6.
- Gomes-Pereira M, Monckton DG: Chemical modifiers of unstable expanded simple sequence repeats: what goes up, could come down. *Mutat Res* 2006; **598**: 15–34.
- Kennedy L, Shelbourne PF: Dramatic mutation instability in HD mouse striatum: does polyglutamine load contribute to cell-specific vulnerability in Huntington's disease? *Hum Mol Genet* 2000; **9**: 2539–2544.
- Manley K, Shirley TL, Flaherty L, Messer A: Msh2 deficiency prevents *in vivo* somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nat Genet* 1999; **23**: 471–473.
- Gordenin DA, Kunkel TA, Resnick MA: Repeat expansion – all in a flap? *Nat Genet* 1997; **16**: 116–118.
- van den Broek WJ, Nelen MR, Wansink DG *et al*: Somatic expansion behaviour of the (CTG)_n repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch-repair proteins. PG – 191-8. *Hum Mol Genet* 2002; **11**: 191–198.
- Mirkin SM: Toward a unified theory for repeat expansions. *Nature Structural & Molecular Biology* 2005; **12**: 635–637.
- Peltomaki P: DNA mismatch repair and cancer. *Mutat Res* 2001; **488**: 77–85.
- Brock GJ, Anderson NH, Monckton DG: Cis-acting modifiers of expanded CAG/CTG triplet repeat expandability: associations with flanking GC content and proximity to CpG islands. *Hum Mol Genet* 1999; **8**: 1061–1067.
- Cleary JD, Nichol K, Wang YH, Pearson CE: Evidence of cis-acting factors in replication-mediated trinucleotide repeat instability in primate cells. *Nat Genet* 2002; **31**: 37–46.
- Pearson CE, Nichol Edamura K, Cleary JD: Repeat instability: mechanisms of dynamic mutations. *Nat Rev* 2005; **6**: 729–742.
- Duyao M, Ambrose C, Myers R *et al*: Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nat Genet* 1993; **4**: 387–392.
- Pearson CE, Eichler EE, Lorenzetti D *et al*: Interruptions in the triplet repeats of SCA1 and FRAXA reduce the propensity and complexity of slipped strand DNA (S-DNA) formation. *Biochemistry* 1998; **37**: 2701–2708.
- Frontali M, Novelletto A, Annesi G, Jodice C: CAG repeat instability, cryptic sequence variation and pathogeneticity: evidence from different loci. *Philos Trans R Soc Lond B Biol Sci* 1999; **354**: 1089–1094.
- Choudhry S, Mukerji M, Srivastava AK *et al*: CAG repeat instability at SCA2 locus: anchoring CAA interruptions and linked single nucleotide polymorphisms. *Hum Mol Genet* 2001; **10**: 2437–2446.
- Nakamura K, Jeong SY, Uchihara T *et al*: SCA17, a novel autosomal dominant cerebellar ataxia caused by an expanded polyglutamine in TATA-binding protein. *Hum Mol Genet* 2001; **10**: 1441–1448.
- Zuhlke C, Hellenbroich Y, Dalski A *et al*: Different types of repeat expansion in the TATA-binding protein gene are associated with a new form of inherited ataxia. *Eur J Hum Genet* 2001; **9**: 160–164.
- Fujigasaki H, Martin JJ, De Deyn PP *et al*: CAG repeat expansion in the TATA box-binding protein gene causes autosomal dominant cerebellar ataxia. *Brain* 2001; **124**: 1939–1947.
- Silveira I, Miranda C, Guimaraes L *et al*: Trinucleotide repeats in 202 families with ataxia: a small expanded (CAG)_n allele at the SCA17 locus. *Arch Neurol* 2002; **59**: 623–629.
- Zuhlke C, Dalski A, Schwinger E, Finckh U: Spinocerebellar ataxia type 17: report of a family with reduced penetrance of an unstable Gln49 TBP allele, haplotype analysis supporting a founder effect for unstable alleles and comparative analysis of SCA17 genotypes. *BMC Medical Genetics* 2005; **6**: 27.
- Oda M, Maruyama H, Komure O *et al*: Possible reduced penetrance of expansion of 44 to 47 CAG/CAA repeats in the TATA-binding protein gene in spinocerebellar ataxia type 17. *Arch Neurol* 2004; **61**: 209–212.
- Cummings CJ, Zoghbi HY: Fourteen and counting: unraveling trinucleotide repeat diseases. *Hum Mol Genet* 2000; **9**: 909–916.
- Koide R, Kobayashi S, Shimohata T *et al*: A neurological disease caused by an expanded CAG trinucleotide repeat in the TATA-binding protein gene: a new polyglutamine disease? *Hum Mol Genet* 1999; **8**: 2047–2053.
- Maltecca F, Filla A, Castaldo I *et al*: Intergenerational instability and marked anticipation in SCA-17. *Neurology* 2003; **61**: 1441–1443.
- Coolbaugh-Murphy MI, Xu J, Ramagli LS *et al*: Microsatellite instability (MSI) increases with age in normal somatic cells. *Mech Ageing Dev* 2005; **126**: 1051–1059.
- Coolbaugh-Murphy M, Maleki A, Ramagli L *et al*: Estimating mutant microsatellite allele frequencies in somatic cells by small-pool PCR. *Genomics* 2004; **84**: 419–430.
- Leefflang EP, Zhang L, Tavare S *et al*: Single sperm analysis of the trinucleotide repeats in the Huntington's disease gene: quantification of the mutation frequency spectrum. *Hum Mol Genet* 1995; **4**: 1519–1526.
- Martorell L, Monckton DG, Gamez J, Baiget M: Complex patterns of male germline instability and somatic mosaicism in myotonic dystrophy type 1. *Eur J Hum Genet* 2000; **8**: 423–430.
- Dorsman JC, Bremmer-Bout M, Pepers B *et al*: Interruption of perfect CAG repeats by CAA triplets improves the stability of glutamine-encoding repeat sequences. *Biotechniques* 2002; **33**: 976–978.
- Chong SS, McCall AE, Cota J *et al*: Gametic and somatic tissue-specific heterogeneity of the expanded SCA1 CAG repeat in spinocerebellar ataxia type 1. *Nat Genet* 1995; **10**: 344–350.
- Costanzi-Porrini S, Tessarolo D, Abbruzzese C *et al*: An interrupted 34-CAG repeat SCA-2 allele in patients with sporadic spinocerebellar ataxia. *Neurology* 2000; **54**: 491–493.
- Moseley ML, Schut LJ, Bird TD *et al*: SCA8 CTG repeat: en masse contractions in sperm and intergenerational sequence changes may play a role in reduced penetrance. *Hum Mol Genet* 2000; **9**: 2125–2130.
- Silveira I, Alonso I, Guimaraes L *et al*: High germinal instability of the (CTG)_n at the SCA8 locus of both expanded and normal alleles. *Am J Hum Genet* 2000; **66**: 830–840.
- Tomiuk J, Bachmann L, Bauer C *et al*: Repeat expansion in spinocerebellar ataxia type 17 alleles of the TATA-box binding protein gene: an evolutionary approach. *Eur J Hum Genet* 2007; **15**: 81–87.
- Li YC, Korol AB, Fahima T *et al*: Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol Ecol* 2002; **11**: 2453–2465.
- Ellegren H: Microsatellites: simple sequences with complex evolution. *Nat Rev* 2004; **5**: 435–445.

Supplementary information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)