6352–6361 *Nucleic Acids Research, 2006, Vol. 34, No. 21* *Published online 16 November 2006*
doi:10.1093/nar/gkl846

Replication in mammalian cells recapitulates the locus-specific differences in somatic instability of genomic GAA triplet-repeats

Paul M. Rindler[1], Rhonda M. Clark[1], Laura M. Pollard[1], Irene De Biase[1] and Sanjay I. Bidichandani[1,2,*]

[1]Department of Biochemistry and Molecular Biology and [2]Department of Pediatrics, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104, USA

Received July 24, 2006; Revised and Accepted October 9, 2006

ABSTRACT

Friedreich ataxia is caused by an expanded $(GAA{\cdot}TTC)_n$ sequence in intron 1 of the *FXN* gene. Small pool PCR analysis showed that pure $(GAA{\cdot}TTC)_{44+}$ sequences at the *FXN* locus are unstable in somatic cells *in vivo*, displaying both expansions and contractions. On searching the entire human and mouse genomes we identified three other genomic loci with pure $(GAA{\cdot}TTC)_{44+}$ sequences. Alleles at these loci showed mutation loads of <1% compared with 6.3–30% for *FXN* alleles of similar length, indicating that somatic instability *in vivo* is regulated by locus-specific factors. Since distance between the origin of replication and the $(CTG{\cdot}CAG)_n$ sequence modulates repeat instability in mammalian cells, we tested if this could also recapitulate the locus-specific differences for genomic $(GAA{\cdot}TTC)_n$ sequences. Repeat instability was evaluated following replication of a $(GAA{\cdot}TTC)_{115}$ sequence in transfected COS1 cells under the control of the SV40 origin of replication located at one of five different distances from the repeat. Indeed, depending on the location of the SV40 origin relative to the $(GAA{\cdot}TTC)_n$ sequence, we noted either no instability, predominant expansion or both expansion and contraction. These data suggest that mammalian DNA replication is a possible mechanism underlying locus-specific differences in instability of GAA triplet-repeat sequences.

INTRODUCTION

Several inherited neuromuscular diseases are caused by abnormal expansion of triplet-repeat sequences (1). Whereas the majority of them involve expansion of the $(CAG{\cdot}CTG)_n$ sequence, Friedreich ataxia (FRDA) is so far the only disease associated with expansion of the $(GAA{\cdot}TTC)_n$ sequence. FRDA is an autosomal recessive disease. Normal individuals have <30 triplets and most patients are homozygous for alleles with 66–1700 triplets (E alleles) in intron 1 of the *FXN* gene on chromosome 9q21 (2). A minority of patients have borderline alleles, with ~44–66 triplets, in addition to a conventional E allele (3).

Using a sensitive technique called small pool PCR (SP-PCR) to measure the repeat length in individual *FXN* genes, we have shown that $(GAA{\cdot}TTC)_{44+}$ alleles are unstable in human somatic cells *in vivo* (4,5). Long E alleles (>500 triplets) showed a marked contraction bias and short E alleles (<500 triplets) and borderline alleles showed an expansion bias (3,5). It is clearly important to understand what controls repeat instability *in vivo*. Not only does the length of the disease-associated allele determine the severity and rate of progression of disease (6–9), but also in dorsal root ganglia, the region of the nervous system showing the earliest and most significant pathology in FRDA (10), the disease-associated alleles are known to undergo further large expansions in an age-dependent manner (11). Furthermore, in carriers of borderline alleles, somatic instability *per se* was necessary for the development of FRDA (3).

The mechanism of $(GAA{\cdot}TTC)_n$ repeat instability remains poorly understood. We, along with others, have shown that in simple replication model systems in *Escherichia coli* (4,12) and *Saccharonyces cerevisiae* (13), the $(GAA{\cdot}TTC)_n$ sequence is more unstable when GAA serves as the template for lagging strand synthesis. However, the resulting instability comprised mainly contractions, and the expansions seen *in vivo* with borderline and short E alleles were not seen. Interestingly, the tissue-specificity and bias for expansion seen in human tissues, was reproduced in a transgenic mouse model containing either $(GAA{\cdot}TTC)_{82}$ or $(GAA{\cdot}TTC)_{190}$ sequences within the appropriate sequence context of the entire 'human' *FXN* locus (14,15). This indicated that the

*To whom correspondence should be addressed at Department of Biochemistry and Molecular Biology, University of Oklahoma Health Sciences Center, 975 NE, 10th, BRC458, Oklahoma City, OK 73104, USA. Tel: +1 405 271 1360; Fax: +1 405 271 3910; Email: Sanjay-Bidichandani@ouhsc.edu

sequence context of the human *FXN* locus and perhaps also the mammalian cellular milieu are required for somatic instability *in vivo*.

Here we report that $(GAA{\cdot}TTC)_{44+}$ alleles situated at other loci in the human and mouse genomes are much less unstable compared with alleles at the human *FXN* locus. We also show that altering the orientation of replication and the distance from the eukaryotic origin of replication within transfected mammalian cells is able to reproduce the locus-specific differences seen in $(GAA{\cdot}TTC)_n$ repeat instability. Specifically, depending on the conditions, replication of the $(GAA{\cdot}TTC)_n$ sequence in mammalian cells may either result in increased frequency of expansions, increase of both expansions and contractions or even the absence of instability. Our data indicate that local differences in DNA replication can explain both the instability seen at the *FXN* locus and the stability seen at other genomic loci.

MATERIALS AND METHODS

Genomic DNA samples

Human DNA was previously obtained from blood samples from a panel of 100 unrelated Caucasian adults. DNA from FRDA patients was obtained from blood samples using an IRB approved protocol. Mouse genomic DNA was obtained from blood and cerebellum of a 12-month-old mouse (C57BL/6J background). Blood samples were initially treated with 1% Triton X-100 and the pelleted leukocytes were resuspended in PBS. Genomic DNA was purified using the DNeasy tissue kit (Qiagen).

Genome analysis

Homo sapiens v34a and *Mus musculus* v32 complete genomes were downloaded from the NCBI website. A custom program in C, that identifies all 10 non-redundant triplet motifs, as previously described (16,17), was used to identify $(GAA{\cdot}TTC)_n$ sequences. Sequences of desired length (see Results) were extracted along with flanking non-repeat sequence in order to design primers for PCR amplification.

Small pool PCR

This was performed as described previously (5,18). Briefly, serial dilutions of genomic DNA, ranging from 6 to 600 pg, were prepared in siliconized microfuge tubes. Primers for PCR amplification of $(GAA{\cdot}TTC)_n$ sequences at *FXN*, 3q13, 5q23 and 10q24 were as previously described (4,17). The following primer pairs were used for amplification of $(GAA{\cdot}TTC)_n$ sequences at the three mouse loci:

1e2.3 (5′-CAGTTCTCTGTGAGACCT-3′ and 5′-GCCAGGATGTAAGGAGAATCT-3′); 8b3.3 (5′-GGGATAGCATTGAAAATGTAATT-3′ and 5′-TTTGCATGGACCAGCCTTGTG-3′); 8b3.3b (5′-GACGGTGGATTTCTGAGTTTA-3′ and 5′-CACTTGCCACACACACAGTAT-3′). PCR products were resolved by electrophoresis on 1.5% agarose gels and bands detected by Southern blotting using an end-labeled $(TTC)_{11}$ oligonucleotide probe. Calculation of the number of individual molecules per reaction was performed by Poisson analysis as described previously (18). For each genomic DNA sample multiple reactions were performed using 'small pools' of 2.5–25 individual molecules (typically 5–10) per reaction to detect mutations. Mutation load was calculated as the proportion of amplified molecules that differed by >5% in length from the constitutional (most common) allele determined by conventional PCR.

Plasmid construction

The $(GAA{\cdot}TTC)_{120}$ repeat sequence was amplified from genomic DNA of a FRDA patient with E alleles of 120 and 880 triplets in intron 1 of the *FXN* gene. DNA was isolated from whole blood and PCR was performed using the following primers:

GAA-104F (5′-GGCTTAAACTTCCCACACGTGTT-3′) and GAA-629R (5′-AGGACCATCATGGCCACACTT-3′), followed by nested PCR using the following primers:

ttcpst1-F (5′-GCTCCGCTGCAGCGCGCGACACCACGCCCGGCTAAC-3′) and
ttcxba1-R (5′-GATGCGTCTAGACCCAGTATCTACTAAAAAATAC-3′).

Purified PCR products were digested with PstI and XbaI, which recognize sequences located at the 5′ ends of the forward and reverse primers, respectively. The fragment containing the $(GAA{\cdot}TTC)_{120}$ sequence along with minimal flanking sequence from intron 1 of the human *FXN* gene (38 bp 5′ and 35 bp 3′ to the repeat) was cloned into the PstI and XbaI sites of vector pSEA to generate pSEA-115. pSEA is a derivative of pUC19, modified by replacing the SphI site with a synthetic polylinker containing SacII, EcoRV and AflII restriction sites. The $(GAA{\cdot}TTC)_{120}$ sequence was cloned in the 'stable orientation' relative to the *E.coli* ColE1 origin of replication, such that TTC would serve as the lagging strand template (4). Repeat length (determined to be 115 triplets) and orientation were confirmed by sequencing.

The Hind III–Sph I fragment (viral nucleotide positions 5171 to 128) of the SV40 viral genome, which contains the SV40 origin of replication (SV40 *ori*), was amplified by PCR. Primer pairs were designed with one of five different restriction enzyme recognition sequences at the 5′ ends (designated by XYZ) as follows (note that other unique sites were also included for future manipulation):

Fwd: 5′-XYZ-(NheI, EagI, BglII)-TTTTGCAAAAGCCTAGGCC-3′ and
Rev: 5′-XYZ-(NotI, SpeI, XhoI)-ATCTCAATTAGTCAGCAACC-3′.

Purified PCR products were digested with the appropriate enzyme and cloned into either the NdeI, KpnI, HindIII, SapI or AflIII site of pSEA-115 to generate the final five constructs. To minimize any potential effect of transcription from the SV40 viral early promoter, which overlaps with the bidirectional SV40 *ori*, the fragment used in pSEA-115 contained a truncated (<20 bp) enhancer sequence and was subcloned in the orientation so transcription, if any, would proceed away from the repeat tract. The actual repeat lengths in the five constructs ('115' used for nomenclature), as determined by sequencing, were as follows: pSEA-N115 = 116 triplets; pSEA-K115 = 115 triplets; pSEA-H115 = 115

triplets; pSEA-S115 = 113 triplets; and pSEA-A115 = 114 triplets.

Cell culture and transfection

COS1 cells were grown in DMEM (Invitrogen) plus 10% fetal bovine serum (FBS). Transfections were performed in triplicate as follows: cells were plated in six-well tissue culture plates and when ~40% confluent they were starved in 2 ml of serum-free medium for 1 h. Cells were then incubated for 4 h in 750 µl serum-free medium containing 1 µg of plasmid DNA and 10 µl of Lipotaxi transfection reagent (Stratagene). After 4 h, 750 µl of DMEM plus 20% FBS was added and cells were incubated for 16 h, after which the medium in each well was replaced with fresh DMEM plus 10% FBS. Cells were incubated for 48 h and DNA was extracted using the DNeasy tissue kit (Qiagen). CV1 cells were grown in minimum essential medium (ATCC) and transfection was performed as described for COS1 cells.

Plasmid replication in COS1 cells and analysis of repeat instability

This was essentially performed using the STRIP assay as previously described (19). Briefly, DNA extracted from transfected COS1 cells was digested with 4 U of DpnI for 3 h at 37°C to digest all unreplicated plasmids used in the initial transfection. Replication efficiency was determined by analyzing equal quantities of either DpnI or mock digested products, which were further digested with EcoRI and Eco0109I to linearize the DpnI resistant plasmids. The products were analyzed by Southern blotting and hybridization with an end-labeled $(TTC)_{11}$ oligonucleotide probe. The relative proportion of DpnI resistant plasmid was measured by densitometry, which showed that the efficiency of replication of all five constructs was similar (mean ± SEM; 0.53 ± 0.07). For measurement of instability, DpnI-resistant plasmids, representing products of *de novo* replication in mammalian cells, were purified and transformed into *E.coli* (DH5α). Individual colonies, each representing a single product of replication, were analyzed by colony PCR using primers GS-F and GS-R and resolved on 2.5% agarose gels as described previously (4). Some of the mutant inserts were sequenced to confirm that they represented variability in repeat length. The level of instability was measured as the ratio of plasmids with altered repeat lengths to the total number of successful PCRs (full-length plus mutant), and was represented as a percentage. Background instability was determined for each construct before transfection by direct transformation of the DNA derived from the same plasmid preparation into DH5α (i.e. without transfection into COS1 cells). Background instability was measured by colony PCR exactly as described above. Statistical significance of instability observed upon replication in COS1 cells was determined by comparison with the background level in *E.coli* and was calculated by Chi square analysis (19). In the case of CV1 cells, the level of repeat instability was determined as described, however, post-transfection DNA was transformed directly into DH5α without DpnI treatment.

RESULTS

Locus-specific differences in somatic instability of GAA triplet-repeat sequences in the mammalian genome

To test if there are genomic locus-specific differences in the level of somatic instability of $(GAA{\cdot}TTC)_n$ sequences *in vivo*, we searched for other naturally occurring long and pure $(GAA{\cdot}TTC)_n$ sequences in the human and mouse genomes. We decided to use a combination of SP-PCR and the previously determined 'threshold length' for the initiation of instability at the *FXN* locus [≥44 pure GAA triplets (4,5)] as a quantitative indicator of the level of somatic instability. The underlying rationale was that loci with greater instability would manifest instability at a lower threshold length, and comparison of level of instability among alleles of similar length would allow a quantitative analysis of the instability at distinct genomic loci. We initially excluded $(GAA{\cdot}TTC)_n$ sequences with flanking G/A-rich (non-GAA) sequences because of their ability to dramatically stabilize adjacent $(GAA{\cdot}TTC)_n$ sequences [(4); see below]. Following a comprehensive search of the human genome we had previously identified 25 genomic loci with long $(GAA{\cdot}TTC)_n$ tracts (17). However, sequencing of multiple alleles at all of these loci revealed that only one locus (5q23) contained suitable alleles for our analysis (i.e. ≥44 pure GAA triplets, without any flanking G/A-island sequence). We then screened blood DNA from a panel of 100 unrelated human controls and identified five alleles at 5q23 that ranged from 59 to 71 pure GAA triplets. Similarly, in the mouse genome we identified only two out of the 107 loci with $(GAA{\cdot}TTC)_{40+}$ sequences to have ≥44 pure GAA triplets without any flanking G/A-rich sequences. We sequenced both these loci and confirmed that 1e2.3 and 8b3.3 had pure tracts of $(GAA{\cdot}TTC)_{48}$ and $(GAA{\cdot}TTC)_{53}$, respectively. None of the three loci with alleles containing ≥44 pure GAA triplets (human 5q23 and the two murine loci) mapped within a confirmed or predicted transcriptional unit.

SP-PCR analysis of 1701 individual molecules of the five human alleles at 5q23 and 688 individual molecules of the two mouse alleles showed either absence or very low levels of somatic instability (Figure 1 and Table 1). The human alleles at 5q23 were derived from blood samples of anonymous adults, and in the case of the mouse, we tested DNA from 12-month-old blood and cerebellum. In comparison, analysis of alleles of similar length at the human *FXN* locus (also from human blood) showed that somatic instability was significantly greater. At the *FXN* locus, mutation loads were 6.3% for $(GAA{\cdot}TTC)_{44}$ (45.3% expansions, 54.7% contractions; range: 7–113 triplets) and 30% for $(GAA{\cdot}TTC)_{66}$ (59.8% expansions, 40.2% contractions; range: 2–227 triplets) versus only 0–1% at all other loci ($P < 0.001$; Table 1). Moreover, the threshold for instability at the human *FXN* locus was lower (≥44 triplets) compared with the other genomic loci (>48–71 triplets).

We then tested four $(GAA{\cdot}TTC)_n$ alleles (23–51 triplets; 1467 individual molecules) which had flanking G/A-island sequences at two additional loci in the human genome (3q13 and 10q24) and a pure $(GAA{\cdot}TTC)_{36}$ allele at another locus in the mouse genome (8b3.3b; 464 individual molecules). No somatic instability was detected at each of these loci (data not shown).

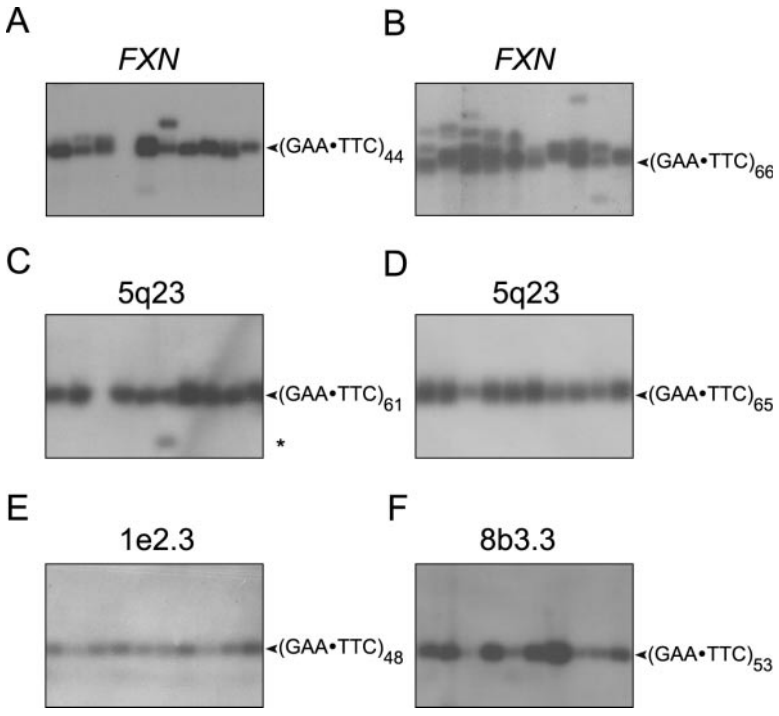

Figure 1. Somatic instability of the GAA triplet-repeat sequence is locus-specific. Representative gels are shown for small pool PCR analysis of pure $(GAA{\cdot}TTC)_n$ sequences at various genomic loci in the human [(**A** and **B**) *FXN* locus 9q21 and (**C** and **D**) another locus at 5q23] and mouse [(**E**) 1e2.3 and (**F**) 8b3.3] genome. The locations of the progenitor allele size, as determined by conventional PCR, are indicated by arrowheads. Each lane contains 5–10 individual molecules (see Materials and methods). Note that the $(GAA{\cdot}TTC)_n$ sequence at the human *FXN* locus shows higher levels of instability than alleles of similar sequence and length at all other loci. A rare mutation at the 5q23 locus is indicated by an asterisk.

Table 1. Locus-specific instability of the GAA triplet-repeat sequence

Locus	GAA length	Mutation load (%)	Molecules analyzed
FXN[a,b]	39	0	1150
	44	6.3	2304
	66	30	1230
5q23[b,c]	59	<1	555
	61	<1	291
	61	<1	334
	65	<1	223
	71	<1	298
1e2.3[c,d]	48	0	392
8b3.3[c,d]	53	0	296

[a]Previously published data from (4).
[b]Human locus.
[c]Novel data; potentially unstable GAA repeats in human and mouse genomes.
[d]Mouse locus.

Design of constructs to test the effect of DNA replication in mammalian cells on GAA triplet-repeat instability

We reasoned that alteration of the orientation of replication and the distance between the origin of replication and the repeat tract could potentially simulate locus-specific differences in the genome and thereby result in differential instability in mammalian cells. Furthermore, alteration of the orientation of replication and the distance between the SV40 *ori* and the $(CAG{\cdot}CTG)_n$ sequence is known to modulate repeat instability in mammalian cells (19). To test this hypothesis we constructed five plasmid vectors that would allow replication of the $(GAA{\cdot}TTC)_{115}$ sequence in transfected COS1 cells (Figure 2; see Materials and Methods). In order to test the effect of orientation of replication, the SV40 *ori* was cloned in such a way as to result in GAA (pSEA-N115 and pSEA-K115) or TTC (pSEA-H115, pSEA-S115 and pSEA-A115) as the lagging strand template. The five locations selected around the $(GAA{\cdot}TTC)_{115}$ sequence would allow testing of the effect of varying the distance from SV40 *ori*. The two equidistant pairs (pSEA-N115 [458 nt]/pSEA-S115 [467 nt] and pSEA-K115 [233 nt]/pSEA-H115 [227 nt]) would allow testing of the effect of orientation of replication independent of the distance between the repeat and the SV40 *ori*.

Since the $(GAA{\cdot}TTC)_n$ sequence is unstable in *E.coli*, generating a high frequency of contractions when GAA is the lagging strand template (4,12), two steps were taken to ensure accurate analysis of instability in COS1 cells. The $(GAA{\cdot}TTC)_{115}$ tract was cloned in the stable orientation with respect to the ColE1 origin of replication to minimize mutations in *E.coli*. Furthermore, to determine if the level of instability seen after replication in COS1 cells was significantly enhanced, results were compared by simultaneously assessing instability in *E.coli* (using the same DNA preparation) without transfection.

Mammalian DNA replication results in differential GAA triplet-repeat instability depending on the distance from the SV40 *ori* and the orientation of replication

All five constructs were transfected into COS1 cells and individual products of mammalian replication were analyzed for repeat instability using the previously described STRIP

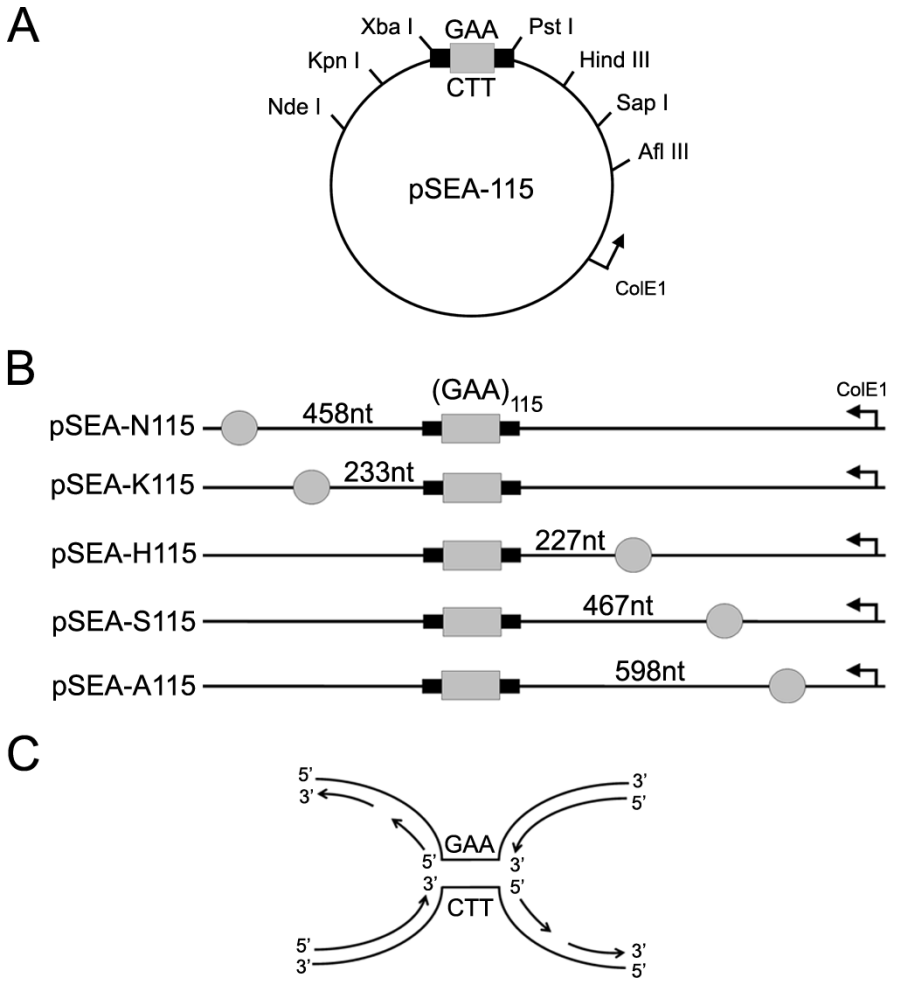

Figure 2. Constructs for replication of $(GAA{\cdot}TTC)_n$ sequences in COS1 cells. (**A**) pSEA-115 is a pUC19 derivative with a patient-derived $(GAA{\cdot}TTC)_{115}$ sequence (gray rectangles), flanked by minimal genomic sequence (black boxes), subcloned into the PstI and XbaI sites. The repeat is in the stable orientation with respect to the ColE1 origin of replication (black arrow) in order to minimize mutagenesis during propagation in *E.coli*. (**B**) The eukaryotic SV40 *ori* was subcloned in one of five locations (NdeI, KpnI, HindIII, SapI or AflIII) flanking the $(GAA{\cdot}TTC)_{115}$ sequence. The distance in nucleotides of the SV40 origin of bidirectional replication (gray circles) from the proximal edge of the repeat tract is indicated for each construct. (**C**) Schematic diagram of replication depicting the DNA strand that will serve as the lagging strand template during replication in COS1 cells (*note*: both directions are shown in the same diagram). Since the SV40 *ori* is a bidirectional origin of replication, its placement determines which strand will serve as the lagging strand template. For example, GAA and TTC will serve as lagging strand templates in pSEA-K115 and pSEA-H115, respectively.

assay [(19); Figure 3)]. Comparison of the overall mutational frequency in all five constructs revealed that mammalian replication produced a unique profile of mutations compared with *E.coli*. In contrast to the very low prevalence of expansions noted in *E.coli*, replication in COS1 cells resulted in an overall higher frequency of expansions (0.8% versus 3.9%, $P < 0.001$). Some of the expansions were quite large (Figure 3); expansions ranged from gains of 6 to 57 triplets [5–49%] with a median expansion size of 18 triplets [gain of 15.7%]. Whereas the size of the contractions were comparable, the overall number of contractions was also significantly higher following replication in COS1 cells (16.1% versus 34.2%, $P < 0.001$).

The presence and type of repeat instability was found to depend on the location of the SV40 *ori* relative to the $(GAA{\cdot}TTC)_{115}$ sequence. The repeat tract was either stable (pSEA-N115 and pSEA-H115), showed predominant expansions (pSEA-S115 and pSEA-A115), or an overall increase in instability as reflected by both enhanced expansions and contractions (pSEA-K115) (Figure 4). The general mutational pattern in *E.coli*, irrespective of the construct used, consisted of ~15% contractions and ⩽1% expansions (with the exception of pSEA-H115, which had an overall higher frequency of mutations). In contrast, while the frequency of contractions was the same in almost all but one construct (pSEA-K115), replication in COS1 cells resulted in a significantly higher prevalence of expansions in three of the five constructs (Figure 4).

The presence and type of instability was also dependent on the orientation of replication in COS1 cells. The pSEA-K115 construct, which showed an overall increase in instability had the GAA sequence as the template for lagging strand synthesis, and the two constructs showing significantly higher prevalence of expansions (pSEA-S115 and pSEA-A115) utilized TTC as the lagging strand template. Comparing the two pairs of constructs in which the SV40 *ori* was

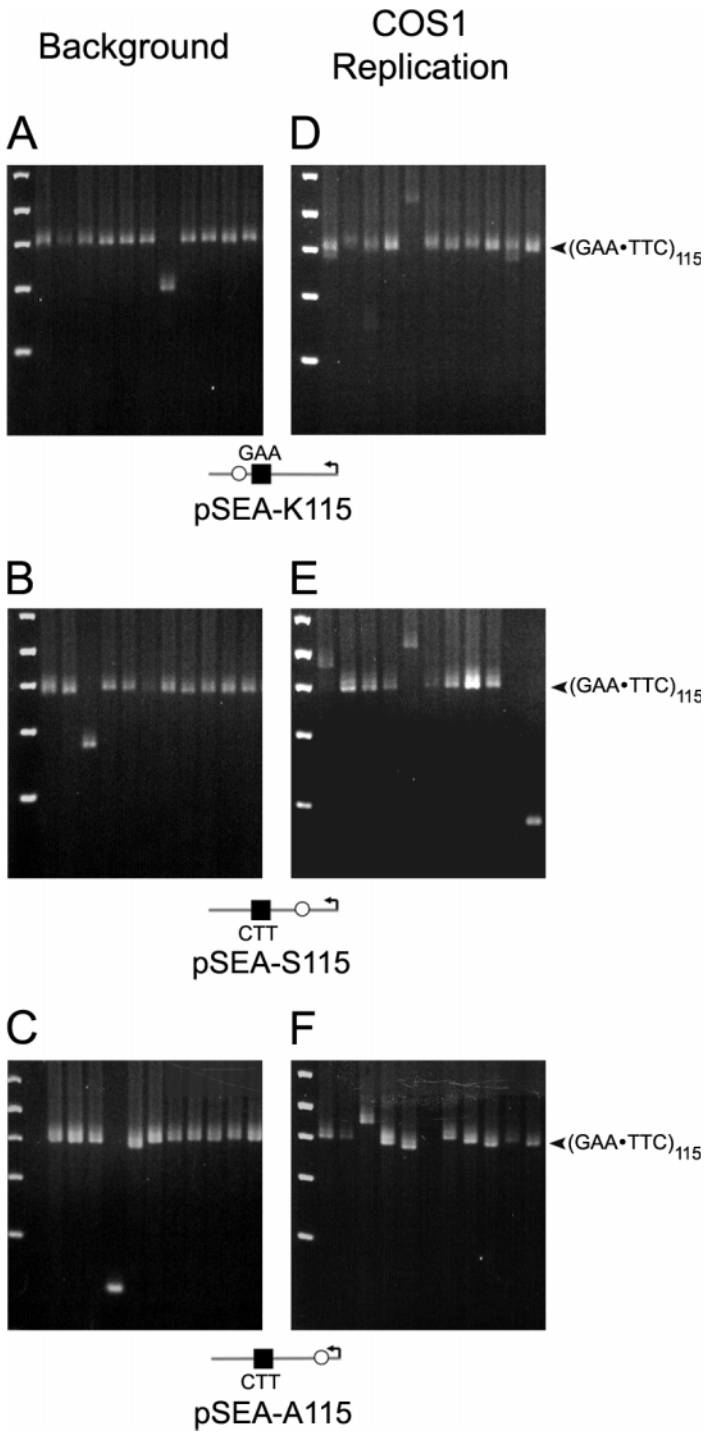

Figure 3. Expansions are induced by replication in COS1 cells. Representative results of colony PCR visualized by agarose gel electrophoresis. (**A–C**) Colony PCR of constructs transformed directly into *E.coli* (without transfection in COS1 cells) to measure the background level of instability for each construct. (**D–F**) Individual products of replication, detected by colony PCR, of the corresponding constructs in (**A)–(C)** after transfection and replication in COS1 cells. Note the presence of expansions following replication in COS1 cells that were only rarely visible after propagation in *E.coli*. Schematic drawings below each pair of gels indicate the relative locations of the SV40 *ori* (open circles) and ColE1 origin of replication (black arrows) with respect to the repeat tract (black boxes) as well as whether GAA or TTC serves as the template for lagging strand synthesis.

located at the same distance on either side of the $(GAA{\cdot}TTC)_{115}$ tract, indicated that the direction of replication clearly influenced repeat instability. While pSEA-N115 (located 458 nt 5′ of the repeat) showed no increase in mutations, pSEA-S115 (located 467 nt 3′ of the repeat) showed a significant increase in expansions, and while pSEA-H115 (located 227 nt 3′ of the repeat) showed no increase in mutations, pSEA-K115 (located 233 nt 5′ of the repeat) showed a significant increase in overall instability.

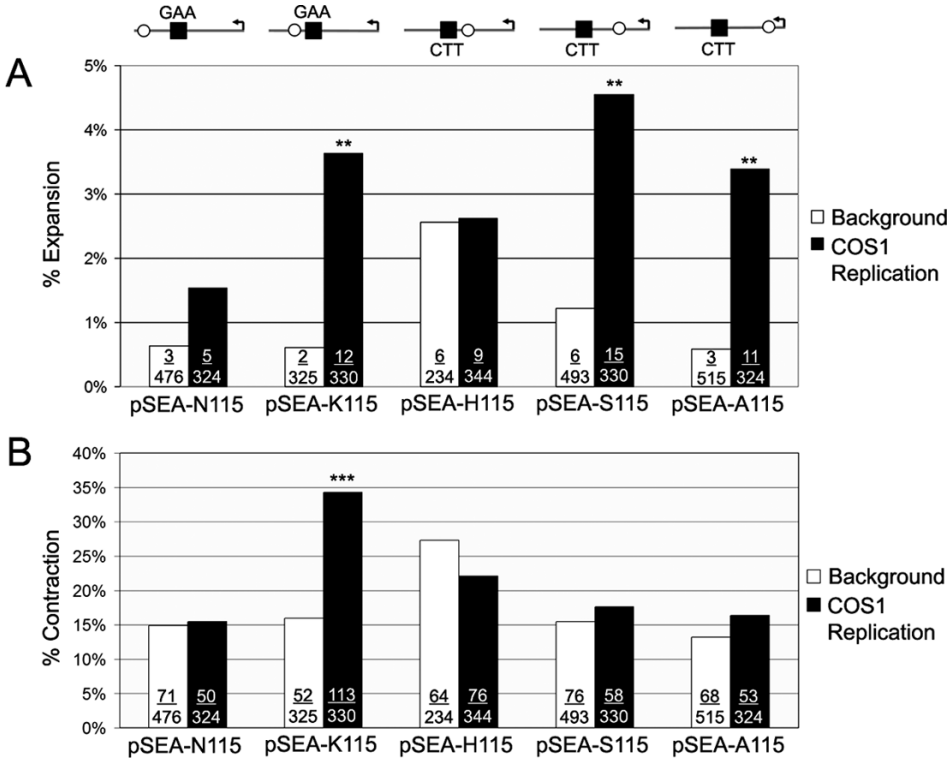

Figure 4. Replication in COS1 cells results in significant levels of $(GAA{\cdot}TTC)_n$ instability. Repeat instability, calculated as the ratio of the total number of length changes (mutant) and the total number of successful PCRs (normal and mutant), is displayed as a percentage. Contractions and expansions were analyzed separately. (**A**) Expansions induced via replication in COS1 cells. (**B**) Contractions induced via replication in COS1 cells. White bars represent the background instability of each construct, i.e. without transfection in COS1 cells. Black bars represent repeat instability following transfection and replication in COS1 cells. The fractions within each bar indicate the total number of mutations (expansions or contractions) over the total number of successful PCRs. The schematic drawings above each data set show the relative location of the SV40 *ori* (open circles), whether GAA or TTC served as the lagging strand template, and the location of the ColE1 origin (black arrows). Asterisks represent statistically significant data determined by the χ^2-test; $^{**}P < 0.01$ and $^{***}P < 0.001$.

DNA replication was required for instability of the GAA triplet-repeat sequence

In order to determine that the mutations arose via DNA replication in COS1 cells, two of the three constructs showing enhanced mutations in COS1 cells (pSEA-K115 and pSEA-A115) were also transfected in the CV1 cell line, a precursor of the COS1 cell line, which does not support SV40 viral replication. As shown in Figure 5A, the frequencies of expansion and contraction after transfection in CV1 cells was identical to that seen in the DNA propagated in *E.coli*, indicating that replication was necessary for the mutations generated by transfection in COS1 cells.

A further control experiment was performed to determine if replication was required where pSEA-115 (i.e. the vector containing the $(GAA{\cdot}TTC)_{115}$ sequence and no SV40 *ori*) was transfected into COS1 cells. As seen in Figure 5B, the frequencies of expansion and contraction after transfection in COS1 cells (determined without DpnI digestion) was identical to that seen in the DNA propagated in *E.coli*, indicating that replication was necessary for the mutations generated by transfection in COS1 cells.

Expansions in COS1 cells are similar to those seen with a $(GAA{\cdot}TTC)_{120}$ allele in somatic cells *in vivo*

The $(GAA{\cdot}TTC)_{115}$ repeat tract used for making the constructs was derived from a FRDA patient with a $(GAA{\cdot}TTC)_{120}$ allele (see Materials and Methods). SP-PCR analysis of blood DNA from this patient was performed to compare the mutation profile of the $(GAA{\cdot}TTC)_{120}$ allele in somatic cells *in vivo* with what we had seen upon replication in COS1 cells (Figure 6A). A total of 33 altered bands were detected out of 139 individual *FXN* molecules analyzed. As we have previously noted for borderline alleles and short E alleles, the $(GAA{\cdot}TTC)_{120}$ allele showed a slight bias for expansions; 14.4% expansions versus 9.3% contractions. Since it was not possible to determine which of the contractions seen after COS1 transfection were specifically due to replication in COS1 cells (due to the high background frequency of contractions in *E.coli*), we compared the expansions seen in somatic cells *in vivo* with those seen after COS1 transfection (this was done for the three constructs which showed a significantly higher frequency of expansion). Even though the range of expansions seems larger after COS1 replication, the magnitudes of expansion were comparable in both systems; the median expansion size for the human *FXN* locus and products of COS1 replication was 16 and 18 triplets, respectively (13.3% versus 15.6%; Mann–Whitney test, $P = 0.66$) (Figure 6B).

DISCUSSION

The high level of instability of $(GAA{\cdot}TTC)_{44+}$ alleles at the human *FXN* locus and the relative stability of similar or

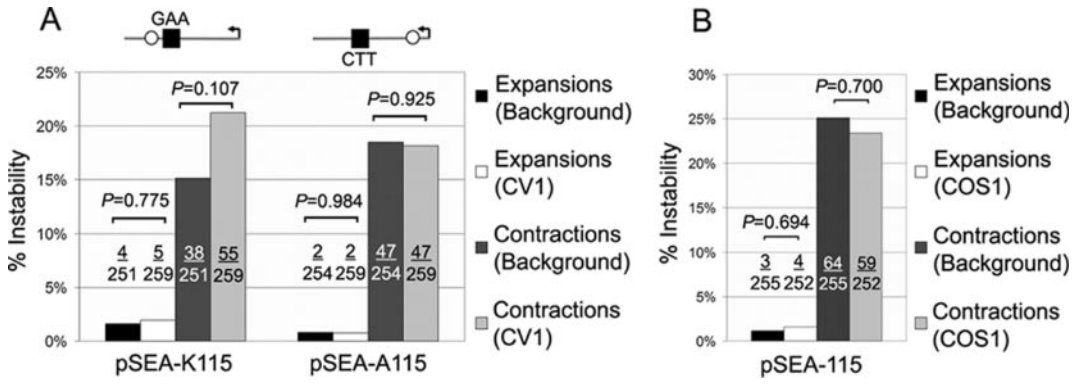

Figure 5. $(GAA{\cdot}TTC)_n$ instability is due to replication in COS1 cells. (**A**) Constructs pSEA-K115 and pSEA-A115, which showed enhanced instability upon transfection in COS1 cells, were transiently transfected into CV1 cells which do not support SV40 viral replication. Background instability (expansion or contraction) was determined for each construct before transfection. Post-transfection DNA extracted from CV1 cells was directly transformed into *E.coli*, without DpnI treatment and analyzed by colony PCR to determine repeat instability (expansion or contraction). No significant difference was found between the prevalence of background mutations and upon transient transfection in CV1 cells, indicating that replication in COS1 cells was essential for repeat instability. The data shown within each bar indicate the total number of length changes over the total number of successful PCRs. The schematic drawings above each dataset show the relative location of the SV40 *ori* (open circles) to the repeat tract (black boxes). (**B**) The pSEA-115 construct, which does not contain the SV40 *ori* sequence, was transfected into COS1 cells, and the resulting instability was compared with the same construct after direct transformation into *E.coli*. No significant difference was found between the prevalence of background mutations and upon transient transfection in COS1 cells, indicating that replication in COS1 cells was required for repeat instability.

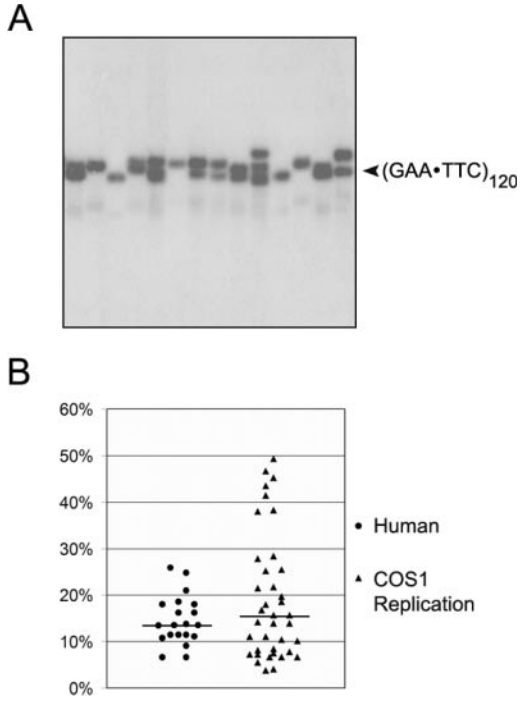

Figure 6. Similar distribution of expansions in COS1 replication of $(GAA{\cdot}TTC)_{115}$ and somatic instability of a $(GAA{\cdot}TTC)_{120}$ sequence at the *FXN* locus in a FRDA patient. (**A**) Small pool PCR analysis of blood DNA from a FRDA patient with a $(GAA{\cdot}TTC)_{120}$ allele at the *FXN* locus (only one allele is shown). (**B**) The magnitude of expansion (%) seen at the human *FXN* locus (circles) is similar to those seen after replication in COS1 cells (triangles; Mann–Whitney test, $P = 0.66$). All expansions seen following replication of pSEA-S115, pSEA-A115 and pSEA-K115 are included. Black lines indicate the median size of expansion at the human *FXN* locus (16 triplets) and in products of COS1 replication (18 triplets).

even longer alleles at other human and murine loci indicates that somatic instability is modulated via locus-specific factors. Clearly, there is not one defined threshold length for the initiation of somatic instability at disparate mammalian genomic loci. Theoretically, locus-specificity may be mediated via several mechanisms (1), including variability in local DNA metabolism (replication, transcription, repair, recombination, other interactions with *trans*-acting factors) or genomic organization (genetic, epigenetic, chromatin, chromosome topology, sub-nuclear localization/interaction). Indeed, one or more of these factors could control local repeat instability, thus determining the characteristic mutability at each genomic locus.

Previous observations indicated that the direction of replication through the $(GAA{\cdot}TTC)_n$ repeat tract could alter the mutation frequency in simple model systems (4,12,13). However, since instability occurred in both directions (albeit with different frequency) this in itself cannot explain the stability we see at certain genomic loci. Moreover, replication in both bacteria and yeast did not reproduce the expansions seen in somatic cells *in vivo* (4,12,13). The distance between the origin of replication and the $(CAG{\cdot}CTG)_n$ sequence resulted in differential instability in mammalian cells (19). For these reasons we investigated both the direction of replication and the distance from the origin of replication as potential mediators for the locus-specific variability we observed at endogenous $(GAA{\cdot}TTC)_n$ sequences. Indeed, replication of the pSEA-K115 construct in COS1 cells reproduced the characteristic instability of alleles of comparable size (borderline and short E alleles) seen at the human *FXN* locus in somatic cells *in vivo* (3,5). Moreover, replication of the pSEA-N115 and pSEA-H115 constructs resulted in no significant instability, thereby mimicking the alleles at the other human and murine genomic loci. That the human *FXN* locus carries some specific mediators of somatic instability is further supported by the observation of similar levels and type of somatic instability in a transgenic mouse model carrying the $(GAA{\cdot}TTC)_n$ repeat within the context of the entire human *FXN* gene (14,15). Similar instability was noted in two transgenic mouse lines with independent insertion sites indicating that it was the human *FXN* locus itself and not the murine sequence flanking the insertion site that supported

the instability. Although it is possible that other factors may cause the locus-specific differences in somatic instability, our data demonstrate that the mere alteration of the conditions of replication was able to reproduce the differences seen *in vivo*.

The role of DNA replication *per se* in mediating the differential instability associated with varying location of the SV40 *ori* is supported by the control transfection experiment in CV1 cells. The instability seen in replication-competent COS1 cells were not reproduced in the replication-incompetent CV1 cells, cell lines that are otherwise similar to each other. Moreover, presence of the SV40 *ori* sequence was required for the repeat instability to manifest. Whereas the $(GAA{\cdot}TTC)_n$ sequence is known to stimulate recombination (20), we do not believe that this is the cause of the differential instability we observed with the various constructs, since they were not reproduced in CV1 cells. Even if recombination plays a role in mediating instability, it would have to be downstream of DNA replication. Indeed, that the distance from the origin of replication influences microsatellite repeat instability, has now been noted for most disease-associated motifs, including $(CTG{\cdot}CAG)_n$, $(CGG{\cdot}CCG)_n$, $(CCTG{\cdot}CAGG)_n$ and now $(GAA{\cdot}TTC)_n$ [(19,21,22) and the present study)]. Cleary *et al*. (19) proposed the 'fork-shift' model [also for a review see (1)] to explain the differences seen in instability wherein the location of the repeat tract in the replication fork during Okazaki priming/processing is critical in mediating repeat instability. However, in our series of constructs we only had one pair of constructs (pSEA-K115 and pSEA-H115) where the origin of replication was located <350 bp from the repeat tract (i.e. those allowing relatively accurate prediction of the location of the Okazaki initiation zone) and therefore we do not have enough data to support or refute this model. It is pertinent to note also that while all of the microsatellite repeats analyzed to date are unstable in *E.coli*, the assay we used to determine if instability arose specifically in mammalian cells involved parallel analyses of instability following isolation of newly replicated plasmids from transfected cells versus direct propagation in *E.coli* and was identical to what has been successfully employed in all previous studies (19,21,22).

The use of the 'threshold length' of instability as a surrogate indicator of genetic instability has been previously employed for the analysis of the $(CAG{\cdot}CTG)_n$ sequence (23,24). The sequence of the repeat but not the flanking GC-content seemed to alter the instability threshold (23). However, interruptions within the repeat sequence stabilized the sequence via a mechanism that was independent of alteration of the instability threshold (23). The latter property is similar to our previous observation that interruptions also stabilize $(GAA{\cdot}TTC)_n$ sequences in human cells and in *E.coli* (4), thus justifying the removal of such sequences from our analysis. However, we did analyze a few long but interrupted sequences and found that they were stable. Pelletier *et al*. (24) also made the observation that $(CAG{\cdot}CTG)_n$ repeat instability beyond a well-defined threshold length required DNA replication in mammalian cells.

Interestingly, expansions were seen in all three constructs that showed instability. The magnitude of the expansions obtained by COS1 transfection seemed to match what was seen in patient cells *in vivo*. One potential caveat of our assay is that large expansions may not be faithfully transmitted via the detection step that required shuttling through *E.coli*. However, it should be noted that we did identify a greater range of expansions, including some large expansions, compared with what was detected by SP-PCR of the endogenous $(GAA{\cdot}TTC)_{120}$ repeat in patient cells. In two of the constructs, both of which utilized TTC as the lagging strand template, we found predominantly expansions. Expansion of the $(GAA{\cdot}TTC)_n$ repeat in somatic cells of patients is likely to be phenotypically significant. For example, further expansion of E alleles noted in the dorsal root ganglia of patients possibly influences the specific pathology seen in FRDA (11), and instability and expansion of borderline alleles possibly determines their likelihood of phenotypic expression (3). It is interesting to note that the rare large expansions seen in our transfection experiments mimicked those seen in DRG and cerebellum in human and mouse tissues in terms of magnitude and frequency (11,15). Our assay will serve as a useful model to decipher the underlying mechanism of expansion in somatic cells.

ACKNOWLEDGEMENTS

This research was made possible in part by grants from the NIH/NINDS, MDA, OCAST and FARA to S.I.B. I.D.-B. is supported by a postdoctoral fellowship from the National Ataxia Foundation. Funding to pay the Open Access publication charges for this article was provided by NIH/NINDS.

Conflict of interest statement. None declared.

REFERENCES

1. Pearson,C.E., Nichol,E.K. and Cleary,J.D. (2005) Repeat instability: mechanisms of dynamic mutations. *Nature Rev. Genet*., **6**, 729–742.
2. Campuzano,V., Montermini,L., Moltó,M.D., Pianese,L., Cossée,M., Cavalcanti,F., Monrós,E., Rodius,F., Duclos,F., Monticelli,A. *et al*. (1996) Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science*, **271**, 1423–1427.
3. Sharma,R., De Biase,I., Gómez,M., Delatycki,M.B., Ashizawa,T. and Bidichandani,S.I. (2004) Friedreich ataxia in carriers of unstable borderline GAA triplet-repeat alleles. *Ann. Neurol*., **56**, 898–901.
4. Pollard,L.M., Sharma,R., Gómez,M., Shah,S., Delatycki,M.B., Pianese,L., Monticelli,A., Keats,B.J. and Bidichandani,S.I. (2004) Replication-mediated instability of the GAA triplet repeat mutation in Friedreich ataxia. *Nucleic Acids Res*., **32**, 5962–5971.
5. Sharma,R., Bhatti,S., Gómez,M., Clark,R.M., Murray,C., Ashizawa,T. and Bidichandani,S.I. (2002) The GAA triplet-repeat sequence in Friedreich ataxia shows a high level of somatic instability *in vivo*, with a significant predilection for large contractions. *Hum. Mol. Genet*., **11**, 2175–2187.
6. Dürr,A., Cossée,M., Agid,Y., Campuzano,V., Mignard,C., Penet,C., Mandel,J.L., Brice,A. and Koenig,M. (1996) Clinical and genetic abnormalities in patients with Friedreich's ataxia. *N. Engl. J. Med*., **335**, 1169–1175.
7. Filla,A., De Michele,G., Cavalcanti,F., Pianese,L., Monticelli,A., Campanella,G. and Cocozza,S. (1996) The relationship between trinucleotide (GAA) repeat length and clinical features in Friedreich ataxia. *Am. J. Hum. Genet*., **59**, 554–560.
8. Montermini,L., Richter,A., Morgan,K., Justice,C.M., Julien,D., Castellotti,B., Mercier,J., Poirier,J., Capozzoli,F., Bouchard,J.P. *et al*. (1997) Phenotypic variability in Friedreich ataxia: role of the associated GAA triplet repeat expansion. *Ann. Neurol*., **41**, 675–682.
9. Monrós,E., Moltó,M.D., Martinez,F., Cañizares,J., Blanca,J., Vílchez,J.J., Prieto,F., de Frutos,R. and Palau,F. (1997) Phenotype

correlation and intergenerational dynamics of the Friedreich ataxia GAA trinucleotide repeat. *Am. J. Hum. Genet*., **61**, 101–110.
10. Pandolfo,M. and Koenig,M. (1998) Clinical and Pathological Aspects of Freidreich Ataxia. In Wells,R., Warren,S. and Sarmiento,M. (eds), *Genetic Instabilities and Hereditary Neurological Diseases*. Academic Press, pp. 373–398.
11. De Biase,I., Rasmussen,A., Endres,D., Al-Mahdawi,S., Monticelli,A., Cocozza,S., Pook,M. and Bidichandani,S.I. Progressive GAA expansions in dorsal root ganglia of Friedreich ataxia patients. *Ann. Neurol*. (In press).
12. Ohshima,K., Montermini,L., Wells,R.D. and Pandolfo,M. (1998) Inhibitory effects of expanded GAA.TTC triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication *in vivo*. *J. Biol. Chem*., **273**, 14588–14595.
13. Krasilnikova,M.M. and Mirkin,S.M. (2004) Replication stalling at Friedreich's ataxia (GAA)n repeats *in vivo*. *Mol. Cell. Biol*., **24**, 2286–2295.
14. Al-Mahdawi,S., Pinto,R.M., Ruddle,P., Carroll,C., Webster,Z. and Pook,M. (2004) GAA repeat instability in Friedreich ataxia YAC transgenic mice. *Genomics*, **84**, 301–310.
15. Clark,R.M., De Biase,I., Malykhina,A., Al-Mahdawi,S., Pook,M. and Bidichandani,S.I. (2006) The GAA triplet-repeat is unstable in the context of the human *FXN* locus and displays age-dependent expansions in cerebellum and DRG in a transgenic mouse model. *Hum. Genet*., in press.
16. Clark,R.M., Dalgliesh,G.L., Endres,D., Gómez,M., Taylor,J. and Bidichandani,S.I. (2004) Expansion of GAA triplet repeats in the human genome: unique origin of the FRDA mutation at the center of an Alu. *Genomics*, **83**, 373–383.
17. Clark,R.M., Bhaskar,S.S., Miyahara,M., Dalgliesh,G.L. and Bidichandani,S.I. (2006) Expansion of GAA trinucleotide repeats in mammals. *Genomics*, **87**, 57–67.
18. 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.
19. Cleary,J.D., Nichol,K., Wang,Y.H. and Pearson,C.E. (2002) Evidence of *cis*-acting factors in replication-mediated trinucleotide repeat instability in primate cells. *Nature Genet*., **31**, 37–46.
20. Napierala,M., Dere,R., Vetcher,A. and Wells,R.D. (2004) Structure-dependent recombination hot spot activity of GAA.TTC sequences from intron 1 of the Friedreich's ataxia gene. *J. Biol. Chem*., **279**, 6444–6454.
21. Dere,R., Napierala,M., Ranum,L.P. and Wells,R.D. (2004) Hairpin structure-forming propensity of the (CCTG.CAGG) tetranucleotide repeats contributes to the genetic instability associated with myotonic dystrophy type 2. *J. Biol. Chem*., **279**, 41715–41726.
22. Nichol Edamura,K., Leonard,M.R. and Pearson,C.E. (2005) Role of replication and CpG methylation in fragile X syndrome CGG deletions in primate cells. *Am. J. Hum. Genet*., **76**, 302–311.
23. Dixon,M.J. and Lahue,R.S. (2004) DNA elements important for CAG·CTG repeat thresholds in *Saccharomyces cerevisiae*. *Nucleic Acids Res*., **32**, 1289–1297.
24. Pelletier,R., Farrell,B.T., Miret,J.J. and Lahue,R.S. (2005) Mechanistic features of CAG·CTG repeat contractions in cultured cells revealed by a novel genetic assay. *Nucleic Acids Res*., **33**, 5667–5676.