

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

Rhonda M. Clark¹, Irene De Biase¹, Anna P. Malykhina², Sahar Al-Mahdawi⁴, Mark Pook⁴ and Sanjay I. Bidichandani^{1,3,*}

Department of Biochemistry & Molecular Biology¹, Physiology² and Pediatrics³, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA; Department of Biosciences⁴, School of Health Sciences and Social Care, Brunel University, Uxbridge, UK

* Address all correspondence to:

Sanjay I. Bidichandani, MBBS, PhD

975 NE, 10th, BRC458

Oklahoma City, OK 73104

USA

Tel: (405) 271-1360

Fax: (405) 271-3910

Email: Sanjay-Bidichandani@ouhsc.edu

ABSTRACT

Friedreich ataxia (FRDA) is caused by homozygosity for *FXN* alleles containing an expanded GAA triplet-repeat (GAA-TR) sequence. This expanded GAA-TR sequence is unstable in somatic cells of FRDA patients, showing age-dependent expansions in dorsal root ganglia (DRG), the tissue where pathology occurs earliest and is most significant. This is thought to be the basis for the progressive, tissue-specific pathology seen in FRDA, but the mechanism(s) for this somatic instability is unknown. We show that transgenic mice containing the expanded GAA-TR sequence (190 or 82 triplets) in the context of the human *FXN* locus show tissue-specific and age-dependent somatic instability that mimics the human condition. Small pool PCR analysis, which allows quantitative analysis of instability by assaying individual transgenes *in vivo*, showed age-dependent expansions specifically in the cerebellum and DRG. The (GAA)₁₉₀ allele showed some instability by 2 months, progressed at about 0.3 – 0.4 triplets/week, resulting in a significant number of expansions by 12 months. Repeat length determined the age of onset of somatic instability, and the rate and magnitude of expansion. Whereas the GAA-TR was unstable in the context of the human *FXN* locus, pure GAA-TR sequences at other genetic loci in the human and murine genomes showed no instability. These data indicate that somatic instability of the GAA-TR sequence in the human *FXN* gene is determined by a combination of unique *cis* and *trans*-acting factors. This mouse model will serve as a useful tool to delineate the mechanism(s) of disease-specific somatic instability in FRDA.

INTRODUCTION

Friedreich ataxia (FRDA) is characterized by progressive sensory ataxia, dysarthria, reduced tendon reflexes, loss of position and vibration senses, and extensor plantar responses (1, 2). These neurological manifestations result from primary degeneration of the dorsal root ganglia (DRG), associated with axonal degeneration in the posterior columns, spinocerebellar tracts, and the corticospinal tracts of the spinal cord, and large myelinated fibers in the peripheral nerves (3). In later stages, the cerebellum is affected, but other regions of the nervous system remain unaffected. FRDA is an autosomal recessive disease caused by inheriting two copies of the *FXN* gene containing an expanded GAA triplet-repeat (GAA-TR) sequence in intron 1 (E alleles). Approximately 1% of the Indo-European population are heterozygous carriers of E alleles, and since *de novo* expansions of premutation to E alleles are rare, this results in an incidence of 1-2 / 50,000. E alleles are associated with a length-dependent deficiency of *FXN* transcript, and this manifests clinically with an earlier age of onset and rapid progression of disease in patients with longer E alleles.

E alleles show a remarkable degree of somatic instability *in vivo*. Small pool PCR analysis of tissues from FRDA patients has shown that somatic instability is length-dependent, tissue-specific and age-dependent. Somatic instability of GAA-TR alleles in peripheral leukocytes initiates when the length is ≥ 44 triplets, with shorter E alleles (<250 triplets) showing a slight expansion bias, and long E alleles (>500 triplets) showing a strong tendency to contract. We recently showed that the mutation load in peripheral leukocytes from patients increases with age. We also noted a tendency for E alleles to undergo further large expansions (>20% over the original length) specifically in

the DRG, a tissue that shows the earliest and most significant pathology in FRDA. More significantly, we detected an age-dependent increase in the frequency of these large expansions in DRG of patients. Similar large expansions, albeit at lower frequency, were also seen in the cerebellum. Thus, it seems that this progressive expansion of E alleles in the DRG and cerebella of FRDA patients may contribute to the specific and progressive tissue pathology seen in FRDA. Therefore, determining the mechanisms of the age-dependent and tissue-specific instability would be important to the understanding of disease pathogenesis, and perhaps even for identifying potential strategies to slow the progression of disease.

Whereas FRDA is the only disease caused by expansion of a GAA repeat sequence, several diseases, including myotonic dystrophy (DM) and Huntington disease (HD) are caused by abnormal expansion of CTG/CAG repeats at their respective genetic loci. Expanded CTG/CAG repeats also show tissue-specific and age-dependent expansions in tissues affected in the disease process. Indeed, several transgenic mouse models have been created to reproduce the somatic instability of CTG/CAG repeats seen in patients. With some exceptions, these mouse models of instability also showed age-dependent, tissue-specific, expansion-biased repeat instability. Somatic instability did not correlate with the proliferative capacity of tissues, but was dependent on the flanking genomic sequence context, and the genetic background of the host.

These transgenic mice are clearly important for the analysis of CTG/CAG instability, but given the differences in the type of somatic instability displayed by expanded GAA-TR sequences in humans, we previously created two transgenic mouse

lines with expanded GAA-TR alleles in the context of the human *FXN* locus (YG8 & YG22 []). They were previously reported to have germline instability and increasing “smearing” of the GAA repeat upon conventional PCR analysis of cerebellar DNA []. Here, we performed a detailed quantitative analysis of multiple tissues from both these mouse lines using SP-PCR and found that they recapitulate the age-dependent expansions in cerebellum and DRG. We also present data that support the existence of *cis*-acting sequence elements at the human *FXN* locus which are essential for somatic instability.

RESULTS

Tissue-specific instability characterized by expansions in cerebellum and DRG

The YG8 and YG22 mice were made by randomly integrating a YAC transgene containing the entire human *FXN* locus with a (GAA)₁₉₀ sequence []. Whereas the YG22 line has a single copy (GAA)₁₉₀ insert, the YG8 line has the (GAA)₁₉₀ transgene but also contains another YAC in tandem with a (GAA)₈₂ repeat tract. These transgenic mice were used to test tissue-specific instability of two lengths of GAA-TR sequences in the context of the human *FXN* locus.

Small pool PCR (SP-PCR) is a powerful technique that allows the study of repeat lengths in individual genes (molecules) represented in an appropriately diluted sample of genomic DNA. We used SP-PCR to analyze somatic instability of the (GAA)₁₉₀ tract in multiple tissues from twelve-month old mice derived from both transgenic lines (Fig. 1). Over 2000 molecules, representing individual somatic transgenes, were analyzed from each transgenic line [a total of 4315 individual (GAA)₁₉₀ molecules]. Mutation load and the frequencies of expansion and contraction were measured for each allele in several tissues (see Materials & Methods). While we did see some instances of expansion and contraction in most tissues, the (GAA)₁₉₀ sequence was most unstable in cerebellum and DRG (Fig. 1). The mutation load in cerebellum was 20.4% and 39.4% in YG8 and YG22, respectively, and in DRG of YG8 it was 8.3% (DRG was not analyzed in YG22). By comparison, the combined mutation loads of the other tissues (i.e., excluding cerebellum and DRG) were only 1.4% and 4.3% in the YG8 and YG22 lines, respectively ($P < 0.001$ in both lines). Somatic instability in cells derived from actively

proliferating tissues such as peripheral blood and sperm was also much lower than in the cerebellum and DRG ($P < 0.001$; Fig. 1).

A significant expansion bias was noticed in all tissues, with at least an 8-fold greater frequency of expansions over contractions in both lines ($P < 0.001$ in both lines; Fig. 2). Again, most of the expansions were noted in the cerebellum and DRG, which accounted for 75% and 44% (92 / 122 and 68 / 154) of all expansions seen in the YG8 and YG22 lines, respectively. It is noteworthy that the same frequency of expansions was noted for the $(GAA)_{190}$ transgene in both YG8 and YG22 mice ($P = 0.7$; Fig. 2). The magnitude of expansions was comparable for the two lines (Fig. 3), with maximum expansions in cerebellum of 50+ triplets noted in both lines [representing an increase of >30%]. Similar instances of expansion by 50+ triplets, although less frequent than in cerebellum, were also noted for the $(GAA)_{190}$ transgene in DRG.

Somatic instability is age-dependent

Next we compared the levels of somatic instability in tissues derived from young (2 month old; 2400 individual molecules) versus old (12 month old; 2000 individual molecules) littermates of YG8 transgenic mice (Table 1; DRG was analyzed at 3 months and 14 months). Through 2 months of age, the $(GAA)_{190}$ allele showed low levels of somatic instability ($\leq 1\%$) including in the cerebellum and DRG (Table 1, Fig. 4A). The $(GAA)_{82}$ allele was completely stable through 2 months, indicating that repeat length may determine the age of onset of somatic instability (Fig. 5). In older littermates we noted a significant increase in mutation load, which was mainly due to the accumulation of expansions. The $(GAA)_{190}$ allele showed a significant age-dependent increase in the

frequency of expansions in DRG and cerebellum ($P < 0.001$ for each tissue; Table 1). The (GAA)₈₂ allele showed a similar age-dependent increase in mutation load, which at 12 months was indistinguishable from the (GAA)₁₉₀ allele in terms of the proportion of mutant molecules ($P = 0.12$ for all tissues [Fig. 5], and $P = 0.36$ for cerebellum only). No appreciable increase in the mutation load was noted up to 12 months in blood and sperm (proliferative cells), spinal cord (another region of the central nervous system), and kidney (the tissue showing maximal instability for CTG/CAG repeats in transgenic mouse models).

We then tested if the age-dependent increase in expansions seen in the cerebellum of YG8 mice was reproduced in the YG22 mice. The (GAA)₁₉₀ allele, which was only slightly unstable at 2 months (data not shown) and 5 months (4.3% of 325 individual molecules), showed a significant increase in mutation load (39.3% of 211 individual molecules; $P < 0.001$ compared with 2 or 5 months), and an expansion bias (4.5-fold greater frequency of expansions over contractions; $P < 0.001$) by 12 months (Fig. 4B).

As a semi-quantitative measure of the dynamics of expansions, we determined the “median rate of expansion” as previously described by Gomes-Pereira et al. The (GAA)₁₉₀ allele showed comparable median rates of expansion in YG8 cerebellum, YG8 DRG, and YG22 cerebellum (+0.33, +0.44, and +0.41 triplet/week, respectively). The cerebellar expansion rate of the (GAA)₈₂ allele was +0.21 triplet/week, i.e. ~40 – 60% less than the cerebellar expansion rate of the (GAA)₁₉₀ allele. Thus, despite similar frequencies of mutation, the slower rate of mutation resulted in smaller magnitude of cerebellar expansions of (GAA)₈₂ [median: 10 triplets; range: 4 – 25] versus (GAA)₁₉₀

[median: 16 triplets; range: 2 – 64] by 12 months of age ($P < .$). These data indicate that the length of the repeat determines the age of onset of somatic instability, the rate of and magnitude of expansion, but not the total number of molecules with mutations (indicated by the proportion of mutant molecules; Fig. 5).

Locus-specific somatic instability of GAA triplet-repeat sequences

The YG8 and YG22 mice have a significant amount of flanking sequence from the human *FXN* locus, and it is therefore not surprising that there was no difference in the pattern of somatic instability in the two transgenic lines despite the random integration of their individual transgenes. To test if flanking DNA sequence alters the level of somatic instability of the GAA-TR sequence *in vivo*, thus acting as a *cis*-acting modifier, we sought to identify other naturally occurring long and pure GAA-TR sequences in the human and mouse genomes. We decided to use the previously determined “minimum threshold length” for the initiation of instability at the *FXN* locus (≥ 44 pure GAA triplets, without any flanking G/A-island sequence) as a quantitative measure of the level of somatic instability. Following a comprehensive search of both genomes we identified several human loci with long GAA tracts [], however, sequencing of all potential loci to identify suitable alleles (≥ 44 pure GAA triplets, without any flanking G/A-island sequence, which we have shown to result in stabilization of adjacent GAA-TR sequences) led to the identification of only one human locus (5q23). In the mouse genome we identified only two out of 107 loci with (GAA)₄₀₊ sequences to have ≥ 44 pure GAA triplets without any flanking G/A-island sequence (1e2.3 and 8b3.3). Both human and murine loci did not map within known or predicted transcriptional units. On

screening of ~100 “normal” human controls we identified suitable alleles at 5q23 that ranged from 59 to 71 pure GAA triplets. Sequencing of the two murine loci showed allele lengths of 48 and 53 pure GAA triplets.

SP-PCR analysis of five human alleles at 5q23 (1701 individual molecules) and the two mouse alleles (688 individual molecules) showed negligible somatic instability (Figure 6; Table 2). The human DNA samples were derived from blood samples from an anonymous panel of normal controls and so we do not know their individual ages, except that they were all adults (≥ 18 years). However, in the case of the mouse, we tested 12-month old cerebellum (shown in Fig. 6E,F) which was completely stable. In comparison, alleles of similar length at the human *FXN* locus (also from blood DNA) were significantly more unstable ($P < 0.001$; Table 2). We similarly tested several long GAA-TR sequences which had flanking G/A-island sequences at other loci the human genome, and some additional shorter pure alleles in the mouse genome, and did not detect any somatic instability (data not shown). Taken together, these data suggest the existence of *cis*-acting determinants of somatic instability at the human *FXN* locus which function in human and transgenic mouse cells.

DISCUSSION

We previously showed that FRDA patients have age-dependent expansions in DRG and to a lesser extent in cerebellum, which we believe to be the basis for the selective and progressive pathology in DRG, and the later involvement of cerebellum. Our transgenic mice recapitulate the age-dependent accumulation of GAA expansions seen in DRG and cerebellum of FRDA patients, although they show greater cerebellar-specific expansions. Despite this subtle quantitative difference, the specific involvement of sites of nervous system pathology in FRDA indicate that these mice may serve as useful models to investigate the mechanistic basis of tissue-specific expansions, and perhaps also for testing interventional strategies in the context of a complex organism. The absence of large contractions in peripheral blood and sperm commonly seen in FRDA patients is most likely due to the smaller size of the GAA-TR sequence in our transgenic mice (190 versus >500 triplets in typical E alleles).

Indeed, the lower level of somatic mutations seen blood and sperm versus non-proliferative tissues such as cerebellum and DRG indicate that DNA replication *per se* is unlikely to be a major cause for the age-dependent expansions. This is consistent with similar observations indicating that DNA replication alone is unlikely to explain the age-dependent, tissue-specific expansions seen in transgenic mice carrying expanded CTG/CAG tracts. It is still unknown if components of the mismatch repair pathway will play a critical role in the genesis of somatic instability in our mice as is known for instability of CTG/CAG repeats. What is striking is the remarkable difference in the tissue distribution of somatic instability displayed in mice carrying transgenes with expanded GAA or CTG/CAG repeats. The CTG/CAG repeat, in several genomic

contexts, consistently shows the highest instability in kidney and the striatum, but very low levels of instability in the cerebellum. In contrast, our mice showed the highest level of instability in the cerebellum (striatum was not analyzed), and kidney showed the least instability (in YG8 we found no detectable instability even at 12 months). The cause(s) for this difference remains unknown, but the specific involvement of disease-specific tissues (DRG and cerebellum in FRDA; muscle, brain and eyes in DM; striatum in HD) is provocative, and suggests the role of cell type-specific, *trans*-acting factors that modify specific triplet-repeat sequences. In fact, Fortune et al. showed that specific fractions of cells within the kidney have a tendency to remain stable or to expand with age. In our mice we observed that the length of the GAA repeat determines the age of onset of somatic instability and the rate of expansions, but not the proportion of mutant molecules (expanded transgenes). This supports the notion that only a specific fraction of cells initially develops expansions, which then subsequently continue to expand as the mouse ages, with longer repeats undergoing faster expansion. It should be noted that while we saw maximal expansions amounting to increases in length by 30 – 40%, we did not observe the very large changes (300 – 500%) that others have seen in transgenic mice with long CTG/CAG repeats. However, this is consistent with what is seen in the corresponding human disease states; very large expansions are seen in somatic tissues of DM and HD patients, and in contrast the large expansions we have seen in the DRG of FRDA patients typically represent length gains of 20 – 40%.

We also found that the genomic context flanking the GAA-TR sequence can modify its somatic instability, and therefore there must be *cis*-acting elements in addition to the above mentioned *trans*-acting factors. Despite a comprehensive search, we were

unsuccessful in identifying any naturally occurring GAA-TR sequences in the human or mouse genomes that is as unstable as the *FXN* locus. Indeed, integrating the entire human *FXN* locus along with the expanded GAA-TR sequence into the mouse genome at two random loci recapitulated the tissue-specific instability seen in human tissues. Furthermore, the two transgenic lines showed a very similar distribution of expansions versus contractions, tissue-specificity of somatic instability, and the magnitude and rate of expansions, indicating that the genomic context afforded by the human *FXN* locus presents the necessary *cis*-acting elements for somatic instability in the appropriate cell type (and is not / less dependent on the mouse genomic sequence flanking the site of transgenic integration). Our data concerning the GAA repeat are consistent with the observation that the appropriate flanking genomic context is also required for somatic instability of CTG/CAG repeats. Indeed, somatic instability was either absent or highly dependent on the transgenic insertion site when very little or no human flanking genomic sequence was used. Somatic instability of CTG/CAG repeats was consistently observed when either a large amount of human flanking sequence was used or the repeat tract was specifically “knocked-in” into the homologous murine locus. It is nevertheless intriguing, notwithstanding the obvious role of *cis*-acting elements, that the sequence motif (GAA versus CTG/CAG) itself seems to also determine the type of instability in specific tissues.

In summary, we describe a transgenic mouse model that recapitulates the disease-related, tissue-specific expansions of GAA repeats seen in tissues from FRDA patients. We suggest that both *cis* and *trans*-acting factors regulate the tissue-specific and age-dependent somatic instability. This mouse model will serve as a useful

resource to delineate the mechanism(s) of somatic instability seen in FRDA patients, and perhaps also as an animal model to test potential interventional strategies.

MATERIALS AND METHODS

DNA Purification: Genomic DNA from all tissues (except sperm) was purified using the Qiagen tissue extraction kit. Sperm DNA was purified using a protocol [] that exploits the differential lysis of sperm tails and non-sperm cells (discarded) and sperm heads (used), in order to ensure that we were analyzing instability in germline cells. Multiple tissues were analyzed in YG8 (at 12 months: blood, sperm, kidney, spinal cord, brain stem, cerebellum; at 2 months: same tissues except kidney; DRG was analyzed at 3 and 14 months) and YG22 (at 12 months: pancreas, heart, skeletal muscle, cerebrum, and cerebellum; at 2 and 5 months: only cerebellum).

Small pool PCR analysis: This was performed as described previously []. Briefly, serial dilutions of genomic DNA, ranging from 6 – 600 pg, were prepared in siliconized microfuge tubes. PCR was performed using primers 147F (5'-GAAGAACTTTGGGATTGGTTGC-3') and 602R (5'-AGGACCATCATGGCCACACTT-3'), which allowed accurate sizing of alleles used in the present study. PCR products were resolved by electrophoresis on 1% agarose gels, and bands detected by Southern blotting using an end-labeled (TTC)₉ oligonucleotide probe. The calculation of the average number of individual molecules per reaction was performed by Poisson analysis as described previously []. 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. Although no significant inter-tissue variability was noted in the size of the constitutional alleles within the same individual,

all calculations of mutation load and measurement of altered allele sizes were performed following accurate sizing of the constitutional alleles in each tissue.

ACKNOWLEDGEMENTS

This research was made possible in part by grants from the National Institutes of Health (NINDS), Muscular Dystrophy Association, OCAST, and FARA to S.I.B.. M.P. acknowledges support from The Wellcome Trust. I. D-B. is supported by a postdoctoral fellowship from the National Ataxia Foundation.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

REFERENCES

FIGURE LEGENDS

Figure 1. SP-PCR analysis showing somatic instability of the GAA triplet-repeat sequence in multiple tissues from YG8 and YG22 transgenic mice. Representative gels are shown for the indicated tissues. The locations of the original (GAA)₁₉₀ and (GAA)₈₂ alleles, as determined by conventional PCR, are indicated by arrowheads. Each lane typically contains 5 – 10 individual molecules (transgenes; see Material & Methods). Note that cerebellum and DRG show higher levels of instability.

Figure 2. Somatic instability of the (GAA)₁₉₀ allele in tissues from YG8 and YG22 transgenic mice shows a significant bias for expansions. Note that both transgenic mice show a >8-fold excess of expansions over contractions.

Figure 3. Similar magnitude of cerebellar expansions of the (GAA)₁₉₀ allele in YG8 and YG22 transgenic mice. Bars represent the frequency of expansions with the indicated increase in repeat length (measured in triplets).

Figure 4. Somatic instability of the (GAA)₁₉₀ and (GAA)₈₂ alleles is age-dependent. Representative gels are shown for cerebellar DNA, and the locations of the original (GAA)₁₉₀ and (GAA)₈₂ alleles, as determined by conventional PCR, are indicated by arrowheads. Each lane contains a comparable number (5 – 10) of individual molecules (transgenes; see Material & Methods). Note the increase in instability and bias for expansion, seen in 12 month versus 2 or 5 month old cerebellum.

Figure 5. Somatic instability of the (GAA)₁₉₀ and (GAA)₈₂ alleles is age-dependent. The graph shows a significant increase in somatic instability (mutation load; see Materials and Methods) of the (GAA)₁₉₀ and (GAA)₈₂ alleles in (all tissues of) 12 month versus 2 month old littermates of YG8 transgenic mice (DRG was collected 13 and 3 months instead).

Figure 6. Somatic instability of the GAA triplet-repeat sequence is seen in the context of the human *FXN* locus. Representative gels are shown for pure GAA triplet-repeat sequences at various genetic loci in the human [(A,B) *FXN* locus [9q21] and (C,D) an extragenic sequence at 5q23) and mouse [(E) 1e2.3 and (F) 8b3.3] genome. The locations of the original (GAA)₄₀₊ alleles, as determined by conventional PCR, are indicated by arrowheads. Each lane typically contains 5 – 10 individual molecules (genes). Note that the GAA triplet-repeat at the human *FXN* locus (A,B) shows higher levels of instability than alleles of similar sequence and length at other loci (C-F). A rare mutation at 5q23 is indicated by an asterisk (C).

TABLES

Table 1. Age-dependent and expansion biased somatic instability of the (GAA)₁₉₀ allele in the YG8 transgenic mouse.

n/a = not analyzed; ^aDRG was analyzed at 3 and 14 months instead.

Table 2. Locus-specific instability of the GAA triplet-repeat sequence.

^aPreviously published data from Refs X and Y

^bHuman genomic locus

^cNovel data; potentially unstable GAA repeats in human and mouse genomes

^dMouse genomic locus

Figure 1

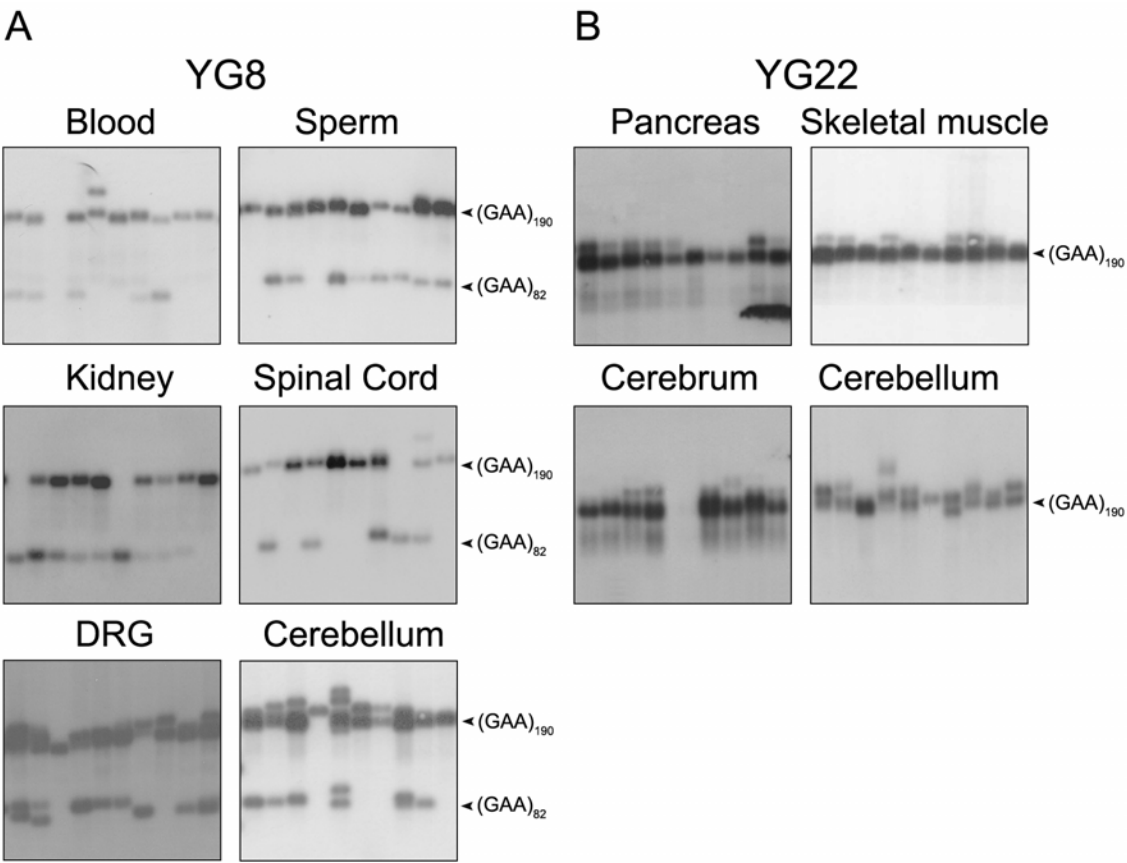


Figure 2

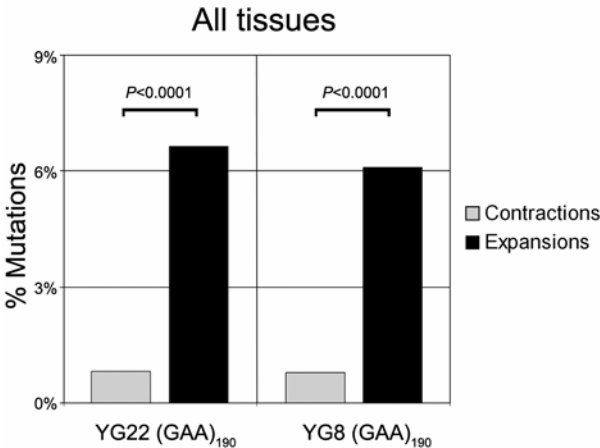


Figure 3

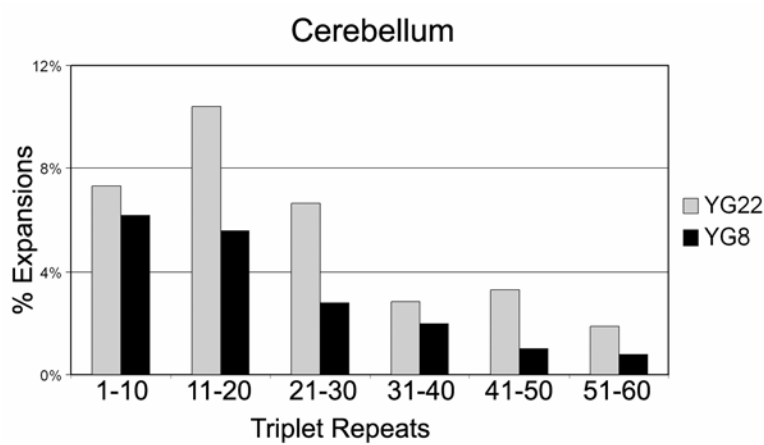


Figure 4

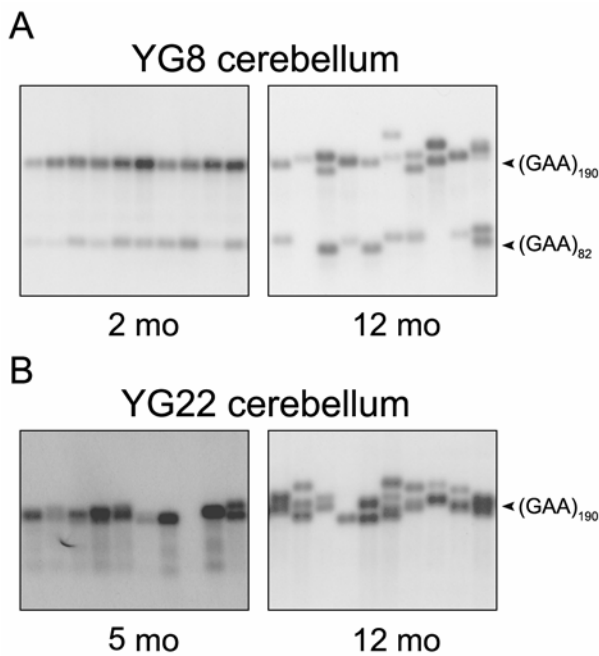


Figure 5

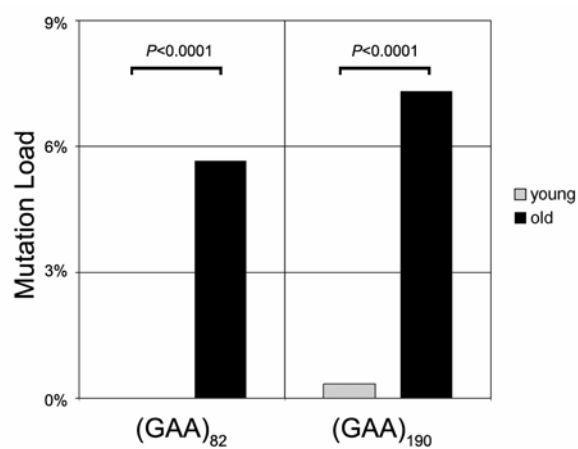


Figure 6

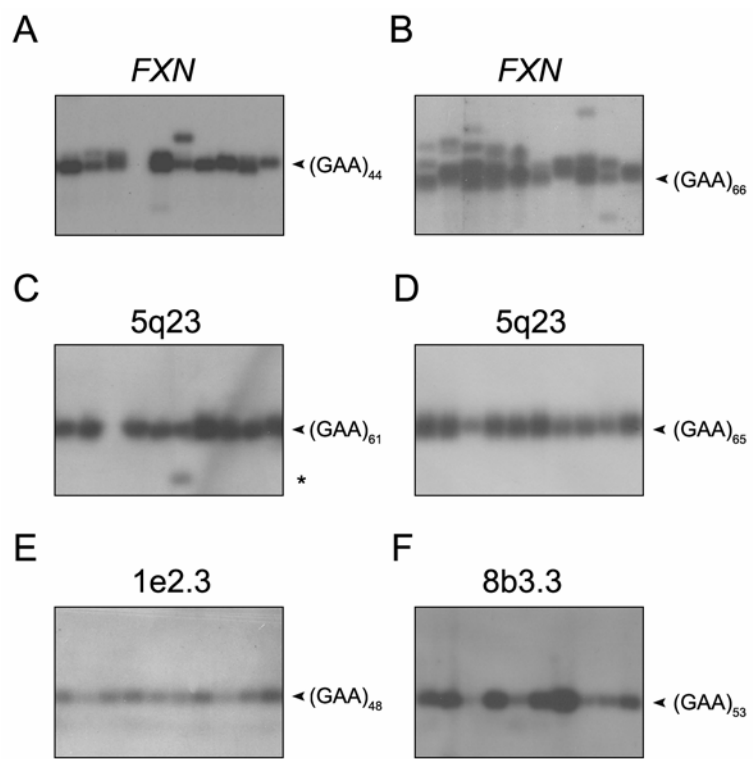


Table 1.

Tissues	2 month				12 month			
	ML (%)	Molecules analyzed	Con.	Esp.	ML (%)	Molecules analyzed	Con.	Esp.
Blood	0	898	0	0	0.6	711	0	4
Sperm	0	121	0	0	2.3	216	0	5
Kidney	N/A	N/A	N/A	N/A	0	189	0	0
Spinal Cord	1	100	0	1	2.4	42	0	1
Brain Stem	0.65	306	0	2	5.5	161	2	7
DRG	0.9	112	1	0	8.3	180	2	13
Cerebellum	0.35	862	0	3	20.4	501	10	92

Table 2.

Locus	GAA size	Mutation Load (%)	Molecules analyzed
<i>FXN</i> 5q23	39	0	1150
	44	6.3	2304
	66	30	1230
	59	<1	555
	61	<1	291
	61	<1	334
	65	<1	223
	71	<1	298
1e2.3	48	0	392
8b3.3	53	0	296