ARTICLE

Androgen receptor YAC transgenic mice carrying CAG 45 alleles show trinucleotide repeat instability

A. R. La Spada1,2,3,*, K. R. Peterson3,4, S. A. Meadows1, M. E. McClain1, G. Jeng1, R. S. Chmelar1, H. A. Haugen6, K. Chen7, M. J. Singer5, D. Moore2, B. J. Trask5, K. H. Fischbeck7, C. H. Clegg6 and G. S. McKnight1

Departments of 1Pharmacology, 2Laboratory Medicine, 3Medicine (Division of Medical Genetics), 4Genetics and 5Molecular Biotechnology, University of Washington, Seattle, WA 98195, USA, 6ZymoGenetics, Seattle, WA 98121, USA and 7Department of Neurology, University of Pennsylvania, Philadelphia, PA 19104, USA

Received February 2, 1998; Revised and Accepted March 11, 1998

X-linked spinal and bulbar muscular atrophy (SBMA) is caused by a CAG repeat expansion in the first exon of the androgen receptor (AR) gene. Disease-associated alleles (37–66 CAGs) change in length when transmitted from parents to offspring, with a significantly greater tendency to shift size when inherited paternally. As transgenic mice carrying human AR cDNAs with 45 and 66 CAG repeats do not display repeat instability, we attempted to model trinucleotide repeat instability by generating transgenic mice with yeast artificial chromosomes (YACs) carrying AR CAG repeat expansions in their genomic context. Studies of independent lines of AR YAC transgenic mice with CAG 45 alleles reveal intergenerational instability at an overall rate of ∼10%. We also find that the 45 CAG repeat tracts are significantly more unstable with maternal transmission and as the transmitting mother ages. Of all the CAG/CTG repeat transgenic mice produced to date the AR YAC CAG 45 mice are unstable with the smallest trinucleotide repeat mutations, suggesting that the length threshold for repeat instability in the mouse may be lowered by including the appropriate flanking human DNA sequences. By sequence-tagged site content analysis and long range mapping we determined that one unstable transgenic line has integrated a ∼70 kb segment of the AR locus due to fragmentation of the AR YAC. Identification of the cis-acting elements that permit CAG tract instability and the trans-acting factors that modulate repeat instability in the AR YAC CAG 45 mice may provide insights into the molecular basis of trinucleotide repeat instability in humans.

INTRODUCTION

X-linked spinal and bulbar muscular atrophy (SBMA, Kennedy’s disease) is caused by a CAG repeat expansion in the first exon of the androgen receptor (AR) gene (1). Disease-associated alleles (37–66 CAGs) undergo repeat length alteration when transmitted from parents to offspring with a significantly greater tendency to change in size when inherited paternally (2,3). Most repeat length changes are small at the AR locus (–4 to +6 CAGs), yet there is a trend toward further expansion with male transmission (4). CAG repeat expansions have been identified as the cause of at least seven other inherited neurodegenerative diseases in addition to SBMA (5,6). For most of the CAG/polyglutamine tract diseases, disease length thresholds are comparable (6). Furthermore, these disorders display high rates of intergenerational instability, a tendency toward repeat expansion and parent-of-origin effects upon repeat instability (7). Although a number of models have been proposed to explain trinucleotide repeat instability (7–10), none satisfactorily accounts for its unique genetic features or has experimental evidence to support it.

To determine the mechanisms that underlie trinucleotide repeat instability investigators have sought to reproduce it in model organisms. The mouse is an ideal model organism because of its short generation time, genetic similarity to humans and ability to be genetically manipulated. Investigators first attempted to create murine models of neurodegeneration and trinucleotide repeat instability by transgenic approaches—introducing disease gene cDNAs with CAG tract expansions into mice. Although some of these studies did yield mice with the expected neurodegenerative phenotypes, initial characterization of these mice failed to detect trinucleotide repeat instability (11–14). Indeed, expanded CAG tracts were stable in mice at repeat lengths that show significant rates of intergenerational instability in humans. Among these stable mice were AR cDNA transgenic lines carrying 45 and

*To whom correspondence should be addressed at: Box 357110, Room NW 120, University of Washington Medical Center, 1959 NE Pacific Street, Seattle, WA 98195-7110, USA. Tel: +1 206 548 6131; Fax: +1 206 548 6189; Email: laspada@mail.labmed.washington.edu
66 CAG repeats (11,15). To produce mice that show intergenerational instability and somatic mosaicism workers turned to extremely long CAG/CTG tracts or maintained moderately sized repeat expansions in their native genomic DNA context (16–18). Although most of the repeat length changes observed in these studies were small, the sex of the transmitting individual and the age at the time of breeding were found to influence repeat instability (17,18). Recent studies of spinocerebellar ataxia type 1 (SCA1) mice carrying 82 CAGs on cDNA constructs now indicate that this repeat is unstable with maternal transmission and that the frequency and magnitude of repeat instability increases with age in these mice (19).

Given the absence of intergenerational instability in mice carrying AR cDNAs with 45 and 66 CAG repeats and the evidence for cis-elements acting to produce instability of intermediate alleles at the Huntington's disease (HD) locus (20), we attempted to model disease pathogenesis and repeat instability at the SBMA locus by generating YAC transgenic mice carrying CAG repeat expansions in the human AR gene. Studies of independent lines of AR YAC transgenic mice with 45 CAG repeats reveal that each line shows an ~10% rate of repeat length instability in transgene-positive progeny. We also find that CAG 45 alleles are significantly more unstable with maternal transmission and as the transmitting female ages. Our results suggest that human locus-specific sequences are required to produce trinucleotide repeat instability in mice. As one unstable YAC transgenic line has integrated an ~70 kb segment of the AR locus, we suggest that cis-acting instability element(s) at this locus may be present within this 70 kb domain.

RESULTS

AR YAC constructs

A 450 kb YAC containing the human AR gene was isolated from a human genomic YAC library. As chimerism of genomic DNA inserts can be a problem in YAC libraries, we performed fluorescence in situ hybridization (FISH) analysis using the AR YAC as the probe. Metaphase spreads of human lymphocytes probed with gel-purified AR YAC DNA yielded a single signal on the proximal long arm of the X chromosome (data not shown), indicating that the AR YAC is not chimeric and does carry genomic DNA from the proper chromosomal region. To determine the location of the ~100 kb AR gene within the 450 kb YAC and to ensure that the entire AR gene was contained within this YAC we performed Southern blot analysis of agarose plugs containing YAC DNA after digestion with rare-cutter restriction endonucleases and fractionation by pulsed field gel electrophoresis (PFGE). We found that the eight exon AR gene is centrally located within this YAC, with ~150 kb of DNA 5' of the start site of transcription and ~200 kb of DNA 3' of the last exon.

Sequencing of the repeat region in the AR YAC showed that it contains 20 CAGs (AR YAC CAG 20). The CAG repeat in the AR gene is located at the beginning of the coding region in the 2.8 kb first exon. We targeted CAG repeat expansions into AR YAC CAG 20 by a two-step gene replacement method, using the LYS2 gene in a positive selection step followed by a negative selection step (Fig. 1). From the transformation with the AR exon 1 fragment with 45 CAGs we obtained a YAC carrying 45 CAG repeats (AR YAC CAG 45). Southern blot analysis of exon 1 and sequencing of the repeat region and of the previous site of LYS2 integration revealed no rearrangements, deletions, insertions or other sequence alterations (data not shown).

Generation of AR YAC transgenic mice

We isolated intact AR YAC CAG 45 DNA in a high salt buffer (21) and microinjected it at low concentrations (0.5–1.0 ng/µl) into C57BL/6J x C3H hybrid mouse oocytes that were then transplanted into foster mothers. We screened tail DNA from 72 progeny using primers specific for the human AR CAG repeat and identified four positive founder mice (5.6%). Three of the four founder mice were able to transmit the AR YAC transgene, but one female founder (TG61) failed to breed. Transmission of the AR YAC transgene occurred at the expected 50% frequency in the F1, N2 and N3 mice derived from the breeding founders.

Assessment of AR YAC integrity and copy number

To ensure that the AR YAC had not rearranged during integration we performed Southern blot analysis using probes from the 5' and 3' regions of the AR gene. A human AR exon 1 probe yielded restriction endonuclease fragments of expected lengths in transgenic founders, confirming that the region containing the AR CAG repeat expansion is unrearranged in all four lines of mice.

![Figure 1. Generation of the AR YAC CAG 45 construct by two-step gene replacement. The selectable gene, LYS2, was cloned into the NruI site of exon 1 of the AR gene on a plasmid carrying the AR cDNA with 45 CAG repeats (pCMV-AR45). An Eagl-KpnI fragment containing the LYS2 gene with a 45 CAG repeat was isolated and transformed into AR YAC CAG 20 yeast spheroplasts that were plated out on medium lacking lysine, uracil and tryptophan. Yeast colonies carrying AR YACs with the LYS2 gene and a 45 CAG repeat integrated at exon 1 of the AR gene were identified. To recombine out the LYS2 gene and obtain the final AR YAC CAG 45 construct an Eagl-KpnI fragment with a 45 CAG repeat but without the LYS2 gene was then transformed into AR YAC CAG 45–LYS2 yeast spheroplasts that were plated out on medium containing α-amino adipic acid. The box representing the CAG repeat domain is not drawn to scale.](image-url)
Figure 2. (a) Long range mapping of the human AR gene and flanking DNA regions. Rare-cutter restriction endonuclease sites are as follows: B, BssHII; C*, PacI; E, EagI; F, SfiI; M, MluI; P*, PmeI; S, SacII; W*, SwaI. Those symbols followed by an asterisk indicate rare-cutter restriction endonucleases for which additional unmapped sites within the 3′-flanking region DNA may be present. The order and relative locations of 10 PCR markers used for screening are included. The map locations of the anonymous STS markers 611 and 3357 are as indicated. (b) STS content analysis of the AR YAC CAG 45 transgenic lines. For each of the transgenic lines results of the PCR screening are given as follows: +, present; −, absent.

(design not shown). We compared the intensity of Southern blot bands seen in our transgene-positive mice with those seen in normal human males and females and thus determined that each founder had integrated one or two copies of the AR YAC transgene. Transgene-negative mice did not show any cross-hybridizing bands with the AR exon 1 probe. The observation of PCR products of a single size repeat length in progeny mice showing repeat instability (see below) also suggests that all breeding founders integrated a single copy of the AR YAC CAG 45 construct.

Sequence-tagged site (STS) content mapping of the AR YAC transgenic mice

To determine the extent of the human AR locus present in each transgenic line we generated PCR primer pairs specific for the left and right arms of the AR YAC, as well as for various exons of the AR gene. To map the DNA regions flanking the human AR gene in the AR YAC transgenic mice we synthesized oligonucleotides generated by the Washington University chromosome X mapping project and were kindly provided with primer sequences for Xq11–12 STS markers by M. Schueler and H.F. Willard (Case Western Reserve University). Although the order of the STS markers on proximal Xq is established, relative distances between these markers are not known. To localize these anonymous Xq11–12 STS markers we probed pulsed field gel blots of AR YAC agarose plugs digested with various rare-cutter restriction endonucleases (data not shown). In this way we were able to generate a long range map of the AR gene and the DNA 5′ of the start site of the AR gene (Fig. 2a).

STS PCR analysis indicated that fragmentation of AR YAC CAG 45 had occurred during isolation or microinjection and that each breeding line had integrated different segments of the human AR locus. Transgene-positive F1 progeny of TG1 were positive for the YAC left arm PCR primers as well as for all intragenic AR exon primers tested, but were negative for the YAC right arm primers, indicating that the AR gene and ~150 kb of DNA 5′ of the beginning of the AR gene are present (Fig. 2b). STS PCR analysis further indicated that TG1 transgene-positive mice contain all 5′-flanking sequence markers and all eight AR exons, but are negative for 3362, the first STS marker 3′ of the final AR exon (Fig. 2b). Founder TG61 was positive for all primer pairs tested (Fig. 2b), suggesting intact integration of the entire AR YAC CAG 45 construct. We could not confirm this by testing for co-segregation of STS PCR markers, however, as TG61 did not produce any progeny.

Although STS PCR analysis of the TG33 founder suggested that the entire AR gene had integrated as a single unit, STS PCR screens of TG33 F1 individuals showed that two segments of the AR locus had integrated at unlinked sites and were segregating independently. We therefore selected TG33 F1 individuals carrying the segment of the AR locus containing the CAG repeat expansion for further study. These TG33 F1 individuals, as well as F1 individuals from line TG55, were positive for exon 1 primers but not exon 2 primers or YAC left arm primers (Fig. 2b), indicating that for both of these lineages the 3′ boundary of integrated AR YAC DNA falls within an at least 20 kb intron 1 and the 5′ boundary of the integrated AR locus DNA is upstream to the start of the AR gene. STS 3357 represents the 5′ boundary of integrated AR YAC DNA for line TG33, while STS 611 is also
positive in progeny of TG55 (Fig. 2b), indicating that TG55 mice carry more upstream DNA than TG33 mice but less upstream DNA than TG1 mice. Pulsed field gel blot mapping of STS markers 611 and 3357 showed that STS 3357 is ∼50 kb upstream of the AR gene and that STS 611 is ∼85 kb 5’ of the beginning of the AR gene (Fig. 2a).

**AR YAC CAG 45 mice show intergenerational repeat instability**

We extensively bred lines TG1 and TG33 and studied transmission of the CAG repeat expansion in transgene-positive progeny. Both lines of AR YAC transgenic mice showed significant rates of intergenerational CAG repeat instability (Fig. 3). For line TG1 we observed nine repeat length changes in 101 transmissions and for line TG33 we detected eight repeat length changes in 77 transmissions (Table 1). The rates of instability for lines TG1 and TG33 were similar (8.9 versus 10.4%). These rates of instability for transmissions and for line TG33 we detected eight repeat length changes in 101 representations of the above, TG55 and TG33 progeny were found to carry similar inherited repeat length changes were mostly small contractions, while maternally inherited repeat length changes were mostly small contractions, with occasional small expansions.

### Table 1. CAG repeat instability in the AR YAC CAG 45 mice

<table>
<thead>
<tr>
<th>TG line</th>
<th>Alterations/transmissions (rate of instability)</th>
<th>[repeat length changes]</th>
<th>Female</th>
<th>Male</th>
<th>Total</th>
<th>Female</th>
<th>Male</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>TG line 1</td>
<td>7/49 (14.3%)</td>
<td>5/24 (19.2%)</td>
<td>12/73 (16.4%)a</td>
<td>[-1 (5), -2 (2)]</td>
<td>[+1, -1 (2), -2, -3]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TG line 33</td>
<td>5/2/4 (9.8%)</td>
<td>4/105 (4.8%)a</td>
<td>5/4/105 (4.8%)</td>
<td>[+1, -18]</td>
<td>[+1 (2), -20]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both</td>
<td>9/101 (8.9%)</td>
<td>8/77 (10.4%)</td>
<td>17/178 (9.6%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aStatistically significant by \( \chi^2 \) analysis, \( P < 0.01 \).

### Table 2. Effect of parental age on CAG repeat instability in the AR YAC CAG 45 mice

<table>
<thead>
<tr>
<th>Parental age</th>
<th>Alterations/transmissions (rate of instability)</th>
<th>[repeat length changes]</th>
<th>Female</th>
<th>Male</th>
<th>Female</th>
<th>Male</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 32 weeks</td>
<td>0/17 (0%)</td>
<td>1/22 (4.6%)</td>
<td>1/30 (3.3%)</td>
<td>[+1 (2), –2, –3]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 32 weeks</td>
<td>1/12 (8.3%)</td>
<td>1/13 (7.7%)</td>
<td>4/12 (33.3%)</td>
<td>2/40 (5.0%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1/29 (3.4%)b</td>
<td>2/35 (5.7%)c</td>
<td>11/44 (25.0%)b</td>
<td>3/70 (4.3%)c</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

bStatistically significant by \( \chi^2 \) analysis, \( P < 0.05 \).
c\( \chi^2 = 0.10, \ P = 0.74 \).

### Effect of parental sex and parental age upon repeat instability

We compared the frequency of repeat length alteration in the 178 offspring of transgene-positive males and females and found that female mice are three times more likely to transmit an altered repeat than male mice (Table 1). Although males transmit their repeat expansions more stably, the two large contractions detected in this study were inherited from transgene-positive male mice. We also considered the effect of parental age and found that rates of trinucleotide repeat instability significantly increase with age only for transgene-positive female parents (Table 2). About one quarter of all offspring of transgenic female parents older than 32 weeks of age at the time of birth showed shifts in trinucleotide repeat length. For offspring of female parents 32 weeks of age or younger at the time of birth the rate of instability in transgenic progeny averages only ∼3%. Rates of repeat instability were ∼5% and were equivalent for younger and older transgenic male parents, however.

### AR YAC CAG 45 mice do not show somatic mosaicism

The absence of multiple PCR products in the instability studies suggests that each AR YAC transgenic mouse carries a single CAG repeat expansion of uniform length in tail DNA. Somatic instability of moderately sized CAG/CTG repeat expansions, however, can occur in various other tissues in transgenic mice and affected humans (17,18,22,23). We therefore selected two sets of age-matched and gender-matched transgene-positive mice from lines TG1 and TG33 and isolated DNAs from liver, lung, testes or ovaries, kidney, brain and spinal cord for small pool PCR analysis of the CAG repeat region (24). PCR amplification of ∼6 or ∼25 genome copies of DNA from each of these tissues yielded...
identically sized repeat length products (Fig. 4), indicating that expanded CAG repeats are somatically stable in the AR YAC CAG 45 transgenic mice.

Expression analysis

Of the three breeding lines of AR YAC CAG 45 mice generated only line TG1 appears to have integrated the entire human AR gene, based on STS content mapping and Southern blot analysis. Lines TG33 and TG55 have integrated the first exon of the AR gene and various amounts of upstream DNA. As the AR CAG repeat is translated at the beginning of a 1.2 kb coding sequence in the first AR exon and previous studies have shown that expression of truncated huntingtin or truncated ataxin-3 is sufficient to produce neurodegenerative phenotypes in mice (14,25), we screened TG33 and TG55 mice for expression of a truncated AR gene product in addition to probing for full-length human AR protein in TG1 mice. We incubated western blots of protein lysates isolated from brain, spinal cord, testes, seminal vesicles, kidney, liver and lung with human-specific or non-specific anti-AR antibodies raised against peptides from the N-terminus. Although we detected human AR protein in control samples, we did not find truncated human AR protein products that were specific to TG33 or TG55 tissues (Fig. 5a). We also observed full-length murine AR protein in our transgenic and non-transgenic samples, but did not see expression of human AR protein with 45 glutamines in TG1 mice (Fig. 5b). Reverse transcription-PCR amplification of RNA samples from transgene-positive tissues similarly showed lack of expression of the human AR gene at the RNA level for all three breeding lines (data not shown).

DISCUSSION

Numerous lines of transgenic mice carrying trinucleotide repeat expansions have been created to reproduce disease pathology and model intergenerational repeat instability. Among these are the AR CAG 45 and 66 mice (11,15), HD CAG 44 mice (13), SCA3 CAG 79 mice (14), SCA1 CAG 82 mice (19), HD CAG 113–156 mice (17), myotonic dystrophy (DM) CAG 162 mice (18) and DM CAG 55 mice (16). While the first four of these studies found expanded repeats to be stable in their transmission, the latter four studies observed significant rates of intergenerational instability and/or somatic mosaicism. A reasonable interpretation of all this work is that trinucleotide repeat instability may be achieved in mice by using extra long repeat tracts (i.e. >80 triplets) and/or by including flanking genomic DNA in cis.

Studies of humans reveal that sequences in cis affect trinucleotide repeat instability at CAG/polyglutamine tract loci. Two types of cis-acting elements likely exist: (i) sequences within or adjacent to the CAG repeat itself; and (ii) flanking sequences at some distance from the CAG repeat. In SCA1 and SCA2 loss of CAT or CAA interruptions of the CAG repeat appear to be a prerequisite for instability of the CAG tract (26,27). Sequences immediately adjacent to the SCA3 CAG repeat appear to influence its stability (28,29). At the HD locus PCR analysis of single sperm has shown that intermediate alleles (IAs) of identical sizes can have markedly different rates of expansion to the disease size range (30). Differences in expansion tendencies are even seen for comparably sized IAs within the same individual (20). Although a very small percentage of these unstable IAs show changes in adjacent sequences, most do not, strongly arguing for cis -information within the haplotype that shows the expansion tendency. Other evidence for locus-specific sequences includes identification of an unusually stable CAG repeat expansion in HD patients from Crete. Affected individuals from this isolated population have a late onset form of HD, with zero to one repeat unit differences between generations (31). The SCA7 CAG repeat is highly unstable upon transmission at tract lengths that are only moderately unstable at the SBMA and HD loci (5). Additional evidence for the existence of cis-acting elements comes from the recently identified ERDA1 locus (32). Intergenerational stability of CAG/CTG alleles numbering 66–87 repeats at this locus likely indicates the importance of flanking...
sequences in producing size changes, as these CAG/CTG repeats are stable at tract lengths that well exceed the recognized size threshold for repeat instability.

As AR transgenic mice with 45 or 66 CAG repeats introduced on cDNA constructs do not show instability (11,15) we examined repeat instability at this locus by generating mice with AR YACs carrying targeted CAG repeat expansions. In lines of transgenic mice generated with the AR YAC CAG 45 construct we observed equivalent rates of intergenerational instability in transgene-positive progeny, regardless of lineage or generation. In the present study we used C57BL/6j × C3H hybrids as parental founders and backcrossed to C57BL/6j individuals to produce F1, N2 and N3 progeny. Although the AR cDNA CAG 45 and CAG 66 mice were produced as C57BL/6j × SJL/J hybrids, these mice were similarly backcrossed to the C57BL/6j strain (11,15). The AR YAC transgenic mice and the AR cDNA transgenic mice thus share 50–95% of the C57BL/6j genetic background, depending on their generation. Absence of instability in the AR cDNA N3 generation transgenic mice makes strain background an unlikely explanation for the repeat instability that we observe in the AR YAC transgenic mice. Other differences between the AR YAC transgenic mice and the AR cDNA transgenic mice are the size of the integrated construct and the expression status of the AR gene. Although instability due to disruption of the mouse genome by integration of a large piece of DNA cannot be excluded, CAG 44 tracts in HD transgenic lines carrying insertions of up to 120 kb were found to be intergenerationally stable in one study (13). Differences in chromatin structure due to the absence of AR gene expression might also play a role, though intergenerational repeat instability both in the presence and absence of gene expression is well documented (16–19). Instability of CAG 45 alleles in AR YAC transgenic mice suggests that flanking DNA from the human AR locus causes AR CAG repeat instability in the mouse. Of all the CAG/CTG repeat mice found to show instability to date, the AR YAC CAG 45 mice are unstable with the smallest trinucleotide repeat mutations, indicating that the length threshold for repeat instability in the mouse may be lowered by including the appropriate sequence information from the human locus. The only other mice to show repeat instability with <60 triplets are the DM CTG 55 mice that were generated with a 45 kb genomic fragment (16). Interestingly, of all the murine instability models produced the DM CTG 55 mice show intergenerational instability frequencies that are most comparable with those observed in the AR YAC CAG 45 mice (16). Repeat instability in the AR YAC CAG 45 mice and the DM CTG 55 mice suggests that human flanking DNA is necessary if instability is to be obtained in mice with moderately sized repeat tracts.

The AR YAC mice that we produced contain different segments of the AR locus, because fragmentation of the 450 kb AR YAC construct occurred during isolation or microinjection. Integration of unique fragments of the AR YAC construct in our breeding lines of transgenic mice has allowed us to delineate the extent of flanking DNA needed to produce intergenerational instability at the human AR locus in mice. By STS content and mapping analysis we determined that line TG33 has integrated an ~70 kb segment of the human AR locus containing the 45 CAG repeat. We conclude that cis-acting sequences (‘instability elements’) necessary for CAG repeat instability at the human AR locus are likely contained within this domain. We speculate that these instability elements might be binding sites for proteins that regulate DNA conformation or might themselves participate in formation of specific DNA structures during replication (33). Characterization of the DNA sequences that permit or modulate repeat instability at the human AR locus in mice may thus provide an insight into the basic mechanisms that underlie trinucleotide repeat instability in humans.

As the AR YAC CAG 45 mice show intergenerational instability, we looked for repeat instability within neural and non-neural tissues in transgene-positive mice. To assess our mice for somatic instability we performed small pool PCR analysis. All of the tissues from the transgene-positive mice that we studied showed a single repeat length product, indicating somatic stability. This result, though inconsistent with the somatic instability observed in the DM CTG 55, DM CTG 162 and HD CAG 113–156 mice (16–18), does agree with data from SBMA patients who show no somatic mosaicism in either neural or non-neural tissues (34,35).

Absence of mRNA and protein expression in our transgenic mice likely reflects fragmentation of the AR gene in lines TG33 and TG55. Although the AR gene appears intact in line TG1, the 3′boundary of the integrated fragment is between the final exon and STS 3362, suggesting that 3′sequences essential for production of stable transcript are lacking. In the HD CAG 113–156 mice absence of repeat instability was noted for one transgenic line that failed to show expression of the construct, implying that transcription of the CAG repeat expansion may be correlated with instability (17). Repeat instability in both lines of AR YAC CAG 45 mice in the absence of gene expression, however, suggests that transcription is not required for instability to occur at the AR CAG repeat locus. Generation of AR YAC CAG 45 mice that show expression will allow us to determine if expression status affects trinucleotide repeat instability at this locus.

We identified two factors, parental sex and parental age, that influence intergenerational repeat instability in AR YAC transgenic mice. We found that maternally inherited CAG repeats are more unstable than paternally inherited CAG tracts and that older transmitting females show higher rates of CAG repeat length alteration. These results are consistent with previous studies of CAG/CTG tract instability in transgenic mice (17–19). Indeed, the SCA1 CAG 82 mice only show repeat instability with maternal transmission of the expanded CAG tract and demonstrate a pronounced effect of age upon repeat instability (19). Thus three murine models of CAG tract instability (the HD CAG 113–156 mice, the SCA1 CAG 82 mice and the AR YAC CAG 45 mice) indicate that as transgene-positive mice age they transmit expanded CAG tracts less stably. It is also interesting that parent-of-origin differences in repeat instability are observed in the DM CTG 162 mice, the HD CAG 113–156 mice, the SCA1 CAG 82 mice and the AR YAC CAG 45 mice. Maternal transmission of expanded CAG tracts in mice appears associated with higher rates of intergenerational instability and/or a tendency toward repeat contraction. For the AR YAC CAG 45 transgenic mice paternal instability changes consisted of either large contractions or small expansions. Therefore, in the AR YAC CAG 45 mice and the other murine models frequency and direction of repeat length change vary with respect to the sex of the transmitting parent. Whether the germline factors producing these differences in mice will be relevant to the repeat instability process in humans remains to be seen.

How does the trinucleotide repeat instability seen in the AR YAC transgenic mice and other transgenic mice carrying
CAG/CTG tract expansions compare with the trinucleotide repeat instability seen in humans? Important differences between the murine models and the human repeat disease loci include: (i) lower rates of repeat instability in mice; (ii) absence of large repeat expansions in mice; (iii) tendency of repeats to contract rather than expand in mice; and (iv) reversal of the parent-of-origin effect in mice. The basis of these differences remains unclear, although it is tempting to speculate that unique molecular pathways operate during germ cell differentiation or early development in humans. Certain molecular events and molecular interactions during these processes may thus make humans susceptible to trinucleotide repeat expansion and instability. Despite the differences noted above, the trinucleotide repeat instability in the murine models does retain some features seen in humans: (i) there is a length threshold effect in mice, albeit at longer repeat lengths; (ii) parental sex does affect repeat instability in mice; and (iii) flanking genomic sequences can modulate repeat instability in mice. Although the repeat instability patterns observed in the murine models do not perfectly reproduce those seen in humans, these common features suggest that mice are potentially useful to approach the problem of trinucleotide repeat instability.

In summary, we have produced trinucleotide repeat instability in mice using YAC segments of the human AR gene into which a 45 CAG repeat has been targeted. Instability in unique AR YAC transgenic lines with fragments of this construct lowers the size threshold for repeat instability in mice to 45 CAG repeats. We conclude that repeat instability in our AR YAC CAG 45 mice and in DM CTG 55 mice, together with the results of studies in humans, suggest that locus-specific sequences are required for repeat instability. We propose that cis-acting DNA sequences (‘instability elements’) likely exist at the human AR locus and operate at other trinucleotide repeat loci. As parental sex and age significantly affect CAG/CTG tract instability in the AR YAC CAG 45 mice and in other murine models of repeat instability, we predict that germline-specific trans-acting factors and DNA repair pathways influenced by the aging process are involved in trinucleotide repeat instability in mice. Elucidation of the mechanisms producing trinucleotide repeat instability in mice may point us toward the pathways that underlie trinucleotide repeat instability in humans.

MATERIALS AND METHODS

YAC clone isolation

Primers specific for the first exon of the androgen receptor (AR) gene were used for PCR screening of DNA pools of the CEPH human genomic YAC library (36). This work was performed by A. Craig Chinnault (Baylor College of Medicine, Houston, TX), who identified clone 550 D1, which we refer to hereafter as AR YAC CAG 20.

Fluorescence in situ hybridization (FISH)

Slides containing metaphase spreads of lymphocytes from normal XY and XX individuals were prepared as described previously (37). AR YAC CAG 20 DNA, purified by PFGE, was labeled with biotin–labeled primer by a universal primed PCR procedure (DOP-PCR) and used to probe these slides (38).

YAC targeting and recombination

We obtained the pLNT plasmid clone of the LYS2 gene (39). The LYS2 gene was subcloned into a blunt-ended NotI site in exon 1 of an AR cDNA clone carrying 45 CAG repeats (pCMVAR45, kindly provided by Elizabeth Wilson). Spheroplasts of AR YAC CAG 20 were prepared following standard protocols (40). Aliquots of 300 ng gel-purified Eagl–KpnI exon 1 fragment containing the LYS2 gene inserted into it was used to transform the AR YAC CAG 20 spheroplasts with 4.7 µg plasmid DNA as carrier. Transformants were selected for on plates lacking lysine, uracil and tryptophan. In this way we generated YAC 2-41, carrying 45 CAG repeats and the LYS2 gene inserted 3′ of the repeat locus. We deleted the LYS2 selectable marker by transforming an Eagl–KpnI exon 1 fragment carrying 45 CAG repeats into YAC 2-41 spheroplasts that were plated out on selective medium lacking uracil and tryptophan, but containing α-amino adipic acid, a compound that the enzyme Lys2 converts into a toxic metabolite. Thus this second round of gene replacement negatively selects for homologous recombination of the LYS2 gene out of exon 1 of the AR YAC. Transformant colonies were screened by PCR and Southern blot analysis.

Generation of AR YAC CAG 45 mice

AR YAC CAG 45 DNA was isolated intact in a high salt buffer as previously described (21). Gel-purified AR YAC CAG 45 DNA was then microinjected at low concentrations (0.5–1.0 ng/µl) into mouse pronuclei in (C57Bl/6J × C3H) oocytes that were then transplanted into foster mothers. We screened tail DNA from the 72 progeny using primers specific for the human AR CAG repeat (see below) and thereby identified four founder mice (5.6%). Transgene-positive mice were backcrossed to C57Bl/6J individuals to produce F1, N2 and N3 progeny. DNA was isolated from post-natal mouse tail samples by proteinase K digestion, phenol/chloroform extraction and ethanol precipitation (41).

Identification of transgene-positive mice and STS PCR mapping

To 250 ng mouse genomic DNA we added PCR reaction components to the following final concentrations in 50 µl reactions: dNTPs, 200 µM; 10X Stratagene PCR buffer #9 (1X; Stratagene); oligonucleotide primers, 500 nM; BSA, 0.1 mg/ml; DTT, 5 mM; formamide, 2.5% v/v; Amplitaq, 0.05 U/µl (Perkin-Elmer Cetus). For identification of transgene-positive mice we used primers specific for the human AR CAG repeat (1). Primers for PCR analysis of the left arm and the right arm of the YAC have been published (42). The following STS PCR primer sets from within and around the human AR gene were also used to map the extent of integrated YAC DNA: 611, 1967, 3356, 3357, 3358, 3360, 3361, 3362 and 3363. The oligonucleotide sequences for STS 611 and STS 1967 are available from the Washington University X chromosome web site (genome.wustl.edu/cgm/cgm.html). All of the other STS primer sequences were obtained from M. Schueler and H.F. Willard (Case Western Reserve University). PCR amplifications with all these primer sets were performed as described above, except that different annealing temperatures (52–54°C) were employed. All PCR reactions were run on a Perkin-Elmer Cetus DNA thermocycler. Aliquots of 15 µl each reaction product was electrophoresed in 4% agarose gels stained with ethidium bromide.
non-denaturing 1× TBE acrylamide minigels for 40 min at 35 mA in a Bio-Rad Mini-Protean apparatus.

**Southern blot hybridization and PFGE mapping of STS markers**

Total yeast DNAs were digested with restriction endonucleases according to the buffer and temperature conditions specified by the manufacturer. Agarose plugs (1 × 1 cm) containing total genomic yeast DNAs were equilibrated in 200 µl specified enzyme digestion buffer for 1 h, then the enzyme buffer was replaced and 20 U rare-cutter restriction endonuclease were added and incubated overnight at room temperature. The next day the enzyme buffer was replaced, 40 U rare-cutter restriction endonuclease were added and the agarose plugs were digested at the specified temperature for 4–6 h. PFGE analysis was performed in 1% agarose gels in 0.5× TBE using the CHEF DR III apparatus (Bio-Rad).

After acid hydrolysis, alkali denaturation and neutralization fractionated DNAs were transferred to nylon membranes (Hybond-N) in 20× SSC and incubated in prehybridization solution (5× SSC, 50 mM sodium phosphate, pH 7.0, 50% formamide, 0.5% SDS, 5× Denhardt’s solution, 100 µg/ml bovine thymus DNA) at 45°C for at least 2 h. The following DNA fragments were used as probes: a 144 bp NciI–EagI fragment from pBR322 (left arm YAC vector); the 347 bp exon 8 PCR product (3’ AR probe) (36); a 1550 bp Scal–Ndel fragment from pBR322 (right arm YAC vector); a 807 bp SplI–SstI fragment from pBR322 (right arm YAC vector); various STS PCR reaction products. Probes were labeled with [α-32P]dATP to a specific activity of ∼1 × 109 c.p.m./µg DNA using a random hexamer labeling kit (Boehringer Mannheim), denatured and added to the prehybridization solution for overnight incubation at 45°C. The nylon filters were first washed in 2× SSC, 0.5% SDS, 8 mM sodium phosphate, 2 mM sodium pyrophosphate at 60–65°C for 30 min and then washed in 0.5% SDS, 2.5 M EDTA, 5 mM Tris base, 2 mM sodium pyrophosphate at 50–55°C for 15 min. Blots were exposed overnight to X-omat autoradiography film.

For determination of copy number in our lines of transgenic mice we electrophoresed 8 µg EcoRI-digested DNA from transgenic and non-transgenic mice and included samples from human males and females as dosage standards. For copy number determination we used a 187 bp fragment from the exon 1–intron 1 boundary of the human AR gene as probe (36).

**PCR analysis of repeat instability**

For determination of CAG repeat lengths in transgene-positive mice we performed PCR as described above except that we included 0.5 µl [α-32P]dATP/reaction, substituted internal primers (3A, 5’-GTTCGAGACGGTCGGCGGAAA-3’, 4A, 5’-TACGATGGGGCTTGGG GAGA-3’) and used an annealing temperature of 52°C. Sizing of PCR products was performed by denaturing PAGE as previously described (1). For small pool PCR analysis we obtained 300 mg tissue samples from the liver, lung, testes or ovaries, kidney, brain and spinal cord of four 40-week-old mice (one set of males and one set of females). DNA was isolated by proteinase K digestion, phenol/chloroform extraction and ethanol precipitation and prepared as described (24). We performed PCR analysis of repeat instability using either 6 or 2 genome copies as template for the reaction.

**Western blot analysis**

Protein lysates were prepared in 2× loading buffer (200 mM DTT, 4% SDS, 100 mM Tris–HCl, pH 6.9, 20% glycerol and 0.04% bromophenol blue), resolved by 8% SDS–PAGE and transferred to nitrocellulose membranes at 35 V overnight in 20% methanol, 192 mM glycine, 25 mM Trizma base, pH 8.2, using the Bio-Rad Mini-Protean II apparatus. Immunoblotting was performed with the mouse monoclonal human-specific anti-AR antibody F3934.1 (Biogenex) at 1:5000 dilution in TBS containing 0.1% Tween 20, 3% non-fat dry milk or with the rabbit polyclonal non-species-specific anti-AR antibody PG-21 (Affinity Bioreagents) at 1:1000 dilution in PBS containing 0.1% Tween 20, 5% BSA. Primary antibodies were visualized with horseradish peroxidase-coupled goat anti-mouse or anti-rabbit antibodies (Amersham) at 1:10 000 dilution and enhanced chemiluminescence (Amersham).

**Statistical analysis**

All statistical tests and comparisons were performed using non-parametric applications from Statistica v.4.1 (Statsoft) for the Macintosh.

**ACKNOWLEDGEMENTS**

We wish to thank D.E. Merry, B.P. Brooks, D. Robins and A. Roy for supplying us with antibodies, protein lysates and tissue samples. We are grateful to T. Birkebak, M. Cunningham and J.A. Zebala for their assistance. A.R.L.S. is a Howard Hughes Medical Institute Physician Post-doctoral Fellow. This work was supported by grants from the Muscular Dystrophy Association (G.S.M. and A.R.L.S.) and the National Institutes of Health (K.H.F.).

**ABBREVIATIONS**

AR, androgen receptor; HD, Huntington’s disease; SBMA, spinal and bulbar muscular atrophy; SCA, spinocerebellar ataxia; STS, sequence-tagged site; YAC, yeast artificial chromosome.

**REFERENCES**
