Human Molecular Genetics, 2007, Vol. 16, No. 24 ***3174–3187***
doi:10.1093/hmg/ddm293
Advance Access published on October 6, 2007

An antisense transcript spanning the CGG repeat region of *FMR1* is upregulated in premutation carriers but silenced in full mutation individuals

Paula D. Ladd[1], Leslie E. Smith[1], Natalia A. Rabaia[1], James M. Moore[1], Sara A. Georges[1], R. Scott Hansen[3], Randi J. Hagerman[4,5], Flora Tassone[4,6], Stephen J. Tapscott[1,2,*] and Galina N. Filippova[1,*]

[1]Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA, [2]Department of Neurology, University of Washington Medical School, Seattle, WA 98195, USA, [3]Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA 98195, USA, [4]Medical Investigation of Neurodevelopmental Disorders Institute, School of Medicine, University of California–Davis, Sacramento, CA 95817, USA and [5]Department of Pediatrics, School of Medicine and [6]Department of Biochemistry and Molecular Medicine, School of Medicine, University of California–Davis, Sacramento, CA 95817, USA

Received August 5, 2007; Revised and Accepted October 1, 2007 GenBank accession nos EU48200–EU48204

Expansion of the polymorphic CGG repeats within the 5′-UTR of the *FMR1* gene is associated with variable transcriptional regulation of *FMR1*. Here we report a novel gene, *ASFMR1*, overlapping the CGG repeat region of *FMR1* and transcribed in the antisense orientation. The *ASFMR1* transcript is spliced, polyadenylated and exported to the cytoplasm. Similar to *FMR1*, *ASFMR1* is upregulated in individuals with premutation alleles and is not expressed from full mutation alleles. Moreover, it exhibits premutation-specific alternative splicing. Taken together, these observations suggest that in addition to *FMR1*, *ASFMR1* may contribute to the variable phenotypes associated with the CGG repeat expansion.

INTRODUCTION

The CGG repeat expansion in the 5′-UTR of the fragile X mental retardation gene (*FMR1*) has been implicated in the pathogenesis of two distinct disorders, fragile X syndrome (FXS), a neurodevelopmental disorder (1,2), and fragile X-associated tremor and ataxia syndrome (FXTAS), a progressive neurodegenerative disease (3,4). In the general population, the *FMR1* 5′-UTR contains five to 54 CGG repeats, whereas expansions of this trinucleotide repeat outside of the normal range fall into two distinct categories, the premutation range with 55–200 repeats and the full mutation range with greater than 200 and up to thousands of repeats (3,5).

Full mutation expansions coupled with cytosine methylation result in transcriptional silencing of the *FMR1* gene, loss of expression of FMR1 protein (FMRP) and FXS, the most common form of heritable X-linked mental retardation. The FXS phenotype is complex and highly variable, with mental impairment ranging from mild learning disabilities and emotional problems to severe mental retardation (1,2). The severity of mental retardation is correlated with the degree of cytosine methylation of the *FMR1* promoter and repeat region (6,7).

Premutation alleles are associated with FXTAS, one of the most common single-gene forms of gait ataxia and tremor in older males (3,4,8,9). Although the full mutation alleles are transcriptionally silent, premutation alleles demonstrate a 2–10-fold increase in *FMR1* mRNA levels (10) but normal or reduced amounts of FMRP (11–13). Since FXTAS is restricted to premutation carriers and not found in full

*To whom correspondence should be addressed at: Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Mail Stop C2-023, PO Box 19024, Seattle, WA 98109, USA. Tel: +1 2066674468; Fax: +1 2066676523; Email: gfilippo@fhcrc.org (G.N.F.) or Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Mail Stop C3-168, PO Box 19024, Seattle, WA 98109, USA. Tel: +1 2066674499; Fax: +1 2066676524; Email: stapscot@fhcrc.org (S.J.T.)

mutation individuals, increased *FMR1* transcript level rather than the reduced FMRP is thought to be the underlying cause of this disorder. Accordingly, postmortem examination of the brains of FXTAS individuals revealed intranuclear inclusions containing the *FMR1* transcript, ubiquitin, molecular chaperones and components of the proteasome (14,15). The current molecular model for FXTAS is that the transcript containing the expanded premutation size CGG repeat sequesters or results in misfolding of cellular proteins, such as CGG-binding proteins, leading to the formation of intranuclear inclusions (16,17). This model is consistent with the RNA gain-of-function model proposed for another trinucleotide-repeat associated disease, myotonic dystrophy (DM1), in which the CUG repeat in the 3′-UTR of *DMPK* sequesters RNA-binding proteins, such as the members of the muscleblind family (18,19).

Although deregulation or mutations of *FMR1* have been implicated in the pathogenesis of both FXS and FXTAS (1,3,20,21), it is important to note that there have been individuals with phenotypic manifestation of FXS with no mutations found in the *FMR1* gene (22,23). In addition, only one-third of male premutation carriers develop FXTAS (9). These observations together with the wide variability of FXS and FXTAS phenotypes suggest that in addition to *FMR1*, there are other genes potentially involved in pathogenesis of these disorders.

In this report, we identify a novel gene, *ASFMR1*, overlapping the CGG repeat region of the *FMR1* gene in the antisense orientation. Similar to *FMR1*, the *ASFMR1* transcript is elevated in lymphoblastoid cells and peripheral blood leukocytes of individuals with premutation alleles relative to normal and is not expressed from full mutation alleles, suggesting that the antisense transcription across the expanded CGG•CCG repeat may contribute to the pathogenesis of FXTAS and FXS. In addition, the *ASFMR1* transcript exhibits premutation-specific alternative splicing, providing a qualitative molecular abnormality associated with FXTAS.

RESULTS

Identification of an antisense transcript at the human *FMR1* locus

Identification of antisense transcripts at several disease-associated trinucleotide repeat loci, including *DM1* and *SCA8* (24,25), suggests that bidirectional expression might result in RNA-mediated transcriptional silencing of the expanded repeats and that the sense and/or antisense transcript might contribute to a clinical phenotype. To test whether similar mechanisms apply to the human *FMR1* locus, we analyzed the locus for the presence of antisense transcripts. Human *FMR1* is located in a gene poor region; however, the genome database indicates that a number of spliced and unspliced expressed sequence tags (ESTs) overlap the *FMR1* gene locus. Although many of the ESTs reflect portions of the *FMR1* transcript, there are a number of ESTs in the antisense direction (BX482783 and AA714549; Fig. 1A). Using strand-specific RT-PCR (Fig. 1B) on human lymphoblastoid cells derived from individuals with characterized CGG repeat expansions, we mapped an antisense transcript at the position −500 to −1000 bp relative to the *FMR1* major transcription start site (26) and at the position +337 to +210 bp overlapping *FMR1* at exon 1 downstream of the CGG repeat (Fig. 1C, panels 1 and 5). Both of these regions of antisense transcription were identified in the cell lines with normal or premutation alleles, H930-1 and FX-TF, respectively, but not in a full mutation cell line, FX-GM (Fig. 1C, panels 1 and 5), similar to the expression pattern of the *FMR1* transcript in these cell lines (Fig. 1C, panel 7). No sense transcription was detected upstream of the known *FMR1* transcription start sites (26,27) (Fig. 1C, panel 3).

The identification of antisense transcripts on both sides of the CGG repeat suggested that an antisense transcript overlaps the *FMR1* CGG repeat region. Northern blot analysis to identify the full-length antisense transcript was unproductive, likely due to the low abundance of the transcript. However, using multiple strand-specific RT–PCR, including a betaine-based protocol for the amplification of CGG•CCG repeats (28) (Fig. 2A), we identified overlapping regions of antisense transcription (Fig. 2B–F), indicating a continuous transcript spanning the *FMR1* CGG repeat region in the antisense orientation. We named this transcript the AntiSense transcript at the *FMR1* locus, *ASFMR1*.

Identification of multiple splice forms

Mapping of *ASFMR1* revealed multiple alternative splice forms of the transcript, as depicted in Figure 2A. Using primers from −1000 to −196, we identified the unprocessed form and three different spliced forms of the *ASFMR1* transcript (Fig. 2C and Supplementary Material, Fig. S1). Removal of 381 nt from −538 to −921 results in transcript C (Fig. 2D). An intermediate form B splices a total of 542 nt of sequence from −274 to −530 and from −635 to −921, resulting in an additional short exon ∼100 nt in size (Fig. 2C, transcript B). The largest splice removes 647 nt of the sequence from −274 to −921 (Fig. 2C, transcript A). The spliced transcripts A–C follow the consensus splice donor/acceptor recognition sequence, gt to ag (Supplementary Material, Fig. S1). Both spliced and unspliced forms of the transcript were detected in normal and premutation cell lines, H930-1 and FX-TF, respectively, but not in the full mutation line, FX-GM. Although the stoichiometry of the bands suggests that the major splice form for both cell lines is transcript A, there are noticeably different ratios of the intermediate transcripts, B and C, to the unspliced transcript in the premutation cell line, FX-TF, relative to the normal cell line, H930-1 (Fig. 2C). In addition, in premutation cell line, FX-TF, primer set +10243 and +210 revealed another alternative splice form of the *ASFMR1* transcript with a small intron from +10155 to +10070 that uses a non-consensus CT-AC splice site (Fig. 2F and Supplementary Material, Fig. S1).

The *ASFMR1* transcript is polyadenylated

To test whether the *ASFMR1* transcript is polyadenylated, we used oligo-dT beads (Invitrogen) to pull-down poly(A) RNA from the normal cell line, H930-1. Both oligo-dT-bound and unbound fractions were analyzed for the presence of the antisense transcript at −500 to −1000 and +337 to +210, as

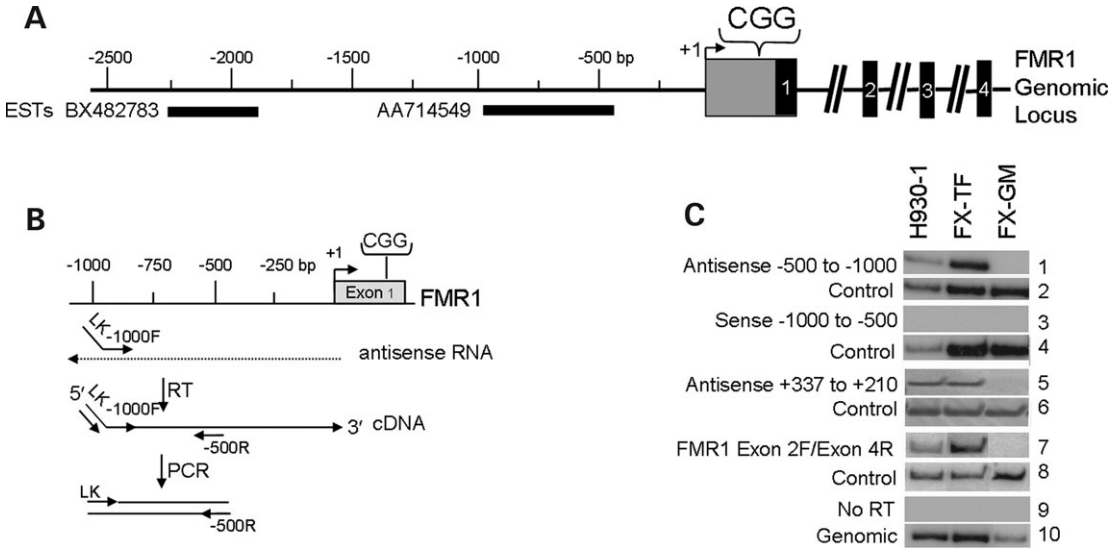

Figure 1. Identification of an antisense transcript at the *FMR1* locus. (**A**) Diagram of the FMR1 locus. Exons 1–4 of *FMR1* are indicated by rectangles, shaded black for coding regions and gray for the 5′-UTR, with the CGG repeat region noted. Direction of transcription is indicated with an arrow. Thick horizontal bars and corresponding accession numbers indicate the location of reported ESTs. Nucleotide numbering is relative to the *FMR1* major transcription initiation site according to the published sequence (26). (**B**) Diagram of the strand-specific RT–PCR method. To identify cDNA generated from a particular strand, a linker sequence is attached to a gene-specific primer. The linker sequence is used as a primer for subsequent PCR amplification. (**C**) Strand-specific RT–PCR analysis of human lymphoblastoid cell lines identified an antisense transcript at the position −500 to −1000 bp upstream of *FMR1* and at the position +337 to +210 bp overlapping exon 1 of *FMR1* in normal, H930-1, and premutation, FX-TF, cells, but not in full mutation, FX-GM (panels 1 and 5). No sense transcript was detected upstream of *FMR1* (panel 3). Panels 9 and 10 represent no-RT and genomic DNA controls, respectively. The quality of cDNA synthesis was verified by amplification of the control transcript, CTCF, panels 2, 4, 6 and 8.

well as two control transcripts, *FMR1*, a known polyadenylated transcript, and non-polyadenylated 18S rRNA. Similar to *FMR1*, the *ASFMR1* transcript was enriched in the poly(A) fraction relative to the unbound, whereas 18S was enriched in the unbound fraction (Fig. 2G).

To verify that the *ASFMR1* transcript is polyadenylated, we performed 3′-RACE using an oligo-dT primer with an attached linker sequence. We identified several 3′ ends at −1037, −1807 and −1846 bp (Fig. 3B, 3′-RACE). Although the 3′ end at −1037 bp is consistent with the sequence for EST AA714549, examination of the genomic sequence indicated a poly(A) stretch in the antisense orientation at −1040 bp, suggesting that this 3′ end and EST AA714549 are artifacts of oligo-dT cDNA generation. The −1807 and −1846 bp 3′ ends did not correspond to A-rich genomic sequence, indicating that these RACE products represent real 3′ ends of the transcript (Fig. 3B, 3′-RACE). In addition to these two polyadenylation regions, strand-specific RT–PCR indicated that at least some transcripts extend farther to −2490 as depicted in Figure 3A. Although a consensus polyadenylation site is present at −2463 (Supplementary Material, Fig. S1), this site has not been validated by 3′-RACE.

Identification of the 5′ ends of *ASFMR1*

To identify the 5′ end of the *ASFMR1* transcript, we performed 5′-RLM-RACE using two different approaches. We took advantage of an important feature of 5′-RLM-RACE that allows for the ligation of an RNA oligo to full-length capped transcripts, but not truncated or non-mRNA transcripts, and, in a separate analysis, we modified the 5′-RLM-RACE technique to allow detection of predominately uncapped transcripts. Accordingly, we identified three distinct regions of 5′ ends at +10243 and +305 bp and clustered at −99 to −208 bp, relative to FMR1 (Fig. 3B, 5′-RLM-RACE). The +10243 and −99 to −208 5′ ends were detected by RLM-RACE enriched for capped transcripts, whereas the +305 bp 5′ end was preferentially identified by the 'uncapped' RACE, suggesting that it might be secondary to RNA cleavage or processing. This is consistent with the identification of an ∼100–150 nt sized fragment using northern blot analysis for short processed RNA fragments with a probe from +205 to +295 corresponding to the antisense sequence (data not shown). It is also important to note that the +10243 5′ end of the *ASFMR1* transcript was identified in both normal and premutation lymphoblastoid cells, H930-1 and FX-TF, respectively, whereas the −99 to −208 cluster of 5′ ends was detected only in normal cells, H930-1, suggesting that similar to *FMR1* (27) repeat expansion may be accompanied by the shift in the transcription start site usage by *ASFMR1*. Both regions of 5′ ends were identified in normal human brain tissue. Together, these data suggest the presence of two alternative promoters driving transcription of *ASFMR1* as depicted in Figure 3A.

The −99 to −208 putative promoter has multiple transcription initiation sites identified by 5′-RLM-RACE, which is consistent with GC-rich, TATA-less promoters (Fig. 3C). This promoter overlaps the *FMR1* promoter and is consistent with the transcription start site reported for the mouse antisense transcript, *AK148387*. The +10243 putative promoter appears to have multiple basal promoter elements, including an upstream weak TATA box (29) and overlapping canonical initiator elements with downstream promoter elements

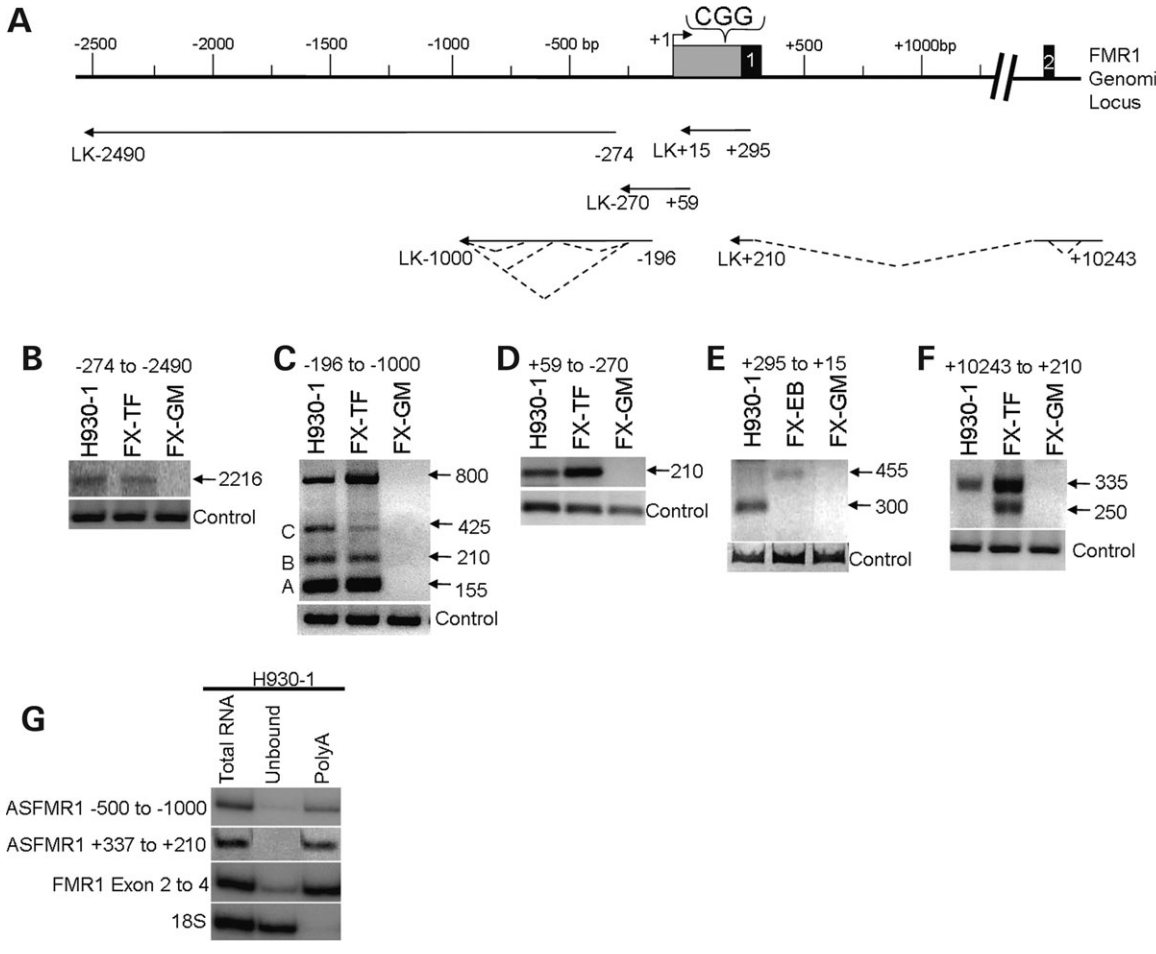

Figure 2. A continuous antisense transcript at the *FMR1* locus spans CGG repeat in the antisense orientation. (**A**) The *FMR1* genomic locus is depicted as in Figure 1A. Multiple overlapping strand-specific RT–PCR regions, spanning the AntiSense transcript at the *FMR1* locus, *ASFMR1* from −2490 to +10243, are depicted by arrows. Numeric values indicate the location of each linker containing primer used for strand-specific cDNA generation, as well as the reverse primer used for PCR amplification, relative to *FMR1*. Primer sequences are listed in Supplementary Material, Table S1. (**B**–**D**) Strand-specific RT–PCR analysis of human lymphoblastoid cell lines identified an antisense transcript in normal, H930-1, and premutation, FX-TF, cells, but not in full mutation, FX-GM, at the positions (B) −274 to −2490 bp, (C) −196 to −1000, revealing multiple splice forms A, B and C and the unspliced transcript, and (D) +59 to −270. (**E**) Betaine-based strand-specific RT–PCR analysis to amplify CGG-rich sequences identified an antisense transcript spanning the CCG repeat region from +295 to +15 in normal, H930-1, and premutation, FX-EB, cells containing 82 CGG repeats, but not in full mutation, FX-GM. (**F**) Strand-specific RT–PCR analysis of human lymphoblastoid cell lines identified an antisense transcript at the position +10243 to +210 bp in normal H930-1, and premutation, FX-TF, cells, but a splice form in premutation, FX-TF, only. The synthesis of cDNA was verified by amplification of the control transcript, *CTCF*. The −196 to −1000, +59 to −270 and +10243 to +210 strand-specific RT–PCR products and the control transcript were amplified in a linear range. The *ASFMR1* expression was normalized to the control transcript. The semi-quantitative strand-specific RT–PCR was repeated for each pair of primers at a minimum of two to three times. (**G**) Strand-specific RT–PCR analysis of the oligo-dT bound and unbound fractions of total RNA demonstrated that, similar to *FMR1*, the antisense transcript at −500 to −1000 and +337 to +210 is enriched in poly(A) fraction. 18S transcript was used as a non-polyadenylated control.

(Fig. 3D). The transcript initiated at +10243 spans the *FMR1* CGG repeat in the CCG orientation and contains a 9.7 kb intron corresponding to the *FMR1* intron (Fig. 3A and B) that uses the complementary splice donor and acceptor to *FMR1*, representing a non-consensus CT to AC splice site (Supplementary Material, Fig. S1). Using strand-specific RT–PCR, we confirmed that the *ASFMR1* transcript initiated at +10243 is expressed in normal and premutation cell lines, but not full mutation (Fig. 2F).

CTCF-binding sites flank the CGG repeat

To map the location of CTCF-binding sites (30,31), which are associated with other triplet repeat loci (24,32), we used a contiguous set of 15 overlapping DNA probes for electrophoretic mobility shift assay (EMSA) (Fig. 4A) with *in vitro* translated 11ZF, a truncated version of the CTCF protein containing the zinc finger DNA-binding domain. Four CTCF-binding sites were identified by EMSA (Fig. 4B and C) and confirmed by chromatin immunoprecipitation (ChIP) (Fig. 4D). The identified CTCF-binding sites flanking the CGG repeat region and the +10243 promoter region are consistent with the peaks of CTCF-binding identified in a recent whole genome analysis (depicted in Fig. 4A) (33). Together, these results suggest that CTCF might have a role in regulating *FMR1* and *ASFMR1* transcription or regional chromatin loop formation (31).

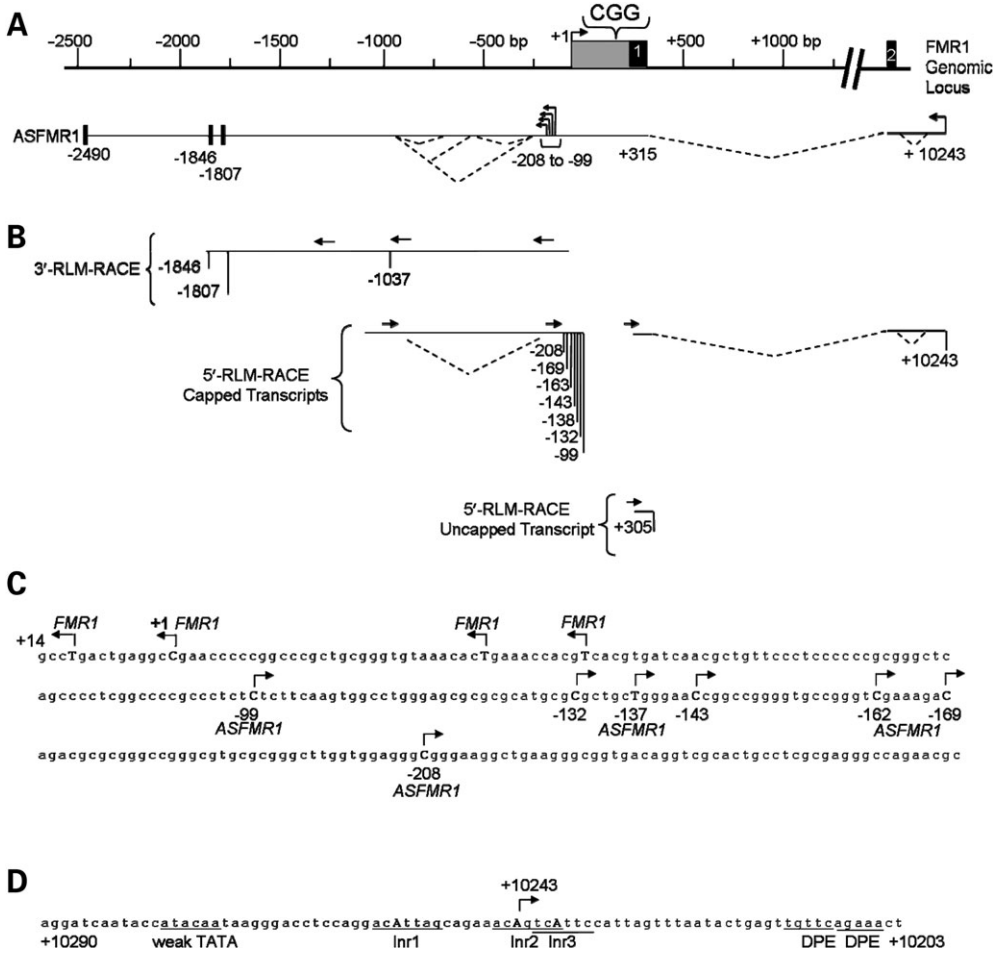

Figure 3. General organization of the *ASFMR1* transcript. (**A**) The *FMR1* genomic locus is depicted as in Figure 1A. The *ASFMR1* transcript summary is indicated below the *FMR1* genomic locus, arrows indicate the transcription initiation sites identified by 5′-RACE and location of putative promoters. Characterized splice sites are indicated by dotted lines. Thick black vertical bars depict polyadenylated 3′ ends of the *ASMFR1* transcript. (**B**) Identification of the 3′ and 5′ ends of the *ASFMR1* transcript by RLM-RACE using total and poly(A) RNA from human brain and normal and premutation human lymphoblastoid cell lines, H930-1 and FX-TF, respectively. Location of primers used for RACE is indicated by arrows. Identified 3′ and 5′ ends of the *ASFMR1* transcript are indicated numerically. 5′-RACE results for capped and uncapped transcripts are depicted separately. (**C**) The putative bi-directional promoter region for the *ASFMR1* transcript initiated at −99 to −208, as indicated numerically and with arrows. Previously identified sites for *FMR1* transcription initiation, including the major site indicated with a +1, are shown by arrows. (**D**) Putative promoter region for the transcript initiated at +10243. Core promoter elements are indicated, a previously characterized weak TATA box, Initiator (Inr) elements and downstream promoter elements are underlined.

ASFMR1 is subject to X chromosome inactivation

Analysis of the previously characterized human:hamster somatic cell X hybrids (34) revealed that *ASFMR1* is exclusively transcribed from the active X chromosome and is subject to X-inactivation (Fig. 5A). Our observation that the *ASFMR1* transcript is not expressed by cells containing a full mutation CGG repeat expansion, suggested that cytosine methylation is an important factor in the transcriptional silencing of *ASFMR1*. To determine whether *ASFMR1* expression can be reactivated by treatment with 5azaC, an inhibitor to cytosine methyltransferases, we analyzed a hybrid cell line, which was derived by 5azaC treatment of the inactive X chromosome hybrid and then clonally selected for *HPRT* expression (34). Consistent with published results, this cell line continued to express *XIST* and demonstrated reactivation of *HPRT* and *FMR1* (Fig. 5A). Furthermore, *ASFMR1* expression was also reactivated, supporting the hypothesis that cytosine methylation plays a similar role in the transcriptional regulation of *FMR1* and *ASFMR1*.

ASFMR1 is widely expressed in human tissues with relatively high expression in brain

To determine the pattern of *ASFMR1* expression relative to *FMR1* expression in human tissues, we performed semi-quantitative multiplex RT–PCR using a commercially derived RNA survey panel of twenty different human tissues (Ambion) (Fig. 5B). Under the same PCR conditions, we observed that *FMR1* expression levels vary from relative high abundance in tissues such as brain, kidney, ovary, testes and thyroid to barely detectable levels in skeletal muscle (Fig. 5B, lanes 3, 8, 11, 14, 17 and 19) consistent with the previous

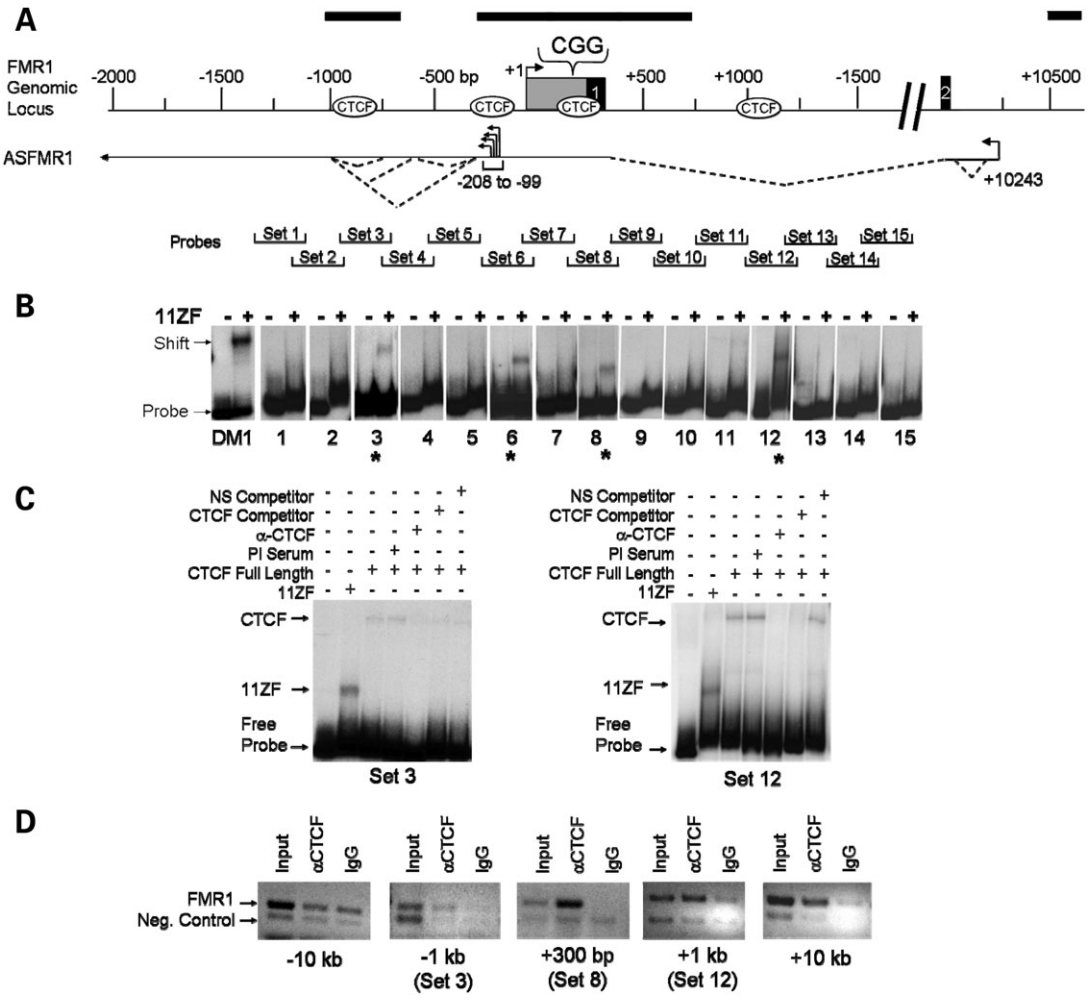

Figure 4. CTCF-binding sites flank CGG repeat at the *FMR1* locus. (**A**) Diagram of 15 overlapping DNA probes used in EMSA relative to the *FMR1* genomic locus. CTCF-binding sites identified by EMSA are depicted as ovals, whereas the CTCF peaks identified by ChIP-Seq (33) are indicated as black bars above the genomic locus. (**B**) EMSA with *in vitro* translated CTCF-DNA-binding domain (11ZF). Each probe, including DM1 control, is shown without (−) and with (+) 11ZF. Horizontal arrows indicate free probe and mobility shift for DM1 control. Asterisks denote probes, which shifted when incubated with 11ZF protein. (**C**) EMSA using probe sets 3 and 12 with *in vitro* translated full-length CTCF. Horizontal arrows indicate free probe, mobility shift for 11ZF control, as well as mobility shift for the full-length CTCF. Incubation with CTCF-specific and non-specific antibodies or competitors is indicated by +/− above each lane. (**D**) ChIP assay confirmed that CTCF binds to the human *FMR1* locus *in vivo*. ChIP was performed with anti-CTCF antibody (Upstate Biotech) on the human lymphoblastoid cell line, H930-1, and subjected to semi-quantitative multiplex PCR analysis using primer pairs for *FMR1* (upper band) and *KIAA522* (lower band), as a negative control previously described (53). *FMR1* locus was amplified at the position −10134 to −9895 (−10 kb), −1000 to −738 (−1000), +210 to −426 (+300), +950 to +1232 (+1000) and +10924 to +11184 (+11 kb) relative to the *FMR1* transcriptional start site. All *FMR1* amplicons, except the −10 kb region, showed enrichment for CTCF binding, indicating that CTCF-binding sites flank the CGG repeat and the +10243 promoter of *ASFMR1*.

report (35). When we compare *ASFMR1* expression with *FMR1*, we see that *ASFMR1* is expressed in all of the tissues examined, with relatively high expression in brain and kidney (Fig. 5B, lanes 3 and 8) and barely detectable in heart, placenta, prostate, skeletal muscle, thyroid and trachea (Fig. 5B, lanes 7, 12–14, 19 and 20). Real-time PCR analysis confirmed that the expression of *ASFMR1* normalized to *FMR1* is consistently higher in brain than in other tissues: 5.6 ± 1.5 times higher than in colon ($P = 1.6\text{E}-07$), 6.4 ± 1.1 times higher than in kidney ($P = 1.0\text{E}-08$) and 46.5 ± 12.5 times higher than in heart ($P = 6.1\text{E}-10$). This expression profile suggests a potential role for *ASFMR1* in the neurological phenotype of both FXTAS and FXS.

Effect of CGG repeat expansion on *ASFMR1* expression and alternative splicing

It has been well documented that *FMR1* expression increases with increasing CGG repeat length in the premutation range (11,12). To determine the effect of CGG repeat expansion on the expression levels of the *ASFMR1* gene, we performed semi-quantitative multiplex RT–PCR (Fig. 6A) and real-time RT–PCR analyses (Fig. 6B) of the human lymphoblastoid cell lines derived from individuals with characterized CGG repeat expansions (Table 1). Similar to *FMR1*, *ASFMR1* expression increases as the CGG repeats expand within the premutation range ($P < 1.0\text{E}-05$). Whereas *FMR1* and *ASFMR1*

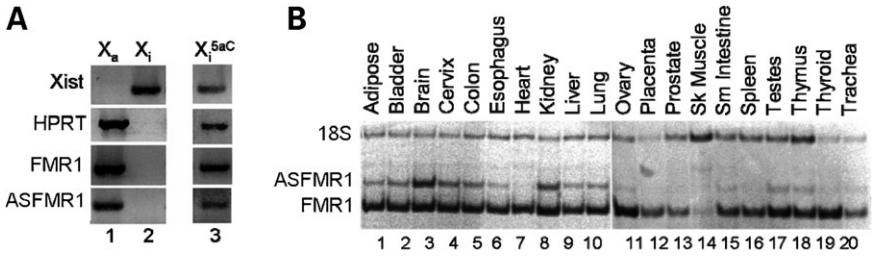

Figure 5. Expression profile of the *ASFMR1* transcript. (**A**) RT–PCR of human:hamster hybrid cells using primers to human *XIST*, *HPRT*, *FMR1* and *ASFMR1*. X_a and X_i contain the active and inactive X chromosome, respectively. X_i^{5aC} indicates the X_i cell line after 5aC reactivation treatment. (**B**) Semi-quantitative multiplex RT–PCR analysis of human tissues using primers to *ASFMR1*, *FMR1* and *18S*, as described in Materials and Methods. Each tissue type is indicated and reflects a pool of RNA derived from three different individuals (Ambion).

expression levels are reduced in the partially methylated cell lines ($P < 0.003$), neither *FMR1* nor *ASFMR1* expression is seen in the full mutation cell lines ($P < 1.2E-10$) (Fig. 6B). Analysis of RNA isolated from the peripheral blood leukocytes of six individuals with characterized CGG repeats (Table 1) showed that *ASFMR1* expression in primary tissue is consistent with the expression profile observed in lymphoblastoid cell lines. The premutation sample 97-06, with 122 CGG repeats, demonstrated significantly increased transcript levels of both *FMR1* and *ASFMR1* genes compared with the normal samples (Table 1) (Fig. 6C), whereas very little to no expression of *FMR1* and *ASFMR1* was found in individuals with hypermethylated full mutation alleles, 139-06 and 46-06. Interestingly, *FMR1* expression level increases slightly, but not *ASFMR1*, for premutation sample 89-06, carrying an allele of 102 CGG repeats (Fig. 6C).

It is important to note that the PCR primers used to quantify the *ASFMR1* expression levels by both semi-quantitative and real-time RT–PCR detected the total of the *ASFMR1* transcripts initiated at both putative promoters: the −99 to −208 promoter and the +10243 promoter (Fig. 6A–C). Analysis of multiple regions of the *ASFMR1* transcript by semi-quantitative strand-specific RT–PCR (Fig. 2C, D and F) demonstrated that, similar to the total levels of the *ASFMR1* transcript, the transcript initiated at the +10243 promoter is also upregulated in premutation cells in comparison to normal. Together with the 5′-RACE results that identified the +10243 site as a major transcriptional start site in the premutation cell line, FX-TF, this suggests that the increased *ASFMR1* expression in premutation cells results in mutant transcripts containing expanded CCG repeats.

The *ASFMR1* transcript with the splice site from +10155 to +10070 was detected in cells with the premutation and partially methylated full mutation alleles, but not in cells with normal or hypermethylated full mutation alleles (Fig. 6D and E). Notably, the premutation sample 89-06 with 102 repeats and lower *ASFMR1* expression demonstrated lower levels of expression of this splice form relative to the unspliced transcript (Fig. 6E). Moreover, the spliced transcript was not present in the premutation cells FX-EB carrying the shortest premutation sized allele of 82 CGG repeats (Fig. 6D), suggesting that CGG repeat expansion may influence *ASFMR1* transcript processing and/or its stability. Analysis of a larger collection of RNA samples isolated from the peripheral blood leukocytes of 11 individuals with normal alleles and 12 individuals with premutation alleles confirmed that this splice form is specific to the premutation allele (Fig. 6F).

The *ASFMR1* transcript is transported to the cytoplasm and contains a polyproline ORF

To determine whether the *ASFMR1* transcript is retained in the nucleus or transported to the cytoplasm, we analyzed the cytosolic and nuclear fractions of normal H930-1 and premutation FX-TF cell lines for the presence of unprocessed and spliced *ASFMR1* transcripts as depicted in Figure 7A. Similar to *FMR1* used as a control transcript, the unspliced *ASFMR1* transcript was enriched in the nucleus, whereas the spliced transcript was evenly distributed between the cytoplasm and nucleus of H930-1 and FX-TF cells (Fig. 7B). This indicates that after processing in the nucleus, the mature *ASFMR1* transcript is exported to the cytoplasm. In addition, no differences were detected in the cellular localization of either the *ASFMR1* or *FMR1* transcripts between cells with normal or premutation sized CGG repeat regions.

Analysis of both the unspliced and spliced *ASFMR1* transcripts for the presence of putative open-reading frames (ORFs) identified an ORF encoding a polyproline peptide. This ORF is found in the transcript initiated from +10243. The stretch of proline is a consequence of the repeat, which is CCG with interspersed CCU for *ASFMR1* (Fig. 7C). Both CCG and CCU code for proline, thus the ORF encodes for a polyproline peptide whether or not AGG•CCU interspersion is lost during CGG•CCG repeat expansion. Accordingly, CGG expansion would result in an even longer polyproline stretch. The entire ORF for the putative protein is contained in the region of the *ASFMR1* transcript that overlaps *FMR1*. Interestingly, there are several potential translational start sites located upstream of this ORF, which could interfere with translation of the polyproline peptide. In this context, it is important to note that the small splice site from +10155 to +10070 identified in the premutation cells (Figs 2F and 6D–F) would remove one of the potentially interfering upstream AUGs associated with the stronger Kozak sequence and therefore likely enhance translation of the polyproline peptide.

ASFMR1 is present in mouse

Strand-specific RT–PCR analysis identified two regions of the antisense transcription at the murine *Fmr1* locus: at the

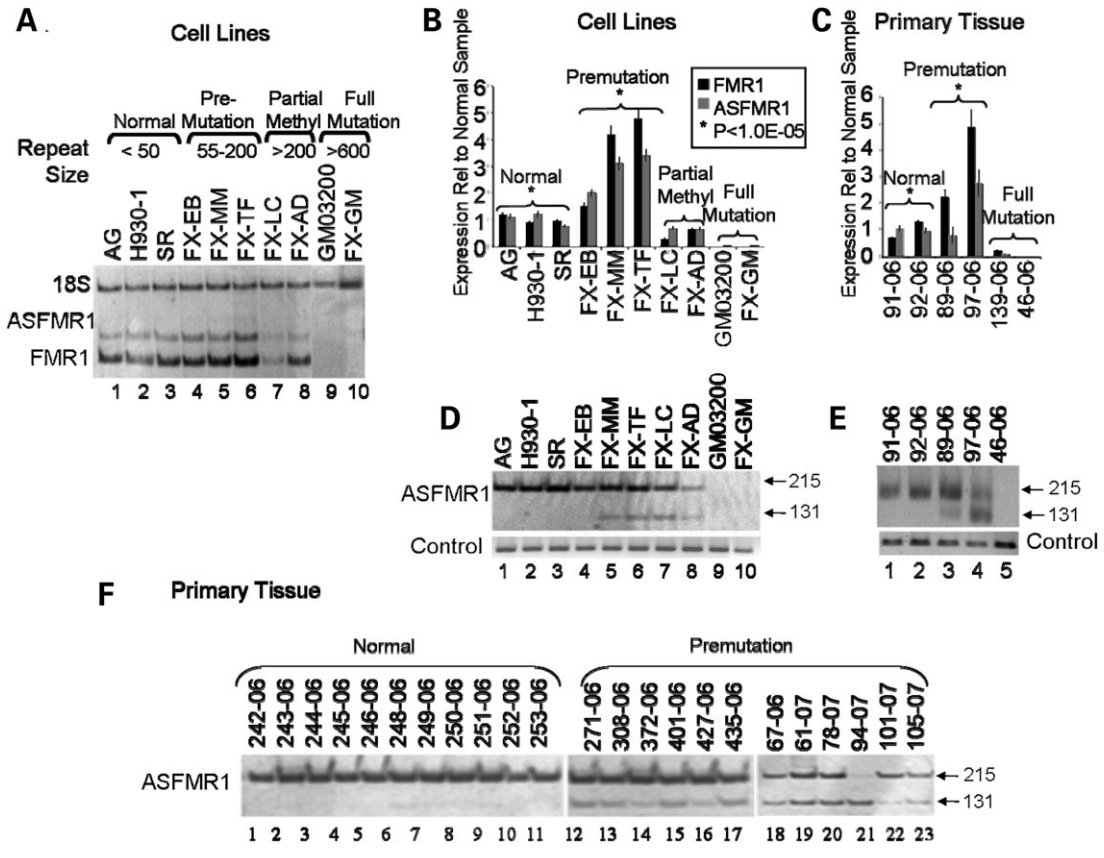

Figure 6. Effect of CGG repeat expansion on *ASFMR1* expression and alternative splicing. (**A**) Semi-quantitative multiplex RT–PCR analysis of human lymphoblastoid cell lines with characterized CGG repeats. (**B**) and (**C**) Real-time RT–PCR analysis of the lymphoblastoid cell lines and the peripheral blood leukocytes from individuals with characterized CGG repeats indicated as cell lines and primary tissue, respectively, showed that, similar to *FMR1*, *ASFMR1* is upregulated in individuals with premutation alleles relative to normal and is not expressed from full mutation alleles. Semi-quantitative RT–PCR and real-time RT–PCR analyses of *ASFMR1* expression were performed with the primers sets to −1427/−1168 and to −1427/−1185, respectively, that allow identification of the spliced and unspliced forms of the *ASFMR1* transcripts initiated at both putative promoters. Black and gray bars indicate *FMR1* and *ASFMR1* expression, respectively. Expression levels for *FMR1* and *ASFMR1* relative to the average of the normal samples and the corresponding *P*-values are summarized in Table 1. (**D–F**) RT–PCR analysis of the human lymphoblastoid cell lines and the peripheral blood leukocytes for the presence of the alternative splice form of the *ASFMR1* transcript using primers to +10028 and +10243 relative to *FMR1*. For primary tissues depicted in (F), normal alleles range from 23 to 54 CGGs, and premutation alleles range from 64 to 124 CGGs.

position −156 to −379 bp, relative to the *Fmr1* transcriptional start site, which is consistent with the previously reported murine antisense transcript *AK148387*, and at the position +204 to +312 bp, overlapping first coding exon of *Fmr1* (Supplementary Material, Fig. S2). Together, these observations suggest that similar to the human *FMR1* locus, there is an antisense transcript *Asfmr1* spanning the CGG-rich region of murine *Fmr1*. Moreover, similar to *ASFMR1*, *Asfmr1* contains a potential proline-rich ORF, conserved between human and mouse.

DISCUSSION

In this study, we have identified a novel gene, *ASFMR1*, transcribed in the antisense orientation to *FMR1*. Similar to *FMR1*, the *ASFMR1* transcript is silenced in full-mutation individuals and upregulated in carriers of premutation alleles, suggesting that, in addition to *FMR1*, *ASFMR1* might be involved in the pathogenesis of FXTAS and FXS. Consistent with the neurological phenotypes of both FXTAS and FXS, *ASFMR*1 exhibits relatively high expression levels in human brain. The *ASFMR1* transcript is alternatively spliced, polyadenylated and exported to the cytoplasm. *ASFMR1* transcription appears to be driven by two alternative promoters: the *FMR1* bidirectional promoter and the promoter located in the second intron of *FMR1*. Both promoters are flanked by CTCF-binding sites. Notably, the latter promoter has been identified as a major promoter in premutation cells. The transcript initiated at this promoter spans the CGG repeat of the *FMR1* gene in the CCG orientation, exhibits premutation-specific alternative splicing and contains an ORF with the CCG repeat encoding a polyproline peptide. Interestingly, we have identified an antisense transcript in the mouse that overlaps the murine *FMR1* gene and has a potential proline-rich ORF, suggesting a conserved cellular function for *ASFMR1*. Taken together, these findings strongly suggest that *ASFMR1* expression from the expanded allele might contribute to the variable clinical phenotypes associated with the CGG repeat expansion by either an RNA and/or protein-mediated mechanism.

Table 1. Patient sample summary*

	Gender	Repeat size	Category	FMR1 expression Mean ± SE	*n*	FMR1 *P*-value	ASFMR1 expression Mean ± SE	*n*
Cell line								
AG	Female	16/29	Normal	1.16 ± 0.06	6	na	1.07 ± 0.11	6
H930-1	Male	nd	Normal	0.88 ± 0.08	6	na	1.17 ± 0.11	6
SR	Male	nd	Normal	0.95 ± 0.07	6	na	0.75 ± 0.06	6
FX-EB	Male	82	Premutation	1.51 ± 0.14	6	1.0E−02	2.03 ± 0.14	6
FX-MM	Male	170	Premutation	4.20 ± 0.34	6	1.5E−04	3.11 ± 0.26	6
FX-TF	Male	195	Premutation	4.80 ± 0.35	6	7.0E−05	3.41 ± 0.22	6
FX-LC	Male	250	Partial	0.28 ± 0.06	6	3.0E−10	0.68 ± 0.06	6
FX-AD	Male	293/950	Partial	0.64 ± 0.03	6	7.8E−06	0.65 ± 0.08	6
GM03200	Male	>600	Full mutation	0	6	2.3E−13	0.02 ± 0.00	6
FX-GM	Male	610	Full mutation	0	6	2.3E−13	0.03 ± 0.01	6
Primary cell								
91-06	Male	30	Normal	0.71 ± 0.04	12	na	1.04 ± 0.14	5
92-06	Male	30	Normal	1.32 ± 0.08	11	na	0.98 ± 0.11	11
89-06	Male	102	Premutation	2.25 ± 0.31	12	1.4E−03	0.81 ± 0.33	6
97-06	Male	122	Premutation	4.87 ± 0.69	12	1.1E−04	2.77 ± 0.53	6
139-06	Male	331/572/806	Full mutation	0.25 ± 0.02	6	8.5E−10	0.12 ± 0.01	6
46-06	Male	271/662	Full mutation	0	12	6.2E−08	0.02 ± 0.01	6

nd, not determined; na, not applicable.
*Expression levels for FMR1 and ASFMR1 relative to the average of the normal samples and the corresponding *P*-values were determined by real-time RT–PCR. Two repeat sizes are denoted for two alleles for normal female and multiple repeat sizes are indicated for mosaic male samples.

Elevated levels of *FMR1* mRNA containing the expanded CGG repeats have been implicated in the formation of intranuclear inclusions through the sequestering and/or misfolding of cellular proteins, including RNA-binding proteins, in brains of the premutation individuals with FXTAS (36,37). Since elevated *ASFMR1* expression from the premutation alleles also results in accumulation of the mutant transcript containing an expanded CCG repeat, it seems reasonable to propose a similar role for the *ASFMR1* transcript in the inclusion formation. Previous studies have not identified an antisense transcript in the inclusions (15); however, the probe used for *in situ* hybridization overlapped the region of *ASFMR1* that we find difficult to amplify, thus it will be important to examine the inclusions for other regions of *ASFMR1* transcript. In this context, it is also important to mention that the analysis of the protein composition of the nuclear inclusions in FXTAS patients revealed the presence of the muscleblind-like protein 1, MBNL1 (14), which was proposed to mediate the toxic RNA gain of function in DM (38). Interestingly, the study of RNA-binding properties of MBNL1 demonstrated that, in addition to CUG and CCUG, it specifically interacts with CCG but not CGG repeats (39), supporting a possible role for the antisense transcription across the expanded CGG•CCG repeat in the inclusion formation in FXTAS. Indeed, a recent report demonstrated that expression of the CCG premutation length repeat is pathogenic and leads to an RNA-mediated neurodegenerative phenotype in *Drosophila* (40).

If the *ASFMR1* transcript is translated to protein, the expansion of the CCG repeat would increase the size of the polyproline region, similar to the increased size of the polyglutamine stretch in CAG repeat diseases, such as Huntington's disease and many of the spino-cerebellar ataxias (41). In this regard, the nuclear inclusions seen in *FMR1* premutation individuals who develop FXTAS may be similar to the nuclear inclusions of patients with polyglutamine-induced diseases (16,41). The recent discovery of the polyglutamine-containing nuclear inclusions as a result of bidirectional expression of the CTG•CAG repeats in SCA8 (25), and possibly in HDL2 (42), strengthens the idea that both protein and RNA-mediated mechanisms may be involved in the pathogenesis of the trinucleotide repeat disorders. It will be interesting to see if the *ASFMR1* encoded polyproline peptide can be identified in the nuclear inclusions of premutation individuals with FXTAS. Furthermore, if the expression of the polyproline expansion is involved in the inclusion formation in FXTAS individuals, then the alternative splice form of the *ASFMR1* transcript identified in premutation individuals (Figs 2F and 6D–F) might affect translation efficiency of the polyproline peptide and serve as a predictive marker for FXTAS development. It is worth noting that the splice sites in the 5'-end of the *ASFMR1* are non-consensus CT-AC sites, which have been reported for other transcripts (43–45). Although these are non-consensus sites, they were present in both strand-specific RT-PCR and 5'-RLM RACE products.

Epigenetic profiling of repetitive elements in the mammalian genome revealed a strong correlation between tandem repeats, double-stranded RNA (dsRNA), histone H3 lysine 9 methylation and DNA methylation, suggesting that bidirectional transcription across tandem repeats may play a role of a primary trigger for stable repeat-associated repressive chromatin imprints (46). A model for heterochromatin formation at repetitive elements involves the processing of bidirectional RNA transcripts into small RNAs, which then recruit histone methyltransferases, HP1 and DNA methyltransferases to the region resulting in heterochromatin formation and spreading (46–48). Although the molecular events that induce methylation of the expanded FMR1 allele are not known, our demonstration of bidirectional transcription suggests that dsRNA might be generated if *FMR1* and *ASFMR1* are transcribed simultaneously from the same allele. This suggests that transcriptional silencing of the full mutation *FMR1* locus might

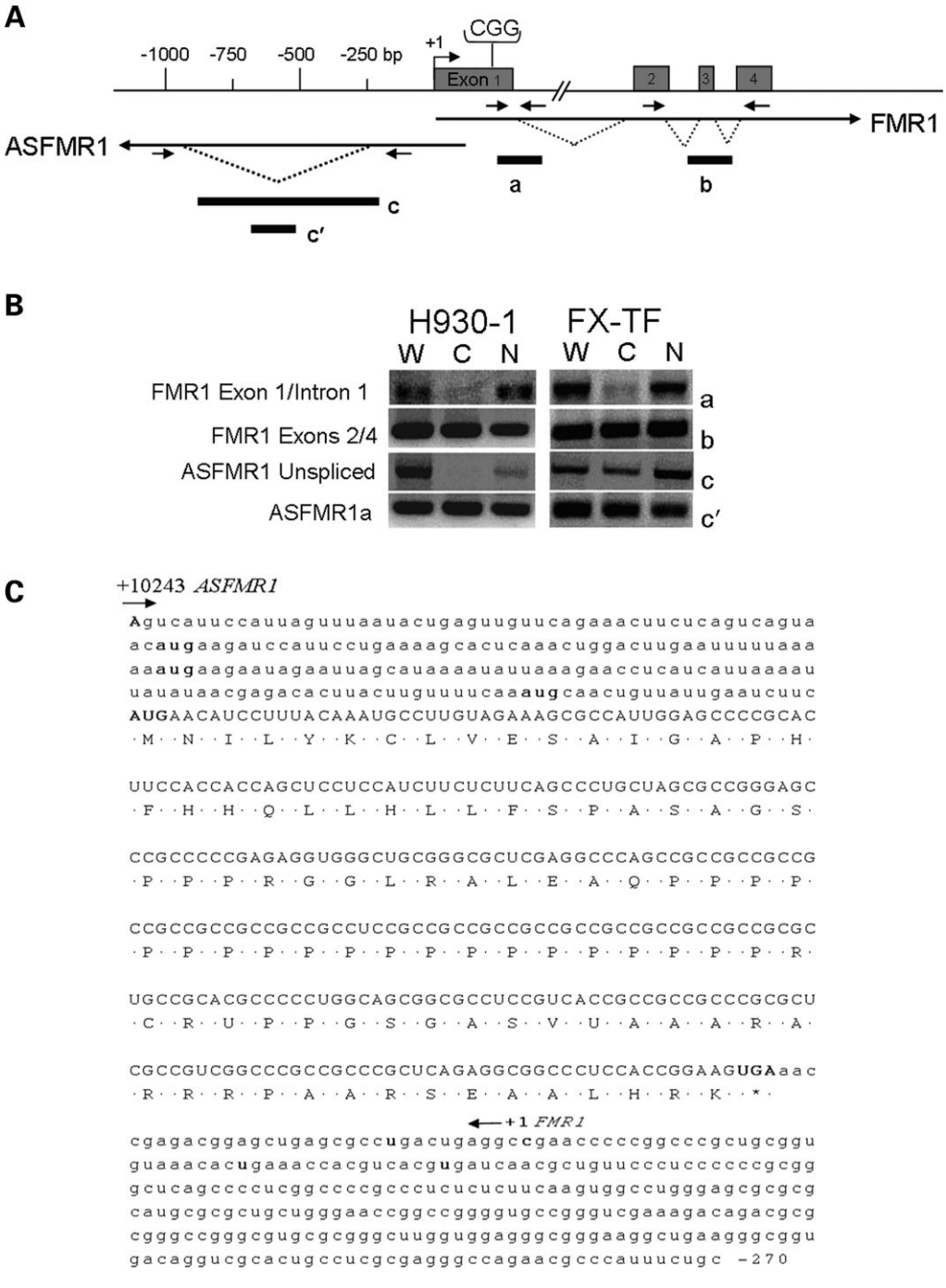

Figure 7. The *ASFMR1* transcript is transported to the cytoplasm and contains a putative ORF. (**A**) Diagram of the FMR1 locus depicting *ASFMR1* and exons 1–4 of *FMR1*. Spliced *ASFMR1* and *FMR1* transcripts are shown with dotted lines. PCR primer sets used for analysis are indicated as small black arrows, whereas the bars a, b, c and c′ demonstrate PCR products. (**B**) Semi-quantitative RT–PCR analysis of RNA isolated from whole cells (W), cytoplasmic fraction (C) and nuclear fraction (N). Amplification with the primer sets to *FMR1*, exon 1/intron 1 (unspliced, a) and exons 2/4 (spliced, b), and to ASFMR1, −1000F/−196R (unspliced, c, and spliced, c′), showed that similar to FMR1, the unspliced *ASFMR1* transcript is enriched in the nucleus, while the spliced *ASFMR1* transcript is evenly distributed between the cytoplasm and nucleus of both normal, H930-1, and premutation, FX-TF, cells. (**C**) The *ASFMR1* transcript sequence depicting the amino acid sequence for the ORF encoding a polyproline peptide is shown. The *ASFMR1* and FMR1 transcription initiation sites are indicated with arrows.

be under the influence of RNA directed recruitment of chromatin modifying complexes, a mechanism similar to what is proposed at the CTG repeat region of the *DM1* locus (24). Similar to DM1, CTCF may prevent heterochromatin spreading at the CGG repeat region in normal and premutation *FMR1* alleles. In addition, since it has been demonstrated that RNA derived from CGG repeat expansions larger than 200 repeats is able to form stable hairpins that can be processed into small RNAs by the RNase III enzyme Dicer (49–51), bidirectional transcription might interfere with the formation

of stable hairpin RNA. This could either modulate Dicer-mediated post-transcriptional silencing or alter the translation of the hairpin containing RNA.

In summary, we have identified *ASFMR1*, an antisense transcript at the *FMR1* locus, that might contribute to the pathogenesis of FXS and FXTAS either through an RNA and/or protein-mediated mechanism. Bidirectional transcription has now been described at several trinucleotide repeat loci, including *DM1*, *SCA8* and now *FMR1*. Understanding of the complex transcriptional regulation at the trinucleotide repeat loci and relative contribution of the bidirectional expression of the expanded repeats may provide further insight to the variable clinical phenotypes associated with FXTAS and FXS, as well as other triplet repeat disorders.

MATERIALS AND METHODS

Cell cultures

Human lymphoblastoid cell lines from individuals with the characterized CGG repeat expansions were established by Epstein–Barr virus transformation and characterized previously (13). Human:hamster hybrid cell lines, GM06318 and X8-6T2S1, retaining either an active or an inactive human X chromosome, respectively, as well as 19AS2, derived by 5′-azacytidine (5aC) treatment of X8-6T2S1, were described previously (34). Cells were maintained in RPMI 1640 medium supplemented with 10% fetal bovine serum (HyClone, UT, USA), glutamine and penicillin–streptomycin (Invitrogen) at 37°C in a humidified atmosphere of 5% CO_2.

Nuclear isolation

Lymphoblastoid cells ($4.0–6.0 \times 10^5$) were harvested by centrifugation at 700*g* for 5 min at 4°C. The cells were washed one time with ice-cold PBS and pelleted by centrifugation at 700*g* for 5 min at 4°C. The cells were resuspended in a 5× volume of ice-cold buffer A (10 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, supplemented with 1 mM DTT), incubated on ice for 10 min to swell and then pelleted at 500*g* for 10 min at 4°C, and the supernatant was discarded. The cell pellet was resuspended in 2× volume of buffer A and the outer cell membranes were disrupted by Dounce homogenization, 15 strokes with pestle 'B' or 'tight', on ice. The homogenate was centrifuged at 500*g* for 10 min at 4°C. The upper, less turbid phase containing the cytoplasmic fraction was separated from the lower, nuclei fraction. RNA was extracted from each fraction.

RNA isolation

Total RNA from lymphoblastoid cell lines was isolated using standard methods (TRIZOL, Invitrogen). The isolated RNA was treated with TURBO DNA-free DNase (Ambion) to remove contaminating DNA. For poly(A) RNA enrichment, total RNA was incubated with magnetic oligo-dT linked beads (Invitrogen) and eluted with 10 mM Tris–HCl. As a control for non-poly(A) RNA, the unbound fraction was collected and the remaining RNA was harvested with TRIZOL.

Blood samples were collected directly into Tempus Blood RNA Tubes (Applied Biosystems, Foster City, CA, USA) following informed consent and according to an approved Institutional Review Board protocol. Total RNA from peripheral blood leukocytes from individuals carrying either a normal, premutation or full mutation allele was isolated using the ABI PRISM™ 6100 Nucleic Acid PrepStation according to manufacturer's protocol.

cDNA generation

cDNA was generated with Superscript III (Invitrogen) under the following conditions. The appropriate primer (random hexamers, oligo-dT, gene-specific or strand-specific) was added to 1–3 μg of RNA and incubated at 70°C for 5 min in a thermocycler. The temperature was cycled to 50°C and the RNA/primer mixture equilibrated for 5 min. Next, the reaction mixture, nuclease free H_2O, 5× buffer, RNase inhibitor, dNTPs, DTT and RT (except for no-RT control), pre-equilibrated at 50°C, was added to the RNA/primer mixture and incubated for 2 h at 50°C. The RT was inactivated at 70°C for 5 min. Synthesis of cDNA was verified by amplification of the control transcript, *CTCF*, using the primers *CTCF* For: 5′-TGATGAGAGACCACACAAGTGCCA-3′ and *CTC*F Rev: 5′-CTGCACAAACTGCACTGAAACGGA-3′. RNA was tested for DNA contamination using a no-RT control and primers were verified by amplification of genomic DNA (genomic DNA control).

Strand-specific RT–PCR

For strand-specific RT–PCR, cDNA was generated using gene-specific primers which had a linker (LK) sequence, LK 5′-CGACTGGAGCACGAGGACACTGA-3′ attached to the 5′ end. Primer sequences are listed in Supplementary Material, Table S1. cDNA was generated with 1–3 μg of RNA and Superscript III (Invitrogen) at 50°C. PCR amplification of the strand-specific transcripts was performed using the LK sequence alone as a primer and a gene-specific reverse primer (35 cycles at 94°C for 30 s, 55°C for 30 s and 68°C for 1 min). PCR amplification of the strand-specific transcripts across the CGG repeat region was performed using 1.7 M betaine essentially as described previously (28). The PCR products were cloned into the pCR®4-TOPO vector (Invitrogen) and sequenced using an ABI Prism 3730xl DNA analyzer (Applied Biosystems).

Semi-quantitative RT–PCR

To establish the linear range for each gene-specific and control primer set, aliquots of the PCR reaction were collected every two cycles ranging from 26 to 40 cycles. The linear range for the *FMR1* primer set, Exon 2Forward, 5′-CATGAAGATTCAATAACAGTTGC-3′, with Exon 4Reverse, 5′-CACTTTAGCTAACCACCAACAG –3′, was established at 30–36 cycles, whereas the linear range for the *ASFMR1* primer set, −1427Forward, 5′-CATGTGACTACTCCAAAGACCCTAGTCC-3′, with −1168Reverse, 5′-CTTCCATGATGGCGGAACATGACCTAGTC-3′, was at 34–40 cycles. The *ASFMR1* primers were designed in an area common to the

spliced and unspliced transcripts. As a control, an adjustable 18S primer and competimer set (Ambion) was used to allow PCR amplification of 18S in the same linear range as the gene-specific primer set. The *FMR1/ASFMR1* multiplex RT–PCR was performed with an 18S primer to competimer ratio of 3:17 for 35 cycles. For strand-specific semi-quantitative RT–PCR, *CTCF* was used as the control transcript. The linear range for the *CTCF* was established at 28–34 cycles using the primer set listed in cDNA generation. The semi-quantitative strand-specific RT–PCR was repeated for each pair of primers at a minimum of two to three times.

Real time RT–PCR

FMR1 and *ASFMR1* transcript levels were quantified using real-time PCR performed on the automated ABI 7900 PCR machine (Applied Biosystems) using Fast Start SYBR Green Master Mix (ROCHE) with ROX passive reference dye added. cDNA was generated with random decamers using 1.5 μg RNA, then diluted 1:1. Two microliters of cDNA was used as template for the real-time reaction. *ASFMR1* was amplified with primers −1427Forward, 5′-CATGTGACTACTCCAAAGACCCTAGTCC-3′, and −1185Reverse, 5′-CATGACCTAGTCTGGGGTGGAG-3′, the *FMR1* primers were the same as those used for semi-quantitative RT–PCR. The standard curve assay as described by Applied Biosystems was used for absolute quantification. The values calculated for *FMR1* and *ASFMR1* transcript levels were normalized to β-*Glucoronidase* (*GUS*) expression as a control, with primers GUS Forward: 5′-CTCATTTGGAATTTTGCCGATT-3′ and GUS Reverse: 5′-CCGAGTGAAGATCCCCTTTTA-3′ as described previously (12). Approximate threshold cycle (C_t) values with these primers for normal lymphoblastoid cells were as follows: *FMR1* 23, *ASFMR1* 29 and *GUS* 28 cycles. PCR cycling was performed at [94°/2 min (94°/30 s, 60°/30 s, 72°/30s) 40] with an additional ramping step added after cycling to calculate the dissociation curves and confirm that fluorescence detected was due to full size PCR product and not PCR artifacts. *P*-values were determined by standard Student's *t*-test.

RLM-RACE

RNA ligase-mediated rapid amplification of 5′ and 3′ cDNA ends (RLM-RACE) was performed using the GeneRacer™ Kit (Invitrogen) according to manufacturer's protocol. This protocol allows obtaining full-length ends by eliminating uncapped transcripts. Briefly, for 5′-RACE, either total or poly(A) RNA from commercially derived human brain tissue (Ambion) and lymphoblastoid cell lines with a normal or a premutation allele, H930-1 and FX-TF, respectively, was treated with calf intestinal phosphatase to remove the 5′-phosphates of incomplete transcripts. The 5′-cap structure was removed with tobacco acid pyrophosphatase to leave a 5′-phosphate available for subsequent ligation of the GeneRacer™ RNA oligonucleotide. The ligated mRNA was reversed transcribed using Superscript III and random hexamers. In a separate analysis, phosphatase treatment and/or removal of the 5′-cap structures was omitted to allow amplification of both capped and uncapped transcripts or to specifically enrich for uncapped transcripts only. For 3′-RACE, total RNA was reverse-transcribed using Superscript III and the GeneRacer™ oligo-dT primer, containing a linker sequence attached to the 5′ end of the oligo-dT primer. For PCR amplification of the first-strand cDNA, each primer pair contained a gene-specific primer complementary to the *ASFMR1* sequence (or to the control *β-Actin* sequence) and the appropriate GeneRacer™ primer for the 5′ or 3′ end. A second round of PCR amplification was performed using 1 μl of the previous PCR product as template and a 5′- or 3′-nested GeneRacer™ primer and a corresponding gene-specific nested primer (by either touch-down PCR as described in the RLM-RACE kit or 35 cycles at 94°C for 30 s, 55°C for 30 s and 68°C for 5 min). Sequences of gene-specific primers used for 5′- and 3′-RACE are listed in Supplementary Material, Table S1. The PCR products were cloned into the pCR®4-TOPO vector (Invitrogen) and sequenced using an ABI Prism 3730xl DNA analyzer (Applied Biosystems). The identified 5′ and 3′ ends were confirmed by strand-specific RT–PCR.

Northern blot analysis

Northern blot analysis was performed according to standard protocol. The *ASFMR1* and 18S probes were radioactively labeled using Ready-to-Go DNA labeling beads (Amersham). For small RNA northern blot analysis, 30 μg RNA from each cell line were separated electrophoretically in a 10% polyacrylamide/8 M urea/1× TBE gel. RNA was electro-blotted onto Nytran SPC nylon membrane (Whatman) in 1× TBE at 250 mA for 45 min and UV cross-linked. Blots were hybridized overnight at 35°C in Ultrahybe Oligo buffer (Ambion) with radiolabeled oligonucleotide probes complementary to the *FMR1* and *ASFMR1* sequences of interest.

Analysis of the CGG repeat size and methylation status

Southern blot and PCR-based analyses were performed as described previously (28). Genomic DNA was isolated from peripheral blood leukocytes (5 ml of whole blood) using standard methods (Puregene and Purescript Kits, Gentra Inc.). For Southern blot analysis, 5–10 μg of isolated DNA was digested with *Eco*RI and *Nru*I. Hybridization was performed using the StB12.3 probe specific for the *FMR1* gene. For PCR-based analysis of genomic DNA, hybridization was performed with an oligonucleotide probe (CGG)10. Analysis of trinucleotide expansion allele size was conducted using an Alpha Innotech FluorChem 8800 Image Detection System.

Electrophoretic mobility shift assay

Fifteen 200–300 bp DNA fragments were ^{32}P-labeled, gel purified, and used as DNA probes for gel mobility shift assays with equal amounts of the *in vitro* translated 11-ZF DNA binding domain and full-length human CTCF proteins as described (32). A known CTCF-binding site in the *DM1* locus was used as a positive control (32). Binding reactions were carried out in the buffer containing standard PBS with 5 mM $MgCl_2$, 0.1 mM $ZnSO_4$, 1 mM DTT, 0.1% NP-40 and 10% glycerol in the presence of poly(deoxyinosinic-deoxycytidylic acid). Reaction mixtures of 20 μl of final

volume were incubated for 30 min at room temperature and then analyzed on 5% non-denaturing PAGE run in 0.5× Tris-borate-EDTA buffer.

Chromatin immunoprecipitation

The ChIP assays were performed as previously described (52), using commercially available antibody against CTCF (Upstate). Briefly, formaldehyde cross-linked chromatin after sonication was incubated with the CTCF antibody, then precipitated by the addition of protein-G Sepharose. After reverse cross-linking, the IP products were analyzed by semi-quantitative PCR. A region of the *DM1* locus, which is positive for CTCF-binding (32) was used as a positive control to verify CTCF-specific ChIP. A region of the *KIAA522* locus, which is negative for CTCF-binding (53), was used for normalization.

Accession codes

GenBank: *Homo sapiens* ASFMR1d mRNA, from +10243 to −2490 with splice from +10018 to +315, EU48200; *Homo sapiens* ASFMR1c mRNA, from +10243 to −2490 with splice from +10018 to +315 and from −538 to −919, EU48201; *Homo sapiens* ASFMR1b mRNA, from +10243 to −2490 with splice from +10018 to +315, −274 to −530 and −635 to −921, EU48202; *Homo sapiens* ASFMR1a mRNA, from +10243 to −2490 with splice from +10018 to +315 and from −274 to −921, EU48203; *Homo sapiens* ASFMR1e mRNA, from +10243 to −2490 with splice from +10018 to +315 and from +10155 to +10070, EU48203.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

ACKNOWLEDGEMENTS

We acknowledge Stan Gartler, Charles Laird, Paul Hagerman, Adam Geballe, Stephanie Child and Steve Collins for helpful discussions and technical advice.

Conflict of Interest statement: The authors have no conflict of interest.

FUNDING

This work was supported by NIH/NICHD P30 HD002274-35S1 (S.J.T. and G.N.F.), NIH/NCI R01 CA068360 (G.N.F.) and NIH Chromosome Metabolism and Cancer Training grant T32 CA09657-14 and NIH NIAMS grant F32 AR052581 (S.A.G.).

REFERENCES

1. Neri, G. and Chiurazzi, P. (1999) X-linked mental retardation (Review). *Adv. Genet.*, **41**, 55–94.
2. O'Donnell, W.T. and Warren, S.T. (2002) A decade of molecular studies of fragile X syndrome (Review). *Annu. Rev. Neurosci.*, **25**, 315–338.
3. Hagerman, P.J. and Hagerman, R.J. (2004) The fragile-X premutation: a maturing perspective. *Am. J. Hum. Genet.*, **74**, 805–816.
4. Willemsen, R., Mientjes, E. and Oostra, B.A. (2005) FXTAS: a progressive neurologic syndrome associated with Fragile X premutation (Review). *Curr. Neurol. Neurosci. Rep.*, **5**, 405–410.
5. Oostra, B.A. and Willemsen, R. (2003) A fragile balance: FMR1 expression levels. *Hum. Mol. Genet.*, **12**, R249–R257.
6. Hansen, R.S., Gartler, S.M., Scott, C.R., Chen, S.H. and Laird, C.D. (1992) Methylation analysis of CGG sites in the CpG island of the human FMR1 gene. *Hum. Mol. Genet.*, **1**, 571–578.
7. Stoger, R., Kajimura, T.M., Brown, W.T. and Laird, C.D. (1997) Epigenetic variation illustrated by DNA methylation patterns of the fragile-X gene FMR1. *Hum. Mol. Genet.*, **6**, 1791–1801.
8. Dombrowski, C., Levesque, S., Morel, M.L., Rouillard, P., Morgan, K. and Rousseau, F. (2002) Premutation and intermediate-size FMR1 alleles in 10572 males from the general population: loss of an AGG interruption is a late event in the generation of fragile X syndrome alleles. *Hum. Mol. Genet.*, **11**, 371–378.
9. Jacquemont, S., Hagerman, R.J., Leehey, M.A., Hall, D.A., Levine, R.A., Brunberg, J.A., Zhang, L., Jardini, T., Gane, L.W., Harris, S.W. *et al.* (2004) Penetrance of the fragile X-associated tremor/ataxia syndrome in a premutation carrier population. *JAMA*, **291**, 460–469.
10. Tassone, F., Beilina, A., Carosi, C., Albertosi, S., Bagni, C., Li, L., Glover, K., Bentley, D. and Hagerman, P.J. (2007) Elevated FMR1 mRNA in premutation carriers is due to increased transcription. *RNA*, **13**, 555–562.
11. Kenneson, A., Zhang, F., Hagedorn, C.J. and Warren, S.T. (2001) Reduced FMRP and increased FMR1 transcription is proportionally associated with CGG repeat number in intermediate-length and premutation carriers. *Hum. Mol. Genet.*, **10**, 1449–1454.
12. Tassone, F., Hagerman, R.J., Taylor, A.K., Gane, L.W., Godfrey, T.E. and Hagerman, P.J. (2000) Elevated levels of FMR1 mRNA in carrier males: A new mechanism of involvement in fragile X syndrome. *Am. J. Hum. Genet.*, **66**, 6–15.
13. Primerano, B., Tassone, F., Hagerman, R.J., Hagerman, P., Amaldi, F. and Bagni, C. (2002) Reduced FMR1 mRNA translation efficiency in fragile x patients with premutations. *RNA*, **8**, 1482–1488.
14. Iwahashi, C.K., Yasui, D.H., An, H.J., Greco, C.M., Tassone, F., Nannen, K., Babineau, B., Lebrilla, C.B., Hagerman, R.J. and Hagerman, P.J. (2006) Protein composition of the intranuclear inclusions of FXTAS. *Brain*, **129**, 256–271.
15. Tassone, F., Iwashashi, C. and Hagerman, P.J. (2004) FMR1 RNA within the intranuclear inclusions of fragile X-associated tremor/ataxia syndrome (FXTAS). *RNA Biol.*, **1**, 103–105.
16. Greco, C., Hagerman, R.J., Tassone, F., Chudley, A., Del Bigio, M.R., Jacquemont, S., Leehey, M. and Hagerman, P.J. (2002) Neuronal intranuclear inclusions in a new cerebellar tremor/ataxia syndrome among fragile X carriers. *Brain*, **125**, 1760–1771.
17. Greco, C.M., Berman, R.F., Martin, R.M., Tassone, F., Schwartz, P.H., Chang, A., Trapp, B.D., Iwahashi, C., Brunberg, J., Grigsby, J. *et al.* (2006) Neuropathology of fragile X-associated tremor/ataxia syndrome (FXTAS). *Brain*, **129**, 243–255.
18. Ranum, L.P. and Cooper, T.A. (2006) RNA-mediated neuromuscular disorders. *Annu. Rev. Neurosci.*, **29**, 259–277.
19. Tapscott, S.J. (2000) Deconstructing myotonic dystrophy. *Science*, **289**, 1701–1702.
20. Lugenbeel, K.A., Peier, A.M., Carson, N.L., Chudley, A.E. and Nelson, D.L. (1995) Intragenic loss of function mutations demonstrate the primary role of FMR1 in fragile X syndrome. *Nat. Genet.*, **10**, 483–485.
21. Meijer, H., de Graaff, E., Merckx, D.M., Jongbloed, R.J., de Die-Smulders, C.E., Engelen, J.J., Fryns, J.P., Curfs, P.M. and Oostra, B.A. (1994) A deletion of 1.6 kb proximal to the CGG repeat of the FMR1 gene causes the clinical phenotype of the fragile X syndrome. *Hum. Mol. Genet.*, **3**, 615–620.
22. Chiurazzi, P., de Graaff, E., Ng, J., Verkerk, A.J., Wolfson, S., Fisch, G.S., Kozak, L., Neri, G. and Oostra, B. (1994) No apparent involvement of the FMR1 gene in five patients with phenotypic manifestations of the fragile X syndrome. *Am. J. Med. Genet.*, **51**, 309–314.
23. Reyniers, E., Wolff, G., Tariverdian, G., De Boulle, K., Storm, K., Kooy, R.F. and Willems, P.J. (1996) Severe mental retardation and macroorchidism without mutation in the FMR1 gene (Review). *Am. J. Med. Genet.*, **64**, 408–412.

24. Cho, D.H., Thienes, C.P., Mahoney, S.E., Analau, E., Filippova, G.N. and Tapscott, S.J. (2005) Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF. *Mol. Cell*, **20**, 483–489.
25. Moseley, M.L., Zu, T., Ikeda, Y., Gao, W., Mosemiller, A.K., Daughters, R.S., Chen, G., Weatherspoon, M.R., Clark, H.B., Ebner, T.J. *et al.* (2006) Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8. *Nat. Genet.*, **38**, 758–769.
26. Drouin, R., Angers, M., Dallaire, N., Rose, T.M., Khandjian, E.W. and Rousseau, F. (1997) Structural and functional characterization of the human FMR1 promoter reveals similarities with the hnRNP-A2 promoter region. *Hum. Mol. Genet.*, **6**, 2051–2060.
27. Beilina, A., Tassone, F., Schwartz, P.H., Sahota, P. and Hagerman, P.J. (2004) Redistribution of transcription start sites within the FMR1 promoter region with expansion of the downstream CGG-repeat element. *Hum. Mol. Genet.*, **13**, 543–549.
28. Saluto, A., Brussino, A., Tassone, F., Arduino, C., Cagnoli, C., Pappi, P., Hagerman, P., Migone, N. and Brusco, A. (2005) An enhanced polymerase chain reaction assay to detect pre- and full mutation alleles of the fragile X mental retardation 1 gene. *J. Mol. Diagn.*, **7**, 605–612.
29. Meyer, T., Carlstedt-Duke, J. and Starr, D.B. (1997) A weak TATA box is a prerequisite for glucocorticoid-dependent repression of the osteocalcin gene. *J. Biol. Chem.*, **272**, 30709–30714.
30. Bell, A.C., West, A.G. and Felsenfeld, G. (2001) Insulators and boundaries: versatile regulatory elements in the eukaryotic genome (Review). *Science*, **291**, 447–450.
31. Filippova, G.N. (2007) Genetics and epigenetics of the mulitfunctional protein CTCF. *Curr. Top. Dev. Biol.*, **80**, 337–362.
32. Filippova, G.N., Thienes, C.P., Penn, B.H., Cho, D.H., Hu, Y.J., Moore, J.M., Klesert, T.R., Lobanenkov, V.V. and Tapscott, S.J. (2001) CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. *Nat. Genet.*, **28**, 335–343.
33. Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I. and Zhao, K. (2007) High-resolution profiling of histone methylations in the human genome. *Cell*, **129**, 823–837.
34. Hansen, R.S., Canfield, T.K., Fjeld, A.D. and Gartler, S.M. (1996) Role of late replication timing in the silencing of X-linked genes. *Hum. Mol. Genet.*, **5**, 1345–1353.
35. Hinds, H.L., Ashley, C.T., Sutcliffe, J.S., Nelson, D.L., Warren, S.T., Housman, D.E. and Schalling, M. (1993) Tissue specific expression of FMR-1 provides evidence for a functional role in fragile X syndrome. *Nat. Genet.*, **3**, 36–43.
36. Arocena, D.G., Iwahashi, C.K., Won, N., Beilina, A., Ludwig, A.L.T., , F., Schwartz, P.H. and Hagerman, P.J. (2005) Induction of inclusion formation and disruption of lamin A/C structure by premutation CGG-repeat RNA in human cultured neural cells. *Hum. Mol. Genet.*, **14**, 3661–3671.
37. Tassone, F., Hagerman, R.J., Garcia-Arocena, D., Khandjian, E.W., Greco, C.M. and Hagerman, P.J. (2004) Intranuclear inclusions in neural cells with premutation alleles in fragile X associated tremor/ataxia syndrome. *J. Med. Genet.*, **41**, e43.
38. Fardaei, M., Rogers, M.T., Thorpe, H.M., Larkin, K., Hamshere, M.G., Harper, P.S. and Brook, J.D. (2002) Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Hum. Mol. Genet.*, **11**, 805–814.
39. Kino, Y., Mori, D., Oma, Y., Takeshita, Y., Sasagawa, N. and Ishiura, S. (2004) Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats. *Hum. Mol. Genet.*, **13**, 495–507.
40. Sofola, O.A., Jin, P., Botas, J. and Nelson, D.L. (2007) Argonaute-2 dependent rescue of a *Drosophila* model of FXTAS by FRAXE premutation repeat. *Hum. Mol. Genet.*, 2326–2332.
41. Zoghbi, H.Y. and Or, H.T. (2000) Glutamine repeats and neurodegeneration. *Annu. Rev. Neurosci.*, **23**, 217–247.
42. Margolis, R.L., Holmes, S.E., Rosenblatt, A., Gourley, L., O'Hearn, E., Ross, C.A., Seltzer, W.K., Walker, R.H., Ashizawa, T., Rasmussen, A. *et al.* (2004) Huntington's Disease-like 2 (HDL2) in North America and Japan. *Ann. Neurol.*, **56**, 670–674.
43. French, C., Menegazzi, P., Nicholson, L., Macaulay, H., DiLuca, D. and Gompels, U.A. (1999) Novel, nonconsensus cellular splicing regulates expression of a gene encoding a chemokine-like protein that shows high variation and is specific for human herpesvirus 6. *Virology*, **262**, 139–151.
44. Haeger, P., Cuevas, R., Forray, M.I., Rojas, R., Daza, C., Rivadeneira, J. and Gysling, K. (2005) Natural expression of immature Ucn antisense RNA in the rat brain. Evidence favoring bidirectional transcription of the Ucn gene locus. *Brain Res. Mol. Brain Res.*, **139**, 115–128.
45. Kurose, K., Koyano, S., Ikeda, S., Tohkin, M., Hasegawa, R. and Sawada, J. (2005) 5′ diversity of human hepatic PXR (NR1I2) transcripts and identification of the major transcription initiation site. *Mol. Cell Biochem.*, **273**, 79–85.
46. Martens, J.H., O'Sullivan, R.J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P. and Jenuwein, T. (2005) The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J.*, **24**, 800–812.
47. Grewal, S.I. and Jia, S. (2007) Heterochromatin revisted. *Nat. Rev. Genet.*, **8**, 35–46.
48. Talbert, P.B. and Henikoff, S. (2006) Spreading of silent chromatin: inaction at a distance. *Nat. Rev. Genet.*, **7**, 793–803.
49. Handa, V., Saha, T. and Usdin, K. (2003) The fragile X syndrome repeats form RNA hairpins that do not activate the interferon-inducible protein kinase, PKR, but are cut by Dicer. *Nucleic Acids Res.*, **31**, 6243–6248.
50. Krol, J., Fiszer, A., Mykowska, A., Sobczak, K., de Mezer, M. and Krzyzosiak, W.J. (2007) Ribonuclease dicer cleaves triplet repeat hairpins into shorter repeats that silence specific targets. *Mol. Cell*, **25**, 575–586.
51. Napierala, M., Michalowski, D., de Mezer, M. and Krzyzosiak, W.J. (2005) Facile FMR1 mRNA structure regulation by interruptions in CGG repeats. *Nucleic Acids Res.*, **33**, 451–463.
52. Nelson, J.D., Denisenko, O. and Bomsztyk, K. (2006) Protocol for the fast chromatin immunoprecipitation (ChIP) method. *Nat. Protoc.*, **1**, 179–185.
53. Filippova, G.N., Cheng, M.K., Moore, J.M., Truong, J.P., Hu, Y.J., Nguyen, D.K., Tsuchiya, K.D. and Disteche, C.M. (2005) Boundaries between chromosomal domains of X inactivation and escape bind CTCF and lack CpG methylation during early development. *Dev. Cell*, **8**, 31–42.