ARTICLE

A novel gene, FAM11A, associated with the FRAXF CpG island is transcriptionally silent in FRAXF full mutation

Marie A Shaw¹, Pietro Chiurazzi², Dennis R Romain³, Giovanni Neri⁴ and Jozef Gécz^{*,1,5}

¹Department of Cytogenetics and Molecular Genetics, Women's and Children's Hospital, Adelaide, SA, Australia; ²Department of Paediatrics, University Hospital of Messina, Italy; ³Cytogenetics Laboratory, Department of Laboratory Services, Wellington Hospital, Private Bag 7902, Wellington South, New Zealand; ⁴Istituto di Genetica Medica, Universita' Cattolica, Roma, Italy; ⁵ Department of Paediatrics, University of Adelaide, Adelaide, SA, Australia

The cytogenetic expression of the *FRAXF* fragile site is due to an expanded, hypermethylated and unstable CGG repeat in Xq28. Normal individuals have 6-38 triplet repeats while individuals expressing the fragile site have expansions of greater than 300 triplets. Through analysis of the region adjacent to the fragile site, we have identified a ~2.6 kb cDNA originating from the *FRAXF* fragile site associated CpG island, and containing the unstable *FRAXF* CGG repeat in its 5' UTR region. This gene, *FAM11A*, comprises at least seven exons, shows alternative splicing, and extends over 35 kb of genomic DNA distal to the *FRAXF* fragile site. Analysis of the *FAM11A* cDNA sequence has identified a 1050 bp open reading frame encoding a 350 amino acid protein. We have also identified *FAM11B* a highly conserved (88% at the protein level) transcribed chromosome 2 retropseudogene. We show that the novel *FRAXF* fragile site associated *FAM11A* transcriptionally silenced in a normal individual with a cytogenetically and molecularly detectable *FRAXF* CGG full mutation (fragile site). Finally, we were able to reactivate *FAM11A* transcription by treatment of a *FRAXF* lymphoblastoid cell line with the demethylating agent 5-azadeoxycytidine, thus demonstrating the critical role of *FRAXF* methylation in *FAM11A* silencing. *European Journal of Human Genetics* (2002) **10**, 767–772. doi:10.1038/sj.ejhg.5200881

Keywords: FRAXF; FAM11A; fragile sites; transcriptional silencing; transcribed retropseudogene; FAM11B

Introduction

FRAXF is a folate sensitive fragile site in Xq28. It was originally identified in 1993 in a family with developmental delay.¹ Since then, it has been seen in only a small number of additional families tested cytogenetically because of MR or developmental delay.¹⁻⁴ In addition to *FRAXF* there are two other folate sensitive fragile sites in Xq27-28, *FRAXA* and *FRAXE*. The molecular basis of all three Xq27-28 fragile sites is an expanded, unstable, and hypermethylated CGG repeat.³⁻⁹ While *FRAXA* and *FRAXE* are associated with mental retardation, *FRAXF* was consid-

Received 26 April 2002; revised 22 July 2002; accepted 24 July 2002

ered to be benign. A study by Holden *et al*¹⁰ suggested that the high incidence of 12 and 14 repeat alleles, with only 1/501 alleles having 13 repeats, indicated a functional significance for the *FRAXF* repeat region. *FRAXF* has been investigated as a causative factor in diseases such as mental retardation with multiple congenital anomalies,¹¹ autism and pervasive developmental disorder,¹² and PPM-X (Psychosis, Pyramidal signs, Macroordism),¹³ however association with any of these phenotypes is yet to be confirmed. Based on this and previous work on *FRAXF* ^{3,4} it has been assumed that the *FRAXF* fragile site does not disturb expression of any gene, at least not at any phenotypically discernible level.

We report on the identification of a novel *FRAXF* fragile site associated gene *FAM11A*. The *FAM11A* gene is characterised in terms of its full length cDNA sequence,

^{*}Correspondence: J Gécz; Department of Cytogenetics and Molecular Genetics, Women's and Children's Hospital, 72 King William Road, North Adelaide, SA 5006, Australia. Tel: 61 8 8161 6339; Fax: 61 8 8161 7342; E-mail: jozef.gecz@adelaide.edu.au



genomic structure, conservation, and transcriptional silencing in *FRAXF* full mutation.

Materials and methods

Family information

We present data from a male who expresses the *FRAXF* fragile site in Xq28. This individual (II:3) was initially described by Romain *et al*,² in a family without mental retardation. Parrish *et al*,³ later showed that this individual (II:3, family 2) exhibited methylated *FRAXF* allele of 1.6 kb. An EBV virus transformed lymphoblast cell line from this individual was established and cultured.

Database searches and gene structure determination

Genbank sequence searches were carried out with the BLAST[®] search algorithm (Basic Local Alignment Search Tool)¹⁴ at the NCBI website (http://www.ncbi.nlm.nih.gov). An EST cluster, mapping to the *FRAXF* region, was identified from the Unigene database at NCBI. Comparison between the human Unigene cluster Hs.11522 and *FRAXF* genomic sequence (U71148) was used to determine the genomic structure of the *FAM11A* gene. DNA sequences, ESTs and Unigene clusters were downloaded and further manipulated using the Lasergene software package (DNA Star).

Expression analysis of FRAXF full mutation

Lymphoblast cells were harvested and RNA isolated from approximately 5×10^6 cells using TRIZOL reagent (GIBCO BRL). Approximately $1 \mu g$ of purified total RNA was random primed for first-strand cDNA synthesis, using SuperScript II RNAse H⁻ reverse transcriptase (GIBCO BRL). Reverse transcription was carried out for 1 h at 42°C in a thermal cycler (Hybaid). Negative controls containing no reverse transcriptase were run at the same time. PCR was carried out using primers designed to amplify the region from exon 2 to 5 of FAM11A (E2F1: 5'-GGA TGG CAT CAT ACA GTG GAG-3', E5R1: 5'-CAA GAC GGA CCA CAC AAT GTA G-3'). EsteraseD (ESD) specific primers were used as a control for RT-PCR efficiency.¹⁵ Standard conditions for PCR were 30 s at 94°C, 30 s at 60°C, and 30 s at 72°C for 35 cycles in a final 50 μ l volume containing 100 mM Tris-HCl, pH=8.3; 50 mM KCl; 1.5 mM MgCl₂; 0.2 mM of each dNTP; 50 pmol of each primer; and 2 U of Taq DNA polymerase (Roche). PCR products were analysed on a 1.5% agarose gel with 0.5 mg ml⁻¹ ethidium bromide in $1 \times TBE$.

Reactivation of *FAM11A* transcriptional silencing with 5-azadeoxycytidine

For treatment with 5-azadeoxycytidine, patient and control lymphoblast cell lines were grown in a 50 ml flask with RPMI1640 culture medium with or without 5-azadeoxycytidine (1 μ M) for 7 days as previously described.¹⁶ Primers and RT–PCR conditions were as described above.

Alternative splicing

Lymphoblast cDNA was used for PCR to determine the extent of alternative splicing. PCR product was extracted from the agarose gel using the Concert Nucleic Acid Purification Kit (Invitrogen) and sequenced using BigDye Terminator ready reaction kit, according to the manufacturer's specifications (Applied Biosystems).

RT-PCR and Northern blot hybridisation

In order to confirm that the *FRAXF* CGG repeat was part of the *FAM11A* transcript, the following flanking oligonucleotide primers were designed (Figure 1), F_{CGG} 5'-CAG GGG GCG GTG GCT CAG GTT TC-3', and R_{CGG} 5'-CTC AGG TT<u>C AT</u>G GCG GAG AAC-3' (ATG translation start codon is underlined). These two primers amplify a 210 bp product from exon 1 of *FAM11A*. The 3' UTR of *FAM11A* was amplified from lymphoblast cDNA using the primers; F: 5'-CAC CCC GCA GAC CAG AAC CAG-3', R: 5'-TCA GCA TGC CCA AAG AGA AAC TAC-3'. This 500 bp product was then used as a probe for Northern blot. A human 12 lane Multi-Tissue Northern blot (Clontech), containing 1 μ g of polyA⁺ RNA isolated from various human adult tissues, was hybridised with this probe using ExpressHyb solution (Clontech) as per manufacturer's instructions.

Results

Identification of a novel gene near FRAXF

Database searches using genomic sequence encompassing FRAXF fragile site (U71148) were used to identify transcribed sequences in the region around the FRAXF CGG repeat. Two Unigene clusters (Hs.352168 and Hs.11522) were identified which span the region distal to triplet. These two clusters corresponded to an as yet uncharacterised gene, which was assigned the symbol FAM11A (Family with sequence similarity 11, member A, Genbank accession no. AF530473). Northern blot analysis indicates that FAM11A is expressed nearly ubiquitously (also supported by analysis of ESTs) with predominant expression in heart, skeletal muscle, kidney and placenta (Figure 2a). It is transcribed as a 2554 bp message, composed of at least seven exons, encompassing about 35 kb of genomic sequence distal to the FRAXF CpG island. All intron/exon boundaries conform to the standard splice acceptor/splice donor consensus sequences¹⁷ (Table 1). The open reading frame (ORF) is 1050 bp long and encodes a protein of 350 amino acids. Analysis of the cDNA sequence has identified a translation start codon within exon 1 (Figure 1). A putative transcription start site was located at position -372 with respect to the ATG start codon (Figure 1). RT-PCR analysis with primers flanking the FRAXF CGG repeat showed that the CGG repeat is transcribed as part of the FAM11A transcript (results not shown). This is also supported by the analysis of available ESTs one of which does contain the CGG repeat in its sequence (BI910833) and two others

Exon	Exon size (bp)	3' splice site intron/exon	5' splice site exon/intron	Intron size (bp)	
1	410	5′ UTR	TCAACCCCGAG/ gt gaggcggc	20 077	
2	177	tttctcctag/TAAATTCCTC	CTCAATATCG/ gt aatactgc	2453	
3	159	ccttttttag/AGCAGAAGGA	GTTTCCCCGGT/gtctgttgca	(4)4587/(4a)4 5	
3a	208	ccttttttag/AGCAGAAGGA	GTCACTAGAG/ gt gagatttc	(4)4538/(4a)4 5	
4	80	ttttagttag/AAATCCTGTG	GCCCTGGCTT/ gt atgtaact	3507	
4a	84	tcatttttag/TTAGAAATCC	GCCCTGGCTT/ gt atgtaact	3507	
5	176	ccacccccag/GTTGTGTGTG	TACATTTGAG/ gt aagcgttc	625	
6	124	ctttctttag/ATTCTGCTGG	GGAAACCACT/ qt atqtact	1298	
7	1701	tatttttcag/GGTGGTTTGG	3' UTR	-	

 Table 1
 Splice sites of the FAM11A gene^a

^aIntronic sequences are shown in lower case, exonic sequences are shown in upper case.

```
aaaagggaac ccaggagggg aaggcggttg gcacgggaag tgactttggt -663
gctaggtaac ttgagtaggg aatagccagc atatctggac acgctggctt -613
gggaattgga agggaagttg ggacccaaac tttggagtcg ttgacagatg -563
tgacaggtga agcctgggat gacatcgcca aaaatgcaac gtctcactca -513
ttgtcactac tcccagggct cagtcgtcac tggggaaaat ctccagaagg -463
tagcgcgggc caaggtgaca ggtgtctgcc aagatctgcc cgccagactc -413
ccgggcggcg cgctccctcc ctgcaggcct tcagcccgtc
                                           Agcatecet -363
tectogggge cetgeteact eccageetee atceeetge cateteetee -313
gccggtcgcg tgcggacaca aggatgggga cctcccagcg aggagcgctc -263
tgggcggggc tccggacgca tgcgcggccc tcgtacggaa gcccggaagg -213
aggggcaggg ggcggtggct
                     caggtttctc cgggcggcgg cggcggcggc -163
ggcggcgacg gcgacggcga cggcagcggg gacggcagca gtagcgggag -113
cagcagegtg gaegeggetg gegetggege catgaaceeg etgtaaggeg
caggetgtgc ageacggggt gegggggggg aggaggagga egeegeggtg
                                                       -13
aagtteteeg ccATGAACCT
                     GAGGGGCCTC TTCCAGGACT TCAACCCGAG
                N L
                      RGL
                                FODF
                                             Ν
                                                P S
```

```
Figure 1 Genomic sequence around the putative promoter region of the FAM11A gene. The region was analysed in silico for putative promoter and transcription start. A candidate transcription start at position -372 bp (with respect to the translation initiation codon) has been predicted (indicated by a circle). The FRAXF (CGG)<sub>7</sub> repeat is boxed. There is little similarity between human (U71148) and mouse (AL672026) genomic sequence in the promoter region. The only 100% conserved stretch of sequence is that of 20 bp highlighted (in bold and underlined). Two primers F_{CGG} and R_{CGG} used for PCR and RT–PCR across the FRAXF CGG repeat are shown as arrows.
```

(BM011526 and BM469272) which terminate just below the CGG repeat.

Alternative splicing of FAM11A

Two major *FAM11A* isoforms were detected on Northern blot, ~ 2.6 kb and a smaller ~ 2.1 kb. The origin of these two isoforms is not obvious. Originally we speculated that the smaller isoform corresponds to the chromosome 2 transcribed processed retropseudogene, *FAM11B* transcript (see below). However, the probe as used for the Northern blot



Figure 2 Analysis of the expression of the *FAM11A gene*. (a) Northern blot analysis of the human *FAM11A* gene. Human adult 12-lane multiple tissue northern blot was hybridised with a *FAM11A* gene 3' UTR probe. Two bands at ~2.6 and ~2.1 kb (indicated with arrowheads) of various intensity were detected across all tissues tested, however, heart, skeletal muscle, kidney and placenta were among the tissues with the highest *FAM11A* abundance. (b) A diagram showing alternative splicing events in *FAM11A*. Exons 2, 3, 4 and 7 have been found alternatively spliced. Exons (3a, 4a and intron 7) involved in alternative splicing are shaded light grey. The putative ORF is shaded black. I, shows the full length *FAM11A* transcript, while II, indicates the short *FAM11A* transcript truncated after exon 4.

would not cross-hybridise with the *FAM11B* transcript based on sequence alignment (results not shown). As a result the most likely explanation for the existence of the two major isoforms as detected by Northern blot hybridisation is alternative splicing. This was studied on human lymphoblast mRNA. Alternative splicing events were detected for exons



a

2, 3, 4 and 7. An isoform missing exon 2 was detected at a low level, also exons 2 and 3 were missing in a much smaller proportion of transcripts, as determined from the RT-PCR. We have observed that in a small proportion of ESTs exons 3 and 4 show alternative splicing. For exon 3 an additional 49 bp is observed at the 3' end (5' donor splice site of intron 3), and for exon 4 an additional 4 bp at the 5' end (3' acceptor splice site of intron 3; see Table 1, annotated 3a and 4a). An interesting alternative splicing event at the 3' end of the gene occurs within exon 7. Seventy-eight bp within exon 7, which we call intron 7 is excluded in \sim 50% of ESTs. RT-PCR analysis has revealed that the intron is in fact present in a higher proportion of transcripts (data not shown) and is the most predominant form. This results in a splitting of exon 7 so we have annotated this as exons 7a and 7b (see Figure 2b).

Additionally, comparison of the two *FAM11A* associated Unigene clusters Hs.352168 and Hs.11522 revealed that ESTs from cluster Hs.352168 have different 3' end starting from the end of exon 4. In five of these ESTs (BM458996, BG721461, BF734532, AV720618 and BG565086) the donor splice site of exon 4 is ignored. This results in the inclusion of an extra ~600 bp of the *FAM11A* intron 3 sequence in the *FAM11A* transcript and as a consequence premature termination of transcription (see Figure 2).

FAM11A is transcriptionally silenced by *FRAXF* full mutation

To investigate whether *FAM11A* transcription is silenced by the *FRAXF* full mutation, lymphoblast mRNA from an individual expressing the fragile site was analysed by RT–PCR (Figure 3a). The result of this experiment shows that expansion of the *FRAXF* CGG repeat and consequent methylation of the CpG island results in the transcriptional silencing of the *FAM11A* gene.

To confirm that methylation of the *FRAXF* CpG island was directly involved in the transcriptional silencing of the *FAM11A* gene we treated the patient lymphoblastoid cell line with the demethylation agent 5-azadeoxycytidine. Reactivation of the *FAM11A* gene transcription as detected by RT–PCR occurred upon treatment with 5-azadeoxycytidine for 7 days (Figure 3b). This demonstrated that methylation of the *FRAXF* CpG island as a consequence of the CGG repeat expansion has a crucial role in the *FAM11A* gene silencing.

Identification of a chromosome 2 transcribed, processed retropseudogene

dbEST database searches using *FRAXF* genomic sequence revealed additional ESTs (Unigene cluster Hs.44680), which correspond to a highly similar gene to *FAM11A* on chromosome 2q21.2. Comparison of *FAM11A* and the chromosome 2 gene, *FAM11B* (*Family with sequence similarity* <u>11</u>, *member* <u>B</u>, Genbank accession no. AF530474) revealed that the chromosome 2 gene is a transcribed retropseudogene. Genes



Figure 3 (a) RT–PCR analysis of the *FAM11A* gene in lymphoblast cells of a *FRAXF* normal and *FRAXF* full mutation individual. *FAM11A* expression was not detected in the individual with *FRAXF* full mutation. (b) Reactivation of *FAM11A* expression using 5-azadeoxycytidine in a FRAXF normal and full mutation individual. Full mutation cell lines were either treated (T) with 5azadeoxycytidine or not treated (U) before RT–PCR, normal cell line (N) was not treated. *FAM11A* transcription is reactivated in the treated cell line. 1 kb⁺ molecular marker (GIBCO BRL) is indicated. As control *ESD* primer set was used for both experiments.

duplicated by retroposition (whereby RNA is reverse transcribed into a DNA copy and inserted into the genome) which remain active, often with amino acid changes or altered tissue expression patterns, are termed 'retropseudogenes' or 'retroxaptonuons'.¹⁸ *FAM11B* does not contain any of the *FAM11A* introns. The open reading frame (ORF) of the *FAM11B* is also 1050 bp long and encodes a protein of 350 amino acids, 88% identical to the FAM11A (Figure 4). The two genes, *FAM11A* and *FAM11B*, otherwise differ markedly in their 5' and 3' untranslated regions (results not shown).

Conservation of *FAM11A* in *M. musculus*, *D. melanogaster* and *C. elegans*

Database searches of mouse ESTs revealed presence of two clusters, Mm.56701 (87 ESTs) and Mm.24569 (55 ESTs). Comparison with the *FAM11A* gene and *FAM11B* retropseudogene sequences showed that the mouse Unigene cluster Mm.24569 corresponds to the mouse orthologue of *FAM11A* and Mm.56701 to the mouse retropseudogene. This is further supported by matches against the mouse genomic sequences, AL672026 and AC100135 (X-chromosome genomic sequence) and AC100024 (mouse

I.S.FAM11A	MNLRGLFQDF	NPSKFLIYAC	LLLFSVLLAL	RLDGIIQWSY	WAVFAPIWLW	50
1.m.Famlla	MNLRGLFQDF	NPSKFLIYAC	LLLFSVLLAL	RLDGIIQWSY	WAVFAPIWLW	50
I.S.FAM11B	MNPRGLFQDF	NPSKFLIYAC	LLLFSVLLPL	RLDGIIQWSY	WAVFAPIWLW	50
1.m.Faml1b	MNPRGLFQDF	NPSKFLIYAC	LLLFSVLLPL	RLDGIIQWSY	WAVFAPIWLW	50
I.s.FAM11A	KLMVIVGASV	GTGVWARNPQ	YRAEGETCVE	FKAMLIAVGI	HLLLLMFEVL	100
1.m.Famlla	KLMVIVGASV	GTGVWARNPQ	YRAEGETCVE	FKAMLIAVGI	HLLLLMFEVL	100
I.s.FAM11B	KLLVIVGASV	GAGVWARNPR	YRTEGEACVE	FKAMLIAVGI	HLLLLMFEIL	100
1.m.Fam11b	KLLVIVGASV	GAGVWARNPR	YRTEGEACVE	FKAMLIAVGI	HLLLLMFEIL	100
I.s.FAM11A	VCDRIERGSH	FWLLVFMPLF	FVSPVSVAAC	VWGFRHDRSL	ELEILCSVNI	150
1.m.Famlla	VCDRIERGSH	FWLLVFMPLF	FVSPVSVAAC	VWGFRHDRSL	ELEILCSVNI	150
I.s.FAM11B	VCDRVERGTH	FWLLVFMPLF	FVSPVSVAAC	VWGFRHDRSL	ELEILCSVNI	150
1.m.Famllb	VCDRVERGTH	FWLLVFMPLF	FVSPVSVAAC	VWGFRHDRSL	ELEILCSVNI	150
I.s.FAM11A	LQFIFIALRL	DKIIHWPWLV	VCVPLWILMS	FLCLVVLYYI	VWSVLFLRSM	200
1.m.Famlla	LQFIFIALRL	DKIIHWPWLV	VCVPLWILMS	FLCLVVLYYI	VWSVLFLRSM	200
I.s.FAM11B	LQFIFIALRL	DRIIHWPWLV	VFVPLWILMS	FLCLVVLYYI	VWSLLFLRSL	200
1.m.Fam11b	LQFIFIALRL	DRIIHWPWLV	VEVPLWILMS	FLCLVVLYYI	VWSLLFLRSL	200
I.S.FAM11A	DVIAEQRRTH	ITMALSWMTI	VVPLLTFEIL	LVHKLDGHNA	FSCIPIFVPL	250
1.m.Famlla	DVIAEQRRTH	ITMALSWMTI	VVPLLTFEIL	LVHKLDGHNA	FSCIPIFVPL	250
I.S.FAM11B	DVVAEQRRTH	VTMAISWITI	VVPLLIFEVL	LVHRLDDHNT	FSYISIFIPL	250
1.m.Fam11b	DVVAEQRRTH	VTMAISWITI	VVPLLIFEVL	LVHRLDDHNT	FSYISIFIPL	250
I.s.FAM11A	WLSLITLMAT	TFGQKGGNHW	WFGIRKDFCQ	FLLEIFPFLR	EYGNISYDLH	300
1.m.Famlla	WLSLITLMAT	TFGQKGGNHW	WFGIRKDFCQ	FLLEIFPFLR	EYGNISYDLH	300
I.s.FAM11B	WLSLLTLMAT	TFRRKGGNHW	WFGIRRDFCQ	FLLEVFPFLR	EYGNISYDLH	300
1.m.Famllb	WLSLLTLMAT	TFRRKGGNHW	WFGIRRDFCQ	FLLEVFPFLR	EYGNISYDLH	300
I.s.FAM11A	hed n eeteet	PVPEPPKIAP	MFRKKARVVI	TQSPGKYVLP	PPKLNIEMPD	350
1.m.Famlla	HEDSEETEET	PVPEPPKIAP	MFRKKARVVI	TQSPGKYVLP	PPKLNIEMPD	350
I.S.FAM11B	HEDSEDAEDA	SVSEAPKIAP	MFGKKARVVI	TQSPGKYVPP	PPKLNIDMPD	350
1.m.Famllb	HEDSEDAEDA	SVSEAPKIAP	MFGKKARVVI	TQSPGKYVPP	PPKLNIDMPD	350

Figure 4 ClustalW multiple protein alignment of the human and mouse FAM11A and FAM11B proteins. Amino acids shaded in black are those that differ from the consensus. Exon/intron boundaries are indicated by arrowheads. All exon/exon boundaries were 100% conserved between the human and mouse genes.

chromosome 1, unplaced), respectively. Interestingly, in mouse (likewise in man), both gene *Fam11a* and retropseudogene *Fam11b* are transcribed. ClustalW alignment of the putative human and mouse gene and retropseudogene proteins is shown in Figure 4. While the human and mouse proteins differ in only one amino acid, the retropseudo-gene-encoded proteins are 100% identical. This high conservation of both gene and retropseudogene encoded proteins between man and mouse provides an indication of an important, conserved function. *FAM11A* appears to be highly conserved, with orthologues in other species including *D. melanogaster* (AAF48931, 36% similarity) and *C. elegans* (CAB55145, 26% similarity).

1

I N I

Discussion

Expansion of *FRAXF* was initially described in a family with developmental delay.¹ Since then it was reported only in a small number of additional cases.³ In contrary to the other two more proximal Xq27.3-q28 folate sensitive fragile sites, *FRAXA* and *FRAXE*, *FRAXF* was considered benign. This was primarily because of the absence of any *FRAXF* associated phenotype, but also because of a failure to identify any *FRAXF* associated gene(s).^{3,4}

We report here on the identification of a novel gene, *FAM11A*, originating from the *FRAXF* CpG island and

containing the *FRAXF* CGG repeat in its 5' untranslated region. Similar to the other folate sensitive fragile site associated genes, ie *FMR1/FRAXA*,¹⁹ *FMR2/FRAXE*^{8,9} and *FMR3/FRAXE*,²⁰ the *FRAXF* CGG repeat is transcribed (but not translated). We have also demonstrated that the paradigm of expansion-methylation-transcription silencing seen for the other folate sensitive fragile site associated genes is seen also for the *FAM11A* gene at *FRAXF*. However, as only one *FRAXF* full mutation was available for testing this paradigm needs to be further confirmed on additional cases. It also remains to be tested whether *FRAXF* CGG expansion and transcriptional silencing of *FAM11A* can be associated with developmental delay as seen in some *FRAXF* families studied.

At least for *FRAXE* and *FMR2*, there are now several families known where full mutation does not cause developmental delay (IQ < 70) (for review see Gécz^{21}). When this model is applied for *FRAXF*, we may speculate that absence of *FAM11A* does not necessarily lead to developmental delay, but might in combination with other, yet unknown factors. Otherwise, the lack of the FAM11A protein is either disposable (redundant) or its function can be complemented by the highly similar chromosome 2 retropseudogene product, FAM11B. Comparison of FAM11A and FAM11B proteins shows 88% identity. More

interestingly, when respective gene and retropseudogene proteins are compared between man and mouse, 99.7% (one amino acid difference) and 100% conservation is detected. There are numerous retropseudogenes in the human genome, some of which are transcribed.^{22,23} Their function is not known, but it has been suggested that they may be acquired during evolution as novel genes.^{18,23,24} Their protein products may even be functionally indistinguishable from the original gene.

In summary we have identified the gene *FAM11A* associated with the *FRAXF* fragile site. Its transcription is extinguished by the *FRAXF* full mutation. In the absence of any *FRAXF* associated phenotype we can only speculate about the normal role of this gene.

Acknowledgements

We thank the members of the family studied for participation, and C Derwas and S MacDonald for help with cell culture. This work was supported by the National Health and Medical Research Council of Australia.

References

- 1 Hirst MC, Barnicoat A, Flynn G *et al*: The identification of a third fragile site, *FRAXF* in Xq27–q28 distal to both *FRAXA* and *FRAXE*. *Hum Mol Genet* 1993; **2**: 197–200.
- 2 Romain DR, Chapman CJ: Fragile site Xq27.3 in a family without mental retardation. *Clin Genet* 1992; **41**: 33–35.
- 3 Parrish JE, Oostra BA, Verkerk AJ *et al*: Isolation of a GCC repeat showing expansion in *FRAXF*, a fragile site distal to *FRAXA* and *FRAXE*. *Nature Genet* 1994; 8: 229–235.
- 4 Ritchie RJ, Knight SJ, Hirst MC *et al*: The cloning of *FRAXF*: trinucleotide repeat expansion and methylation at a third fragile site in distal Xqter. *Hum Mol Genet* 1994; **3**: 2115–2121.
- 5 Kremer EJ, Pritchard M, Lynch M *et al*: Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. *Science* 1991; **252**: 1711–1714.
- 6 Oberle I, Rousseau F, Heitz D *et al*: Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science* 1991; **252**: 1097–1102.
- 7 Yu S, Pritchard M, Kremer E *et al*: Fragile X genotype characterized by an unstable region of DNA. *Science* 1991; **252**: 1179–1181.
- 8 Gécz J, Gedeon AK, Sutherland GR, Mulley JC: Identification of the gene *FMR2*, associated with *FRAXE* mental retardation. *Nat Genet* 1996; **13**: 105–108.

- 9 Gu Y, Shen Y, Gibbs RA, Nelson DL: Identification of *FMR2*, a novel gene associated with the *FRAXE* CCG repeat and CpG island. *Nat Genet* 1996; **13**: 109–113.
- 10 Holden JJ, Walker M, Chalifoux M, White BN: Trinucleotide repeats at the *FRAXF* locus: frequency and distribution in the general population. *Am J Med Genet* 1996; **64**: 424–427.
- 11 Vianna-Morgante AM, Mingroni-Netto RC, Barbosa AC, Otto PA, Rosenberg C: *FRAXF* in a patient with chromosome 8 duplication. *J Med Genet* 1996; 33: 611–614.
- 12 Holden JJ, Wing M, Chalifoux M *et al*: Lack of expansion of triplet repeats in the *FMR1*, *FRAXE*, and *FRAXF* loci in male multiplex families with autism and pervasive developmental disorders. *Am J Med Genet* 1996; **64**: 399–403.
- 13 Lindsay S, Splitt M, Edney S *et al*: PPM-X: a new X-linked mental retardation syndrome with psychosis, pyramidal signs, and macroorchidism maps to Xq28. *Am J Hum Genet* 1996; **58**: 1120–1126.
- 14 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. J Mol Biol 1990; 215: 403-410.
- 15 Gécz J, Bielby S, Sutherland GR, Mulley JC: Gene structure and subcellular localization of *FMR2*, a member of a new family of putative transcription activators. *Genomics* 1997; 44: 201–213.
- 16 Chiurazzi P, Pomponi MG, Pietrobono R *et al*: Synergistic effect of histone hyperacetylation and DNA demethylation in the reactivation of the *FMR1* gene. *Hum Mol Genet* 1999; 8: 2317–2323.
- 17 Shapiro MB, Senapathy P: RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* 1987; 15: 7155–7174.
- 18 Brosius J, Gould SJ: On 'genomenclature': a comprehensive (and respectful) taxonomy for pseudogenes and other 'junk DNA'. Proc Natl Acad Sci USA 1992; 89: 10706–10710.
- 19 Verkerk AJ, Pieretti M, Sutcliffe JS *et al*: Identification of a gene (*FMR-1*) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* 1991; **65**: 905–914.
- 20 Gécz J: FMR3 is a novel gene associated with FRAXE CpG island and transcriptionally silent in *FRAXE* full mutations. *J Med Genet* 2000; **37**: 782–784.
- 21 Gécz J: The *FMR2* gene, *FRAXE* and non-specific X-linked mental retardation: clinical and molecular aspects. *Ann Hum Genet* 2000a; **64**: 95–106.
- 22 Venter JC, Adams MD, Myers EW et al: The sequence of the human genome. Science 2001; 291: 1304-1351.
- 23 Brosius J, Venter JC, Adams MD *et al*: Many G-protein-coupled receptors are encoded by retrogenes. *Trends Genet* 1999; **15**: 304–305.
- 24 Kuryshev VY, Skryabin BV, Kremerskothen J *et al*: Birth of a gene: locus of neuronal BC200 snmRNA in three prosimians and human BC200 pseudogenes as archives of change in the Anthropoidea lineage. *J Mol Biol* 2001; **309**: 1049–1066.