J Med Genet 1991; 28: 818–823

Original articles

Fragile X syndrome: genetic localisation by linkage mapping of two microsatellite repeats *FRAXAC1* and *FRAXAC2* which immediately flank the fragile site

R I Richards, K Holman, H Kozman, E Kremer, M Lynch, M Pritchard, S Yu, J Mulley, G R Sutherland

Abstract

We report the genetic localisation of the fragile site at Xq27.3 associated with fragile X syndrome. The position of the fragile site within the multipoint linkage map was determined using two polymorphic microsatellite AC repeat markers *FRAXAC1* and *FRAXAC2*. These markers were physically located within 10 kilobases and on either side of the p(CCG)$_n$ repeat responsible for the fragile site. *FRAXAC1* has five alleles with heterozygosity of 44% and is in strong linkage disequilibrium with *FRAXAC2* which has eight alleles and a heterozygosity of 71%. No recombination was observed either between these markers in 40 normal CEPH pedigrees or with the fragile X in affected pedigrees. These markers provide the means for accurate diagnosis of the fragile X genotype in families by rapid polymerase chain reaction analysis and were used to position the fragile X within the multipoint map of the X chromosome to a position 3·7 cM distal to *DXS297* and 1·2 cM proximal to *DXS296*.

Fragile X syndrome is the most common familial form of mental retardation. It is associated with the expression under certain specific culture conditions of a fragile site at Xq27.3. This cytogenetic association has been used diagnostically; however, incomplete penetrance and the bizarre segregation pattern of the disorder have necessitated the use of linkage analysis to determine accurately carrier status using restriction fragment length polymorphism (RFLP) markers.[1,2] The discovery of length polymorphism of AC repeat microsatellites,[3] which are in general more informative than RFLPs, has enabled the recent characterisation and application of these markers to fragile X diagnosis.[4] AC repeats were identified in the DNA sequences which define the *DXS292* and *DXS297* loci, which are 7 cM and 4 cM, respectively, proximal to the fragile X.

Recently, cloned DNA sequences which span the fragile site at Xq27.3 have been isolated[5] and an unstable region of DNA which segregates with the fragile X genotype has been identified.[6] A DNA probe, pfxa3, adjacent to the unstable p(CCG)$_n$ repeat, is able to detect the fragile X genotype directly in fragile X syndrome pedigrees.[6,7] A correlation between the length of the unstable sequence and phenotype is evident permitting reliable and accurate prenatal diagnosis of fragile X phenotype by Southern analysis.[8]

The exceptionally high GC content of this element and its flanking DNA sequences has prohibited the development of a simple PCR based diagnostic test for amplification of the p(CCG)$_n$ repeat. The Southern blot hybridisation using pfxa3, while accurate in determining genotype, is a relatively slow procedure, particularly for prenatal diagnosis. Genotype can be determined just as accurately by linkage analysis using PCR based markers which show no recombination with the disease locus and which can be determined much more rapidly. Half of the at risk pregnancies will be negative for fragile X genotype and PCR based diagnosis will quickly alleviate parental anxiety in those cases.

We have therefore undertaken characterisation of two AC repeat sequences in the immediate vicinity

Department of Cytogenetics and Molecular Genetics, Adelaide Children's Hospital, 72 King William Road, North Adelaide, South Australia 5006, Australia.
R I Richards, K Holman, H Kozman, E Kremer, M Lynch, M Pritchard, S Yu, J Mulley, G R Sutherland

Correspondence to Dr Richards.

Received for publication 10 July 1991.
Accepted for publication 26 July 1991.

of the fragile X site p(CCG)$_n$ unstable element. These markers allow rapid exclusion of the fragile X genotype and, in conjunction with the pfxa3 analysis, may be incorporated into a rational diagnostic strategy for fragile X syndrome prenatal diagnosis. The remaining pregnancies, positive for fragile X genotype, must undergo diagnosis for phenotype using pfxa3.[68]

Materials and methods

The yeast artificial chromosome XTY26 containing 275 kb of human DNA has previously been shown to span the fragile site at Xq27.3 which segregates with fragile X syndrome.[5] A saturation library of 108 lambda subclones was derived from XTY26 and a contig of overlapping DNA sequences established between the closest known proximal (VK-16) and distal (Alu2) markers to the fragile X site.[6]

Synthetic poly (AC.GT) (Pharmacia) was radioactively labelled with α-^{32}P-dCTP in a random primed reaction (Multiprime, Amersham). AC repeat containing DNA clones was identified by hybridisation to this probe in 0·5 mol/l sodium phosphate, pH 7·0, 7% SDS (without carrier DNA) at 65°C for 16 hours and by washing at 65°C for one hour in 2 × SSC. Clones tested constituted the 108 lambda subclones of the yeast artificial chromosome, XTY26.[56]

DNA from positive clones was digested with either *Hae*III, *Sau*3AI, *Hin*PI, *Hpa*II, *Rsa*I, *Hin*fI, or *Taq*I, electrophoresed on 1·4% agarose gels, blotted onto nylon membranes (GeneScreen Plus, NEN-Dupont), and probed with ^{32}P-poly (AC.GT) as above. Digests which gave a hybridising fragment of less than 600 base pairs were chosen for subcloning into M13mp18 for sequence analysis. Synthetic oligodeoxyribonucleotide primers suitable for PCR were designed from apparently unique sequences flanking the microsatellite AC repeats. Length polymorphism of the AC repeats was typed in a PCR using the reaction conditions previously described,[49] except for the addition of 10 μCi of α-^{32}PdCTP to each reaction.

PCR incubations were performed in 10 μl volumes in a Perkin Elmer-Cetus thermal cycler for 10 cycles at 94°C for 60 seconds, at 60°C for 90 seconds, and at 72°C for 90 seconds, followed by 25 cycles at 94°C for 60 seconds, at 55°C for 90 seconds, and at 72°C for 90 seconds. The volume was adjusted to 40 μl with formamide loading buffer (95% formamide, 1 mmol/l EDTA, 0·01% bromophenol blue, 0·01% xylene cyanol). After denaturation at 90°C for three minutes, 2·5 μl aliquots of each reaction mixture were subjected to electrophoresis in 6% polyacrylamide denaturing (7 mol/l urea) gels. Genotypes were determined after autoradiography for 18 hours.

Multipoint analysis was based on genotypes of each marker in the 40 large kindred pedigrees of the Centre d'Etude du Polymorphisme Humain (CEPH) and was carried out by using the LINKAGE (version 4.9) package for use with CEPH three generation families.

Results

IDENTIFICATION OF AC REPEATS AT *FRAXA*

The library of XTY26 subclones was screened for the presence of AC repeats by hybridisation to ^{32}P labelled synthetic poly (AC.GT). Ten positive clones were assigned to each of three groups by virtue of their position in the XTY26 contig (table 1). That each group contained unique AC repeats was determined by the size of their *Hae*III and *Sau*3AI restriction fragments which hybridised to ^{32}P-poly (AC.GT) (not shown).

DNA from one member of each group was digested with a four base pair enzyme which gave a ^{32}P-poly (AC.GT) hybridising restriction fragment <600 base pairs, suitable for subcloning and sequencing in M13. For lambda 12 this digest was with *Sau*3AI which gave an approximately 280 base pair hybridising fragment (table 2A). For lambda 25 a *Hae*III digest gave an approximately 600 base pair hybridising fragment, part of the sequence of which is shown in table 2B. The initial sequence determined from the universal sequencing primer allowed accurate reading up to and including the AC repeat, but not beyond, to allow for reverse PCR primer design and synthesis. To obtain sequence from this side of the repeat the forward PCR primer was used to prime the sequencing reaction. For lambda 26 an *Rsa*I digest gave a 200 base pair hybridising fragment, the sequence of which is shown in table 2C.

The derived sequences were then used to design synthetic oligodeoxyribonucleotide primers suitable for PCR analysis of length variation in the AC repeat sequences. Sequences were chosen on the basis of their apparent uniqueness, their 50% GC composition, and their lack of consecutive G residues, which appear to interfere with chemical synthesis of oligodeoxyribonucleotides (unpublished observations).

Table 1 Groups of AC repeat containing lambda clones derived from the fragile X containing YAC, XTY26.

	Group		
	I	II	III
Clone no	16	5	26
	24	8	51
	25	12	
	27	80	

Table 2 AC/GT repeat sequences determined from lambda 12 (A), lambda 25 (B), and lambda 26 (C). Sequence recognition sites for the enzymes used to isolate the respective fragments are inferred. Sequences used to design primers for genotyping polymorphism of repeat length are underlined. PCR products are shown in bold.

A

GATCTAATCA ACATCTATAG ACTTTATTGT GTGTGTGTGT GTGTGTGTGT GTATGTGTGT
GTCAGTCTCA CTCTGTCACT CAGGCTTGGA GTGCAGTGGG CAATCTCTGC TCACTGCAAC
100
CTCGCCTCCC AGCTTCAAGT GACTCTCATC ATGCCTCAGC CTCCTGAGTA GCTGGGATTA
CAGGCATGCA CCACCACACC CAGCTAATTT TTTGCATTTT TAGTAGAGTC GGCATTTCAC
200
TATGTTGGCC AGGCTGGTCT CGAACTTCTG GCCTCAAGTG ATC

B

GGCCCTAATC AGATTTCCAC AAATTCTGAC TTAATATTTG CCCGCTTATA TAACAGCTCT
TCTTTAACAA AAACAAGTAC TTTTCTCAAT AGAATTTTAC TAAGAAAGCT CTTTAGTAAA
ACATCGACAT TATACATACA ACATATCTCA GTATCTGCTG ATGAAGAACA CCAAAAAGAA
CCCAGATGTG **ACTGCTCCGG AAGTTGAATC CTCAGTATTT TTGCAAAGTT TGTCTTTCAG**
TATTTTATTT GTGTGTGTGT GTGTGTGTGT GTGTGTGTCT ATATATATAT ATTTTTTTTT
100
TTTTTTTTAA AGACAGGATC TCACTCTGTC ACCTAGGCTG GAGTGCAGTG CATGATCATG

C

GTACTGTATC AGTTATAACC CTATGTGTGT GTGTGCGTGT GTGTGTGTGT GTGTATGCAT
ACCCAAGACT TATCTTATAC AGGTATGCCT TGTTTTATTG CACTTTGCAA ATACTGCATT
100
TTTTTCAAAT TGAAGGTTTG TGGAAACCTT TTTTTTGAGC AATTCTGTAG TGCCATTTTT
TTCAACGGCA TGTGTAC

PHYSICAL LOCATION OF *FRAXA* AC REPEATS

Each of the three AC repeats was shown to represent each of the three groups of lambda subclones of XTY26 DNA by using the synthetic oligodeoxyribonucleotides to prime PCR reactions with each of the lambda DNAs as template. In each case, product of the expected size was synthesised (data not shown).

The location of the group I and group II AC repeats was determined, relative to the unstable element $p(CCG)_n$ which characterises the fragile site, by probing *Eco*RI digests of these clones with ^{32}P labelled poly (AC.GT). The position of the hybridising fragments is indicated in fig 1.

The markers from each microsatellite were subsequently termed *FRAXAC1* (from lambda 12), *FRAXAC2* (from lambda 25), and *FRAXAC3* (from lambda 26).

POLYMORPHISM OF *FRAXA* AC REPEATS

The observed heterozygosity of *FRAXAC3* in 18 unrelated females was only 16% and so the characterisation of this marker was not pursued further. The observed heterozygosities of *FRAXAC1* and *FRAXAC2* were found to be 45% and 80%, respectively, in 40 unrelated females. However, none of the unrelated females who were homozygous for *FRAXAC2* was heterozygous for *FRAXAC1* and so the combined observed heterozygosity was also 80%. This indicates linkage disequilibrium between the two markers. Expected heterozygosities, based on observed allele frequencies, were 44% and 71% for *FRAXAC1* and *FRAXAC2* respectively (table 3). Apart from diagnosis, *FRAXAC2* will be useful as a marker for construction of the index map of the X chromosome which will be based on markers with minimum heterozygosity of 70%. Codominant men-

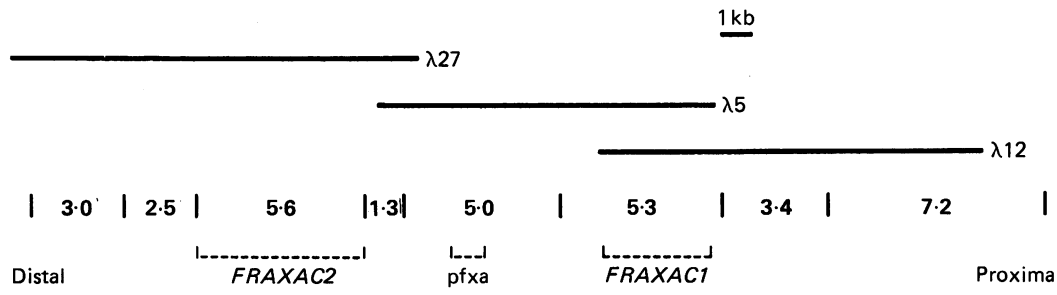

Figure 1 Location of AC repeat sequences relative to the fragile site. Lambda clones from the XTY26 contig across the fragile site[6] are shown as is the location of the fragile site p(CCG)n repeat (pfxa), the exact position of which was previously determined by Kremer et al.[5] The length in kilobases of EcoRI fragments across the contig are indicated and those which hybridise to poly (AC.GT) are identified.

Table 3 Allele frequencies of AC repeat markers at FRAXA.

Marker	Allele	(AC)n	Allele frequency	% heterozygosity	
				Observed	Expected*
FRAXAC1	A	19	0·063	45	44
	B	18	0·012		
	C	17	0·725		
	D	16	0·188		
	E	15	0·012		
FRAXAC2	H	23	0·009	80	71
	G	19	0·018		
	F	18	0·073		
	E	17	0·477		
	D	16	0·193		
	C	15	0·037		
	B	14	0·110		
	A	13	0·083		

*Based on observed allele frequencies.

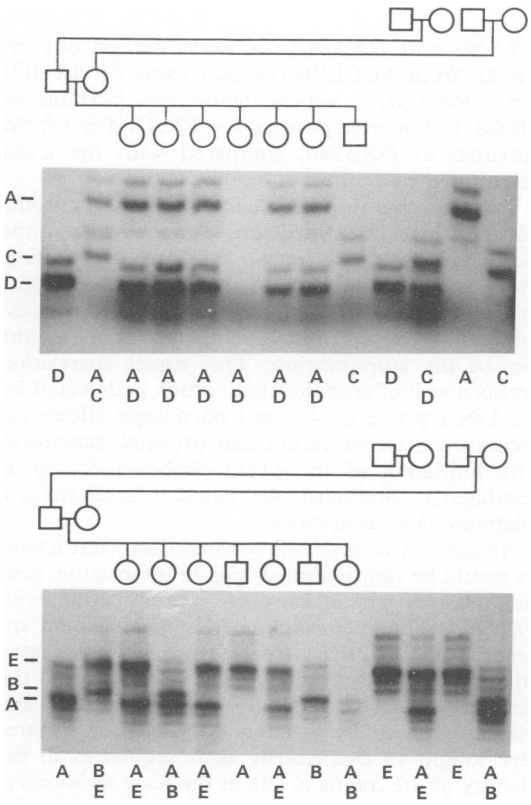

Figure 2 AC repeat genotypes for FRAXAC1 in CEPH pedigree 1420 (top) and FRAXAC2 in CEPH number 12 (bottom). The position of the major band for each allele is indicated alongside the autoradiograph. Allele designations are indicated below each lane.

delian inheritance was observed in each pedigree analysed (examples are shown in fig 2).

FRAXA LOCUS IN FRAGILE X PEDIGREES

Fragile X affected pedigrees, which had previously been shown to have recombinants in the vicinity of the fragile site, were genotyped with *FRAXAC2*. One subject in each of three pedigrees was a recombinant with *IDS* and another in a fourth pedigree was a recombinant with *DXS297* (VK23AC). Three of these pedigrees were informative with *FRAXAC2* and no recombination was found between this marker and the fragile X genotype (as determined by pfxa3 hybridisation). Analysis with *FRAXAC1* was not undertaken because of the high degree of linkage disequilibrium between the two markers. Pedigrees informative for *FRAXAC2* are shown in fig 3.

MULTIPOINT LINKAGE ANALYSIS

In order to place *FRAXA* onto the background genetic map of the X chromosome, the *FRAXAC1* and *FRAXAC2* markers were genotyped from the 40 CEPH pedigrees. The location of *FRAXA* (*FRAXAC1–FRAXAC2*) within the multipoint linkage map is given in table 4.

When genotype analysis of the informative CEPH pedigrees was completed, only one female who was homozygous for *FRAXAC2* was found to be heterozygous for *FRAXAC1*, confirming a high degree of linkage disequilibrium.

Discussion

Suthers *et al*[1] established a background map of genetic markers using the CEPH families. Within the interval of 6 cM on the background map between *DXS296* (VK21) and *DXS297* (VK23), Suthers *et*

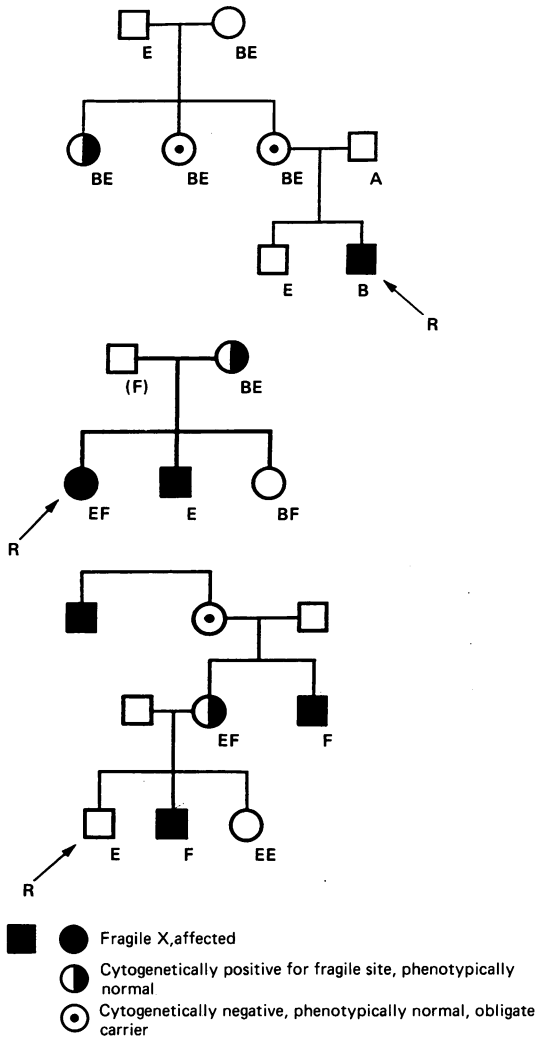

Figure 3 Fragile X pedigrees which had previously been shown to exhibit recombination between the fragile X and the flanking marker at IDS. *In each pedigree the FRAXAC2 marker cosegregates with the fragile site. The member of each pedigree which shows recombination is arrowed (R). Brackets indicate that the genotype was inferred.*

al[2] positioned the *FRAXA* locus 2 cM proximal to *DXS296* using the location score method applied to fragile X families. This location relative to the flanking markers has subsequently been applied to the estimation of risk in fragile X syndrome when diagnosis has been carried out using linked markers. The location of *FRAXA* has now been independently located within the *DXS296* and *DXS297* interval using the CEPH families. This localisation used additional marker data, the AC repeats at

Table 4 Multipoint linkage data around FRAXA.

Marker	Recombination fraction	Odds against inversion
F9		
	0·049	$2·4 \times 10^{1}$
DXS105		
	0·058	1·9
DXS98		
	0·084	$5·4 \times 10^{3}$
DXS292		
	0·032	$1·2 \times 10^{3}$
DXS369		
	0	1·0
DXS297		
	0·037	$2·6 \times 10^{6}$
FRAXAC1		
	0	1·0
FRAXAC2		
	0·012	$1·1 \times 10^{5}$
DXS296		
	0	$4·3 \times 10^{1}$
IDS		
	0·021	$3·2 \times 10^{2}$
DXS304		
	0·097	$7·7 \times 10^{11}$
DXS52		

DXS292 and *DXS297*,[4] and was carried out indirectly from the CEPH database using *FRAXAC1* and *FRAXAC2*, which define the position of *FRAXA*. The new position for *FRAXA* is 1·2 cM proximal to *DXS296*, compared with the 2 cM determined by Suthers *et al.*[2]

The mutation responsible for fragile X syndrome has now been identified and characterised at the molecular level. An unstable element $p(CCG)_n$, which is present at from 15 to 65 copies in normal subjects,[7] is amplified in both carriers and affected subjects, phenotypic status being determined by the size of the amplification. This direct correlation between size of amplification (which is detected by the DNA probe pfxa3) and phenotype allows for accurate carrier detection and prenatal diagnosis.[8] The limitation of the pfxa3 probe is that it is restricted to Southern blot based hybridisation, a relatively slow technology.

An alternative approach to rapid diagnostic analysis would be tightly linked, highly informative genetic markers which show no recombination with *FRAXA*. The microsatellite AC repeats would appear to fulfil both these criteria since they are located on average every 50 kilobases or so throughout the genome, they have generally high PIC values, and they can be rapidly genotyped by PCR. We therefore sought to characterise such sequences in the vicinity of the fragile X site in the hope of identifying any which might make suitable markers for fragile X pedigree analysis. The markers *FRAXAC1* and *FRAXAC2*, with observed heterozygosities of 45% and 80% respectively, fulfil these requirements. Their physical location, flanking and

within 10 kilobases of the fragile X site, make them ideal genetic markers for fragile X genotyping since recombination between either marker and the fragile site ought to be exceptionally rare. The high degree of linkage disequilibrium between the two markers indicates this but makes the less informative of the two (*FRAXAC1*) redundant for linkage analysis. Initial attempts to multiplex the two markers were therefore abandoned. The high level of linkage disequilibrium between these two markers which flank the fragile site confirms that the fragile site is not an area of high meiotic recombination. Verkerk *et al*[10] have recently reported the characterisation of an additional AC repeat marker (*DXS548*) 150 kb proximal to *FRAXA*, with an observed heterozygosity of >80% in fragile X families. Given that *DXS548* and *FRAXAC2* do not exhibit linkage disequilibrium then a multiplex reaction of these two markers will enable application of linkage analysis to virtually all fragile X pedigrees.

Together with the pfxa3 hybridisation probe, the *FRAXAC2* marker can now be incorporated into a rational approach to prenatal diagnosis in fragile X pedigrees. This involves analysis of chorionic villus sample (CVS) DNA with the AC repeat marker at Xq27.3 to determine the *FRAXA* genotype, followed by Southern blots with pfxa3 as probe to detect amplification of the $p(CCG)_n$ repeat in pregnancies positive for fragile X genotype. The initial microsatellite results will allow rapid exclusion in half of the at risk pregnancies whereas the prediction of phenotype for subjects with the *FRAXA* genotype will be subsequently determined by the size of pfxa3 hybridising fragments.[6 8]

RIR wishes to thank Shelley Richards for support and encouragement during the course of these studies. This work was supported in part by grants from the Channel 7 Children's Research Foundation of South Australia Inc and from the National Health and Medical Research Council of Australia.

1 Suthers GK, Oberlé I, Nancarrow J, *et al*. Genetic mapping of new RFLPs at Xq27–q28. *Genomics* 1991;**9**:37–43.
2 Suthers GK, Mulley JC, Voelckel MA, *et al*. Genetic mapping of new DNA probes at Xq27 defines a strategy for DNA studies in the fragile X syndrome. *Am J Hum Genet* 1991;**48**:460–7.
3 Weber JL, May PE. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet* 1989;**44**:388–96.
4 Richards RI, Shen Y, Holman K, *et al*. Fragile X linked mental retardation: diagnosis using highly polymorphic microsatellite markers. *Am J Hum Genet* 1991;**48**:1051–7.
5 Kremer EJ, Yu S, Pritchard M, *et al*. Isolation of a human DNA sequence which spans the fragile X. *Am J Hum Genet* (in press).
6 Yu S, Kremer E, Pritchard M, *et al*. The fragile X genotype is characterized by an unstable region of DNA. *Science* 1991;**252**:1179–81.
7 Kremer EJ, Pritchard M, Lynch M, *et al*. DNA instability at the fragile X maps to a trinucleotide repeat sequence p(CCG)n. *Science* 1991;**252**:1711–4.
8 Yu S, Loesch D, Hay D, *et al*. Fragile X syndrome phenotype correlates with unstable element genotype: a unique genetic phenomenon. (Submitted.)
9 Richards RI, Holman K, Shen Y, Harley H, Brook D, Shaw D. Human glandular kallikrein genes: genetic and physical mapping of the *KLK1* locus using a highly polymorphic microsatellite marker. *Genomics* 1991;**11**:77–82.
10 Verkerk AJMH, 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**:905–14.