The effect of FMR1 CGG repeat interruptions on mutation frequency as measured by sperm typing

Catherine B Kunst, Esther P Leeflang, Jane C Iber, Norman Arnheim, Stephen T Warren

Abstract

Fragile X syndrome results from the unstable expansion of a CGG repeat within the FMR1 gene. Three classes of FMR1 alleles have been identified, normal alleles with 6-60 repeats, premutations with 60-200 repeats, and full mutations with >230 repeats. Premutations are exquisitely unstable upon transmission. Normal alleles, while generally stable upon transmission, are thought to have different intrinsic mutation frequencies, such that some normal alleles may be predisposed towards expansion while others may be more resistant to such change. One variable that may account for this difference is the occurrence of one or more AGG triplets punctuating the normal CGG repeat. The AGG interruptions lead to alleles that have equivalent overall length but different lengths of perfect repeats. To test the influence of the length of perfect repeats on stability, we examined the CGG repeat of single sorted sperm from two males, each with 39 total repeats, but distinct AGG interruption patterns. Sorted sperm of each donor showed ~15% variation in repeat length, consistent with previous studies of sorted sperm at other triplet repeat loci. However, when discounting the majority variation of ±1 repeat, the male with 29 perfect repeats showed 3% expansion changes while the donor with only 19 perfect repeats had none (<0.9%). Moreover, >90% of all variant sperm, including all those observed with expansions, showed expansion or contraction of the 3' end of the repeat array. These data are consistent with the hypothesis that perfect repeat tracts influence the repeat stability and that changes of the FMR1 repeat exhibit polarity.

Medical Institute and Departments of Biochemistry and Pediatrics, Emory University School of Medicine, Atlanta, GA 30322, USA C B Kunst* J C Iber S T Warren

Howard Hughes

Molecular Biology Program, University of Southern California, Los Angeles, CA 90089, USA E P Leeflang N Arnheim

Correspondence to: Dr Warren, Howard Hughes Medical Institute, Emory University School of Medicine, Room 4035 Rollins Research Center, 1510 Clifton Road, Atlanta, Georgia 30322, USA.

*Present address: Eleanor Roosevelt Institute, Denver, CO 80206, USA.

Received 19 February 1996 Revised version accepted for publication 26 March 1997 (J Med Genet 1997;34:627–631)

Keywords: FMR1; CGG repeat; sperm typing

Fragile X syndrome, an X linked dominant disorder with reduced penetrance, is a frequent cause of mental retardation.¹² The molecular basis of fragile X syndrome has been determined to be the loss of function of the FMR1 gene located at Xq27.3.³⁻⁵ The loss of function is unusual as it is the consequence of a CGG repeat expansion in the 5' untranslated portion of FMR1.⁶ Penetrant males with fragile X syndrome have an average of ~800 repeats which are associated with aberrant methylation of the FMR1 gene and transcriptional silencing.⁷⁻⁹ The absence of the encoded nucleocytoplasmic RNA binding protein, FMRP, is responsible for the phenotype.¹⁰⁻¹²

Non-penetrant carriers of fragile X syndrome carry an unmethylated and expressed FMR1 gene with between ~60 and 200 repeats that is referred to as a premutation.^{13 14} Premutations, when transmitted from either sex, are unstable leading to offspring with repeat lengths distinct from the transmitting parent. Although male transmission of a premutation leads to only premutation offspring, female transmission can result in either premutation offspring or children with full mutations, carrying an excess of 230 repeats.^{14 15} The probability of a child having a full mutation rather than one bearing a premutation is directly related to the maternal premutation repeat length.¹³⁻¹⁵ Therefore, it has been proposed that longer tracts of perfect CGG repeats result in a proportionately higher likelihood of DNA slippage or other undefined molecular event expanding the repeat allele.^{16 17}

Studies of the normal FMR1 repeat are consistent with the concept that perfect repeat arrays promote length variation. Normal FMR1 alleles are polymorphic in length as well as content. The overall repeat array varies from 6 to ~60 units with a mean of 30 repeats and can include AGG one to three interruptions.¹⁶ ¹⁸⁻²¹ These interruptions are commonly situated at every 10th repeat from the 5' end of the array. Therefore, tracts of >10 continuous CGG repeats are most often found towards the 3' end of the array, which is the same region that expands in premutations. Studies of nearby flanking polymorphisms have defined haplotypes which exhibit linkage disequilibrium among normal and fragile X chromosomes, suggesting that some normal haplotypes may be predisposed towards instability.22 Indeed, these haplotypes are those in which the normal FMR1 alleles contain fewer interruptions and therefore longer perfect repeat runs.^{16 18} Conversely, haplotypes displaying negative linkage disequilibrium are associated with normal alleles containing two or more interruptions and consequently shorter perfect repeats. Although there is evidence for cis or trans acting elements affecting repeat stability,23 the AGG interruptions and resulting length variation of perfect CGG repeat tracts is the most clearly defined influence to date.

Direct measurement of the effect of the purity of CGG repeat tracts on stability has been difficult and is limited to population or family studies. Here we test the stability of the

Materials and methods

Individual sperm from the chosen donors were sorted into 96 well dishes by published methods.²⁴ Sperm were lysed and neutralised in these dishes and were used without further purification for PCR amplification. After briefly centrifuging the dishes to remove condensation, 3 µl of additional lysis solution (200 mmol/l KOH, 50 mmol/l DTT) and 3 µl of additional neutralisation solution (900 mmol/l Tris, pH 8.3, 300 mmol/l KCl, 200 mmol/l HCl) were added to each well. These samples were transferred to UV irradiated PCR tubes which contained UV irradiated oil. To prevent contamination, all solutions (except primers) were UV irradiated for these experiments. The initial PCR amplification was carried out as previously described²⁵ in a total reaction volume of 30 µl except that a Taq polymerase/Pfu polymerase blend (16:1) was used for the PCR amplifications to minimise repeat slippage and glycerol was added to a final concentration of 10%. Because Tris and KCl are in the lysis and neutralisation buffers, no additional PCR buffer was used in the primary round of amplification, although MgCl₂ was added to a final concentration of 1.25 mmol/l. As described previously,²⁵ 1 µl of the primary amplification product was used for the second round of PCR amplification and samples were separated on a 6% denaturing polyacrylamide gel containing 32% deionised formamide. Repeats were sized against size markers and FMR1 repeats of known sequence. All variant PCR products were directly sequenced by automated fluorescent methods as previously described.16 25

Results

Examination of 155 anonymous semen donors identified two males each with 39 repeats at FMR1 but with distinct AGG interruption patterns (at a frequency similar to previous population screens¹³⁻¹⁶). Donor 727 had a single AGG located at triplet position 10 resulting in a tract of 29 perfect CGG repeats. In contrast, donor 791 had two AGG interruptions at triplet positions 10 and 20 resulting in only 19 perfect CGG repeats. Based upon previous population studies, it has been suggested

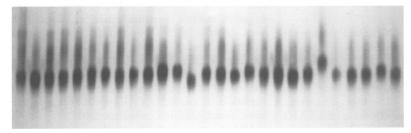


Figure 1 Autoradiograph of the FMR1 repeat of 27 PCR amplified wells, each containing sorted sperm from donor 727. All PCR products with even a slight suggestion of variation were subjected to a second round of secondary amplification, gel analysis, and DNA sequencing.

that FMR1 alleles of various lengths, each containing >24 perfect CGG repeats, may be predisposed towards expansion and comprise a pool of ancestral fragile X syndrome alleles.¹⁶ Thus, donor 727 would harbour such a predisposed allele while donor 791 would not.

Sperm of donors 727 and 791 were sorted into 96 well dishes, each containing eight empty wells as internal negative controls. The individual sperm were lysed and neutralised as previous described²⁴ and used without further purification. Using two sets of primers flanking the FMR1 repeat, a nested PCR amplification was carried out on each well with α^{32} P-dCTP added during the final set of amplifications.²⁵ The PCR products were then resolved by polyacrylamide gel electrophoresis and detected by autoradiography. As shown in fig 1, variation in repeat length of individual sperm products could be easily discerned. Since no signals were ever observed in the empty control wells, the observed bands more than likely emanated from sperm DNA amplification. From subsequent sequence analysis (below), as little as ± 1 triplet could be detected. Although some slippage during PCR is evident by the smearing of the signal, as is typical in microsatellite amplifications, at optimal exposures the predominant product was evident and scored. In selected cases of variation from the expected repeat length, reamplification of the initial PCR reactions with the nested primer set resulted in identical product lengths with the first round. It has already been shown that slippage during PCR amplification does not make a significant contribution to the detection of new allele sizes.24 Although slippage during PCR cannot be totally excluded, the overall trend of the data based on many sperm should reflect the actual germline mutation process.

From 510 wells into which sperm of donor 727 was sorted, 214 wells resulted in PCR products. This 42% well amplification agrees well with the theoretical maximum of 50% X bearing sperm. Amplification of sorted sperm of donor 791 was less efficient resulting in 119 products from the amplification of 531 wells (22% total or 44% of the X chromosomes). The level of PCR efficiency for sperm from donor 791 is within the range seen in other studies and could reflect difficulties in sperm sorting, sperm lysis, or other technical factors.²⁴ ²⁶ ²⁷

As shown in table 1, amplified sperm of both donors showed ~15% variation in length from the constitutional 39 repeats. In donor 791 nearly all the observed variation (94%) resulted from ±1 repeat change while only 63% of the variation for donor 727 resulted from ±1 repeat. Direct sequence analysis of all the variant PCR products confirmed the repeat number and showed that >90% of all variant sperm from either donor maintained the expected AGG interruption pattern. Thus, nearly all variation observed occurred in the 3' portion of the repeat array (fig 2). In donor 727 a loss or gain of a single CGG repeat in the first nine repeats (8+29 and 10+10, table 1) accounted for all changes in the AGG pattern and in donor 791 a similar single repeat loss or

Table 1 Variation of the FMR1 repeat in sorted sperm of two normal males differing in repeat sequence but not length*

Change in repeat length	Donor 727 (39 repeats; 9+29)	Repeat sequence	Donor 791 (39 repeats; 9+9+19)	Repeat sequence
+6	1 (0.5%)	45; 9+35	0	
+2	6 (2.8%)	41; 9+31	0	_
+1	11 (5.1%)	40; 9+30	8 (6.7%)	40; 9+9+20
0	182 (85%)	39; 9+29	99 (83%)	39; 9+9+19
			1 (0.8%)	39; 8+9+21
-1	9 (4.2%)	38; 9+28	8 (6.7%)	38; 9+9+18
	1 (0.5%)	38; 8+29		
-2	1 (0.5%)	37; 9+27	0	_
-3	1 (0.5%)	36; 9+26	0	-
-4	1 (0.5%)	35; 9+25	0	_
-18	1 (0.5%)	21; 10+10	0	_
-19	0	_	1 (0.8%)	20; 10+9

* Repeat sequence designated by total triplet number followed by nomenclature where a + sign represents the position of an AGG within pure tracts of CGG repeat, the length of which is designated by numbers. Thus $(CGG)_{9}AGG(CGG)_{29}$ is 39; 9+29.

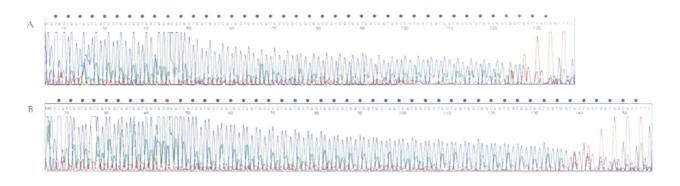


Figure 2 Sequence analysis of the FMR1 repeat from donor 727 sorted sperm. Panel A shows the representative sequence of the predominant allele with 39 repeats and an AGG interruption at position 10. Panel B shows an expanded allele with 45 repeats with retention of the AGG at position 10. Blue dots indicate CGG triplets while red dots indicate AGG interruptions; 5' to 3' polarity is left to right according to the FMR1 reading frame.

Table 2 Sorted sperm results of changes of at least two repeat units*

	Donor 727 (9+29) (n=214)	Donor 791 (9+9+19) (n=117)
Size of expansions		
2	6	0
6	1	0
Total	7	0
Size of contractions		
2	1	0
3	1	0
4	1	0
18	1	0
19	0	1
Total	4	1

* Data taken from table 1 considering changes of +1 to -1 repeat units as no change.

gain in the initial nine triplets was found (8+9+21 and 10+9, table 1). One of these variants in donor 791, the 10+9 sperm, showed the only apparent loss of an AGG interruption.

The frequency of the ± 1 repeat mutation was similar in both donors. When the data were analysed without this class of variant alleles (table 2), a difference was observed between donors 727 and 791 (p=0.05, Fisher's exact test). Seven expansions, with ≥ 2 repeat units, were detected exclusively in donor 727 in 214 sperm analysed (3%). Six expansions of 2 repeat units were observed as well as a single expansion of six triplets. All of this variation occurred within the 3' portion of the repeat. No expansions of ≥ 2 repeat units were observed in donor 791 (<1/117 or <0.9%). Contractions of ≥ 2 repeat units were also more frequent in donor 727 although both donors had one sperm each with a major deletion event of >18repeats. However, differences in the frequency of contraction were not statistically significant.

Discussion

Instability of trinucleotide repeats leading to disease causing repeat expansions remains a poorly understood process. Currently, 11 disorders resulting from trinucleotide repeat expansion have been identified in humans.²⁸²⁹ Fragile X syndrome, one of the first such mutations discovered, is among the best understood of these disorders although few clues exist to explain the mechanisms behind the expansion process. One clue relates to the number and position of AGG interruptions which define tracts of perfect CGG repeats. Kunst and Warren¹⁶ found that the most prevalent haplotypes among fragile X chromosomes are those which are commonly found on normal chromosomes with fewer interrupted FMR1 repeats and therefore longer perfect CGG tracts. Moreover, the long perfect repeats were found at the 3' portion of the array indicating that variation at the FMR1 repeat was highly polar. It was suggested that perfect CGG tracts in excess of 24 triplets may be predisposed towards expansion. Indeed, Eichler et al^{18} found that when the tract of perfect repeats exceeds 34 triplets, instability is frequent enough to observe in small nuclear families. Thus, AGG interruptions, particularly two or more, are believed to stabilise the FMR1 repeat. Loss of one or more AGG triplets, while lengthening the 3' tract of perfect CGG repeats, may enhance instability and result in alleles which may be predisposed towards frag-ile X syndrome.¹⁸⁻²¹ The notion that triplet interruption confers stability has been independently observed with the ataxin gene; the

unstable expansion of a perfect CAG trinucleotide repeat leads to spinocerebellar ataxia type 1 but the normal, stable alleles are interrupted by CAT triplets.³⁶

Sperm analysis has begun to provide insight into trinucleotide repeat instability. Zhang et al^{24} examined the CAG repeat of the androgen receptor (AR) gene in sorted sperm and found a fourfold greater mutation frequency for sperm from donors with 28-31 perfect repeats compared to those donors with only 20-22 repeats. Monckton *et al*³¹ showed, by small pool PCR, that sperm from myotonic dystrophy males with >80 perfect CTG repeats showed a very high mutation frequency with expansions as well as contractions. Similarly, studies with sorted sperm from a patient with Kennedy's disease, with 47 perfect CAG repeats at the AR locus, showed striking instability with contractions as well as more frequent expansions of the repeat tract.²⁶ Similarly, sorted sperm analysis of the CAG repeat of the Huntington's disease (HD) gene showed a clear increase in instability in sperm of donors with increasing repeat length.²⁷ While alleles with only 15-18 perfect repeats had very low sperm mutation frequencies (~0.6%), a male with 30 perfect repeats had a sperm mutation frequency of 11% and a male with 36 repeats had a mutation frequency of 53%. These studies all point to the importance of tracts of lengthy perfect repeats, greater than approximately 24 triplets, in the instability process. Indeed, for fragile X syndrome, Mornet et al³² found an increased rate of instability in sperm DNA of a male with 55 repeats (longest perfect tract of 36 triplets) compared to sperm DNA of a male with 29 repeats (longest perfect tract of nine triplets).32 Although these data support the trend of increased instability with longer perfect repeats, the mutation frequencies were considerably less than those using sorted sperm since the method of small pool PCR cannot reliably distinguish changes of <4 repeats.

The data reported above, while consistent with those of Mornet et al,³² are distinct on several levels. Sorted sperm were used, which allowed small changes of repeat units to be distinguished. This, in turn, leads to mutation frequencies higher than those of Mornet et al,³² but more similar to those of Leeflang et al^{27} using sorted sperm analysis at the HD locus (11% change of 30 perfect HD repeats compared here with 15% change of 29 perfect FMR1 repeats). Variant sperm in our study were subjected to DNA sequencing to verify the exact repeat length and structural variation incurred. The choice of two sperm donors with identical overall repeat lengths but distinct in terms of their longest perfect repeat tract allowed direct examination of the influence of AGG interspersions on FMR1 stability.

Examination of a total of 1041 wells within which individual sperm were sorted resulted in PCR products of the FMR1 repeat in 214 wells from donor 727 and 117 wells from donor 791. Both donors, each with 39 total repeat arrays, showed an overall mutation frequency of ~15%. However, donor 727, who carried 29 perfect CGG repeats owing to a single AGG

interspersion, had more expansions and contractions than donor 791 with only 19 perfect repeats owing to two AGG interspersions. This was more evident when changes of ≥ 2 repeat units were considered. Moreover, when analysed in this way, only donor 727 showed expansions of the repeat array. Although simple slippage mechanisms could account for the observed loss of single 5' CGG triplets from both 727 and 791 and the gain or loss of CGG repeat units from the 3' uninterrupted portion of their repeats, it is difficult to model the mechanism by which the 10+10 and 10+9 alleles were generated. However, such complex contractions of the CGG repeat, although rare (<2% of the sperm analysed), could contribute to the mode of CGG repeat alleles observed at ~20 repeats. Greater than 90% of all variations observed in both donors' sperm involved the lengthening or contraction of the 3' portion of the perfect repeat array without alteration of AGG localisation. These data support the idea that AGG interspersions promote stability of the FMR1 repeat and that variation of the normal repeat involves polarity. It is also possible that variation among people in repeat stability owing to trans factors could account for this difference. However, in conjunction with previous studies using both sorted sperm and population data at a number of loci involving trinucleotide repeat expansion, these data point to the common importance of perfect trinucleotide repeat tracts of greater than approximately 24 triplets influencing repeat instability.

STW is an investigator and CBK is an associate of the Howard Hughes Medical Institute. EPL is a postdoctoral fellow of the Hereditary Disease Foundation. This work was supported in part by NIH grants R37 GM36745 (NA) and R37 HD20521 (STW)

- Warren ST, Nelson DL. Advances in molecular analysis of fragile X syndrome. *JAMA* 1994;27:536-42.
 Warren ST, Ashley CT. Triplet repeat expansion mutations: The example of fragile X syndrome. *Annu Rev Neurosci* 1995;18:77-99.
- 1995;18:77-99.
 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.
 Pieretti M, Zhang F, Fu YH, et al. Absence of expression of the FMR1 gene in fragile X syndrome. Cell 1991;66:817-22.
- 5 Sutcliffe JS, Nelson DL, Zhang F, et al. DNA methylation represses FMR-1 transcription in fragile X syndrome. Hum Mol Genet 1992;1:397-400.
- shley CT, Sutcliffe JS, Kunst CB, et al. Human and murine Assney C1, Sutcline JS, Kunst CB, et al. Human and murine FMR-1: alternative splicing and translational initiation downstream of the CGG repeat. Nat Genet 1993;4:244-51. Kremer EJ, Pritchard M, Lynch M, et al. Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CGCM). Science 1001;221:121.
- Science 1991;252:1711-14. (CGG)n
- p(CGG)n. Science 1991;252:1711-14.
 8 Oberlé I, Rousseau F, Heitz, D, et al. Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. Science 1991;252:1097-102.
 9 Rousseau F, Heitz D, Biancalana V, et al. Direct diagnosis by DNA analysis of the fragile X syndrome of mental retardation. N Engl J Med 1991;325:1673-81.
 10 Ashley CT, Wilkinson KD, Reines D, Warren ST. FMR1 protein: conserved RNP domains and selective RNA binding. Science 1993:262:563-6.
- ing. Science 1993;262:563-6. Khandjian EW, Corbin F, Woerly S, Rousseau F. The fragile
- X mental retardation protein is associated with ribosomes. Nat Genet 1996;12:91-3.
- Eberhart DE, Malter HE, Feng Y, Warren ST. The fragile X mental retardation protein is a ribonucleoprotein contain-
- mentai retardation protein is a ribonucleoprotein contain-ing both nuclear localization and nuclear export signals. *Hum Mol Genet* 1996;5:1083-91.
 Snow K, Doud LK, Hagerman R, Pergolizzi RG, Erster SH, Thibodeau SN. Analysis of a CGG sequence at the FMR-1 locus in fragile X families and in the general population. *Am J Hum Genet* 1993;53:1217-28.
 En VH Kuhl DBA Bieneti A et al Variation of the CCC.
- 14 Fu YH, Kuhl DPA, Pizzuti A, et al. Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. Cell 1991;67:1047-58.

- 15 Heitz D, Devys D, Imbert G, Kretz C, Mandel JL. Inherit-ance of the fragile X syndrome: size of the fragile X premutation is a major determinant of the transition to full muta-tion. J Med Genet 1992;29:794-801.

- tion. J Med Genet 1992;29:794-801.
 16 Kunst CB, Warren ST. Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. Cell 1994;77:853-61.
 17 Richards RI, Sutherland GR. Simple repeat DNA is not replicated simply. Nat Genet 1994;6:114-16.
 18 Eichler EE, Holden JJA, Popovich BW, et al. Length of uninterrupted CGG repeats determines instability in the FMR1 gene. Nat Genet 1994;8:88-94.
 19 Zhong N, Yang W, Dokkin C, Brown WT. Fragile X gene instability: anchoring AGGs and linked microsatellites. Am J Hum Genet 1995;57:351-61.
 20 Hirst MC, Grewal PK. Davies KE. Precursor arrays for tri-
- 20 Hirst MC, Grewal PK, Davies KE. Precursor arrays for triplet repeat expansion at the fragile X locus. *Hum Mol Genet* 1994;3:1553-60.
- 1994;3:1553-60.
 Snow K, Tester DJ, Kruckeberg KE, Schaid DJ, Thibodeau SN. Sequence analysis of the fragile X trinucleotide repeat: implications for the origin of the fragile X mutation. *Hum Mol Genet* 1994;3:1543-51.
 Richards RJ, Holman K, Friend K, et al. Evidence of founder chromosomes in fragile X syndrome. *Nat Genet* 1992;1:257-60.
 Nolin SL, Lewis EA, Ye LL, et al. Emilial transmissions of
- 23 Nolin SL, Lewis FA, Ye LL, et al. Familial transmissions of the FMR1 CGG repeat. Am J Hum Genet 1996;59:1252-61
- 24 Zhang L, Leeflang EP, Yu J, Arnheim N. Studying human mutations by sperm typing: instability of CAG trinucle-

otide repeats in the human androgen receptor gene. Nat Genet 1994;7:531-5.

- 25 Kunst CB, Zerylnick C, Karickhoff L, et al. FMR1 in global
- populations. Am J Hum Genet 1996;58:513-22. 26 Zhang L, Fischbeck H, Arnheim N. CAG repeat length variation in sperm from a patient with Kennedy's disease. Hum Mol Genet 1995;4:303-5.
- 27 Leeflang EP, Zhang L, Tavare S, et al. Single sperm analysis of the trinucleotide repeats in the Huntington's disease gene: quantification of the mutation frequency spectrum. Hum Mol Genet 1995;4:1519-26. 28 Ashley CT, Warren ST. Trinucleotide repeat expansion and
- human disease. Annu Rev Genet 1995;29:703-28.
- 29 Warren ST. The expanding world of trinucleotide repeats. Science 1996;271:1374-5.
- 30 Chung MY, Ranum LPW, Duvic LA, Servadio A, Zoghbi HY. Evidence for a mechanism predisposing to intergen-erational CAG repeat instability in spinocerebellar ataxia type I. Nat Genet 1993;5:254-8.
- 31 Monckton DG, Won LJ, Ashizawa T, Caskey CT. Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: small pool PCR analyses. Hum Mol Genet 1995;4:1-8. 32 Mornet E, Chateau C, Hirst MC, et al. Analysis of germline
- variation at the FMR1 CGG repeat shows variation in the normal-premutated borderline range. Hum Mol Genet 1996;5:821-5.