STABILITY OF THE FRAGILE X CGG REPEAT IN SOMATIC CELLS

by

Robert W. Burman

A DISSERTATION

Presented to the Department of Molecular and Medical Genetics

and the Oregon Health Sciences University

School of Medicine

in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

April 2000

School of Medicine

Oregon Health Sciences University

CERTIFICATE OF APPROVAL

This is to certify that the Ph.D. thesis of

Robert W. Burman

has been approved

[redacted] 4/25/00

Professor in charge of thesis

[redacted] 4/12/00

Member

[redacted] 4/12/00

Member

[redacted] 4/25/2000

Member

[redacted] 4/25/2000

Member

[redacted] /25/

Member

[redacted] 05/01/00

Associate Dean for Graduate Studies

TABLE OF CONTENTS

TABLES & ILLUSTRATIONS

ACKNOWLEGEMENTS

I owe a sincere thank you to many people who have helped to make this thesis possible. Unfortunately, I am not eloquent enough to fully thank each person in a manner that adequately reflects their unique contributions. In lieu of a detailed list of people that have helped me, I would like to issue a general, but nonetheless heartfelt, thank you to everyone that has lended any type of assistance to me during my graduate school adventure (you know who you are). In addition, I would like to offer specific thanks to Brad Popovich and Mitch Turker for acting as my thesis advisors and providing me with two labs to call home for the last 5 years. I would also be remiss not to mention Phil Yates, who has never had a bad idea. In addition, I am indebted to Peter Jacky for his assistance, particularly in procuring samples from fragile X patients that serve as the basis of this thesis. I would also like to thank Mike Liskay, Robb Moses, and the Department of Molecular and Medical Genetics for providing resources, including a place on a training grant, that have supported me throughout my studies at OHSU. Last, but certainly not least, I thank all the members of my thesis committee who have provided important guidance and words of encouragement (Brad Popovich, Mitch Turker, Mike Liskay, Markus Grompe, and Mike Litt).

ABSTRACT

The fragile X syndrome is caused by expansion of a trinucleotide repeat tract and is the most common form of inherited mental retardation with an incidence of ~1:6,000 males. The trinucleotide repeat responsible for the syndrome is a CGG triplet located in the *5'* UTR of the X-linked *FMR1* gene. Studies that have examined the inheritance of the CGG repeat have delineated at least three distinct allele classes that are grouped by length of the repeat tract and defined by the molecular and phenotypic consequences observed in carriers. Within the normal population, the repeat is polymorphic and is composed of ~ 6-50 triplets. These alleles are stable when transmitted between generations and in the somatic cells of an individual. The next larger class of alleles are referred to as premutations and are composed of repeat tracts of ~50-200 triplets. Like normal alleles, premutation alleles are believed to result in no immediate phenotypic consequences in carriers and are stable in somatic cells. However, premutations are potentially unstable when transmitted between generations and can expand to form pathogenic full mutations. Interestingly, the transition from premutation to full mutation occurs only when the premutation is transmitted by a female. Full mutation alleles contain ~200-2,000 triplets and are almost always associated with extensive hypermethylation of the CGG repeat and a surrounding CpG island. Hypermethylation of the CpG island is associated with transcriptional silencing of the *FMR1* gene and the resulting deficit of FMR1 protein is believed to be sufficient to cause the fragile X syndrome phenotype. Most full mutations are highly unstable in

somatic cells. Experimental evidence indicates that this instability is limited to a period during early embryogenesis.

This thesis describes an analysis of repeat length and methylation variability at the *FMR1* locus in somatic cells. I have investigated the relation between repeat expansion and methylation in an unusual individual who carries an unmethylated full mutation. The relation between repeat expansion and methylation is not well understood and in particular it is not known why some full mutations escape the methylation process. I have defined the extent of the *FMR1* methylation deficiency in this patient and have investigated the biological basis of his unmethylated full mutation. I have also utilized unmethylated full mutation alleles derived from this individual to test the hypothesis that methylation serves as a stabilizing influence on CGG repeats. I studied the behavior of methylated and unmethylated full mutations in both primary human fibroblasts and somatic cell hybrids. My analysis demonstrates that methylation correlates with stability in human fibroblasts, but in cell hybrids the determinants of stability are more complex and include repeat length and cellular differentiation. Finally, I have conducted a study to determine if a familial factor(s) influences repeat instability in siblings with the fragile X syndrome. I use a novel comparison strategy to demonstrate that mutation patterns are more similar in siblings than in unrelated patients. This result suggests that somatic mosaicism is generated in a non-random manner and that familial factors may influence this process.

CHAPTER 1

INTRODUCTION

HISTORICAL PERSPECTIVE & CLINICAL FINDINGS

History and Cytogenetics

Fragile X syndrome is a common form of inherited mental retardation caused by mutation of the fragile X mental retardation (*FMR1*) gene located on the distal long arm of the X-chromosome. The syndrome was first described in 1943 by Martin and Bell who published a pedigree showing sex-linked mental retardation in a family in which both males and females were affected (Martin and Bell 1943). Twenty-six years later, Lubs reported a cytogenetic marker on the X-chromosome, *Marker X* or fra(X), that cosegregated with mental retardation in one family that contained four retarded males (Lubs 1969). Despite analysis of additional families displaying X-linked mental retardation (XLMR), confirmation of Lubs' findings was delayed until 1977 when Sutherland demonstrated that induction of the Xq27 marker (figure 1) and other heritable fragile sites depends on culturing cells in specific types of culture medium (Sutherland 1977). A systematic comparison of fra(X) induction in various culture media lead to the seminal discovery that depletion of folate and thymidine were required to elicit induction of the fragile site (Sutherland 1979). These insights suggested a role for folate metabolism and pyrimidine biosynthesis in fragile site expression and prompted a number of studies that sought to establish the biochemical basis of fra(X) expression. Subsequently it was shown that folate antagonists such as methotrexate (Sutherland 1979) and thymidylate synthetase inhibitors such as 5-fluorodeoxycytidine (Glover 1981; Tommerup et al. 1981) are potent inducers of fra(X) expression. Methotrexate is a competitive inhibitor of dihydrofolate reductase, which

catalyzes the reduction of dihydrofolate (DHF) to tetrahydrofolate (THF) (Sutherland et al. 1985). THF is converted by serine hydroxymethyltransferase to N^5,N^{10}-methylene THF that in turn serves as a single carbon donor in the conversion of deoxyuridinemonophosphate (dUMP) to deoxythymidinemonophosphate (dTMP) by thymidylate synthetase (Sutherland et al. 1985). Thus treatment with both methotrexate and 5-fluorodeoxycytidine lead to depletion of the dTMP pool (Sutherland et al. 1985; Jacky 1996). The apparent importance of thymidine suggested that the fragile site might be composed of an adenine (A) and thymidine (T) rich repetitive DNA sequence. However, it is now known that expression of the fragile site correlates with the presence of a large repetitive cytosine (C) and guanine (G) rich repeat (CGG). Although dTMP/dTTP depletion can cause imbalances in other nucleotide pools, it is still unclear how these imbalances are related to fragile site expression in association with a CGG repeat sequence (Jacky 1996).

Clinical Characterization

Prior to publication of Sutherland's studies on fragile site induction, macroorchidism was shown to be associated with XLMR in some families (Escalante et al. 1971; Cantu et al. 1976; Turner et al. 1978). Soon after culture conditions were standardized making induction or the fra(X) more reliable, Turner and colleagues (Turner et al. 1978) demonstrated that macroorchidism cosegregated with the fragile site in some XLMR pedigrees. The original "Martin-Bell family" (Martin and Bell 1943) was reassessed in 1981 and affected males were shown to express the fra(X) marker and macroorchidism as well as other typical clinical features (Richards et al.

1981). The association of XLMR, macroorchidism, and fra(X) expression was originally named the Martin-Bell syndrome (Richards et al. 1981). Today the syndrome is usually referred to as the fragile X syndrome in reference to the cytogenetic fragile site. The term "Martin-Bell phenotype" is sometimes used as an eponym to describe the typical physical features of affected males (de Vries et al. 1998).

A range of phenotypic features has been described in males affected with the fragile X syndrome. Perhaps the most well known feature is mental retardation that on average falls within the moderate to severely retarded range of intelligence (Bennetto and Pennington 1996). Mean full-scale IQ estimates are typically within the range of 22-65 (Bennetto and Pennington 1996) and evidence indicates that cognitive ability declines in an age-dependent manner in many patients (Fisch et al. 1992). Like mental retardation, macroorchidism might also be considered a hallmark of the fragile X syndrome. Macroorchidism affects approximately 80% of adult males and 50% of prepubertal males with the syndrome. A number of investigators have addressed the basis of this macroorchidism and have found interstitial fibrosis, edema, (Cantu et al. 1976; Johannisson et al. 1987) and abnormal tubular morphology in the testes of fragile X patients (Rudelli et al. 1985). In addition, a mouse model of the fragile X syndrome with increased testicular weight was shown to have increased Sertoli cell proliferation from embryonic day 12 to 15 days postnatally (Slegtenhorst-Eegdeman et al. 1998). Despite these testicular abnormalities, males with the fragile X syndrome are fertile and have fathered offspring. In addition to macroorchidism, the most frequent physical features observed in affected males are: long face (70%), prominent ears (70%), high

arched palate (52%), hyperextensible finger joints (67%), double jointed thumbs (53%), and flat feet (71%) (de Vries et al. 1998). Behavioral and emotional abnormalities have also been described and include: hyperactivity, hand flapping, perseveration, shyness, anxiety, gaze avoidance, and tactile defensiveness (Merenstein et al. 1996).

Females heterozygous for the fragile X mutation can be affected with the syndrome, however the phenotypic manifestations tend to be significantly less severe than those observed in affected males. Although approximately 25% of heterozygous females are retarded (Cronister et al. 1991a; Cronister et al. 1991b; Hagerman et al. 1992), their mean IQ is usually in the low average range of 80-91 (Bennetto and Pennington 1996). An important variable that may, theoretically, affect some phenotypic outcomes in heterozygous females is the X-chromosome inactivation ratio. X-inactivation is frequently measured as the "activation ratio", which is the proportion of cells that have the normal non-mutated *FMR1* gene on the active X-chromosome. A number of studies have examined the relation between activation ratio in DNA derived from peripheral blood and cognitive, physical and behavioral findings in females who carry the mutation. The results of these studies with respect to activation ratio and IQ level have been mixed with some showing a weak correlation and others showing no significant correlation (Hagerman 1996). The activation ratio does not seem to be related to physical and behavioral measures except in a subset of non-retarded (IQ > 70) heterozygous females in which it correlates negatively with ear prominence (Hagerman 1996).

Neuropathological studies of brain tissue from a limited number of fragile X males have demonstrated some anatomical differences that may be specific consequences of the syndrome. Hinton studied the brains of three males with fragile X syndrome and found abnormal dendritic spine morphology characterized by immature, thin, and long spines (Hinton et al. 1991). In addition, synaptic length and contact were reduced in one patient relative to controls (Rudelli et al. 1985). Similarly abnormal dendritic spine morphology has also been described in a mouse model of the fragile X syndrome (Comery et al. 1997). Neuroimaging studies using quantitative MRI have been used to study gross anatomical features of the brains of affected males. Compared to normal controls, males with the fragile X syndrome have a reduced size of the posterior cerebellar vermis and increased size of the caudate nucleus, hippocampus, and thalamus (Reiss et al. 1991a; Reiss et al. 1991b; Reiss et al. 1994; Reiss et al. 1995). Some of the features visualized by neuroimaging appear to correlate with the cognitive phenotype. For instance, there is an inverse relation between IQ and caudate nucleus volume in both males and females with the fragile X syndrome (Reiss et al. 1995). The overall significance of the neuroimaging studies in relation to disease pathology is somewhat uncertain and remains a work in progress.

Prevalence and Inheritance

Initial estimates of the prevalence of fragile X syndrome were based on expression of the cytogenetic marker fra(X) and have suffered from, as one author described, a somewhat turbulent history (Turner et al. 1996). Difficulties with the cytogenetic test were based in part on inter-laboratory variation in both the number of

metaphase preparations examined and the cut-off value for the proportion of fra(X) positive cells that was considered diagnostic. Although most labs used a threshold of 2-4% fra(X) positive metaphases as the cut-off for diagnosis, others used limits as low as 1-2% (Turner et al. 1996). As a consequence, these studies often had a high false positive rate. Despite these shortcomings, the inflated prevalence figures are still used in some instances today. The original fra(X)-based prevalence estimates for males were in the range of 1:1,000 to 1:1,500 (Gustavson et al. 1986; Turner et al. 1986; Webb et al. 1986; Kahkonen et al. 1987). The prevalence estimates for females were more variable and were approximately 1:1,700 to 1:5,000 (Turner et al. 1986; Webb et al. 1986; Kahkonen et al. 1987).

Once the *FMR1* gene was cloned, molecular testing for the fragile X mutation became possible and the studies of Turner and Webb (Turner et al. 1986; Webb et al. 1986) were reanalyzed with molecular methods. The molecular test demonstrated the syndrome is less common than previously thought and estimated the true prevalence to be in the range of 1:4,000 - 1:5,700 males and 1:8,000 females (Sherman 1996; Turner et al. 1996; Morton et al. 1997). These values agree with an independent study that estimated prevalence among males to be 1:6,045 (de Vries et al. 1997). Molecular testing has also been used to estimate the prevalence of non-pathogenic premutation alleles, which have the capacity to expand to form disease-causing mutations. A 1995 study by Rousseau and colleagues screened 10,624 unselected French Canadian women for the presence of premutation alleles (Rousseau et al. 1995). The prevalence of the premutation was found to be a surprisingly high 1:259 (Rousseau et al. 1995). It

remains to be determined if this relatively high carrier frequency is unique to the French Canadian population or can be extrapolated to other populations (Sherman 1995).

The inheritance pattern of the fragile X syndrome was the subject of close scrutiny in the years preceding identification and cloning of the *FMR1* gene in 1991. Analysis of penetrance in fragile X pedigrees led to the conclusion that the syndrome is an X-linked dominant condition with reduced penetrance (Sherman et al. 1984; Sherman et al. 1985). Penetrance was estimated to be 80% for males and 35% for females (Sherman et al. 1984). A number of unusual features of fragile X syndrome inheritance set it apart from other X-linked traits. For instance, 20% of males who were inferred to carry the mutation did not express any clinical or cognitive symptoms and were negative for the fra(X) marker but transmitted the mutation to their daughters (Sherman 1996). Furthermore, an individual's risk for expressing the fragile X phenotype was shown to be dependent on that individuals position in the pedigree and the sex of the transmitting parent. Thus daughters of non-expressing transmitting males were rarely, if ever, affected. However, the daughters of non-expressing females had a 30% risk of displaying the phenotype and fra(X) (Sherman 1996). The risk of phenotypic expression showed prominent anticipation such that the risk of having an affected child increased in each successive generation. The unusual inheritance pattern, and in particular the variable penetrance with anticipation, became known as the "Sherman paradox" in recognition of the author. Pembrey and colleagues hypothesized that the existence of a non-pathogenic premutation allele that predisposed female gametes to fra(X) expression could explain the unusual inheritance of the disorder

(Pembrey et al. 1985). Resolution of the Sherman paradox and confirmation of Pembrey's hypothesis occurred in 1991 when the *FMR1* gene was cloned and shown to contain a polymorphic CGG repeat mutation that is prone to intergenerational instability in a parent of origin-dependent manner.

FMR1 GENE STRUCTURE & EXPRESSION

FMR1 Gene and Fragile X Mutations

The gene responsible for the fragile X syndrome, *FMR1*, was identified and cloned in 1991 with a positional cloning strategy that utilized information derived from linkage analysis, physical maps, and somatic cell studies. The first step in isolation and identification of the gene was genetic mapping through linkage analysis that narrowed the region of interest to a 20 Mb interval on Xq27 that was apparently coincident with the cytogenetic fragile site (Goodfellow et al. 1985; Oberle et al. 1986). Warren and colleagues then adopted a somatic cell hybrid approach that relied on selectable markers proximal (hypoxanthinephosphoribosy transferase (*HPRT*)) and distal (glucose-6-phosphatedehydrogenase (*G6PD*)) to the region of interest to facilitate identification of new markers near the fragile site and selection of hybrids that contained exclusively proximal or distal portions of the X-chromosome (Warren and Davidson 1984; Warren et al. 1990). The approach proved successful and allowed new polymorphic loci to be mapped in the region of the fragile site, thereby narrowing the region of interest to a 3 Mb interval (Suthers et al. 1990; Hirst et al. 1991b; Rousseau et al. 1991b; Suthers et al. 1991). Subsequently, yeast artificial chromosomes (YACs) were identified that

contained inserts that spanned the breakpoints identified in the cell hybrids (Dietrich et al. 1991; Heitz et al. 1991; Hirst et al. 1991a; Kremer et al. 1991b; Verkerk et al. 1991) and contained a hypermethylated CpG island (Bell et al. 1991) previously identified in fragile X patients (Vincent et al. 1991). Further characterization of YAC clones containing the CpG island led to the discovery of Southern blot probes that identified apparent insertions (restriction fragments with increased size) in fragile X patients (Oberle et al. 1991). The region of DNA that contained these apparent insertions was successively narrowed to a 1.0 kb *Pst*I fragment that contained a polymorphic CGG trinucleotide repeat that colocalized with the fragile site and was apparently responsible for the restriction fragment length variations observed in fragile X patients (Kremer et al. 1991a; Kremer et al. 1991b; Verkerk et al. 1991; Yu et al. 1991). YAC subclones spanning the CpG island and CGG repeat were used to screen a human cDNA library and identify two cDNA clones (Verkerk et al. 1991) that were used to determine the genomic organization of the *FMR1* gene.

The *FMR1* gene spans 38 kb and contains 17 exons (Eichler et al. 1993) as well as a polymorphic CGG repeat sequence in the *5'* untranslated region (*5'* UTR) of exon 1 (Fu et al. 1991)(figure 2). Variation in the length of the CGG repeat tract is responsible for the large majority of cases of the fragile X syndrome. The fragile X syndrome was the first example of a disease characterized at the molecular level by an expanding trinucleotide repeat sequence. Today at least nine other human disorders are known to be caused by unstable triplet repeats (Timchenko and Caskey 1999).

The fragile X CGG repeat is polymorphic in normal populations, with lengths ranging in size from ~5-54 triplets, and a mode of 30 (Fu et al. 1991; Snow et al. 1993). These normal alleles are stable (e.g. not prone to changes in length) both within the somatic cells of an individual and when transmitted between generations. Carrier females and transmitting males possess a repeat tract of intermediate length, ranging in size from ~55-200 triplets (Fu et al. 1991). Alleles within this intermediate size range are commonly referred to as "premutations". Like normal alleles, premutations are largely stable in the somatic cells of an individual. However, premutations have the ability to expand to form disease causing full mutations with >200 repeats when transmitted intergenerationally (Fu et al. 1991). Interestingly, the premutation to full mutation expansion occurs only when the premutation is transmitted by a female. Furthermore, the probability of intergenerational expansion was demonstrated to be a function of premutation size (Fu et al. 1991; Snow et al. 1993; Nolin et al. 1996). The likelihood of expansion was calculated empirically for premutation alleles grouped by size with the following results: 56-59 repeats (13.4% chance of expansion to full mutation), 60-69 (20.6%), 70-79 (57.8%), 80-89 (72.9%), 90-99 (94.3%), and > 100 (98.9%) (Nolin et al. 1996). Thus with each maternal transmission of the premutation, the risk of producing an affected offspring increased. This correlation between repeat size, penetrance, and position in a pedigree allowed resolution of the Sherman paradox (Fu et al. 1991). Although premutation alleles can expand to from pathogenic full mutations when transmitted between generations, it is generally believed that male and female premutation carriers suffer few if any phenotypic consequences. However, some

authors have argued that some female carriers may manifest mild symptoms such as an increased incidence of anxiety disorder and social phobia (Hagerman 1996). Furthermore, some recent studies have reported an association between premutation alleles and premature ovarian failure (Allingham-Hawkins et al. 1999; Macpherson et al. 1999; Syrrou et al. 1999; Uzielli et al. 1999; Vianna-Morgante 1999; Marozzi et al. 2000).

In addition to variation in repeat length, the fragile X CGG repeat is also polymorphic with respect to the number and location of AGG sequences that can disrupt the continuity of the repeat. In most cases, the AGG interspersions interrupt the repeat with a periodicity of once every nine or ten CGG repeats . Most normal alleles contain two interruptions so that the most common configurations are $(CGG)_{9\text{ or }10}$ AGG $(CGG)_9$ AGG $(CGG)_9$ (Eichler et al. 1994; Kunst and Warren 1994; Zhong et al. 1995). In contrast, premutation alleles usually contain either one or no interspersions, suggesting that loss of the AGG interruptions may predispose alleles to instability (Eichler et al. 1994; Zhong et al. 1995). Consistent with these findings, Eichler and colleagues (1994) demonstrated that the length of the uninterrupted CGG repeat influences stability. The threshold for any degree of instability was estimated to be approximately 34-37 uninterrupted CGG triplets and the threshold for hyperexpansion from premutation to full mutation in a single generation was estimated to be 56-75 uninterrupted repeats (Eichler et al. 1994). Variation in repeat length appears to occur principally at the *3'* end of the repeat (Eichler et al. 1994; Hirst et al. 1994; Kunst and Warren 1994; Snow et al. 1994; Zhong et al. 1995). The prototypical formula for the

FMR1 CGG repeat is thus $(CGG)_{10}$ AGG $(CGG)_{10}$ AGG $(CGG)_x$, where x is equal to the remaining number of repeats (Eichler et al. 1994). The polar variability suggests stability differences between the leading and lagging strand of DNA replication and is similar in nature to that observed at some minisatellite loci (Armour et al. 1993). Because variation seems to occur predominately, if not entirely, at the *3'* end of the repeat, the *5'* end appears to be relatively more stable suggesting that the AGG interspersions may confer some degree of stability (Kunst and Warren 1994). The mechanistic implications of the AGG interspersion pattern is discussed below in the section devoted to "Mechanisms of Triplet Repeat Instability".

The *FMR1* CGG repeat appears to be conserved in mammals. Eichler and colleagues examined the structure of the repeat in the orthologous *fmr1* gene from 44 species of mammals (Eichler et al. 1995). Most non-primate mammals possess small non-interrupted repeats with a mean repeat length of 8 units (Eichler et al. 1995). Among primates, the repeat was longer (mean = 20 units) and more frequently interrupted. In those species that possessed an interrupted repeat, the sequence of the interspersed triplet varied and included AGG, CAG, CGA, and TGG (Eichler et al. 1995). A separate study of the chicken *fmr1* gene identified a complex CCT repeat flanked by dinucleotide repeats in the *5'* UTR (Price et al. 1996).

Although the large majority of patients with the fragile X syndrome have a repeat expansion mutation, other types of pathogenic mutations have been described as well. The most frequent type of non-repeat expansion mutation identified is deletion of all of part of the *FMR1* gene. At least twenty deletions have been reported that vary

with respect to origin (inherited or *de novo*), breakpoints, and phenotypic consequences (Hammond et al. 1997). Atypical clinical features have been described in some males with large deletions extending well beyond the *FMR1* gene. This suggests that the absence of other unidentified loci may have an effect on the phenotypic features in these patients. Ten deletions have been described in mosaic form in conjunction with full mutation alleles thereby demonstrating that these deletions are the product of somatic events. Interestingly, deGraaff and colleagues (de Graaff et al. 1995) identified a ~ 140 bp deletion hotspot upstream of the CGG repeat. The *5'* deletion breakpoint of at least eight separate patients, most of whom are deletion-full mutation mosaics, have been shown to fall within the hotspot (de Graaff et al. 1995; Quan et al. 1995; Mannermaa et al. 1996; Mila et al. 1996; Schmucker et al. 1996). Two point mutations and a small 2 bp deletion have also been identified in fragile X patients (De Boulle et al. 1993; Lugenbeel et al. 1995; Wang et al. 1997). One of the point mutations interrupted an important domain of the FMR1 protein and was reported in a profoundly retarded male with typical fragile X features (De Boulle et al. 1993). Importantly, the point mutations and non-mosaic deletions demonstrate that the absence of FMR1 protein function is sufficient to cause the fragile X syndrome.

Consequences of Repeat Expansion

Hypermethylation of DNA sequences in the vicinity of the CGG repeat was identified in fragile X patients even before the *FMR1* gene was completely characterized (Bell et al. 1991; Vincent et al. 1991). Full mutations (alleles in excess of ~ 200 repeats) are almost always associated with extensive hypermethylation of the

repeat and a surrounding CpG island (Oberle et al. 1991; Hansen et al. 1992; Hornstra et al. 1993). Alleles in the normal and premutation size range appear to be completely free of methylation except when located on an inactive X-chromosome in females (Oberle et al. 1991; Hansen et al. 1992; Hornstra et al. 1993). Although the methylation changes observed on the inactive X-chromosome and in association with full mutation alleles are qualitatively similar, methylation is more extensive and less heterogeneous in association with full mutations (Luo et al. 1993; Stoger et al. 1997). Analysis of methylation patterns at the single molecule level with bisulfite sequencing has demonstrated a bimodal distribution of methylation density in cells of normal females and fragile X patients who are mosaic for premutation and full mutation alleles (Stoger et al. 1997). Thus the CpG sites on a particular DNA molecule were found to have coordinated methylation so that the sites were entirely hypo- or hypermethylated.

Methylation of the fragile CpG island is associated with transcriptional silencing of *FMR1* and expression of the fragile X phenotype (Pieretti et al. 1991; Sutcliffe et al. 1992). Although the mechanism of this silencing has not been fully described, evidence indicates a role for changes in chromatin organization. A recent study by Coffee and colleagues demonstrated that unacetylated histone proteins are associated with the *FMR1* promoter in the cell lines of fragile X patients but not normal controls (Coffee et al. 1999). Furthermore, nuclease sensitivity studies have shown that the *FMR1* promoter is hypersensitive on the normal active X-chromosome but insensitive on both inactive and fragile X chromosomes (Luo et al. 1993; Eberhart and Warren 1996). These findings support a model in which methylated DNA recruits a histone deacetylase

complex, perhaps through interaction with the methyl cytosine binding protein MeCP2, resulting in histone deacetylation, chromatin condensation, and transcriptional silencing (Razin 1998; Coffee et al. 1999). Consistent with this model, *in vivo* footprinting analysis has demonstrated that four protein-DNA interaction sites present in the unmethylated promoter are absent in the methylated promoter of cells that carry a full mutation (Drouin et al. 1997; Schwemmle et al. 1997; Schwemmle 1999). Thus, silencing of *FMR1* transcription may result from a lack of transcription-factor binding secondary to chromatin condensation (Schwemmle et al. 1997).

Although there appears to be a clear demarcation between unmethylated normal and premutation alleles and hypermethylated full mutations, it is unknown why alleles in excess of ~200 CGG repeats serve as targets for methylation. One proposed function for cytosine methylation in the genome is suppression of the transcription of transposable elements such as *L1*, *Alu*, and retroviral elements (Yoder et al. 1997). Therefore, these elements may act as specific targets for *de novo* methylation. When expanded beyond the ~200 repeat threshold, the fragile X repeat might resemble a transposable element thereby becoming a target for *de novo* methylation. This postulated similarity between the CGG repeat and transposable elements could be based on the formation of DNA secondary structures (Bestor and Tycko 1996). Consistent with this hypothesis, some CGG repeat secondary structures have been shown to be efficient substrates for *de novo* methylation *in vivo* (Chen et al. 1995; Chen et al. 1998).

The relation between repeat expansion and methylation is not absolute. A small number of individuals have been described who carry unusual full mutation alleles that

are unmethylated at diagnostic restriction sites (Loesch et al. 1993; McConkie-Rosell et al. 1993; Hagerman et al. 1994; Merenstein et al. 1994; Rousseau et al. 1994b; Smeets et al. 1995; de Vries et al. 1996; Lachiewicz et al. 1996; Wang et al. 1996; Wohrle et al. 1998; Burman et al. 1999b; Taylor et al. 1999). These individuals tend to be described as "high functioning" and typically have cognitive and physical phenotypes in the normal to mild range of affectedness. As might be expected based on these phenotypic findings, FMR1 protein production has been demonstrated in some of these patients, though at somewhat reduced levels relative to normal controls. It is not understood how the full mutation alleles in these individuals have escaped the methylation process. Chapter two of this thesis describes an investigation into the extent and mechanism of methylation failure in a male who carries an unmethylated full mutation (Burman et al. 1999b).

In addition to the well documented changes in *FMR1* methylation that accompany the formation of full mutations, a distinct shift in *FMR1* replication timing has also been described in the cells of affected males. The normal *FMR1* allele is reported to replicate in late S phase (Hansen et al. 1993; Torchia et al. 1994). In contrast, full mutation alleles replicate late in the G2 phase of the cell cycle (Hansen et al. 1993; Torchia et al. 1994). These findings are consistent with studies that have shown transcriptional inactivity can be associated with delayed replication (Goldman et al. 1984; Hatton et al. 1988) and that the inactive X-chromosome replicates late relative to the active X-chromosome (Taylor 1960; Atkins et al. 1962). Interestingly, the region of delayed replication in fragile X cells is quite large and is estimated to extend 400 kb

5' and 600 kb *3'* of *FMR1* (Hansen et al. 1997). This region is thought to contain multiple replicons that are grouped into two zones in which replication timing is coordinated (Hansen et al. 1997). Hansen and colleagues hypothesize that a plausible explanation for the large zones of delayed replication may be a direct effect of the CGG repeat expansion on a master control locus that coordinates the initiation of replication at multiple replicons (Hansen et al. 1997).

Expression of FMR1 mRNA and Protein Production

The primary *FMR1* transcript undergoes alternative splicing events that can theoretically produce up to 12 distinct mRNA products (Ashley et al. 1993a; Verkerk et al. 1993). The largest and smallest splice variants encode predicted proteins of 631 amino acids and 436 amino acids, respectively (Ashley et al. 1993a; Verkerk et al. 1993). *FMR1* is widely expressed both in tissues that apparently contribute to the fragile X syndrome phenotype and in tissues that have not been implicated as part of the clinical spectrum. Northern analysis of human tissue detected high expression of a 4.4 kb *FMR1* message in various regions of the brain and in testes, placenta, lung, and kidney (Hinds et al. 1993). *In situ* hybridization to brain tissue from a 25 week human fetus showed strong *FMR1* expression in cholinergic and pyramidal neurons but only sparse expression in glial cells (Abitbol et al. 1993). In the mouse, the pattern of expression is ubiquitous at day 10 of gestation and then becomes more specific and localized in later embryonic stages until adult expression patterns are formed (Hinds et al. 1993). In the adult mouse, *fmr1* is abundantly expressed in brain, testes, ovary,

esophagus, thymus, eye, and spleen with no expression observed in heart, aorta, or muscle (Hinds et al. 1993).

The *FMR1* gene encodes at least six distinct protein isoforms. Western blot analysis with monoclonal antibody mAb1C3 detects a major 80 kDa tight doublet band and four minor bands with molecular weights of 67-78 kDa in both human and mouse extracts (Devys et al. 1993; Khandjian et al. 1995). These proteins are absent in extracts derived from cells with full mutations as expected based on transcriptional silencing of *FMR1* (Devys et al. 1993). The *FMR1* gene is highly conserved among species and show 97% amino acid homology with the murine homologue (Ashley et al. 1993b). The tissue distribution of *FMR1* protein (FMRP) is nearly identical to the distribution of *fmr1* mRNA in the mouse with high levels of expression of all 67-80 kDa isoforms in the brain and testes (Khandjian et al. 1995; Verheij et al. 1995). In the brain, FMRP is most abundant in neurons of the cortex and Purkinje cells of the granular layer (Devys et al. 1993; Feng et al. 1997b). In the testes, FMRP is expressed primarily in spermatogonia (Devys et al. 1993). Immunohistological studies have demonstrated that FMRP is predominately a cytoplasmic protein (Devys et al. 1993; Verheij et al. 1993). However, transfection experiments with COS cells showed that a small fraction of FMRP can be found in the nucleolus (Willemsen et al. 1996). FMRP contains two transport signals that may control its intracellular distribution. A nuclear localization signal (NLS) has been identified at amino acids 117-184 (Eberhart et al. 1996; Siomi et al. 1996; Sittler et al. 1996) and a nuclear export signal (NES) has been found at amino acids 429-438 (Eberhart et al. 1996; Siomi et al. 1996). Based on these results, it has

been proposed that FMRP shuttles between the nucleus and cytoplasm. Consistent with this model, electron microscopy with immunogold labeled FMRP has revealed that ~ 4-5% of total FMRP is nuclear (Corbin et al. 1997; Feng et al. 1997b). Furthermore, FMRP appeared to be concentrated near nuclear pores in rat brain (Feng et al. 1997b).

FMRP contains motifs, two KH domains and an RGG box, that are found in RNA-binding proteins (Ashley et al. 1993b; Siomi et al. 1993). Binding studies have demonstrated that FMRP can bind a variety of RNA substrates including homopolymers (Siomi et al. 1993), its own message, and ~4% of human fetal brain mRNAs (Ashley et al. 1993b). The importance of the KH domains for FMRP function *in vivo* was demonstrated by identification of a fragile X patient who carries a point mutation that results in an Ile304Asn substitution in one of the KH domains (De Boulle et al. 1993). This mutant FMRP has aberrant RNA binding characteristics under high salt conditions but not under low salt conditions (Siomi et al. 1994; Feng et al. 1997a).

Subcellular fractionation studies have shown that FMRP in the cytoplasm is associated with translating ribosomes (Eberhart et al. 1996; Corbin et al. 1997; Feng et al. 1997a). Confirmation of this ribosomal association was provided by Willemsen and colleagues who used electron microscopy and immunogold labeling to demonstrate that FMRP associates with both free ribosomes and those attached to the endoplasmic reticulum (Willemsen et al. 1996). Disruption of purified polyribosomes with EDTA treatment results in the cosedimentation of FMRP and messenger ribonucleoprotein (mRNP) particles (Eberhart et al. 1996; Corbin et al. 1997; Feng et al. 1997a). Based on the experimental evidence so far available, FMRP may function in selective transport of

mRNAs from the nucleus to the cytoplasm. In the cytoplasm, FMRP-mRNP particles associate with ribosomes and may play some role in the translation of messages to which it is bound.

An animal model of the fragile X syndrome has been created in which the murine *fmr1* gene has been rendered non-functional by insertion of a neomycin cassette in exon 5 (Bakker and Consortium 1994). The knock-out mice do not express FMRP and display mild physical and behavioral abnormalities. The knock-outs show significantly increased testicular weights (Bakker and Consortium 1994) and appear to have abnormal dendritic spine morphology similar to that observed in some fragile X patients (Comery et al. 1997). Using the Morris water maze behavioral test, the knock-outs display some subtle phenotypes including hyperactivity and impaired acquisition, but not retrieval, of spatial information (Bakker and Consortium 1994).

MECHANISM & TIMING OF FRAGILE X REPEAT INSTABILITY

Repeat Instability Definitions and Overview

To facilitate discussion of triplet repeat instability, most authors make a distinction between intergenerational and somatic events. The term intergenerational instability is used to refer exclusively to changes in repeat number that result in a difference in repeat length between parent and offspring. The term is intentionally vague to reflect the lack of knowledge pertaining to the time at which changes in the repeat tract occur. Somatic instability refers to repeat length variability that occurs in an individual's somatic cells. By definition, somatic instability must occur in mitotically

proliferating cells, but not necessarily during mitosis. This instability results in mosaicism, which is variation in length of the repeat tract among different cell populations within an individual. At least four types of mosaicism are recognized in the fragile X literature. The most common type of mosaicism, size mosaicism, is found in the large majority of individuals with a full mutation and can be defined as variation in length of the CGG repeat within the full mutation spectrum of ~200-2,000 repeats. Mutational mosaicism is a special kind of size mosaicism in which repeat length variation occurs in both the premutation (~50-200) and full mutation size ranges. This type of mosaicism results in individuals who carry both unmethylated, and FMRP producing, premutation alleles and completely methylated and silenced full mutation alleles. Deletion mosaicism has been described in a relatively small number of patients and is the co-occurrence of repeat expansion (full and/or premutation) and a deletion that removes all or part of the *FMR1* gene. Methylation mosaicism is rare and refers to a mixture of methylated and unmethylated full mutation alleles.

Very little is known about the mechanism and timing of repeat instability in the fragile X syndrome. Several models that have been proposed to explain repeat length variability are discussed below in the subsection "Current Models of Triplet Repeat Instability". The currently favored model explains repeat instability as a consequence of DNA slippage during replication. This model does not, however, address one of the major unresolved puzzles of the fragile X syndrome; whether the premutation to full mutation transition occurs in the germline of premutation females or after fertilization during early embryogenesis (e.g. pre- or post-zygotic).

The prezygotic model entails expansion to full mutation size at some point during oogenesis. Some authors have casually referred to this possibility as "meiotic expansion" even though the actual mutation event could well occur during either meiosis or one of the estimated 23 mitotic divisions of the primordial germ cells that precedes meiosis (Sato et al. 1999). In either case, the actual allele transmitted in the oocyte would be a full mutation. Somatic repeat instability during early embryogenesis would produce size mosaicism via a process of repeat retraction or possibly both retraction and further expansion. Mutation mosaics (co-occurrence of premutation and full mutation) would be explained by retraction into the premutation size range prior to methylation of the repeat.

The postzygotic expansion model supposes repeat stability throughout oogenesis and transmission of a premutation allele the same size as that contained in the mother's somatic cells. The transition to full mutation would then occur during early embryogenesis. Expansions of different sizes in different cells would generate size mosaicism and in some individuals, a subset of cells would fail to expand thereby generating mutation mosaicism. To explain the strict parent of origin effect observed in the fragile X syndrome, the post-zygotic model must also invoke some kind of imprinting phenomenon that renders only premutations transmitted by a female capable of expansion.

Intergenerational Instability

As previously discussed, there are three major determinants of intergenerational fragile X repeat instability: sex of the transmitting parent, size of the premutation allele,

and number of uninterrupted CGG repeats. The molecular basis of the strict parent of origin effect is not well understood, however, analysis of mutation events in sperm has yielded some interesting clues. In 1993, Reyniers and colleagues analyzed repeat length in the sperm of four full mutation carriers and found only unmethylated premutation alleles (Reyniers et al. 1993). Since this original report, the finding has been confirmed in other fragile X patients as well (de Graaff et al. 1995; Reyniers et al. 1999; Tassone et al. 1999). Two alternative models have been proposed to explain this finding. One model assumes postzygotic expansion and asserts that the germline is somehow uniquely protected from expansion. The competing model invokes a selection process during spermatogenesis that results in the elimination of cells containing full mutations in favor of those with premutations. Support for the latter model was provided in a study that examined the fragile X mutation in gametes of two fragile X fetuses at 13- and 17-weeks gestation (Malter et al. 1997). Analysis of testicular tissue from both fetuses by PCR did not show the presence of premutations. Furthermore, although FMRP expression was observed in the primordial germ cells (PGCs) of a 13-week control fetus, no expression was observed in the fragile X fetus, suggesting that premutation alleles were not present at this stage of development (Malter et al. 1997). The testes of the older 17-week fetus contained a small number of PGCs that expressed FMRP. The authors interpret their data to indicate that male germ cells initially carry full mutation alleles that retract during fetal development to produce only premutation bearing spermatozoa (Malter et al. 1997). Thus the absence of premutation to full mutation transition upon paternal transmission may reflect a selection against full

mutation sperm. This selection is likely not to result from a lack of FMRP since the fragile X knock-out mouse is fertile (Bakker and Consortium 1994) and a family has been reported in which paternal germinal mosaicism of a large *FMR1* deletion resulted in transmission of the deleted allele to three offspring (Meijer et al. 1994). Selection against cells bearing a full mutation is thus more likely to operate at the DNA level, perhaps based on the inability of the cells to faithfully replicate large repeat tracts. To test the selection hypothesis, a recent study by Ashley-Koch (Ashley-Koch et al. 1998) simulated selection against full mutation oocytes by analyzing only female premutation transmissions that resulted in premutation offspring . Previous work by Nolin and colleagues (Nolin et al. 1996) had shown that the frequency of contraction in male premutation transmissions increased with increasing paternal repeat length. Among the subset of maternal premutation to premutation offspring transmissions studied by Ashley-Koch, a similar association was identified. This result favors a prezygotic expansion model and suggests that the difference between male and female transmission may be tolerance of full mutations in oocytes and selection against full mutations in sperm (Ashley-Koch et al. 1998).

Additional evidence in favor of a prezygotic expansion event has come from a very nice study that examined the origin of mutation mosaicism (co-occurrence of premutation and full mutation) (Moutou et al. 1997). The authors of the study noted that the risk of premutation to full mutation expansion is strongly correlated with the size of the maternal allele. They reasoned that if expansion occurs post-zygotically, then one would expect to see a higher proportion of mutation mosaics in the offspring of mothers

who transmit a small premutation. Analysis of 212 full mutation carriers found no effect of maternal premutation size on the incidence of mutation mosaicism (Moutou et al. 1997). The authors concluded that the results of their study provided strong evidence against the postzygotic expansion model.

In addition to the well recognized effects of allele size and sex of transmitting parent on expansion probability, some evidence indicates that a familial factor may influence the size of expansion. A study by Nolin and colleagues examined allele size in families that appeared to show clustering of repeat size in the offspring (Nolin et al. 1996). For example, one family contained three siblings each with a 59 repeat allele while another family had two siblings that both possessed a 114 repeat allele. The study sought to determine if allele sizes were clustered in families in a statistically significant manner. To test the clustering hypothesis, Nolin and colleagues examined variation within and between sibships with analysis of variance (ANOVA). This analysis demonstrated a highly significant clustering of repeat size in the premutation offspring of both males and females (Nolin et al. 1996). Thus allele sizes were more similar within families than between families. This study was deliberately restricted to analysis of offspring with premutations because allele size can be more accurately determined for premutation alleles, which are non-mosaic, than for full mutations that typically display significant size mosaicism. Chapter four of this thesis examines allele clustering in full mutation sibships to determine if familial factors may likewise influence repeat size in full mutation carriers.

Somatic Instability

Somatic instability of the fragile X CGG repeat is observed only in conjunction with full mutation alleles. Instability occurs in the majority of full mutation carriers and can result in a complex hybridization pattern composed of multiple bands and smears when the repeat is visualized by Southern blot analysis (figure 3). The complexity of the hybridization patterns has, to some extent, limited retrospective analysis of these mutations. Moreover, consideration of allele size within the full mutation size range is generally not considered clinically relevant because methylation-associated transcriptional silencing is regarded as a binary variable (either present or absent). For example, methylation of either a 1,100 repeat allele or a 300 repeat allele ultimately results in the same outcome; absence of *FMR1* transcription and FMRP production, and expression of the fragile X syndrome phenotype.

The most commonly reported analysis of full mutation size mosaicism is post-mortem comparison of mutation patterns in multiple tissues of patients. Analysis of fetal (Sutherland et al. 1991; Devys et al. 1992; Wohrle et al. 1992; Wohrle et al. 1993; Wohrle et al. 1995; Moutou et al. 1997) and adult (de Graaff et al. 1995; Reyniers et al. 1999; Tassone et al. 1999) tissues has demonstrated that the mutation patterns tend to be well conserved among multiple tissues within an individual. These findings are significant for two reasons. First, similarity of the mutation patterns in multiple tissues implies that size mosaicism is generated early in development. Notably, several fetuses displayed similar bands among both fetal organs and extra-embryonic tissues such as chorionic villi (Devys et al. 1992; Wohrle et al. 1995; Moutou et al. 1997). This

suggests that size mosaicism is generated before separation of the trophectoderm and inner cell mass (Devys et al. 1992; Moutou et al. 1997). A second significant aspect of these studies is that the conserved mutation patterns imply that the period of instability is limited (Devys et al. 1992). If variability in repeat length occurred continuously, one would expect a smeared mutation pattern with few if any distinct bands and different patterns in different tissues (Devys et al. 1992). Further evidence supporting a window of repeat instability during early development has been provided by analysis of monozygous twins that carry full mutations. Some twins have been shown to possess nearly identical mutation patterns in DNA isolated from peripheral blood (Devys et al. 1992; Kruyer et al. 1994; Antinolo et al. 1996) while others appear to have discordant patterns (Kruyer et al. 1994; Helderman-van den Enden et al. 1999). It has been hypothesized that the timing of twining events relative to the period of instability may account for differences in the discrepant twins (Antinolo et al. 1996; Helderman-van den Enden et al. 1999). At a minimum, twins that maintain similar mutation patterns well after birth, such as the 9-year-old and 30-year-old twins reported by Devys and colleagues (1992), demonstrate the stability of the patterns over time and support the notion of a limited window of instability (Devys et al. 1992). Additional evidence in favor of limited instability has been provided in a study by Wohrle and colleagues (1993) that examined repeat stability in cultured fibroblasts derived from adult and fetal (13-week) full mutation carriers (Wohrle et al. 1993). Repeat size was maintained in clonal derivatives of both fibroblast lines during proliferation *in vitro* (Wohrle et al. 1993). This data indicates that full mutation alleles are stable in fetal and adult cells and

further substantiates a model in which somatic instability is limited to a period during early embryogenesis.

The molecular basis of the apparent limitation on somatic instability is unknown. One hypothesis proposes that methylation has a stabilizing influence on full mutation alleles (Wohrle et al. 1996). According to this model, variation in repeat length would occur in an early developmental window when the genome as a whole is largely undermethylated (Razin and Kafri 1994; Wohrle et al. 1996). Subsequent *de novo* methylation of the repeats at about the time of blastocyst implantation would fix in place any size variability that had occurred. Some experimental evidence in support of this model has been derived from Southern blot analysis of unmethylated alleles that shows these alleles tend to be more diffuse and heterogeneous than methylated counterparts (Wohrle et al. 1998; Glaser et al. 1999). In addition, post-mortem analysis of one full mutation carrier with a partially unmethylated expansion had identified inter-tissue variability in mutation patterns (Taylor et al. 1999). A portion of chapter three of this thesis addresses the proposed stabilizing influence of methylation on full mutation alleles (Burman et al. 1999a).

CGG Repeat Instability in Model Organisms

To study the molecular basis of repeat instability, investigators have in many cases turned to model organisms. The two most frequently utilized organisms are *Escherichia coli* and *Saccharomyces cerevisiae*. Although most of the pioneering work in this field was performed in systems that contained CAG/CTG repeats, more recent work has assessed the behavior of CGG/GCC triplets. These recent investigations have

examined the effect of many of the same variables originally studied in the CAG/CTG systems (e.g. repeat length, replication direction, etc.). One prominent limitation in these systems has been an inability to clone alleles greater than ~74 CGG repeats (Shimizu et al. 1996; Hirst and White 1998). For this reason, analysis has been restricted to relatively small alleles (Hirst and White 1998) or larger CGG tracts interrupted in multiple locations by CAG and AGG interspersions (Shimizu et al. 1996). Because these interruptions occur in regions of the repeat that are uninterrupted in humans (e.g. middle and both ends of repeat tract), some caution is warranted when interpreting the results. Despite these shortcomings, analysis of CGG tracts replicated in *E. coli* have demonstrated that instability was a common event and resulted primarily in deletion products. For both interrupted and non-interrupted repeats, stability varied strongly with length of the repeat tract such that longer repeats showed a greater degree of instability than shorter repeats (Shimizu et al. 1996; Hirst and White 1998). Stability was also demonstrated to depend on the orientation of DNA replication. This suggests that the CGG (G-rich) and CCG (C-rich) strands have a different propensity for mutational events as leading and lagging strand templates (Shimizu et al. 1996; Hirst and White 1998). In two independent studies, the authors concluded that instability is greater when the G-rich strand is the lagging strand template (Shimizu et al. 1996; Hirst and White 1998). The different mutational potential of the C/G-rich strands is commonly explained as a consequence of different propensities for forming DNA secondary structures as discussed below in the subsection "Current Models of Triplet Repeat Instability".

Analysis of normal and premutation (up to 74 repeats) sized repeat arrays in *S. cerevisiae* has yielded results quite similar to those observed in *E. coli*. Instability of the CGG repeat is common in yeast and can produce both expansion and contractions, although contractions were up to ~30 times more common than expansions (Balakumaran et al. 2000). The frequency of instability increased with repeat length and was dependent on the orientation of DNA replication (White et al. 1999; Balakumaran et al. 2000). Contraction events occurred in both orientations but were more frequent when the G-rich strand was the template for lagging strand synthesis (White et al. 1999; Balakumaran et al. 2000). Interestingly, the occurrence of expansions between 5 and 40 repeats were observed when the G-rich strand was the newly synthesized lagging strand (White et al. 1999). The authors believe this may indicate that G-rich secondary structures mediate both expansion and contraction events. To address whether repeat stability was increased during meiosis, hemizygous diploid strains carrying a 58 repeat allele were created and the length of the array in haploid progeny from a single round of meiotic division was assessed (White et al. 1999). The frequency of repeat length changes was found to be the same as that observed in mitotically dividing cells, apparently indicating no meiosis-specific effect on stability. Further analysis in yeast has examined the role or trans-acting factors such as mismatch repair proteins and the Rad27 deoxyribonuclease (White et al. 1999). White and colleagues (1999) reported no increase in repeat length variation in *mlh1*, *msh2*, *msh3*, or *msh6* mutant backgrounds. However, significant destabilization of the CGG repeat was observed in a *rad27* mutant background with an increased frequency of both expansions and contractions (White et

al. 1999). This observation is consistent with a previous report of increased CAG repeat instability in other *rad27* mutant strains (Freudenreich et al. 1998; Schweitzer and Livingston 1998). Because the RAD27 protein is a deoxyribonuclease required for removal of excess nucleotides at the *5'*-ends of Okazaki fragments (Gordenin et al. 1997), these results suggest a role for lagging strand events in triplet repeat instability.

Analysis of repeat dynamics in *E. coli* and *S. cerevisiae* has clearly provided valuable insight on cis- and trans-acting factors that have the potential to influence CGG repeat stability. Nonetheless, investigators have worked to create a mouse model of CGG repeat dynamics that would allow analysis of variables that might be unique to mammals. Three separate reports of transgenic mice that carry premutation sized CGG repeats have been published in the literature (Bontekoe et al. 1997; Lavedan et al. 1997; Lavedan et al. 1998). Despite transmission of the premutation alleles to a combined number of progeny in excess of 500, no intergenerational or somatic variability in repeat length was observed. Two of the transgenic models carried CGG repeats with multiple potentially stabilizing interspersions (Bontekoe et al. 1997; Lavedan et al. 1997). However, the more recently reported transgenic, which is arguably the most likely to exhibit instability, contained a large 120 repeat premutation with 97 uninterrupted CGG repeats at the *3'* end of the repeat tract (Lavedan et al. 1998). Ninety-seven uninterrupted CGG repeats is well in excess of the number estimated to confer both any degree of instability (34-37 repeats) and instability that results in premutation to full mutation expansion in one generation (56-75 repeats) in humans (Eichler et al. 1994). A variety of explanations have been invoked to explain the lack of

instability in the transgenic mice. One possibility is that dynamic mutation is a human specific phenomenon (Lavedan et al. 1998). Although there is a relatively small CGG repeat tract at the murine *fmr1* locus, instability in this repeat has not been reported. Furthermore, dynamic mutations at other loci have not been observed in any high eukaryotes, including fruit fly and mouse (Lavedan et al. 1997). It is possible that species-specific differences in DNA replication and repair could account for dynamic mutation in man but not other organisms. Another possibility that may explain repeat stability in the transgenic mice is that the chromosomal context of the repeat influences mutability (Bontekoe et al. 1997; Lavedan et al. 1997; Lavedan et al. 1998). The large 120 repeat transgene described in the study by Lavedan and colleagues (1998) was flanked by 831 bp of human *FMR1* genomic sequence (Lavedan et al. 1998). The failure of this allele to expand indicates that this amount of sequence context is apparently not sufficient to confer instability. Because the transgenes were integrated in a small number of random sites, other integration sites may exist that have the potential to confer instability (Lavedan et al. 1997).

One concrete conclusion that can be drawn from the transgenic mouse studies is that uninterrupted premutation sized CGG repeats alone are not sufficient to induce instability in mice. This finding indicates the importance of creating a higher eukaryotic model system in which to study CGG repeat dynamics. Chapter three of this thesis describes analysis of fragile X repeat dynamics in an *in vitro* model system based on somatic cell hybrids (Burman et al. 1999a).

Current Models of Triplet Repeat Instability

The molecular mechanism(s) that underlie triplet repeat stability are not well understood. Although the study of repeat variability in model organisms has started to address mechanistic issues, much remains to be learned as demonstrated by the failure of expansion in transgenic mice. Numerous models have been proposed to explain the basis of triplet repeat instability. While each of the models is supported by some experimental evidence, it remains to be determined which most closely reflects events that occur *in vivo*.

Because of the strong parent of origin effect observed in transmission of the fragile X repeat, some models of repeat instability have postulated a role for unequal homologous recombination during meiosis in the female (La Spada 1997). Meiotic recombination seems an unlikely mechanism in the case of the fragile X syndrome because the prediction of equal frequencies of expansion and contraction events is not met and linkage disequilibrium has been demonstrated in the region of the CGG repeat (Chiurazzi et al. 1996). If unequal meiotic recombination was responsible for variation in repeat length, then associations between the expanded repeat and particular alleles at flanking markers would be randomized over evolutionary time thus eliminating the disequilibrium. Although meiotic recombination is probably not the basis of fragile X expansion, it remains possible that mitotic recombination during either gametogenesis or early embryogenesis may be involved in generating repeat length variability (La Spada 1997).

Perhaps the most commonly cited model of repeat instability is that of DNA slippage and misalignment during lagging strand synthesis (Richards and Sutherland 1994) (figure 4). Because lagging strand synthesis results in segments of DNA that are transiently single stranded, the leading and lagging strands may have different mutational potential (Richards and Sutherland 1994). Experimental evidence in support of this unequal potential has been provided in part by the orientation-dependence of repeat stability in *E. coli* and *S. cerevisiae* (Kang et al. 1995; Maurer et al. 1996; Freudenreich et al. 1997; Hirst and White 1998) and the apparent role of the RAD27 protein, which is involved in Okazaki fragment maturation on the lagging strand (Freudenreich et al. 1998; Schweitzer and Livingston 1998; Spiro et al. 1999). Eichler and colleagues (1994) noted that the length of pure CGG repeats sufficient for expansion from premutation to full mutation in one generation (~56-75 repeats) is near the average length of an Okazaki fragment (~150-200 bp) (Eichler et al. 1994). Thus fragile X alleles with greater than ~70 pure CGG repeats risk beginning and ending DNA synthesis within a continuous tract of uninterrupted CGG repeats (Eichler et al. 1994). Newly synthesized fragments without anchor points in unique DNA sequence might be particularly vulnerable to slipped structures and expansion.

Most models of DNA slippage include the formation of DNA secondary structures, such as hairpins, in the slipped DNA. The formation of secondary structure is theorized to be important to stabilize the slipped DNA by minimizing the energy difference between the duplex and slipped states (McMurray 1995). DNA secondary structure might also contribute to instability by stalling the replication fork

(Samadashwily et al. 1997) thereby promoting reiterative DNA synthesis (Sinden 1999). A large number of secondary structures have been reported in CGG/CCG repeats *in vitro* (Darlow and Leach 1998). Significantly, the C-rich and G-rich strand demonstrate different propensities for forming secondary structure (Darlow and Leach 1998), which may contribute to the dependence of repeat instability on the orientation of DNA replication (Kang et al. 1995; Maurer et al. 1996; Freudenreich et al. 1997; Hirst and White 1998). In all model systems so far utilized to study CGG repeat dynamics, a strong relation between the length of the repeat tract and degree of instability has been observed (Hirst and White 1998; White et al. 1999; Balakumaran et al. 2000). Because the likelihood of formation and stability of hairpins increases with the length of the repeat tract (Gacy et al. 1995; Gacy and McMurray 1998), the strong correlation between repeat length and degree of instability may be mediated by hairpin formation (McMurray 1995).

Thesis Rationale

Rationale for Chapter 2: "Hypomethylation of an Expanded *FMR1* Allele is Not Associated with a Global DNA Methylation Defect"

Background

Expansion of the fragile X CGG repeat in excess of ~200 triplets is almost always accompanied by extensive hypermethylation of the repeat and a surrounding CpG island. Hypermethylation is associated with transcriptional silencing of the *FMR1* gene and expression of the fragile X syndrome phenotype. Despite the strong correlation between expansion and methylation, it is not understood why expanded repeats serve as a target for methylation.

The relation between repeat expansion and hypermethylation appears not to be absolute. A small number of non-retarded full mutation carriers have been described who carry expansions that are largely or entirely free of methylation at diagnostic restriction sites. Because methylation in these individuals has been assessed using only conventional methylation-sensitive restriction enzymes, the extent of the methylation abnormality has not been defined. Furthermore, it is unknown why some full mutations appear to escape the methylation process.

Analysis of the Extent of FMR1 Hypomethylation in patient MK

Study Design: In collaboration with Dr. Peter Jacky, I obtained peripheral blood lymphocytes and skin fibroblasts from an adult male who carries a hypomethylated full mutation. I investigated the extent of *FMR1* methylation in these samples with

conventional restriction enzymes and the McrBC endonuclease. **Rationale:** Analysis of *FMR1* methylation with McrBC was a novel approach that had the advantage of assessing methylation at numerous sites in the CpG island, including those located within the CGG repeat.

Analysis of the Biological Basis of FMR1 Hypomethylation in patient MK

Study Design: To investigate the biological basis for the absence of methylation in MK's full mutation, I analyzed methylation at selected repetitive element loci. This analysis was carried out with conventional methylation-sensitive restriction enzymes and novel Southern blot probes generated by PCR. **Rationale:** Methylation of repetitive element loci was studied to gain insight into the maintenance and formation of methylation patterns in patient MK. Methylation of *Alu* elements in the vicinity of the *FMR1* CGG repeat was studied to determine if methylation patterns are maintained normally in the region of the CGG repeat. Methylation of repetitive elements that were likely to have been subject to *de novo* methylation during MK's early development were studied to determine if the absence of *FMR1* methylation was attributable to a *trans* defect that affects *de novo* methylation of other loci.

Study Design: To further investigate the biological basis of MK's hypomethylated full mutation, I used the microcell fusion technique to transfer an unmethylated expansion allele into a mouse embryonal carcinoma (EC) cell line known to be *de novo* methylation competent. **Rationale:** By transferring MK's unmethylated full mutation allele into a cell line capable of *de novo* methylation, I sought to determine if the absence of *FMR1* methylation could be complemented in *trans*. This approach was

suggested by studies that showed the methylation deficiency present in the ICF syndrome could be at least partially complemented in *trans* with whole-cell fusions.

Contributions of this Study to the Fragile X Field

These studies have contributed information to the fragile X field that for the first time addresses the extent and biological basis of hypomethylated full mutations. I have defined the extent of hypomethylation by showing that the entire CpG island, including the CGG repeat, can be free of methylation. This result is important in part because it justifies the use of these alleles in the study of the proposed link between CGG repeat methylation and repeat stability (addressed in Chapter 2 of this thesis). I have also shown that failed maintenance methylation in the *FMR1* region does not explain the existence of MK's hypomethylated full mutation. Furthermore, my results demonstrate that if the methylation abnormality in MK is due to a defect in *trans*, then *de novo* methylation of *Alu* and *L1* elements is likely to be mediated by pathways other than that responsible for *de novo* methylation of the expanded *FMR1* CGG repeats. Finally, I show that the methylation deficiency in patient MK is not amenable to trans-complementation. This result suggests that MK's chromosome may contain a variation in *cis* that renders the CGG repeat resistant to *de novo* methylation, or that expanded CGG repeats do not serve as targets for *de novo* methylation in cultured murine EC cells.

Individuals Who Contributed to this Study

This study was carried out under the tutelage of Drs. Brad Popovich and Mitch Turker. Dr. Peter Jacky contributed to this work by obtaining blood and tissue samples

from patient MK. Dr. Phil Yates contributed to the analysis of *FMR1* methylation by suggesting the use of the McrBC enzyme and engaging in numerous insightful conversations. Lindsay Green contributed to the analysis of repetitive element methylation by performing a Southern blot and generating two probes by PCR. The contribution of all these individuals is acknowledged by authorship on the published manuscript. With the exception of the above-mentioned contributions, Robert Burman carried out all other aspects of this study.

Rationale for Chapter 3: "Fully expanded *FMR1* CGG repeats exhibit a length- and differentiation-dependent instability in cell hybrids that is independent of DNA methylation"

Background

The fragile X CGG repeat is potentially unstable both when transmitted between generations and in the somatic cells of an individual. The molecular basis of repeat instability has not been determined and it is unknown if intergenerational and somatic instability share in common any mechanistic similarities. Two primary determinants of intergenerational instability are the sex of the transmitting parent and the size of the repeat. Variables that influence instability in somatic cells have not been defined, however it is recognized that the large majority of full mutations are somatically unstable. Interestingly, this instability appears to be limited to a period during early embryogenesis. Thus a demarcation appears to exist between embryonic cells in which the repeats are unstable and adult cells where the repeats are maintained with little or no

variation in length. Although the molecular basis of this boundary has not been described, full mutation alleles are believed to be methylated early in development and it has been postulated that methylation may stabilize the repeats.

To describe variables that affect CGG repeat dynamics, investigators have assessed the behavior of the repeat in *E. coli* and *S. cerevisiae*. The main determinants of stability in these organisms were shown to be length of the repeat tract, direction of DNA replication through the repeats, and presence of a functional RAD27 enzyme (involved in processing Okazaki fragments). Investigators have also sought to study repeat dynamics in transgenic mice that carry CGG repeats in the premutation size range. Surprisingly, the CGG repeat was found to be completely stable both intergenerationally and somatically in transgenics produced in three independent studies. Based on these results, it has been hypothesized that CGG repeat instability is a human specific phenomenon or perhaps critically dependent on chromosomal context.

Together, the results mentioned above indicate that stability of the CGG repeat may be mediated by species-specific factors that would best be studied in a mammalian model system in which the repeat is maintained in its native chromosomal context.

Analysis of Repeat Instability in Human Fibroblasts

Study Design: I investigated the stability of the fragile X CGG repeat in cultured primary fibroblasts derived from adult full mutation carriers. I isolated clonal derivatives of the cell lines and therein assessed repeat stability by Southern blot analysis. **Rationale:** I sought to test the hypothesis that unmethylated fragile X CGG repeats are inherently unstable compared to their methylated counterparts. By analyzing

clonal cell lines, I was able to determine if methylated and unmethylated expansions of similar size are maintained with different fidelity.

Analysis of Repeat Instability in Cell Hybrids

Study Design: I studied the stability of the fragile X CGG repeat in human-mouse whole cell hybrids. I assessed stability by Southern blot analysis of genomic DNA samples derived from clonal hybrid cell lines. **Rationale:** The study of repeat dynamics in human fibroblasts is inherently limited by the cell's finite replicative capacity. I sought to create a system in which the behavior of the CGG repeat could be studied over a large number of cell divisions, in different cellular backgrounds, and within the native human chromosomal context. I was able to accomplish these goals by generating whole cell hybrids that contained human fragile X chromosomes.

Contributions of this Study to the Fragile X Field

The aspect of this study that addressed repeat stability in human fibroblasts has contributed to the fragile X field by providing the best evidence so far published that methylation of the CGG repeat correlates with stability. The analysis of repeat stability in cell-hybrids has also made a number of potentially important contributions. In a broad sense, this study has validated the use of cell hybrids as a convenient means of "immortalizing" fragile X chromosomes for the study of repeat dynamics. Contrary to previous predictions, I show that CGG repeats can be unstable in a mouse background and that methylated and unmethylated repeats are similarly unstable. This instability is size-dependent, with a threshold for instability that is remarkably similar to that observed in humans. My finding that methylation is not sufficient to confer repeat

stability in the cell hybrids suggests that the stability of methylated repeats in human cells is not explained by a physical difference between methylated and unmethylated CGG repeat-containing DNA. I also show that an unexpected determinant of repeat stability in the cell hybrids is cellular differentiation. Thus the length of the repeat tract alone does not determine instability in this system. Rather, is seems that a "higher-order" cellular interaction, perhaps between the repeat and components of the DNA metabolizing apparatus, influences repeat dynamics.

Individuals Who Contributed to this Study

This study was carried out under the tutelage of Drs. Mitch Turker and Brad Popovich. Dr. Peter Jacky contributed to this work by obtaining blood and tissue samples from patient MK, which provided a source of unmethylated full mutation alleles. The contribution of all these individuals is acknowledged by authorship on the published manuscript. With the exception of the above-mentioned contributions, Robert Burman carried out all other aspects of this study.

Rationale for Chapter 4: "Fragile X full mutations are more similar in siblings than in unrelated patients: further evidence for a familial factor in CGG repeat dynamics"

Background

The large majority of fragile X full mutation carriers demonstrate repeat size mosaicism due to somatic instability of the CGG repeat during early embryogenesis. This size mosaicism is manifest on Southern blot analysis as a complex smeared and

poly-banded hybridization pattern. It is not known if somatic instability occurs as a consequence of expansion from an inherited premutation allele or retraction of an inherited full mutation. Although analysis of oocytes from a premutation carrier has not yet distinguished definitively between these two possibilities, most experimental evidence favors a model in which a full mutation is inherited and subsequently retracts to varying degrees producing size mosaicism.

Some experimental evidence indicates that a familial factor(s) may influence CGG repeat dynamics. In particular, a 1996 study by Nolin and colleagues demonstrated that within sibships containing only premutation carriers, the size of the premutations are significantly more similar within families than between families. Whether a familial factor may likewise influence repeat dynamics in full mutation carriers has not been determined.

Analysis of Mutation Pattern Similarity in Full Mutation Sibships

Study Design: I systematically studied the mutation pattern of 56 full mutation carriers previously identified in the OHSU DNA Diagnostic lab. Analysis of mutation patterns was carried out by Southern blot analysis. **Rationale:** To test the hypothesis that full mutations are more similar in siblings than in unrelated controls, I compared mutations patterns in 16 sibling pairs and 15 pairs of randomly matched unrelated individuals. I utilized two novel comparison strategies specifically designed to overcome the difficulties inherent in comparing complex mutation patterns.

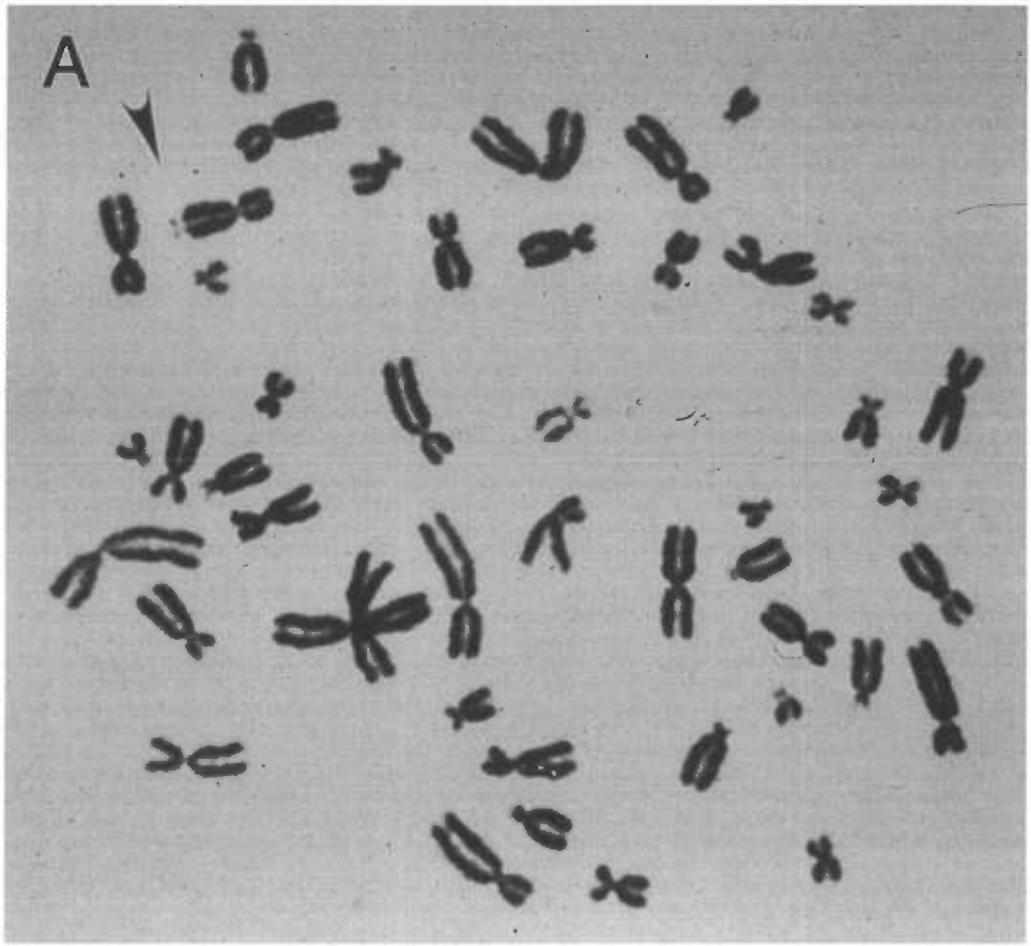
Contributions of this Study to the Fragile X Field

The primary contribution of this study to the fragile X field is the addition of further evidence supporting the hypothesis that familial factors may influence CGG repeat variability. The apparent existence of a familial factor that influences full mutation stability further suggests that the process that generates length variation occurs in a non-random manner. Whether the similarity that results from this process is mediated in *trans* by other genes or perhaps in *cis* by defined DNA secondary structures, remains to be determined. In either case, this data indicates that careful analysis of mutation patterns could potentially provide mechanistically relevant information.

Individuals Who Contributed to this Study

This study was carried out under the tutelage of Dr. Brad Popovich. Kim Anoe contributed to this work by initially suggesting the possibility that allele size appeared to be similar in some sibships. The contribution of both these individuals is acknowledged by authorship on the manuscript. In addition, Dr. Will Bloch contributed to this study by assessing similarity in some of the families with an independent PCR assay (which did not contribute to the data presented here). Dr. Bloch's data provided the impetus for this study by first suggesting the possibility of biologically relevant allele sharing in fragile X siblings. Finally, Drs. Mitch Turker, Bob Wildin, Peter Jacky and Charles Brenner contributed with helpful discussions. The input of these individuals is acknowledged in the acknowledgements section of the manuscript. The DNA samples used in this study were isolated by the staff of the OHSU DNA Diagnostic Laboratory.

With the exception of the above-mentioned contributions, Robert Burman carried out all other aspects of this study.



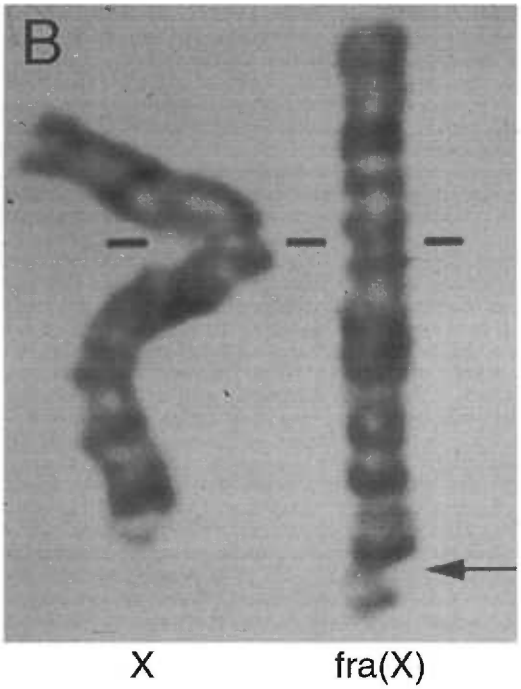


Figure 1. Expression of the fra(X) cytogenetic marker.

A Homogeneously stained metaphase spread with a fra(X) site marked by an arrow head. **B** Closeup of normal X and fra(X) chromosomes. Note the fra(X) constriction marked by an arrow. Both photographs are courtesy of the National Fragile X Foundation.

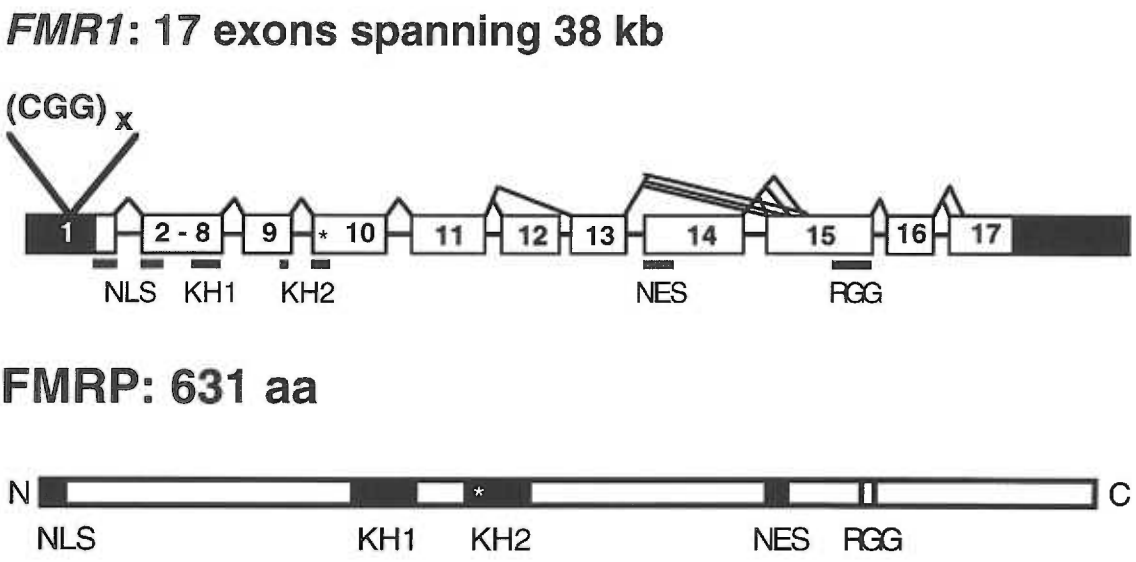


Figure 2. Schematic representations of the *FMR1* gene and protein.

Upper: Exons 1-17 of the *FMR1* gene with alternative splicing indicated by the upper lines. The *5'* and *3'* untranslated regions (UTRs) are shaded. Note the CGG repeat sequence in the *5'* UTR. **Lower:** 631 amino acid *FMR1* protein (FMRP). Regions of the gene thought to encode important functional domains of the protein are indicated. The KH1, KH2, and RGG domains are conserved in some proteins that bind RNA. The functional importance of the second KH domain (KH2) has been demonstrated by a point mutation (*) that results in an Ile304Asn substitution in one severely affected patient with a non-expanded CGG repeat (De Boulle et al. 1993). The positions of a nuclear localization signal (NLS) and a nuclear export signal (NES) are also shown.

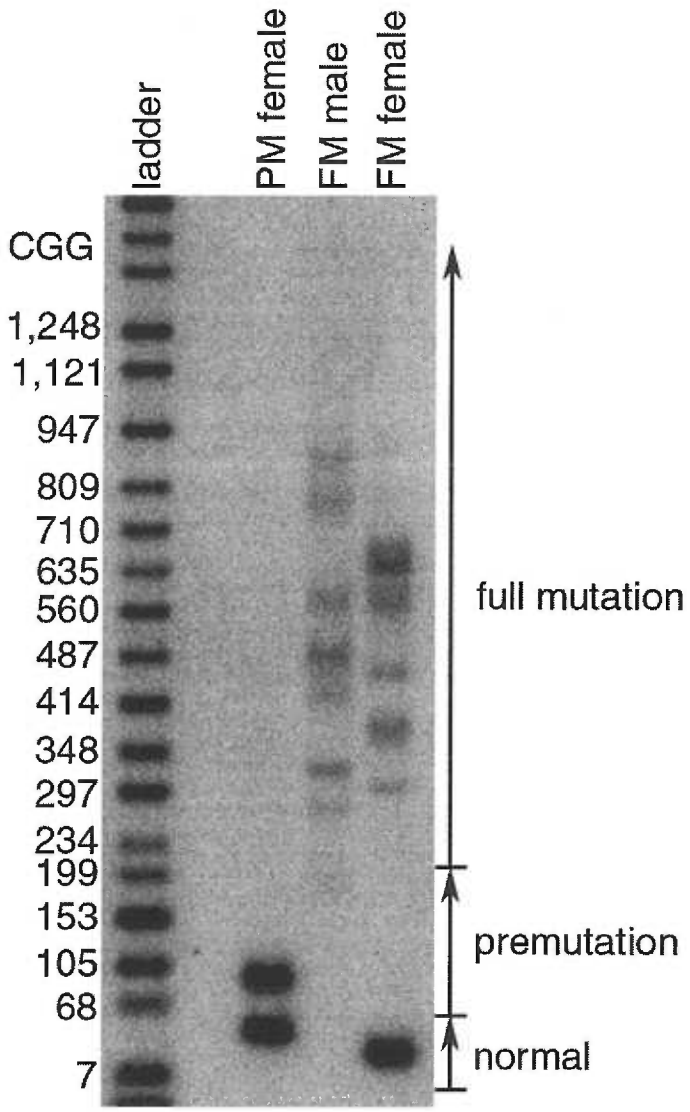


Figure 3. *FMR1* alleles and size mosaicism.

Southern blot of lymphocyte-derived DNA samples digested with *Pst*I and hybridized with probe *pfxa3*. PM female: female premutation carrier with one allele in the normal size range and the other allele a premutation of ~ 90 repeats. FM male: full mutation male with size mosaicism due to somatic instability. FM female: full mutation female with mosaic expansion and a normal allele. Alleles in the normal and premutation size range are stable in somatic cells and consequently do not display size mosaicism.

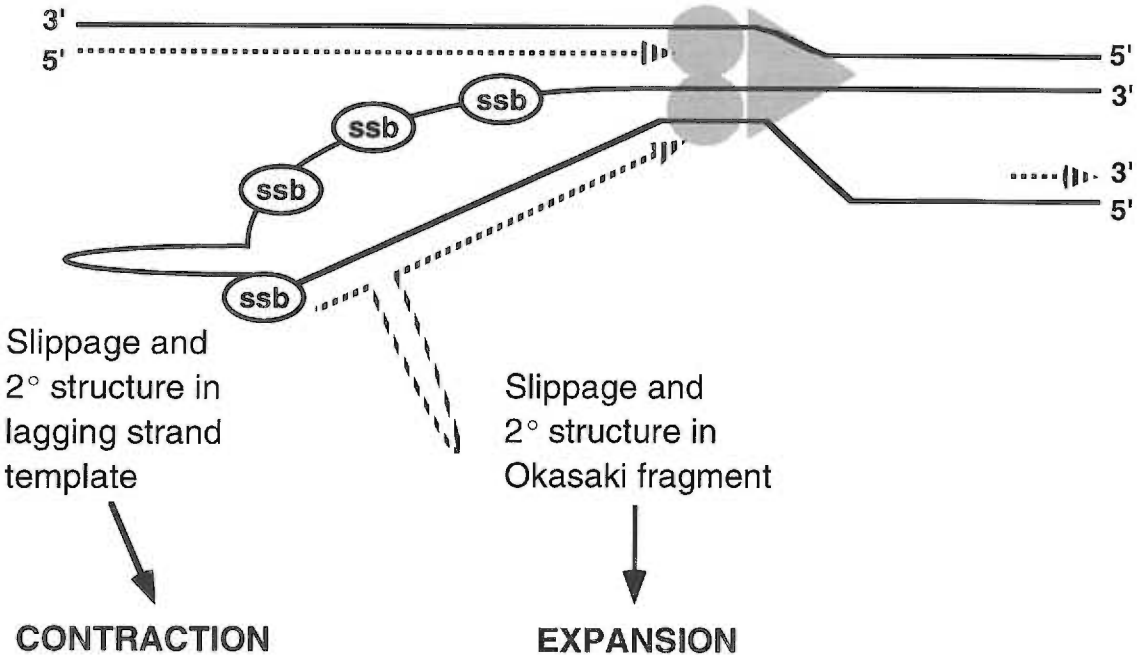


Figure 4. Lagging-strand slippage and misalignment model of repeat instability. The replication fork is shown in the "concurrent model" in which both leading and lagging strands are synthesized simultaneously by a dimeric polymerase complex. The newly synthesized strands are represented by dashed lines. **Contraction:** secondary structure in the template strand will be bypassed by the DNA polymerase resulting in a loss of repeat units equal to the number of repeats extruded in the structure. **Expansion:** slippage and misalignment of the nascent strand, perhaps mediated by secondary structure, results in an increase of repeat units. The formation of secondary structure is believed to stabilize slippage events and increase the lifetime of the slipped state, thereby increasing the likelihood of a change in repeat length (McMurray et al. 1995). This figure is adapted from Gacy and McMurray (1998).

CHAPTER 2

HYPOMETHYLATION OF AN EXPANDED *FMR1* ALLELE IS NOT ASSOCIATED WITH A GLOBAL DNA METHYLATION DEFECT

Robert W. Burman[1,2], Phillip A. Yates[3], Lindsay D. Green[1], Peter B. Jacky[4], Mitchell S. Turker[1,3], and Bradley W. Popovich[1,2*]

[1]Department of Molecular and Medical Genetics, [2]DNA Diagnostic Laboratory, and [3]Center for Research on Occupational and Environmental Toxicology, Oregon Health Sciences University, Portland; and [4]Cytogenetics Laboratory, Kaiser Permanente NW, Clackamas, OR

American Journal of Human Genetics **65**: 1375-1386, 1999

SUMMARY

The vast majority of fragile X full mutations are heavily methylated throughout the expanded CGG repeat and the surrounding CpG island. Hypermethylation initiates and/or stabilizes transcriptional inactivation of the *FMR1* gene, which causes the fragile X syndrome phenotype characterized primarily by mental retardation. The relation between repeat expansion and hypermethylation is not well understood, nor is it absolute as demonstrated by the identification of non-retarded males who carry hypomethylated full mutations. To better characterize the methylation pattern in a patient who carries a hypomethylated full mutation of ~60 to ~700 repeats, we have evaluated methylation with the McrBC endonuclease, which allows analysis of numerous sites in the *FMR1* CpG island including those located within the CGG repeat. We report that the expanded repeat region is completely free of methylation in this full mutation male. Significantly, this lack of methylation appears to be specific to the expanded *FMR1* CGG repeat region because various linked and unlinked repetitive element loci are methylated normally. This result demonstrates that the lack of methylation in the expanded CGG repeat region is not associated with a global defect in methylation of highly repeated DNA sequences. We also report that *de novo* methylation of the expanded CGG repeat region does not occur when it is moved via microcell mediated chromosome transfer into a *de novo* methylation competent mouse embryonal carcinoma cell line.

INTRODUCTION

The fragile X syndrome (MIM 309550) is an X-linked mental retardation disorder with an incidence of approximately one in 6,000 (Turner et al. 1996; de Vries et al. 1997; Morton et al. 1997). In most patients, amplification of an unstable CGG repeat located in the 5' untranslated region of the *FMR1* gene is responsible for the syndrome (Kremer et al. 1991a; Oberle et al. 1991; Verkerk et al. 1991; Yu et al. 1991). In the normal population, *FMR1* alleles contain between 5 and ~55 repeats, are stable on transmission, and are not methylated except when located on an inactive X chromosome in females (Bell et al. 1991; Fu et al. 1991; Hansen et al. 1992; Hornstra et al. 1993). Premutation alleles with ~55 to ~220 repeats are found in unaffected carriers and, like normal alleles, are not methylated unless on an inactive X chromosome. When transmitted by a female, premutation alleles can expand to become disease causing full mutations (Fu et al. 1991; Heitz et al. 1991; Oberle et al. 1991; Snow et al. 1993).

Full mutation alleles contain from ~220 to greater than 1,000 repeats and are almost always associated with extensive hypermethylation of the CGG repeat and a surrounding CpG island (Heitz et al. 1991; Oberle et al. 1991; Pieretti et al. 1991; Hansen et al. 1992; Sutcliffe et al. 1992; Hornstra et al. 1993). Hypermethylation of promoter elements (Pieretti et al. 1991; Sutcliffe et al. 1992) and histone deacetylation (Coffee et al. 1999) are associated with transcriptional silencing of the *FMR1* gene, presumably by interfering with transcription factor binding (Schwemmle et al. 1997). The resulting lack of FMR1 protein (FMRP) is believed sufficient to cause the fragile X syndrome phenotype (Pieretti et al. 1991; Devys et al. 1993; Siomi et al. 1993;

Hammond et al. 1997). Although repeat expansions in excess of ~220 triplets are typically accompanied by hypermethylation, unusual expansions have been described in which full mutation alleles are unmethylated at diagnostic restriction sites (Loesch et al. 1993; McConkie-Rosell et al. 1993; Hagerman et al. 1994; Merenstein et al. 1994; Rousseau et al. 1994b; Smeets et al. 1995; de Vries et al. 1996; Lachiewicz et al. 1996; Wang et al. 1996; Wohrle et al. 1998; Taylor et al. 1999). Male carriers of these mutations are often described as high functioning and can have cognitive and physical phenotypes in the normal to mildly affected range. FMRP expression has been reported in many of these patients, although at levels reduced relative to normal controls. This reduction may be due to translational suppression (Feng et al. 1995). *FMR1* methylation in these individuals has been studied using methylation sensitive restriction endonucleases, which limits the number of methylation sites that can be analyzed. Although FMRP expression suggests that regulatory elements are also unmethylated, the methylation status of most sites in the CpG island including the expanded CGG repeat have not been described.

The association between repeat expansion and methylation is not well understood and in particular it is unknown how some full mutation alleles escape the methylation process. Reports of typically methylated full mutations in the grandsons of males bearing hypomethylated full mutations favor a role for *trans* acting factors rather than a heritable *cis* effect (Smeets et al. 1995; Lachiewicz et al. 1996; Wohrle et al. 1998). The study of methylation variants at the fragile X locus will help define the basic relation between repeat expansion and methylation and may also provide information

about the potential efficacy of therapeutic approaches that utilize demethylating agents (Chiurazzi et al. 1998). Moreover, a description of *FMR1* methylation in hypomethylated full mutation males is a necessary prerequisite for the study of the proposed influence of methylation on repeat stability (Wohrle et al. 1996).

The aim of the present work was to confirm that hypomethylation at the diagnostic *Eag*I site extends into the expanded CGG repeat region and to determine if this lack of methylation is indicative of a *trans* defect in *de novo* and/or maintenance methylation. We have assessed methylation at a large number of sites in the *FMR1* CpG island including the CGG repeat and at various repetitive DNA elements. Our data demonstrate an absence of methylation throughout the CpG island despite normal methylation levels at flanking *Alu* elements and at other repetitive element loci. The methylation deficit in the expanded CGG repeat region was not corrected when the human X chromosome carrying this allele was transferred into a *de novo* methylation competent mouse cell line. These results argue against a model in which hypomethylation of full mutation *FMR1* alleles is due to a *trans* defect in DNA methylation.

SUBJECTS & METHODS

Patient Description

MK is a 39 year-old male who was evaluated for a fragile X mutation based on a positive family history. MK's sister, mother, and maternal aunt are known to be premutation carriers. The maternal aunt has two sons who are both reported to carry

fragile X full mutations. One cousin is reported to be mentally retarded and to carry a typical methylated full mutation. The other cousin, like MK, is reported to carry an unmethylated or partially unmethylated full mutation. In addition, MK has six other siblings all of whom carry *FMR1* alleles in the normal size range.

Cognitively, MK appears to be quite "high-functioning". He graduated from high school and is raising a family with two sons. He has been in the military where he successfully completed specialized training. He is currently employed as a drug and alcohol counselor. MK has good eye contact and comfortably engages in conversation. Based on a short interview that did not include a physical evaluation or formal cognitive testing, he does not appear to be affected with the fragile X syndrome either cognitively or behaviorally.

DNA Isolation and Analysis

Genomic DNA was isolated using either a standard phenol/chloroform extraction or the Puregene DNA Isolation Kit (Gentra Systems). Aliquots (10-30μg) were digested with restriction endonucleases purchased from either Boehringer Manheim: *Eco*RI, *Pst*I, *Msp*I, *Hin*dIII or New England Biolabs: *Eag*I, McrBC, *Hpa*II, *Bst*BI. *Eco*RI/*Eag*I double digests were performed simultaneously with both enzymes. All other double digestions were performed sequentially with an initial digestion using *Pst*I, *Eco*RI, or *Hin*dIII followed by precipitation and then a second digestion with McrBC, *Msp*I, *Hpa*II, or *Bst*BI. Digested DNA was precipitated with sodium acetate and ethanol and then aliqouts of 1.5-2.5 μg were separated by elecrophoresis in 1%

agarose/TAE and transferred to Biodyne B nylon membrane (Gibco BRL) with 5X SSC. Hybridizations with a ^{32}P radiolabelled probe (Boehringer Manheim; Random Prime Labelling Kit) were carried out at 65°C in Church Buffer with BSA (hybridization buffer II) (Strauss 1998) supplemented with 100 µg/ml heat-denatured herring sperm DNA. Probe *pfxa3* (a kind gift from David L. Nelson, Baylor College of Medicine, Houston), a 558 bp *Xho*I-*Pst*I fragment of pE5.1 (Figure 5A) (Fu et al. 1991; Verkerk et al. 1991), was used to detect restriction fragments containing the CGG repeat. Additional probes used to assess methylation at the repetitive element loci are described below. Membranes were washed twice at room temperature in low stringency wash buffer II (Strauss 1998) and then twice at 68.5°C in high stringency wash buffer II (Strauss 1998) diluted to 0.6X. Membranes were exposed sequentially to a phosphorimaging screen (Molecular Dynamics) and then x-ray film (Kodak; X-OMAT) at -70°C.

Analysis of Repetitive DNA Element Methylation

Repetitive element loci were identified in genomic DNA sequence with the Repeat Masker database (Smit and Green, see electronic database information). The *Alu* elements flanking the *FMR*1 CGG repeat as well as the *L1* and *SVA* elements are contained in the "HUMFMR1S DNA sequence" (accession number L29074). The positions of the *Alu*, *L1*, and *SVA* elements within the L29074 sequence are as follows: *Alu* element upstream of *FMR1* CGG repeat (7984-8274), *Alu* element downstream of

CGG repeat (18523-18804), *L1* element (84696-92586), and the *SVA* element (93455-95194). Hybridization probes flanking the elements were generated by PCR using the following primer pairs. Probe for *Alu* elment upstream of CGG repeat: forward primer (9881-9903) 5'-GAATGAGAGGTCATGGTTAAAGG-3', reverse primer (10137-10159) 5'-ATCTGATATTGGAATGATGCTTC-3'. Probe for *Alu* element downstream of CGG repeat: forward primer (18823-18849) 5'-AGGAGGGATATTTTACAATG CTGTAAG-3', reverse primer (19277-19301) 5'-TGTGATGAGAATCTTGGAAT TTGAG-3'. Probe between *L1* and *SVA* elements: forward primer (92687-92710) 5'-CCTCCATCATCTCCTCTCTTAAAG-3', reverse primer (93216-93241) 5'-GAGGT TAGAATTTTGTTAGGGGAGAG-3'. PCR amplifications were carried out in 25 µl reactions with 50 ng genomic DNA, 200µM dNTPs, 1.5 mM $MgCl_2$, 0.8 µM each primer, 1x PCR buffer II (Perkin-Elmer), and 1.25 U *Taq* DNA polymerase (Boehringer Manheim). Thermal cycling parameters were as follows (Stratagene Robocycler): 95°C for 4 min. for one cycle; 95°C for 1 min., 62°C for 1 min., 72°C for 2 min. for 30 cycles; 72°C for 10 min. for one cycle. PCR products were purified from agarose gels (Qiagen; Qiaex II Gel Purification Kit) and cloned (Invitrogen; TA Cloning Kit). The cloned probes were isolated from the vector by standard methods.

The *Alu* element examined on the Y chromosome is contained within the "Homo Sapiens 5' region *ZFY* gene sequence" (accession number U00242). This *Alu* element is located at nucleotides 4669 to 4954. The PCR-generated probe flanking this *Alu* was amplified with the following primer pair: forward primer (3594-3615) 5'-GCAGT

GTCGGCTACGCTTTAGG-3', and reverse primer (4638-4660) 5'-GCTACCTTCT TGATCATCCATCC-3'. PCR amplification conditions were the same as described except the annealing temperature was 66°C. This probe was also gel-purified, cloned, and then isolated from the vector.

Tissue Culture, Cell Lines, and Microcell Mediated Chromosome Transfer

Human fibroblast cell lines were established, with informed consent, from skin biopsies of adult fragile X full mutation carriers (TC38-89, MK) referred for clinical fragile X testing to either the Oregon Health Sciences University DNA Diagnostic Laboratory or the Kaiser Permanente Cytogenetics Laboratory. All human cell cultures were maintained at 37°C/5% CO_2 in α-minimal essential media (α-MEM) supplemented with 20% fetal calf serum (JRH Biosciences), 2mM L-glutamine, and 16 μg/ml gentamicin sulfate.

The DelTG3 mouse cell line is a thioguanine resistant clone isolated form the mouse P19-derived embryonal carcinoma cell line H4D2 (Turker et al. 1989a; Turker et al. 1989b). The Dif6 cell line is a morphologically differentiated and thioguanine resistant derivative of H4D2 (Turker et al. 1991). All mouse cell cultures were maintained at 37°C/5% CO_2 in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 5% fetal calf serum and 5% serum plus (JRH Biosciences).

Whole cell fusions were accomplished by mixing human fragile X fibroblasts and mouse cell line Dif6 to obtain ratios of 1:3, 1:1, and 3:1 with a total of $3x10^6$ cells

per 25-cm^2flask. After mixing, the cells were incubated for approximately six hours and fused by treatment for 60 or 80 seconds with 1 ml of a 50% polyethylene glycol (Sigma; PEG 1,450) solution in α-MEM. The cultures were rinsed three times with PBS, incubated overnight in non-selective media and split the next day to 150 mm dishes at a concentration of 1-2x10^5 cells per dish. Selection for fusion clones was applied approximately 48 hours post fusion with 10 μg/ml hyoxanthine (Sigma), 10 μg/ml azaserine (Sigma), and 1 mg/ml geneticin (Gibco BRL). Individual fusion clones were isolated 11-14 days after initiating selection and were then maintained in media supplemented with hypoxanthine and azaserine to retain the human X chromosome. Microcell fusions were performed essentially as described (Fournier 1981). Microcells were isolated from a whole cell hybrid clone and fused to DelTG3 by treatment for 60 seconds with a 50% PEG solution in α-MEM. Selection for the human X chromosome was accomplished with azaserine and hypoxanthine. Individual microcell hybrid clones were isolated 16 days after fusion and were expanded through approximately 24 population doublings until the cells were harvested for DNA isolation.

RESULTS

A full mutation with no CpG island methylation in DNA isolated from patient MK.

Analysis of methylation at the *FMR1* locus was performed initially by Southern blot with *Eco*RI/*Eag*I digested genomic DNA derived from patient MK's peripheral blood lymphocytes and cultured skin fibroblasts. Hybridization of the membrane with

the probe *pfxa3* resulted in the expected 2.8 kb and 5.2 kb fragments in a control female (Figure 5B, lane 2). The hybridization band at 2.8 kb represents normal unmethylated alleles with approximately 30 repeats. The 5.2 kb band represents the same normal alleles that have a methylated *Eag*I site due to X chromosme inactivation (Figure 5A). Peripheral blood DNA from the patient MK showed an unusual hybridization pattern with a large highly diffuse smear ranging in size from approximately 60 to 700 repeats (Figure 5B, lanes 3,4; also Figure 6A, lane 18). DNA derived from MK's cultured skin fibroblasts produced a hybridization pattern with less smearing, two major bands at the 230 and 330 repeat sizes, and a third faint band at 600 to 650 repeats (Figure 5B, lane 5; also Figure 6A, lane 20). Despite most of the alleles in both tissue types falling well within the full mutation size range, digestion with *Eag*I appeared complete, which is consistent with a complete lack of methylation at this site located 282 bp upstream of the CGG repeat (Figure 5A). Methylation at the *Eag*I site would have produced a hybridization pattern with a minimum size of 5.2 kb. Analysis of DNA samples from 37 clonal fibroblast lines derived from MK's skin culture also revealed no evidence of methylation at the *Eag*I site (data not shown). The allele sizes observed in these clones occured at frequencies proportionate to their relative abundance in the mass culture. Most clones contained alleles of 230 to 330 repeats and some contained large alleles of 600 to 650 repeats.

To assess more fully methylation levels throughout the CGG expansion and flanking regions, DNA preparations were digested with the McrBC endonuclease combination. The recognition sequence of the McrBC enzyme pair consists of two half

sites, each composed of a purine followed by methylcytosine. Since only cytosines within a CpG dinucleotide are eligible for methylation in mammals, an McrBC half site can be either A^{m}CG or G^{m}CG. Digestion of DNA occurs when two half sites are separated by 32 bp to 2 kb, with optimal separation at 55 to 103 bp (Stewart and Raleigh 1998). There are 48 potential half sites and an additional half site at each CGG repeat within the *Pst*I restriction fragment that contains the repeat and most of the CpG island. Analysis of methylation at the McrBC sites was performed by Southern blot of DNA samples digested with either *Pst*I alone or *Pst*I followed by McrBC. Hybridization was again carried out with probe *pfxa3*. DNA samples derived from two control males and two premutation males showed identical hybridization patterns in samples digested with *Pst*I alone or with *Pst*I+McrBC (Figure 6A, lanes 2-9). The failure of McrBC to digest these DNA samples is consistent with an absence of methylation. In contrast, DNA preparations from four full mutation males were digested completely by McrBC as demonstrated by the absence of hybridization bands in these lanes (Figure 6A, lanes 11, 13, 15, 17). Although it is not possible to determine which specific sites are methylated and serving as half sites in these samples, the absence of a hybridization signal in the *Pst*I+McrBC lanes is consistent with substantial methylation. A photograph of the ethidium bromide stained gel taken before Southern blotting shows DNA content in these lanes was indistinguishable from the other *Pst*I+McrBC lanes (figure 6B). DNA samples from the patient MK were not digested by McrBC and produced hybridization patterns identical to those observed in the samples digested with *Pst*I alone (Figure 6A, lanes 18-23). This result indicates that the McrBC half sites throughout the CpG island

and within the CGG repeat are largely or completely free of methylation, though it is possible that a single half site is methylated or that methylation occurs at two or more sites separated by less than 32 bp. We note that some of the *Pst*I+McrBC lanes have a slightly diminished hybridization signal compared to the lanes containing samples digested with *Pst*I alone. We attribute this to underloading of the *Pst*I+McrBC samples as a consequence of GTP in the McrBC digestion buffer, which interferred with spectrophotometric quantitation of the samples post digestion.

Hypomethylation of the expanded CGG repeat region occurs in the presence of normal methylation of repetitive DNA elements.

Most repetitive elements are heavily methylated in mammalian somatic cells, and it has been suggested that methylation of expanded *FMR1* CGG repeats is due to their resemblance to these elements (Bestor and Tycko 1996). To determine if *FMR1* hypomethylation in patient MK is associated with decreased methylation at repetitive DNA elements, methylation was assessed at five repetitive element loci. These loci were selected on the basis of one of the following: proximity to the *FMR1* CGG repeat (two *Alu* elements), high CpG density (an X chromosome *SVA* element), or timing of *de novo* methylation during gametogenesis and early development (a Y chromosome *Alu* element and an X chromosome *L1* element). Methylation at each of the loci was studied by Southern blot analysis of genomic DNA samples digested with a methylation sensitive restriction enzyme. Hybridization probes flanking the elements were generated by PCR as presented in Subjects and Methods.

Alu elements closest to the *FMR1* CGG repeat are located 4.7 kb downstream (in the middle of intron 1) and 5.6 kb upstream. These elements each contain at least one *MspI/Hpa*II restriction site at which methylation can be assessed (Figure 7A, 7C). DNA preparations were digested with *Eco*RI alone, *Eco*RI+*Msp*I, or *Eco*RI+*Hpa*II. Hybridization with a probe flanking the downstream *Alu* showed no apparent *Hpa*II digestion demonstrating a high degree of methylation at these sites in peripheral blood lymphocyte DNA samples from normal controls (Figure 7B, lanes 4,7) and four fragile X carriers (data not shown). The hybridization pattern in MK's lymphocyte and fibroblast DNA samples are indistinguishable from the controls (Figure 7B, lanes 10,13) indicating that this *Alu* element is methylated normally in patient MK. Although most *Alu* elements are heavily methylated in somatic cells (Schmid 1991), these elements are largely hypomethylated in sperm cells (Hellmann-Blumberg et al. 1993; Kochanek et al. 1993). As a control for probe specificity and *Hpa*II digestion, DNA isolated from the sperm of a normal control was analyzed. Hybridization to the sperm DNA *Eco*RI+*Hpa*II lane demonstrated significant *Hpa*II digestion consistent with markedly reduced methylation of both *MspI/Hpa*II sites (Figure 7B, lane 16). Analysis of the upstream *Alu* was accomplished using the same Southern blot membranes which were stripped and hybridized to a probe flanking the upstream element (Figure 7C). The hybridization pattern again showed a high degree of methylation for this element in the peripheral blood lymphocyte controls (Figure 7D, lanes 4,7) and the MK samples (Figure 7D, lanes 10,13). Partial methylation of a *Hpa*II site outside the *Alu* sequence was observed in MK's fibroblast derived DNA (Figure 7D, lane 10), but complete

methylation at this site was observed in his lymphocyte DNA (Figure 7D, lane 13). Unlike the downstream *Alu*, hybridization with the upstream *Alu* probe revealed a high degree of methylation in the sperm-derived DNA (Figure 7D, lane 16).

Alu elements are inherited from the female parent in the methylated state and from the male parent in the unmethylated state (Schmid 1996; Yoder et al. 1997). The paternally inherited *Alu* elements are methylated *de novo* during early embryogenesis (Yoder et al. 1997). To study methylation at an *Alu* element likely to be subject to *de novo* methylation during MK's development, we assessed methylation at a Y chromosome *Alu* located upstream of the *ZFY* gene. DNA samples were digested with *Hin*dIII alone or *Hin*dIII and *Bst*BI (Figure 7E). Digestion by *Bst*BI did not occur in somatic controls and in DNA from MK's lymphocytes and fibroblasts (Figure 7F, lanes 3,5,7,9,11). The inability of *Bst*BI to digest these samples is consistent with a high degree of methylation. Sperm DNA was completely digested by *Bst*BI indicating this site is unmethylated in this tissue (Figure 7F, lane 13).

It is not known if *de novo* methylation of *Alu* elements and fragile X full mutation alleles share in common any mechanistic similarities. Although both can potentially form secondary structures, which may serve as signals for *de novo* methylation, expanded fragile X repeats are significantly larger and more CpG dense than *Alu* elements. To assess methylation at a site that more closely resembles an expanded fragile X repeat, we searched for large repetitive elements with high CpG density near the *FMR1* gene. Approximately 80 kb downstream of the *FMR1* CGG repeat is a region of high CpG density contained within an *SVA* (sindbis virus) element.

The *SVA* spans 1.7 kb and contains 86 CpG dinucleotides (9.9%), of which 21 are located in *MspI/Hpa*II recognition sites. Adjacent to the *SVA* is a large *L1* element that contains two additional *MspI/Hpa*II sites (Figure 8A). Methylation of the *L1* and *SVA* CpG sites was assessed with the same Southern blot membranes used for analysis of the X chromosome *Alu* elements. Hybridization patterns in the normal and fragile X peripheral blood lymphocyte controls were consistent with complete methylation of both sites in the *L1* and near complete methylation at the sites in the *SVA* element (Figure 8B, lanes 4,7). The hybridization pattern in MK's lymphocyte derived DNA was indistinguishable from the controls (Figure 8B, lane 13). MK's fibroblast derived DNA sample was substantially less methylated at sites in both the *L1* and *SVA* elements (Figure 8B, lane 10). However, the same methylation pattern was also observed in fibroblast derived DNA from six controls indicating that the difference between the lymphocyte and fibroblast cells is tissue specific and not attributable to the specific conditions in MK's cells (data not shown). Sperm DNA from a normal control was completely methylated at both *L1* sites and substantially less methylated for at least some of the *SVA* sites compared to the somatic controls (figure 8B, lane 16).

Microcell mediated chromosome transfer into mouse embryonal carcinoma cells does not induce *de novo* methylation of hypomethylated expanded CGG repeats.

To test whether the methylation deficiency in MK's cells could be complemented *in trans*, an X chromosome bearing an unmethylated full mutation was transferred by microcell fusion into a *de novo* methylation competent mouse embryonal

carcinoma (EC) cell. EC cells are capable of methylating certain DNA substrates *de novo*, including proviral genomes (Stewart et al. 1982) and mammalian sequences that direct the formation of methylation patterns (Turker et al. 1991; Mummaneni et al. 1993; Mummaneni et al. 1995). These sequences include *B1* repetitive elements (Yates et al., in preparation) which are homologous to the human *Alu* elements (Schmid 1996). The unmethylated full mutation allele was transferred initially by whole cell fusion to a differentiated and 6-thioguanine resistant (HPRT-) mouse cell line that lacks the capacity for *de novo* methylation of transfected DNA (Turker et al. 1991). Retention of the human X chromosome was selected on the basis of expression of the *HPRT* locus located approximately 19 Mb centromeric of the *FMR1* gene. A whole cell hybrid containing an unmethylated expansion of 350 repeats was selected as the microcell donor. The repeat region remained unmethylated in the differentiated background as indicated by resistance to McrBC digestion (figure 9, lane 9). Microcells were isolated from the donor clone by standard procedures and fused to DelTG3, a thioguanine resistant derivative of the P19 EC cell line. Nine fusion clones that each exhibited the undifferentiated morphology characteristic of EC cells were isolated. All of the microcell clones retained the human *HPRT* locus, but only four of the nine retained the *FMR1* region as indicated by Southern blot and PCR amplification of markers located between the loci (data not shown). Despite propagation of the chromosomes through approximately 24 population doublings, the *FMR1* CpG island remained unmethylated by both *Eco*RI/*Eag*I digestion (data not shown) and *Pst*I+McrBC digestion in each of the microcell clones (Figure 9, lanes 11,13,15,17).

DISCUSSION

The patient examined in this study, MK, is one of a small group of individuals that carry *FMR1* full mutation alleles lacking significant methylation at upstream sites (Loesch et al. 1993; McConkie-Rosell et al. 1993; Hagerman et al. 1994; Merenstein et al. 1994; Rousseau et al. 1994a; Rousseau et al. 1994b; Feng et al. 1995; Smeets et al. 1995; de Vries et al. 1996; Lachiewicz et al. 1996; Wang et al. 1996; Wohrle et al. 1998; Taylor et al. 1999). Characterization of methylation in these individuals has in most cases been based on Southern blot analysis of genomic DNA samples digested with methylation sensitive restriction endonucleases; commonly *Eag*I, *Bss*HII, and *Nru*I. Although each of these enzymes assess methylation at only a small percentage of the CpG dinucleotides that comprise the *FMR1* CpG island, the production of FMRP and normal protein-DNA footprinting interactions (Schwemmle 1999) suggest that sites throughout the promoter are also unmethylated. Furthermore, analysis of epigenetic methylation patterns with bisulfite sequencing has demonstrated that *FMR1* methylation patterns are exclusive, i.e. a particular DNA molecule will be either hypo- or hypermethylated (Stoger et al. 1997). We observed that MK carries a mix of premutation and full mutation alleles that are completely unmethylated at the *Eag*I restriction site. Moreover, analysis with McrBC digested DNA demonstrates that most if not all CpG dinucleotides both within and surrounding the CGG repeats are likewise free of methylation. Although we cannot rule-out a low level of methylation at MK's *FMR1* locus, the large majority of his DNA is clearly resistant to McrBC digestion as detected with probe *pfxa3*. Methylation restricted to the CGG repeat is unlikely to

impact transcription of the *FMR1* gene since the known regulatory elements are located upstream (Hwu et al. 1993; Schwemmle et al. 1997). Methylation of these repeats is however relevant both for the proposed relationship between repeat stability and methylation (Wohrle et al. 1996; Wohrle et al. 1998) and because the expanded repeats are proposed to serve as the initial target for *de novo* methylation (Bestor and Tycko 1996).

Most of the methylated cytosines in human DNA are thought to reside in the 35% of the genome that is comprised of transposable elements such as *Alu*, *L1*, and retroviral elements (Smit 1996). Although the biological roles of cytosine methylation are the subject of debate, one proposed function is to suppress the transcription of transposable or parasitic sequence elements (Liu and Schmid 1993; Yoder et al. 1997). Therefore, these elements may act as specific targets for *de novo* methylation. This model predicts that the fragile X CGG repeat might become a target for *de novo* methylation when expanded sufficiently to resemble a parasitic element. It has been speculated that the *de novo* methylation specificity of the mammalian DNA methyltransferase is dependent on DNA secondary structures (Bestor and Tycko 1996; Bender 1998). Numerous types of secondary structure are observed in disease causing triplet repeats (Sinden 1999) and in the case of CGG repeats, three-way hairpin slippage structures are efficient substrates for *de novo* methylation *in vitro* (Chen et al. 1995; Chen et al. 1998). The demarcation that exists in the fragile X syndrome between unmethylated premutation alleles and methylated full mutations might then represent a threshold at which the CGG repeat reaches a length allowing the formation of

secondary structures resembling parasitic sequence elements. Alternatively, methylation of expanded fragile X repeats may occur randomly or in association with the switch to very late DNA replication that occurs in the *FMR1* region of cells that carry full mutations (Hansen et al. 1997). A selective disadvantage during early development for the cells that contain an unmethylated expansion (Hansen et al. 1997) could explain the appearance of only methylated full mutations in the cells of most adult fragile X patients.

The above discussion suggests that one explanation for rare hypomethylated full mutation alleles is that they have escaped the process of *de novo* methylation. To test this possibility we have examined methylation levels at repetitive sequence elements that were likely to have been subject to *de novo* methylation during MK's early development. Most *Alu* elements inherited from the male parent are initially hypomethylated and achieve the highly methylated state characteristic of adult somatic cells during the wave of *de novo* methylation that occurs after implantation (Yoder et al. 1997). Conversely, *L1* elements are inherited from the female parent in the unmethylated state and, like *Alu* elements, are subject to *de novo* methylation during embryogenesis (Yoder et al. 1997). We observed that both a paternally inherited *Alu* element and a maternally inherited *L1* element are methylated in MK's somatic cells. The presence of methylation in these elements indicates that *de novo* methylation occurred successfully at these sites during MK's early development. Based on these findings, we conclude that if the *FMR1* methylation abnormality in MK is due to a defect in *trans* then *Alu* and *L1* methylation are likely to be mediated by pathways other

than that responsible for methylation of the *FMR1* CGG repeat region. If *de novo* methylation of different genomic sequences occurs with distinct enzymology, repetitive elements that share characteristics in common with the fragile X CGG repeat might be more likely to be methylated with the same pathway. To locate such an element, we searched for regions with high CpG density and identified a 1.7 kb *SVA* element 80kb downstream of the *FMR1* CGG repeat that contains 86 CpG dinucleotides (9.9%). Methylation of this element appears to have occurred normally in MK's cells further indicating that a generalized methylation deficiency is not present at the elements tested.

Another possible explanation for the absence of *FMR1* CpG island methylation in MK is that a failure of maintenance methylation has occured specifically in the *FMR1* region. To assess this possibility we examined methylation levels at *Alu* elements flanking the expanded CGG repeat region. These *Alu* elements were inherited maternally with high levels of methylation that are believed to persist during the early embryonic period characterized by dynamic changes in global methylation patterns. The possibility of perturbed maintenance methylation as an explanation for MK's methylation deficiency is suggested by the *Arabidopsis thaliana* mutant *ddm1* that exhibits a reduction in genomic cytosine methylation that occurs first in repeated sequences (Vongs et al. 1993; Kakutani et al. 1996; Jeddeloh et al. 1998). One model of DDM1 function is that it acts as a part of a nucleosome-remodelling complex that increases accessibility of the DNA to the maintenance methyltransferase (Jeddeloh et al. 1999). This is a particularly attractive model when applied to the fragile X CGG repeats

because it predicts preferential hypomethylation of sequences that exist in highly condensed chromatin, such as fragile X full mutations (Luo et al. 1993; Eberhart and Warren 1996; Godde et al. 1996; Wang and Griffith 1996), can result from a mutation in *trans* . However, it seems unlikely that an explanation for MK's unmethylated full mutation will be found at the level of an *FMR1* region-specific abnormality in maintenance methylation because the two *Alu* elements flanking the CGG repeat at a distance of ~5 kb are maintained with a high level of methylation.

Alternatively, MK's full mutation may have escaped the methylation process not based on a failure of *de novo* or maintenance methylation, but rather due to the timing of repeat expansion during early development. The ability to *de novo* methylate DNA substrates is primarily a characteristic of embryonic cells (Jaenisch 1997). It is possible that during the embryonic period characterized by cellular *de novo* methylation competence, MK carried an allele that failed to serve as a substrate for methylation due to its size in the upper premutation range. If this allele expanded after the period of *de novo* methylation was complete, it might thereafter remain unmethylated despite attaining the size of a typical full mutation (Wohrle et al. 1998). This possibility is supported by studies that have demonstrated unmethylated alleles are unstable in cultured fibroblasts (Wohrle et al. 1998; Burman et al. ; Glaser et al. 1999).

To date, the only inherited defect in genomic methylation identified is the ICF (immunodeficiency, centromeric instability, facial anomalies) syndrome (MIM 242860). The ICF syndrome is a rare autosomal recessive condition which is characterized by chromosomal abnormalities and reduced methylation in repeated

satellite regions and some *Alu* elements (Miniou et al. 1997a; Miniou et al. 1997b). Cell fusion experiments have shown that this defect can be at least partially complemented in *trans* (Schuffenhauer et al. 1995). To directly test if the *FMR1* specific methylation deficiency in patient MK could be complemented in *trans*, we introduced the chromosome bearing this allele into mouse EC cells. These cells have a high capacity to *de novo* methylate transfected DNA (Turker et al. 1991) and also have the ability to methylate target sequences introduced by microcell mediated chromosome transfer (Yates et al. in preparation). One target for *de novo* methylation is the mouse B1 element (Yates et al. 1999), which is homologous to the human *Alu* element (Schmid 1996). We found no evidence of methylation either at the *Eag*I site or with the McrBC assay in DNA prepared from microcell hybrids that had replicated patient MK's X chromosome approximately twenty four times. Several explanations can be invoked to describe this result. One possibility is that expanded CGG repeats do not serve as a methylation signal in the EC cells utilized and/or that they have not formed the secondary structure that is required for *de novo* methylation to occur. Alternatively MK's chromosome may contain some type of variation in *cis* that renders his repeat unrecognizable as a target or otherwise inherently resistant to *de novo* methylation. However, the presence of a cousin with a methylated expansion of presumably the same allele argues against these latter possibilities unless a distinct genetic alteration occured in patient MK. Finally the lack of methylation at the transferred allele may simply reflect an inherent difference in the way mouse and human cells maintain/metabolize large CGG repeats. Additional work will be required to sort through these possibilities.

In summary, we have described a patient with a fully expanded fragile X mutation lacking *Eag*I site methylation. Analysis with the McrBC endonuclease demonstrated further that the CGG repeat and ~1 kb of surrounding DNA are also free of methylation. Methylation patterns are formed normally at *Alu* elements within 5 kb of the repeat and at other repetitive element loci on the X and Y chromosomes. These results suggest that hypomethylation of the expanded *FMR1* CGG repeat region is not due to global or regional defects in *de novo* or maintenance methylation processes. Whether this deficit represents a rare stochastic event or a rare heritable alteration remains to be determined.

ACKNOWLEDGMENTS

This study was supported in part by a National Institutes of Health training grant T32 GM08617 (RWB) and a grant from the Council for Tobacco Research (MST).

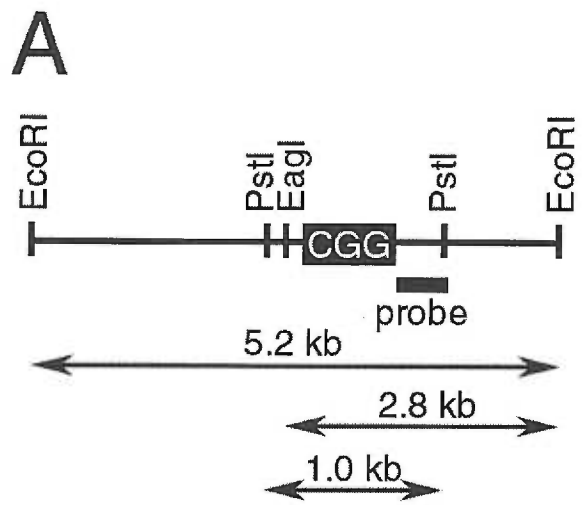
ELECTRONIC DATABASE INFORMATION

RepeatMasker, http://ftp.genome.washington.edu/RM/RepeatMasker.html

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for fragile X syndrome [309550] and ICF syndrome [242860])

Figure 5 legend

ON BACK OF THIS PAGE


A
EcoRI
PstI
EagI
CGG
PstI
EcoRI
probe
5.2 kb
2.8 kb
1.0 kb

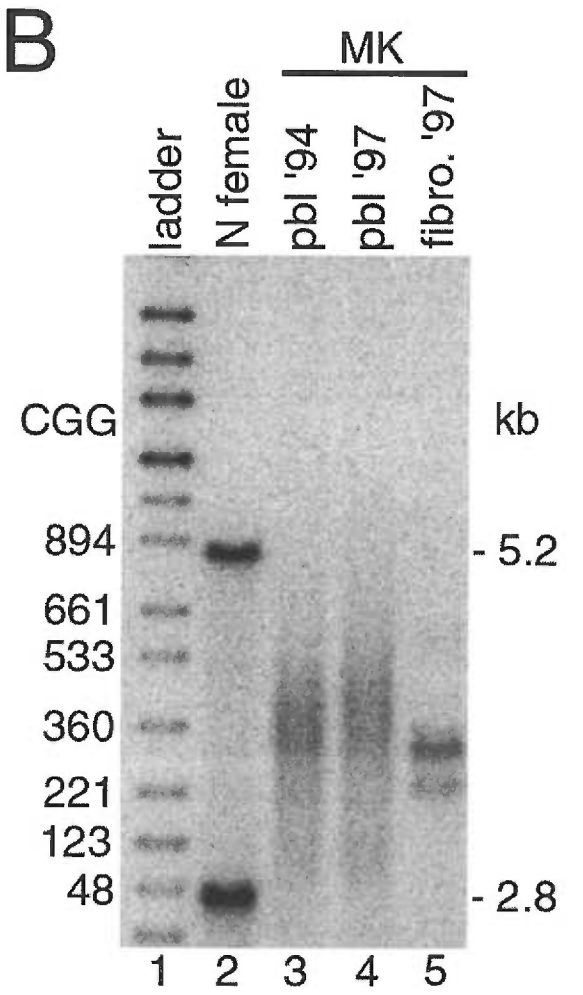
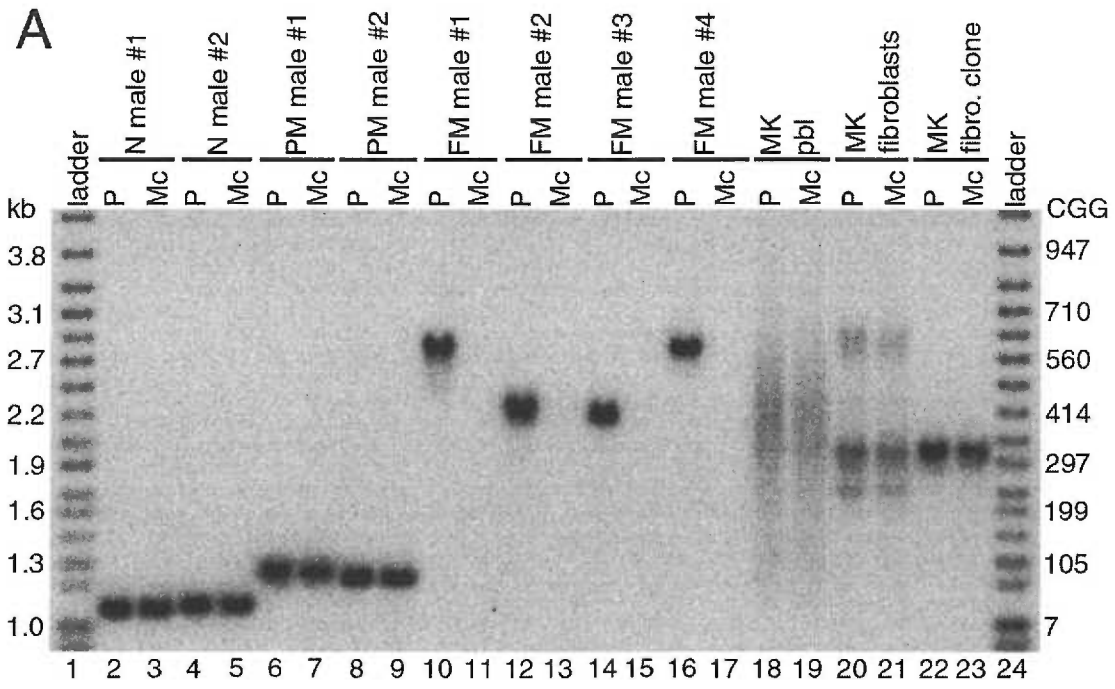

B
MK
ladder
N female
pbl '94
pbl '97
fibro. '97
CGG
kb
894
661
533
360
221
123
48
- 5.2
- 2.8
1 2 3 4 5

Figure 6 legend

ON BACK OF THIS PAGE


A
ladder
N male #1
N male #2
PM male #1
PM male #2
FM male #1
FM male #2
FM male #3
FM male #4
MK pbl
MK fibroblasts
MK fibro. clone
ladder
P
Mc
kb
3.8
3.1
2.7
2.2
1.9
1.6
1.3
1.0
CGG
947
710
560
414
297
199
105
7
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

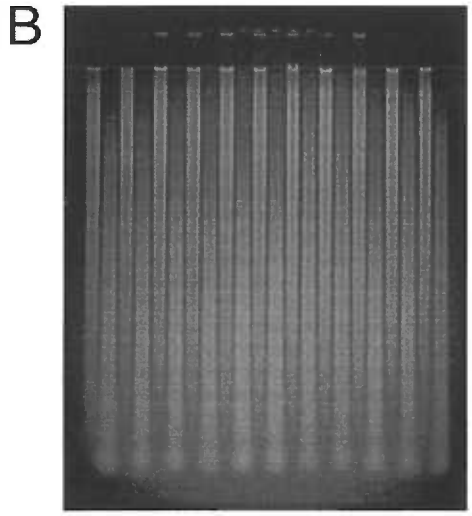

B

Figure 7 legend

ON BACK OF THIS PAGE

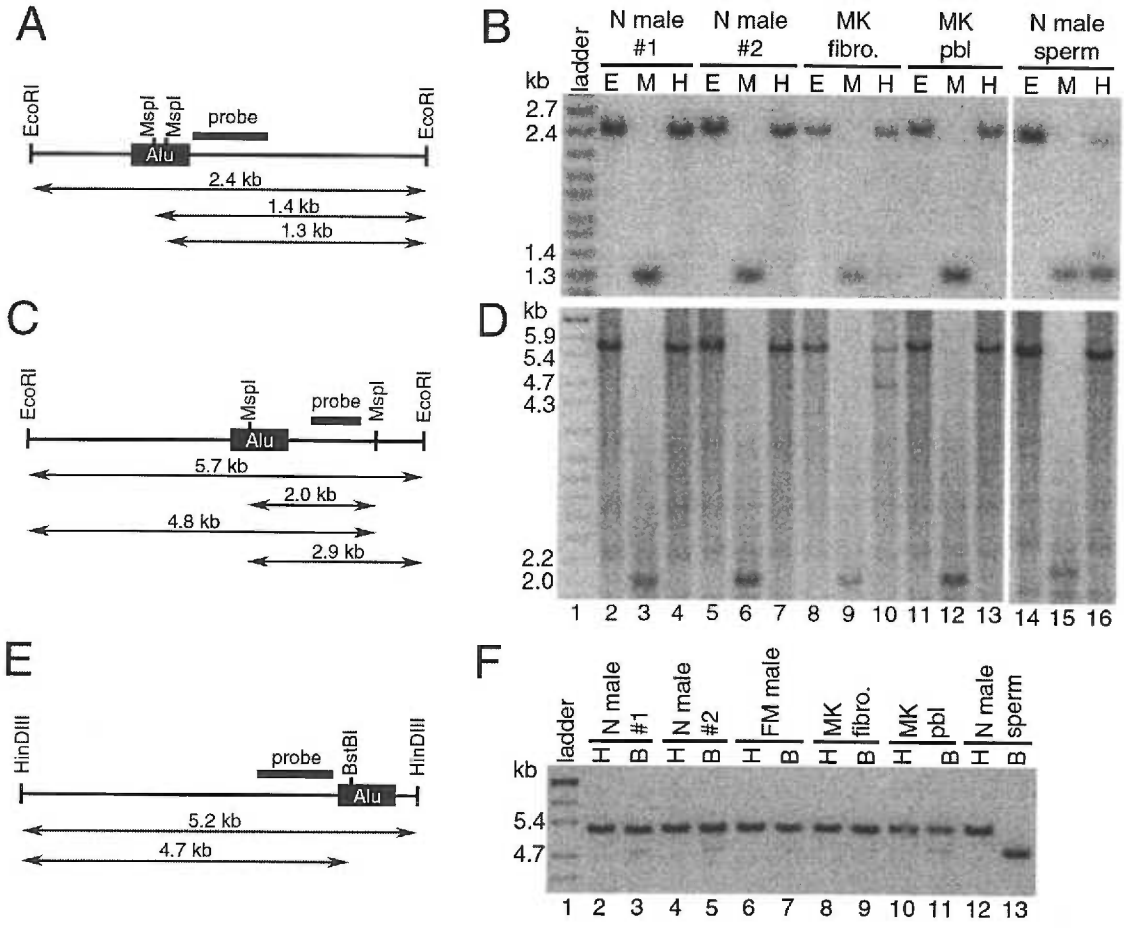

A
EcoRI
MspI
MspI
probe
EcoRI
Alu
2.4 kb
1.4 kb
1.3 kb
B
ladder
N male #1
N male #2
MK fibro.
MK pbl
N male sperm
kb
E M H E M H E M H E M H E M H
2.7
2.4
1.4
1.3
C
EcoRI
MspI
probe
MspI
EcoRI
Alu
5.7 kb
2.0 kb
4.8 kb
2.9 kb
D
kb
5.9
5.4
4.7
4.3
2.2
2.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
E
HinDIII
probe
BstBI
HinDIII
Alu
5.2 kb
4.7 kb
F
ladder
N male #1
N male #2
FM male
MK fibro.
MK pbl
N male sperm
H B H B H B H B H B H B H B
kb
5.4
4.7
1 2 3 4 5 6 7 8 9 10 11 12 13

Figure 8 legend

ON BACK OF THIS PAGE

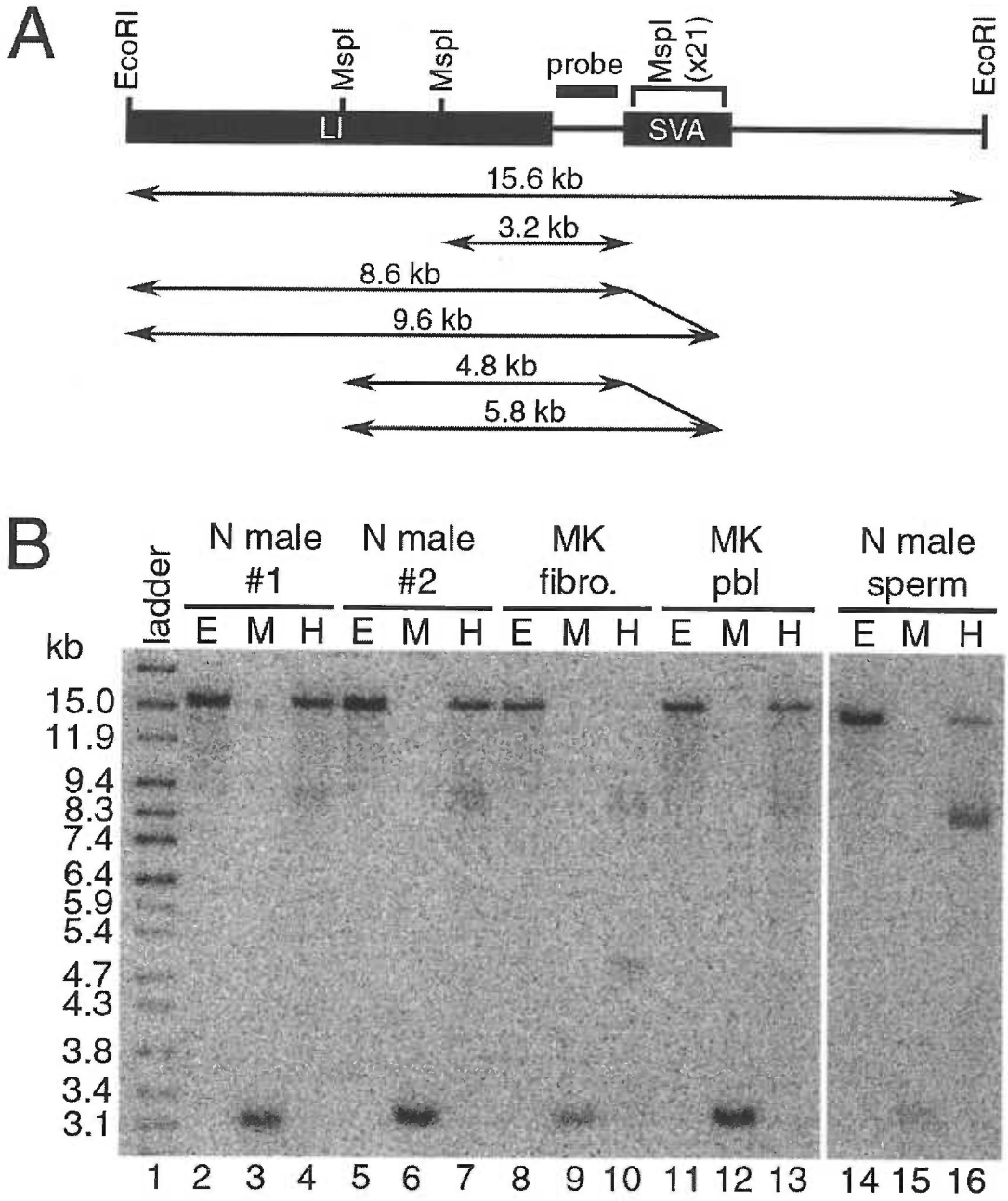
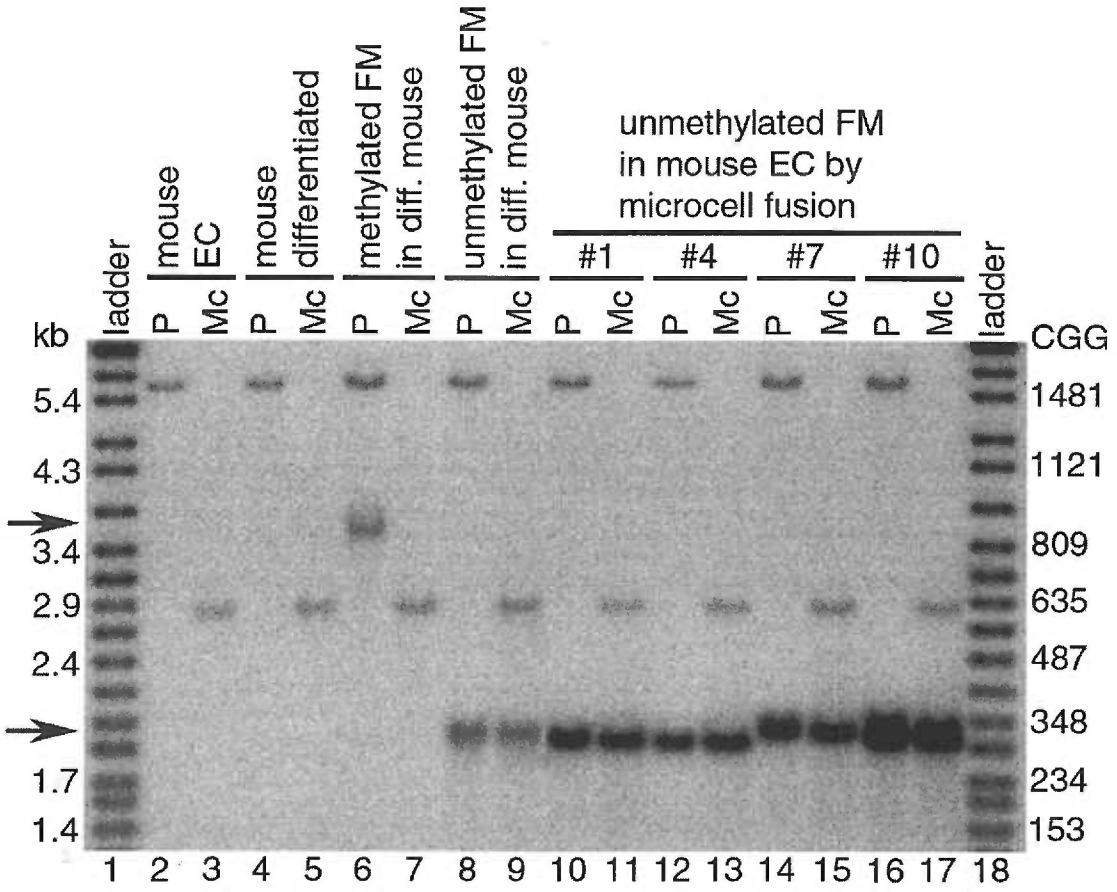

A
EcoRI
MspI
MspI
probe
MspI (x21)
EcoRI
L1
SVA
15.6 kb
3.2 kb
8.6 kb
9.6 kb
4.8 kb
5.8 kb
B
ladder
N male #1
N male #2
MK fibro.
MK pbl
N male sperm
E M H E M H E M H E M H E M H
kb
15.0
11.9
9.4
8.3
7.4
6.4
5.9
5.4
4.7
4.3
3.8
3.4
3.1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Figure 9 legend

ON BACK OF THIS PAGE


ladder
mouse EC
mouse differentiated
methylated FM in diff. mouse
unmethylated FM in diff. mouse
unmethylated FM in mouse EC by microcell fusion
#1
#4
#7
#10
P
Mc
ladder
kb
5.4
4.3
3.4
2.9
2.4
1.7
1.4
CGG
1481
1121
809
635
487
348
234
153
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

CHAPTER 3

FULLY EXPANDED FMR1 CGG REPEATS EXHIBIT A LENGTH- AND DIFFERENTIATION-DEPENDENT INSTABILITY IN CELL HYBRIDS THAT IS INDEPENDENT OF DNA METHYLATION

Robert W. Burman[1,2], Bradley W. Popovich[1,2], Peter B. Jacky[3], and Mitchell S. Turker[1,4]

[1]Department of Molecular and Medical Genetics, [2]DNA Diagnostic Laboratory, and [4]Center for Research on Occupational and Environmental Toxicology, Oregon Health Sciences University, Portland; and [3]Cytogenetics Laboratory, Kaiser Permanente NW, Clackamas, OR, USA

Human Molecular Genetics **8(12)**: 2293-2302, 1999

ABSTRACT

The fragile X syndrome is characterized at the molecular level by expansion and methylation of a CGG trinucleotide repeat located within the *FMR1* locus. The tissues of most full mutation carriers are mosaic for repeat size, but these mutational patterns tend to be well conserved when comparing multiple tissues within an individual. Moreover, full mutation alleles are stable in cultured fibroblasts. These observations have been used to suggest that fragile X CGG repeat instability is normally limited to a period during early embryogenesis. DNA methylation of the repeat region is also believed to occur during early development and some experimental evidence indicates that this modification may stabilize the repeats. To study the behavior of full mutation alleles in mitotic cells, we generated human-mouse somatic cell hybrids that carry both methylated and unmethylated full mutation *FMR1* alleles. We observed considerable repeat instability and analyzed repeat dynamics in the hybrids as a function of DNA methylation, repeat length, and cellular differentiation. Our results indicate that although DNA methylation does correlate with stability in primary human fibroblasts, it does not do so in the cell hybrids. Instead, repeat stability in the hybrids is dependent on repeat length except in an undifferentiated cellular background where large alleles are maintained with a high degree of stability. This stability is lost when the cells undergo differentiation. These results indicate that the determinants of CGG repeat stability are more complex than generally believed and suggest an unexpected role for cellular differentiation in this process.

INTRODUCTION

Expansion of trinucleotide repeat sequences are responsible for a growing list of human diseases including the fragile X syndrome, Huntington's disease, myotonic dystrophy, and a number of ataxias (Ashley and Warren 1995; Reddy and Housman 1997). The sequence of the repeats and the mechanisms by which the expansions cause disease vary among the disorders. Each of the repeats are polymorphic within normal populations, but when expanded beyond critical size thresholds have the ability to undergo large intergenerational changes in repeat number as well as somatic changes within an individual (Ashley and Warren 1995; La Spada 1997; Reddy and Housman 1997). Somatic instability is most pronounced in the fragile X syndrome and myotonic dystrophy, where it results in a high degree of mosaicism in most affected individuals such that the length of the repeat tract can vary widely among different cell populations within an individual.

The molecular basis of repeat instability is not well understood nor is it known if somatic and intergenerational instability share in common any mechanistic similarities. Most information regarding the mechanisms of repeat dynamics has come from the study of cloned repeat tracts replicated in *E. coli* or *S. cerevisiae*. Repeat stability in these organisms is modulated by both *cis* and *trans*-acting factors. Important *cis* parameters include the length and sequence composition of the repeat (Kang et al. 1995; Shimizu et al. 1996; Wierdl et al. 1997; Hirst and White 1998) as well as the direction of replication (Kang et al. 1995; Maurer et al. 1996; Shimizu et al. 1996; Freudenreich et al. 1997; Hirst and White 1998). *Trans* acting factors include mutations in genes that

encode DNA repair proteins (Jaworski et al. 1995; Schweitzer and Livingston 1997; Parniewski et al. 1999), RAD27 (an enzyme involved in processing Okazaki fragments) (Freudenreich et al. 1998; Schweitzer and Livingston 1998), and the SbcC endonuclease (Sarkar et al. 1998). Several mouse models designed to study trinucleotide repeat dynamics in a mammalian context have also been reported. These models have been based on transgenic animals that carry trinucleotide repeat sequences, usually within the context of human genomic or cDNA. The success of these models with respect to repeat instability has been mixed with some CAG transgenics (Bingham et al. 1995; Burright et al. 1995; Goldberg et al. 1996; Ikeda et al. 1996) and all CGG transgenics (Bontekoe et al. 1997; Lavedan et al. 1997; Lavedan et al. 1998) showing no instability. Modest somatic and intergenerational changes in CAG and CTG repeat number have been reported in other transgenic models (Gourdon et al. 1997; Kaytor et al. 1997; Mangiarini et al. 1997; Monckton et al. 1997; La Spada et al. 1998; Sato et al. 1999; Wheeler et al. 1999). However, the changes tend to be small and have not yet reproduced the large intergenerational expansions observed in humans.

The CGG trinucleotide repeat that causes the fragile X syndrome is located within the 5' untranslated region of the *FMR1* gene (Kremer et al. 1991a; Oberle et al. 1991; Verkerk et al. 1991; Yu et al. 1991). In normal populations the repeat tract is stable on transmission and comprised of ~5 to ~55 triplets (Fu et al. 1991). Premutation alleles with ~55 to ~220 repeats occur in unaffected carriers (Fu et al. 1991). When transmitted by a female, premutation alleles can undergo large intergenerational changes in repeat number to form full mutation alleles with greater than ~220 repeats

(Fu et al. 1991; Oberle et al. 1991). Hypermethylation of full mutation alleles correlates with transcriptional silencing of the *FMR1* gene (Pieretti et al. 1991; Sutcliffe et al. 1992). The resulting reduction or absence of the FMR1 protein (FMRP) is believed sufficient to cause the fragile X syndrome phenotype (Pieretti et al. 1991; Devys et al. 1993; Siomi et al. 1993). Somatic instability occurs in the majority of full mutation carriers resulting in repeat length mosaicism, which is manifest as a complex smeared and multi-banded hybridization pattern when examined by Southern blot analysis (Fu et al. 1991; Moutou et al. 1997).

Numerous studies have found that the resulting hybridization patterns are well conserved among multiple tissues within individuals who carry typical methylated full mutations (Sutherland et al. 1991; Devys et al. 1992; Wohrle et al. 1992; Wohrle et al. 1993; de Graaff et al. 1995; Wohrle et al. 1995; Moutou et al. 1997; Reyniers et al. 1999; Tassone et al. 1999) and can be conserved in monozygous twins (Devys et al. 1992; Kruyer et al. 1994; Antinolo et al. 1996). A likely explanation for these observations is that the expanded repeats are maintained with little or no variability after an initial period of instability that presumably occurs during early embryogenesis. The analysis of cultured fibroblasts derived from fetal and adult full mutation carriers has demonstrated that methylated fragile X full mutation alleles are stable in these differentiated cells (Wohrle et al. 1993). Thus a demarcation appears to exist between embryonic cells in which somatic mosaicism is presumably produced by repeat instability and adult cells where the repeats are stable and the mosaic mutational patterns are maintained with little or no variation. Although the molecular basis of this

boundary has not been determined, some experimental evidence suggests DNA methylation stabilizes the repeats (Wohrle et al. 1996). Comparison of methylated and unmethylated alleles in heterogeneous cell populations has shown the mutational patterns to be more smeared and diffuse when the alleles are unmethylated (Wohrle et al. 1998; Glaser et al. 1999). Moreover, post-mortem analysis of one fragile X male with a partially unmethylated expansion has identified intertissue differences in mutation patterns (Taylor et al. 1999). If methylation influences repeat stability, somatic variation might normally occur during the period of global demethylation in early development. According to this hypothesis, *de novo* methylation of the repeats at about the time of blastocyst implantation would fix in place any size variability that had occurred (Wohrle et al. 1996). In the rare individuals that harbour an unmethylated expansion, the period of repeat instability would be extended and result in both intertissue heterogeneity and repeat instability in cultured cells.

We report the dynamic behavior of the fragile X CGG repeat in a tissue culture system. Repeat stability of methylated and unmethylated alleles was examined in both primary human fibroblasts and human-mouse cell hybrids. We show that methylation and repeat stability are correlated in the human cells, but in the cell hybrids both methylated and unmethylated alleles are highly unstable. Instability in the hybrids occurs by both expansion and contraction and is a function of repeat length and cellular differentiation.

MATERIALS & METHODS

Cell Lines and Cell Culture

Human fibroblast cell lines were purchased form the Corriell Human Genetic Mutant Cell Repository (GM04026, GM05847A, GM09497) or were established, with informed consent, from skin biopsies of adult fragile X full mutation carriers (TC38-89, MK) referred for clinical fragile X testing to either the Oregon Health Sciences University DNA Diagnostic Laboratory or the Kaiser Permanente Cytogenetics Laboratory. All human cell cultures were maintained at 37°C/5% CO_2 in α-minimal essential media (α-MEM) supplemented with 20% fetal calf serum (JRH Biosciences), 2mM L-glutamine, and 16 μg/ml gentamicin sulfate. Clonal lines were obtained from the mass cultures by seeding the cells at a concentration of 40-100 cells per 150mm dish. Individual clones were isolated with glass cloning cylinders approximately two weeks after seeding, then expanded until confluent in a 75-cm^2 flask and harvested for DNA isolation.

The Dif6 cell line is a morphologically differentiated and 6-thioguanine resistant clone isolated from the mouse P19-derived embryonal carcinoma cell line H4D2 (Turker et al. 1991). DelTG3 is an undifferentiated and 6-thioguanine resistant clone also isolated from H4D2 (Turker et al. 1989a; Turker et al. 1989b). Fusion experiments utilizing these cell lines were carried out with derivatives that had been transfected with the bacterial neomycin (neo) gene, which confers resistance to geneticin (G418). All mouse cell cultures were maintained at 37°C/5% CO_2 in Dulbecco's Modified Eagle's

Medium (DMEM) supplemented with 5% fetal calf serum and 5% serum plus (JRH Biosciences).

Cell Fusions and Differentiation

Whole cell fusions were accomplished by mixing human fragile X fibroblasts with mouse cell lines Dif6 or DelTG3 at cell ratios of 1:3, 1:1, and 3:1 with a total of $3x10^6$ cells per 25-cm^2 flask. After mixing, the cells were plated and incubated for approximately six hours, and then fused by adding 1 ml of a 50% polyethylene glycol (Sigma; PEG 1,450) solution in α-MEM. The cultures were rinsed three times with PBS, incubated overnight in non-selective media and transferred the next day to 150mm dishes at a concentration of $1\text{-}2x10^5$ cells per dish. Selection for hybrid clones was applied approximately 48 hours post fusion with 10 μg/ml hypoxanthine (Sigma), 10 μg/ml azaserine (Sigma), and 1 mg/ml geneticin (Gibco BRL). Individual hybrid clones were isolated 11-14 days after initiating selection and were then maintained in media supplemented with hypoxanthine and azaserine to retain the human X-chromosome. Clones were expanded to a 75-cm^2 flask for DNA isolation and a 25-cm^2 flask for cryopreservation in DMEM with 10% DMSO (Sigma). Subclones were generated from the parental hybrid clones by seeding the cells at a concentration of 100-200 cells per 150mm dish. Individual subclones were isolated with glass cloning cylinders approximately two weeks after seeding, then expanded until confluent in a 75-cm^2 flask and harvested for DNA isolation. An undifferentiated hybrid clone created with DelTG3

was induced to differentiate by treatment for 7 days in DMEM supplemented with 1.0 μM all-trans retinoic acid (Sigma). Differentiated subclones were isolated as described, with retinoic acid treatment continuing for the first 4 days after the cells were seeded at the cloning dilution.

DNA Isolation and Analysis

Genomic DNA was isolated using either a standard phenol/chloroform extraction or the Puregene DNA Isolation Kit (Gentra Systems). Aliquots (10 μg) were digested with restriction endonucleases *Pst*I (Boehringer Manheim), or *Eco*RI (Boehringer Manheim) and *Eag*I (New England Biolabs). Digested DNA was precipitated with sodium acetate and ethanol and then aliquots of 1.5-2.5 μg were separated by electrophoresis in 1% agarose/TAE and transferred to Biodyne B nylon membrane (Gibco BRL) with 5X SSC. Hybridizations with a ^{32}P radiolabelled probe (Boehringer Manheim; Random Prime Labelling Kit) were carried out at 65°C in Church Buffer with BSA (hybridization buffer II) (Strauss 1998) supplemented with 100 μg/ml heat-denatured herring sperm DNA. Probe *pfxa3* (a kind gift from David L. Nelson, Baylor College of Medicine, Houston), a 558 bp *Xho*I-*Pst*I fragment of pE5.1 (Fu et al. 1991; Verkerk et al. 1991) was used to detect restriction fragments containing the CGG repeat. Membranes were washed twice in low stringency wash buffer II (Strauss 1998) at room temperature and then twice at 68.5°C in high stringency wash buffer II (Strauss 1998) diluted to 0.6X. Membranes were exposed sequentially to a

Molecular Dynamics phosphorimaging screen and then x-ray film (Kodak; X-OMAT) at -70°C. Molecular weight estimations were obtained digitally from film images with the DNA-VIEW program version 21 (Dr. Charles H. Brenner, http://www.ccnet.com /~cbrenner/). The DNA Analysis Marker System (Gibco BRL) was used as the molecular weight standard ("ladder") for all of the Southern blot experiments.

RESULTS

DNA methylation predicts repeat stability in primary human fibroblasts

To assess the proposed relationship between methylation and repeat stability in human cells, clonal fibroblast lines were isolated from primary skin cultures of three males bearing full mutations. Two of the males carry typical full mutations with complete methylation at the diagnostic *Eag*I site (data not shown). The third individual (MK) is a high-functioning male who carries a full mutation that is unmethylated not only at the *Eag*I site, but also throughout the CpG island and the CGG repeat (Burman et al.). For each cell line, clonal isolates were generated by seeding the cells at low density and later harvesting individual clones with glass cylinders. The clonally derived cell cultures were expanded through approximately 24 population doublings until confluent in a 75-cm^2 flask and then harvested for DNA isolation. Repeat stability in the clones was assessed by Southern blot analysis of genomic DNA digested with either *Pst*I alone or *Eco*RI and *Eag*I (Figure 10A). Analysis of DNA digested with *Eco*RI/*Eag*I also allows an assessment of methylation status at the *Eag*I site. Hybridization with probe *pfxa3* revealed discreet bands with no smearing in 14 clones

isolated from the two methylated full mutation carriers. Seven of the clones are shown in figure 10B. The sample in lane 7 contains two hybridization bands that are most likely the result of an impure clone rather than repeat instability. These results are consistent with a previous report of repeat stability in cultured fibroblasts (Wohrle et al. 1993). The clonal cell DNA samples isolated from the hypomethylated full mutation male, MK, showed band patterns with prominent smearing and in some cases samples with multiple allele sizes. The complexity of the patterns is consistent with repeat instability generating extensive repeat length mosaicism during clonal proliferation of the cultures. Thirty seven MK fibroblast clones were analyzed of which 26 contained smearing and/or multiple bands. Seven representative clones are shown in figure 10C. Although it is formally possible that some of these cases of multiple alleles are due to impure clones, a comparison of the hybridization patterns of the methylated alleles (1 of 14 with more than one hybridization band, no smearing) with the unmethylated alleles (26 of 37 with multiple bands and/or smearing) demonstrates that methylation status correlates with repeat stability in human fibroblasts.

Methylated repeats become unstable in differentiated cell hybrids as a function of size

The study of repeat dynamics in primary human fibroblasts is limited by their finite replicative capacity. To better characterize repeat length changes, we sought a system in which repeat stability could be studied in different cellular backgrounds and over a large number of cell divisions. Recognizing that chromosomal context might

have an influence on repeat behavior (La Spada 1997) and that it is not possible at this time to amplify and clone large CGG repeats (Shimizu et al. 1996; Hirst and White 1998), we immortalized human fragile X chromosomes by whole cell fusion with a differentiated mouse cell line (Dif6). Retention of the human X chromosome in the hybrids was selected on the basis of the *HPRT* locus, which is approximately 19 Mb centromeric of the *FMR1* gene.

Human-mouse hybrids were obtained in fusions with fibroblasts from two males bearing full mutations (TC38-89, GM04026) and a female bearing a full mutation and a normal allele (GM05847A). A total of 62 hybrid clones were obtained. Southern blot analysis with probe *pfxa3* demonstrated the presence of the human *FMR1* CGG repeat region in 60 of 62 hybrids, as well as a non-polymorphic mouse specific band at 5.6 kb or 1.2 kb in *Pst*I or *Eco*RI/*Eag*I digested DNA, respectively. All expanded fragile X alleles maintained complete methylation at the *Eag*I site. Despite methylation, hybridization patterns in 40 of the 60 hybrid clones were smeared (data not shown) and thus highly suggestive of instability. The chromosome content of two unstable hybrid clones were evaluated by fluorescent in-situ hybridization (FISH). We examined 20 metaphases from each hybrid and found the mean number of human X-chromosomes and autosomes to be 1 and 15, respectively. This result indicates that the smeared and poly-banded hybridization patterns that demonstrate instability are not explained by the presence of multiple human X-chromosomes in the hybrids.

Because the hybrid clones have a much greater replicative capacity than primary human fibroblasts, repeat length mosaicism that occurs during clonal proliferation can

be demonstrated more clearly by examining allele sizes in subclones. Although repeat instability continues during expansion of the subclones, a random assortment of subclones provides information about the magnitude and direction of repeat length change that has occurred after cell fusion. Repeat stability in 9 hybrid clones with allele sizes ranging from 280 to 1610 repeats was examined further by subcloning. Southern blot analysis of parental and subclone DNA samples showed a large number of contractions and a few expansion products in many subclone lanes (Figure 11). These changes were more frequent and occurred with greater magnitude when the parental hybrid clones began with large alleles (Table 1). The relation between length of the parental repeat and the frequency (Figure 12A) and magnitude (Figure 12B) of change in the subclones is best described with a logarithmic function.

In some subclones, contraction events created small and apparently stabilized alleles with notably sharp and intense hybridization bands. The stability of three contraction products with 320, 270, and 170 repeats was examined by subcloning the cells a second time to create secondary subclones. In each case the small alleles maintained complete methylation of the *Eag*I site and were quite stable with no instability observed in 10 of 10 secondary subclones containing a 170 repeat premutation sized allele (Figure 11 and Table I). These results demonstrate that a large unstable allele can be stabilized by contraction to a size below a stability threshold estimated to be approximately 200-250 CGG repeats (i.e. the size of a large premutation).

Instability of large methylated repeat alleles is a function of cellular differentiation

To assess repeat behavior in an undifferentiated cellular background, chromosomes bearing methylated fragile X full mutation alleles were transferred to the undifferentiated DelTG3 embryonal carcinoma (EC) cell line by whole cell fusion. Both Dif6 and DelTG3 are derivatives of the P19 EC cell line. Fusion to the EC cells was less efficient than to the differentiated cells and generated 3 hybrid clones. Two of the hybrids maintained the undifferentiated morphology characteristic of EC cells (Figure 13A) and the third underwent spontaneous differentiation (Figure 13D). Each hybrid clone was expanded through approximately 24 population doublings and harvested for DNA isolation and cryopreservation. Southern blot analysis showed the presence of large alleles (600-800 repeats) with complete methylation at the *Eag*I site in each of the clones (data not shown). The chromosome content of one clone was studied by FISH analysis of 20 metaphase preparations. The mean number of human X-chromosomes per cell was 1 and the mean number of human autosomes was 10.

Repeat stability in the hybrid clones was assessed with the subcloning strategy described previously. Subclones derived from the undifferentiated hybrid clones maintained the parental repeat size with little variation despite carrying full mutation alleles of sufficient size to be highly unstable in the Dif6 background (Figure 13B,C; Table 1). In contrast to the repeat stability observed in the EC background, subclones derived from the spontaneously differentiated hybrid clone contained a highly unstable repeat (Figure 13E). This result suggested that the human fragile X CGG repeat is maintained with differential stability in differentiated and undifferentiated hybrid cells

with the same genetic background. The destabilizing effect of differentiation was confirmed by treating one of the undifferentiated hybrid clones (containing 740 repeats) with retinoic acid (RA) to induce differentiation. Differentiated subclones were isolated, and an *Eco*RI/*Eag*I Southern blot showed the presence of novel allele sizes in the subclones with an average magnitude of change of 230 repeats, consistent with significant repeat instability (Figure 12 and 13F and Table I). In comparison, the two undifferentiated subclones that exhibited instability revealed contractions of only 40 repeats.

Unmethylated repeats are unstable in differentiated cell hybrids

Forty seven hybrid clones were isolated from fusions between the Dif6 differentiated mouse cells and the MK fibroblasts containing unmethylated full mutation alleles. Southern blot analysis with probe *pfxa3* demonstrated the presence of the human *FMR1* CGG repeats in 45 of the 47 clones. Allele sizes were observed in the hybrids at frequencies proportionate to their relative abundance in the primary fibroblast culture (figure 10C, lane3). Thus most hybrid clones contained alleles with 300-350 repeats and a few contained large alleles of 600-750 repeats. Subclones were generated from four MK hybrid clones with repeat sizes of 740, 730, 350, and 330 triplets. The subclones were expanded through an additional 24 population doublings until DNA was isolated. Repeat stability in the subclones was assessed by Southern blot of *Eco*RI/*Eag*I digested genomic DNA. Comparison of the allele size in the parental hybrid clone to those observed in the subclones again revealed expansion and contraction products in

the subclones that are not visible in the parental clone (Figure 14). Like the methylated alleles in the differentiated cell hybrids, both expansion and contraction products were observed in the subclones and occur with greatest frequency and magnitude when derived from parental clones that contain larger alleles (Table 1).

DISCUSSION

The study of dynamic size changes in the *FMR1* CGG repeat tract has been complicated by several factors. Most notable among these is the inability to isolate large expansions with molecular approaches. We have bypassed this barrier by working with X chromosomes bearing *FMR1* expansions and in doing so have made three novel and unexpected observations. The first is that high levels of CGG repeat instability occur in human-mouse hybrid cells containing a majority of mouse chromosomes. The second is that this instability is independent of DNA methylation status. The third observation is that cellular differentiation triggers repeat instability in the hybrid cells. These results are particularly surprising because work with primary human fibroblasts suggested that DNA methylation status was a prime determinant of repeat stability, i.e. methylated repeats are stable and unmethylated repeats are unstable in these primary cells (Wohrle et al. 1993; Wohrle et al. 1998; Glaser et al. 1999). Others have used similar observations in cultured fibroblasts, and the observation that complex size patterns are conserved between many tissues within typical fragile X patients, to propose that repeat size instability occurs very early in development (Devys et al. 1992; Reyniers et al. 1993; Wohrle et al. 1993) and is restricted by DNA methylation (Wohrle et al. 1996).

This methylation event is presumed to occur at about the time of blastocyst implantation when genome wide methylation is known to occur (Razin and Kafri 1994). Prior to implantation the genome is largely hypomethylated (Razin and Kafri 1994), and it is during this time that repeat instability is predicted to happen.

It is difficult to devise a simple model to reconcile the differences in behavior of the expanded repeats in the hybrid cells versus the fibroblasts. Although it is tempting to simply invoke the difference in genetic backgrounds, i.e. primary human fibroblasts versus immortal hybrid cells, the issue is clearly more complex because instability is only observed in the differentiated hybrids. Those hybrids that retained an embryonic cell morphology maintained the repeats with a high level of stability despite containing the identical genetic background as the differentiated mouse cells in which the repeats are unstable. Therefore, genetic background can not provide the sole explanation for repeat instability in human-mouse hybrids. It is also not possible to use DNA methylation status to explain our results because methylated expansions are stable in the undifferentiated cells yet highly unstable in differentiated cells. Moreover, methylated alleles shifted from stable to unstable upon cellular differentiation *in vitro*. In total, the results with the hybrid cells suggest that *trans*-acting, cell specific factors provide a key role in controlling repeat stability instead of a *cis*-acting factor such as DNA methylation. If so, these factors could be missing from the differentiated mouse cells that were used as well as those rare human cells containing unmethylated full expansions. In this regard, it is important to note that DNA methylation is apparently normal outside of the region immediately surrounding the unmethylated expansion

(Burman et al.) suggesting again that this *cis* modification does not by itself control stability of the region. However, this does not rule out a role for *trans*-acting factors that recognize DNA methylation, such as methylation binding proteins (Ng and Bird 1999).

Another potential, and perhaps related, set of *trans*-acting factors controlling repeat stability could be those involved in determining DNA replication direction. It is known that replication direction has substantial influence on smaller trinucleotide repeat regions propagated in *E. coli* (Kang et al. 1995; Shimizu et al. 1996; Hirst and White 1998) and *S. cerevisiae* (Maurer et al. 1996; Freudenreich et al. 1997). It is also known that *D. melanogaster* embryonic cells initiate replication at closer intervals than adult cells (Coverley and Laskey 1994). Therefore, it is possible that repeat stability in the cells we have examined is a direct result of cell specific factors that influence replication direction through the expanded CGG repeat. A clear and testable prediction is that this direction will switch at the *FMR1* CGG repeat when the undifferentiated cells become differentiated. Replication direction in the hybrid cells could be evaluated by leading strand analysis, which has been utilized to assess the direction of DNA replication at particular loci in higher eukaryotic cells and human-mouse somatic cell hybrids (Aladjem and Wahl 1997).

In addition to the putative *trans*-acting factors discussed above, there is a *cis*-acting factor that clearly plays a major role in controlling repeat stability. Simply stated, repeat instability in the differentiated cell hybrids is proportional to repeat length. Two parameters were used to describe instability in the subclones; the proportion of subclones that maintained the parental repeat length with no variability, and the mean

magnitude of size changes observed in the subclones (see Figure 12). The relation between repeat length and instability is evident when the subclone data is considered as a whole, and is further supported by work with secondary subclones in which large and unstable repeats have contracted to form alleles of relatively small size including one in the premutation size range (170 repeats). This allele maintained complete methylation at the *Eag*I site in contrast to bonafide premutation alleles that are usually unmethylated. The 170 repeat premutation allele and a 270 repeat full mutation allele were the only two alleles that exhibited complete stability in subclones. Based on these observations and the logarithmic line of best fit calculated from the methylated alleles in the Dif6 hybrids, we estimate a threshold for stability at approximately 200-250 CGG repeats. Interestingly, this threshold corresponds to the boundary in human cells between somatically stable premutation alleles and somatically unstable full mutation alleles. It is also consistent with reports describing transgenic mice carrying premutation alleles in which the repeats are stable both somatically and intergenerationally (Bontekoe et al. 1997; Lavedan et al. 1997; Lavedan et al. 1998). In contrast, somatic instability has been observed in transgenic mice carrying human trinucleotide repeats with as few as 55 CTG (Gourdon et al. 1997) and 90 CAG (Wheeler et al. 1999) repeats. These results suggest fundamental differences in size dependent repeat stability for CGG and CAG/CTG trinucleotide repeats.

Most of the repeat length changes observed in the hybrid cells were the result of repeat contraction. Subclones that contained contraction products outnumbered those that contained expansion products by seven to one. A tendency for contraction events

has also been observed when triplet repeats are replicated in *E. coli* (Jaworski et al. 1995; Kang et al. 1995; Shimizu et al. 1996; Hirst and White 1998) and *S. cerevisiae* (Maurer et al. 1996; Freudenreich et al. 1997), and in cell hybrids that contain an expanded myotonic dystrophy repeat (Spring et al. 1998). The bias towards contraction products suggests that the mechanism of repeat instability in the hybrids is probably not unequal homologous exchange because this model predicts expansion and contraction products to be generated with equal frequency (McMurray 1995). Polymerase slippage, perhaps accompanied by the formation of secondary structures (McMurray 1995; Sinden 1999), would be more consistent with these data.

The contraction bias we have observed in the hybrid cells is also consistent with observations that male primordial germ cells of fragile X patients initially contain full mutations that undergo repeat contraction to produce spermatozoa with premutation alleles (Reyniers et al. 1993; Malter et al. 1997). Although it is not known if the mechanisms underlying somatic and germline repeat length variability share in common any mechanistic similarities, the apparent contraction bias in male primordial germ cells may occur mitotically. If the premutation to full mutation transition occurs pre-zygotically as suggested by some experimental evidence (Malter et al. 1997; Moutou et al. 1997), then somatic mosaicism might also be a product of mitotic repeat contraction. Under this model, the final repeat lengths observed in somatic cells and male germ cells might represent different manifestations of the same underlying process where contraction continues unabated to produce premutations in the germline while the contraction process is interrupted in the somatic cells by a so far unidentified stabilizing

influence. To the extent that repeat instability in the cell hybrids is a general model of repeat behavior in mitotically dividing mammalian cells, study of the cell hybrids may allow analysis of pathways relevant to instability in the male germline. Of course the important repeat length changes that might occur in the female germline are less well understood and may occur by distinct mechanisms. Given the contraction bias and stability threshold observed in the cell hybrids, it is not immediately apparent whether analysis of the cell hybrids will yield information that addresses mechanisms of repeat expansion in the female germline.

In conclusion, we have demonstrated that it is possible to generate eukaryotic cells containing unstable fragile X full mutation alleles that undergo both repeat contraction and expansion. In this hybrid cell system, repeat stability is a function of cellular differentiation and repeat size. DNA methylation appears to have no impact on repeat stability in the hybrids as opposed to the correlation of DNA methylation and repeat stability observed in the primary human fibroblasts. These results suggest that the determinants of somatic repeat instability at the *FMR1* CGG repeat are more complicated than generally believed. They also suggest a model to identify novel determinants.

ACKNOWLEDGEMENTS

The authors thank Kimberly J. Bussey for assistance with FISH analysis. This study was supported in part by a National Institutes of Health training grant T32 GM08617 (RWB) and a grant from the Council for Tobacco Research (MST).

Figure 10 legend

ON BACK OF THIS PAGE

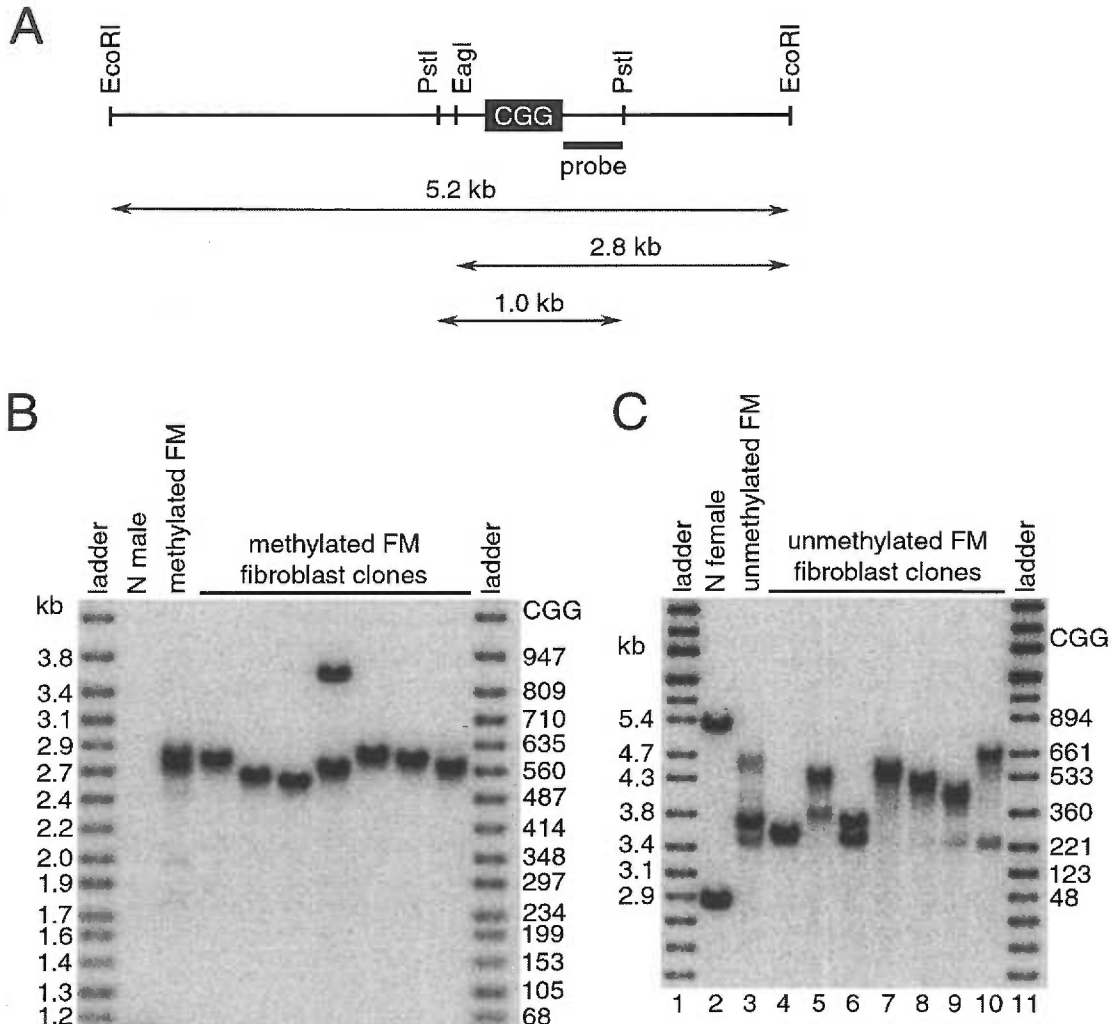

A
EcoRI
PstI
EagI
CGG
PstI
EcoRI
probe
5.2 kb
2.8 kb
1.0 kb
B
ladder
N male
methylated FM
methylated FM fibroblast clones
ladder
kb
CGG
3.8
3.4
3.1
2.9
2.7
2.4
2.2
2.0
1.9
1.7
1.6
1.4
1.3
1.2
1.0
947
809
710
635
560
487
414
348
297
234
199
153
105
68
7
1 2 3 4 5 6 7 8 9 10 11
C
ladder
N female
unmethylated FM
unmethylated FM fibroblast clones
ladder
kb
CGG
5.4
4.7
4.3
3.8
3.4
3.1
2.9
894
661
533
360
221
123
48
1 2 3 4 5 6 7 8 9 10 11

Figure 11 legend

ON BACK OF THIS PAGE

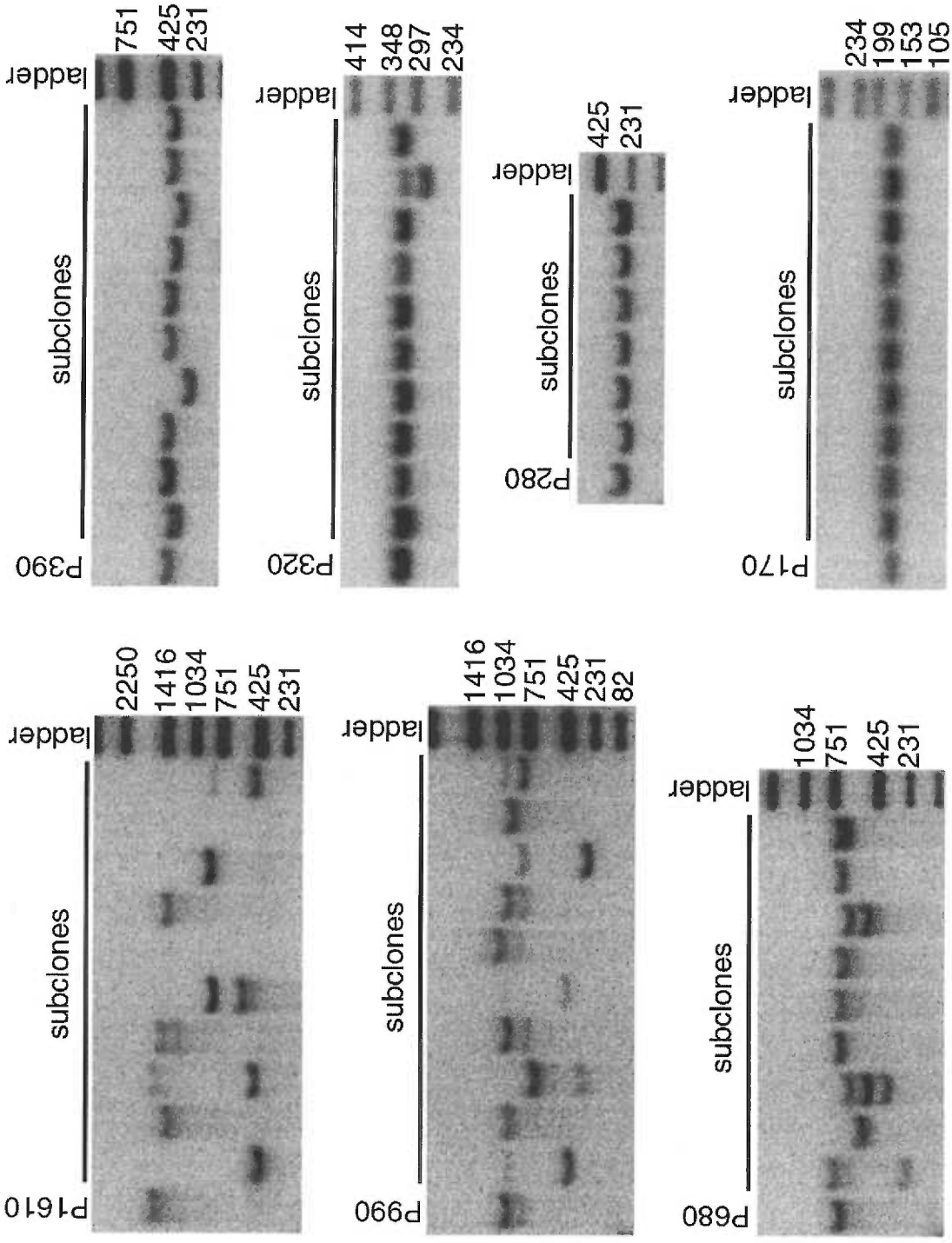
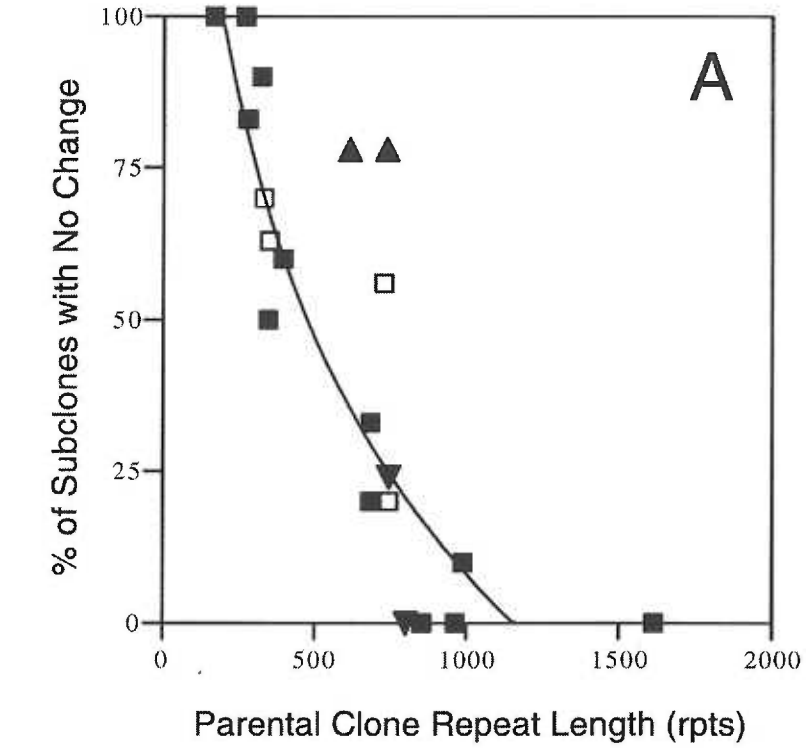

ladder
751
425
231
subclones
P390
ladder
414
348
297
234
subclones
P320
ladder
425
231
subclones
P280
ladder
234
199
153
105
subclones
P170
ladder
2250
1416
1034
751
425
231
subclones
P1610
ladder
1416
1034
751
425
231
82
subclones
P990
ladder
1034
751
425
231
subclones
P680

Figure 12 legend

ON BACK OF THIS PAGE


A
% of Subclones with No Change
100
75
50
25
0
0
500
1000
1500
2000
Parental Clone Repeat Length (rpts)

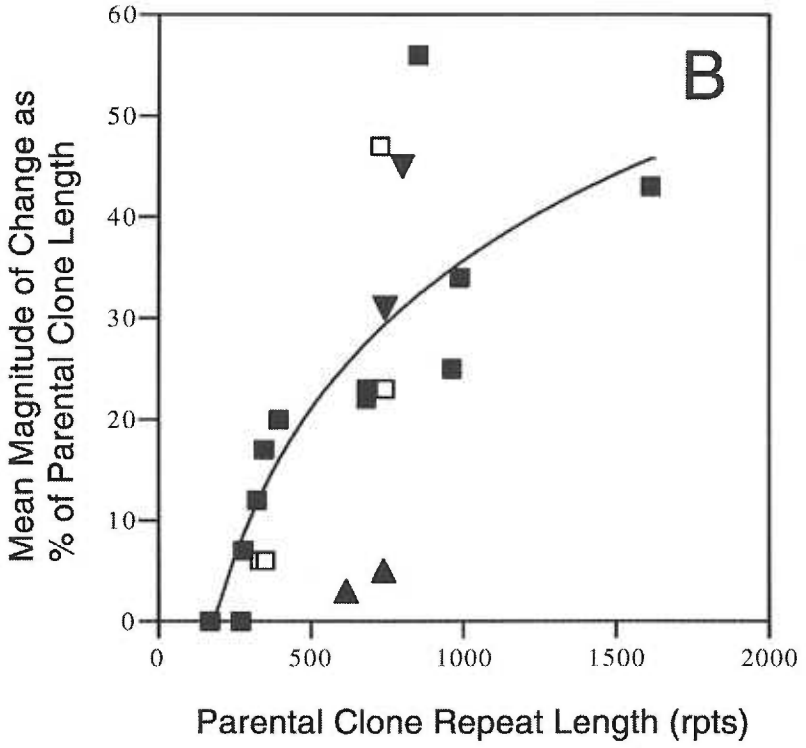

B
Mean Magnitude of Change as
% of Parental Clone Length
60
50
40
30
20
10
0
0
500
1000
1500
2000
Parental Clone Repeat Length (rpts)

Figure 13 legend

ON BACK OF THIS PAGE

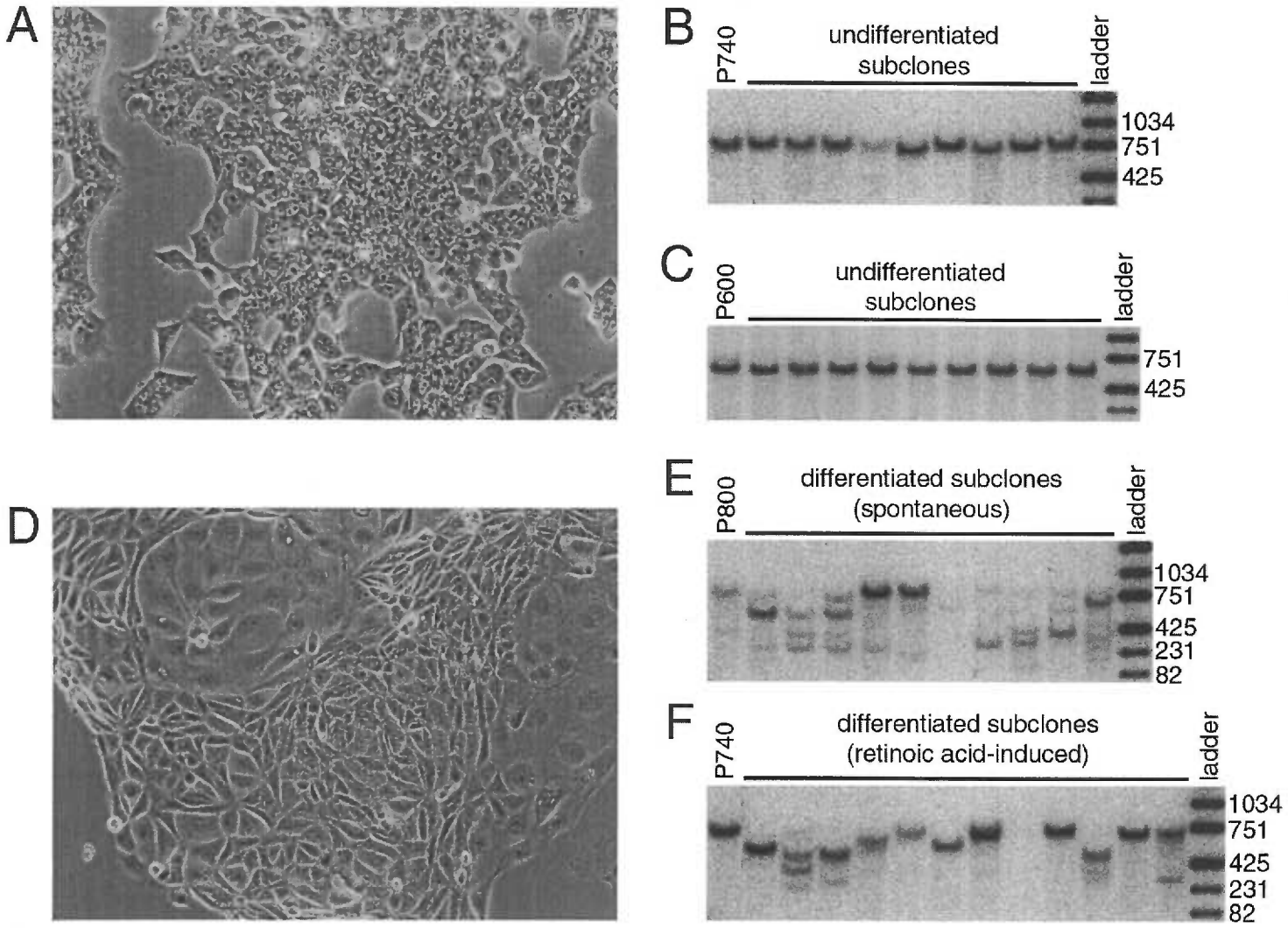

A
B
P740
undifferentiated
subclones
ladder
1034
751
425
C
P600
undifferentiated
subclones
ladder
751
425
D
E
P800
differentiated subclones
(spontaneous)
ladder
1034
751
425
231
82
F
P740
differentiated subclones
(retinoic acid-induced)
ladder
1034
751
425
231
82

Figure 14 legend

ON BACK OF THIS PAGE

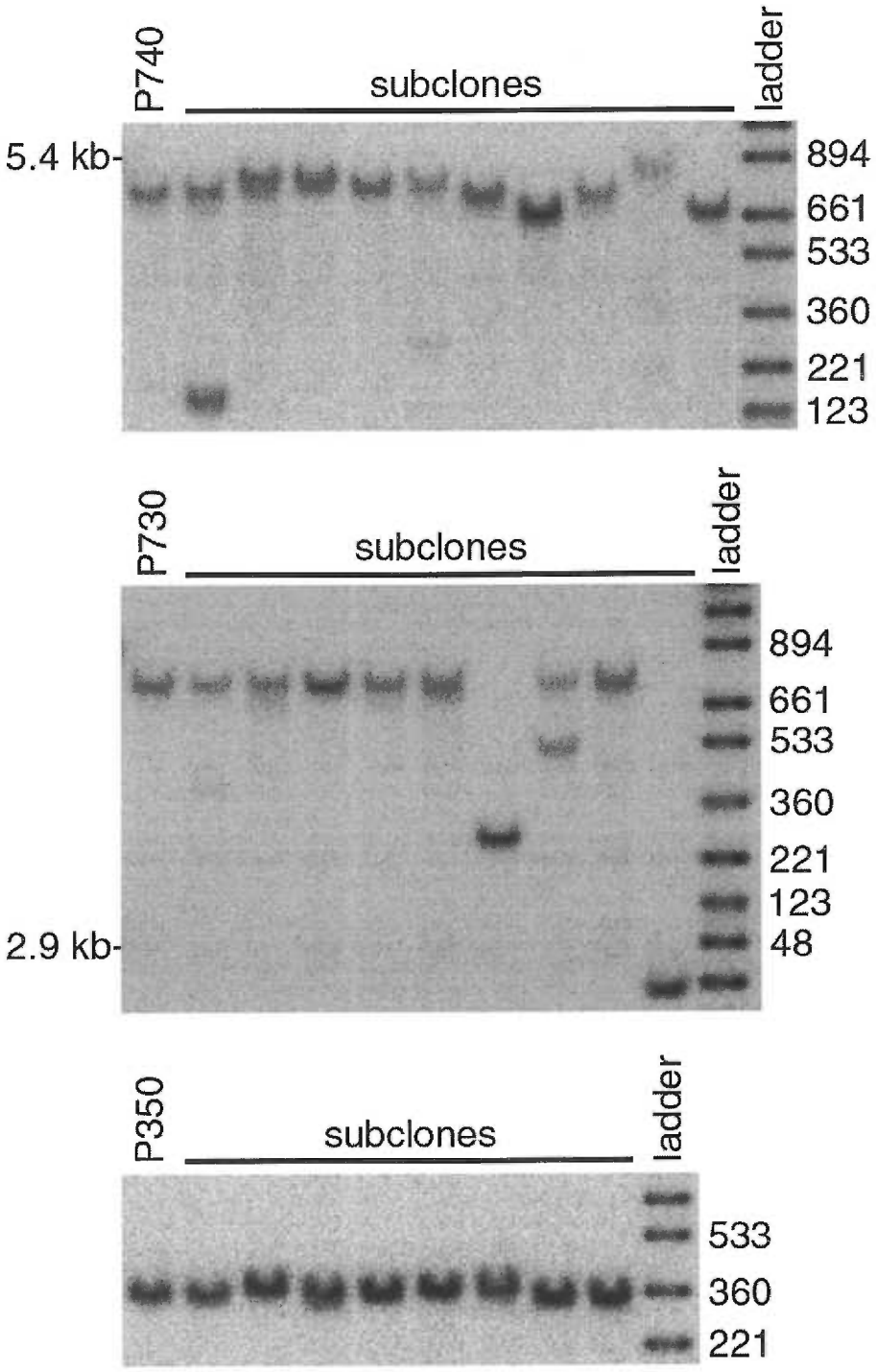

P740
subclones
ladder
5.4 kb-
894
661
533
360
221
123
P730
subclones
ladder
894
661
533
360
221
123
48
2.9 kb-
P350
subclones
ladder
533
360
221

TABLE 1. Analysis of repeat length changes in fusion subclones

Parental Clone		Subclones					
			With Change				
Cell Lines	Parental Clone Length	With No Change	Larger	Smaller	Larger and Smaller	Size Changes (rpts)	Mean Magnitude of Change
Methylated in Dif6							
TC38-89	1610	0	0	8	0	-1180,-1160,-1140,-1060,-780,-780,-720,-360,-200,-180,-120	700
TC38-89	990	1	1	8	0	-720,-680,-600,-560,-300,-220,-180,-180,-180,-80,+80	340
TC38-89	960	0	0	4	0	-660,-360,-360,-240,-200,-120,-100,-100,-20	240
TC38-89	850	0	0	6	0	-680,-680,-580,-580,-220,-140	480
GM04026	680	3	0	5	1	-440,-320,-200,-200,-160,-60,-60,-60,-20,+60	160
TC38-89	680	2	3	4	1	-540,-360,-200,-120,-80,-40,-20,+20,+20,+60,+140	150
GM05847A	390	6	0	4	0	-140,-80,-60,-40	80
GM05847A	350	5	0	5	0	-120,-100,-80,-20,-20,-20	60
TC38-89-sub	320	9	0	1	0	-40	40
GM04026	280	5	0	1	0	-20	20
TC38-89-sub	270	3	0	0	0		
TC38-89-sub	170	10	0	0	0		
Unmethylated in Dif6							
MK (TC43-97)	740	2	4	3	1	-600,-460,-80,-60,+40,+40,+40,+60,+120	170
MK (TC43-97)	730	5	0	4	0	-460,-340,-220, deletion into flanking sequence	340
MK (TC43-97)	350	5	3	0	0	+20,+20,+20	20
MK (TC43-97)	330	7	0	3	0	-20,-20,-20	20
Methylated in DelTG3							
Undifferentiated							
TC38-89	740	7	0	2	0	-40,-40	40
TC38-89	610	7	0	2	0	-20,-20	20
Differentiated							
TC38-89	800	0	0	9	0	-560,-540,-540,-520,-500,-480,-440,-420,-420,-420,-400, -220,-220,-220,-200,-160,-120,-60	360
TC38-89	740	5	0	16	0	-560,-480,-460,-440,-380,-380,-380,-360,-260,-240,-240, -240,-200,-160,-140,-120,-80,-60,-60,-60,-40,-40,-40,-40	230

CHAPTER 4

FRAGILE X FULL MUTATIONS ARE MORE SIMILAR IN SIBLINGS THAN IN UNRELATED PATIENTS: FURTHER EVIDENCE FOR A FAMILIAL FACTOR IN CGG REPEAT DYNAMICS

Robert W. Burman[1], Kim S. Anoe[1], and Bradley W. Popovich[1*]

[1]Department of Molecular and Medical Genetics, DNA Diagnostic Laboratory, Oregon Health Sciences University, Portland, OR

Submitted for publication in: *Genetics in Medicine* (April 2000)

ABSTRACT

The fragile-X CGG repeat is transiently unstable in the somatic cells of most full mutation carriers. Consequently, a heterogeneous mix of allele sizes can exist in different cell populations within an individual. This size mosaicism is visible on Southern blot analysis as a complex pattern composed of distinct bands and smears that tend to be conserved when comparing multiple tissues within an individual. Inter-tissue conservation of mutation patterns suggests that somatic repeat instability is transient and predominantly limited to a period during early embryogenesis. Further support for a limited period of instability is provided by studies that have shown that full mutation alleles are stable in cultured fibroblasts. In the current study, we sought to compare patterns of full mutation size mosaicism to determine if siblings possess mutation patterns more similar than those of unrelated controls. We have utilized novel comparison strategies that are based on overlapping mutation patterns and calculation of weighted mean CGG repeat values. Our results demonstrate that within the population of 56 full mutation carriers analyzed, mutation patterns are more similar in siblings than in unrelated patients. These results indicate that size mosaicism may be generated in a non-random manner and that familial factors influence this process.

INTRODUCTION

Fragile X syndrome [MIM 309550] is a common form of X-linked mental retardation caused by expansion of an unstable CGG trinucleotide repeat in the *5′* UTR of the *FMR1* gene (Kremer et al. 1991a; Oberle et al. 1991; Verkerk et al. 1991; Yu et

al. 1991). In normal populations, the length of the repeat tract is polymorphic and contains ~5-50 triplets. These normal alleles are stable both in the somatic cells of an individual and when transmitted between generations. Premutation alleles with ~50-220 repeats are found in unaffected carriers, and like normal alleles, are usually stable somatically (Fu et al. 1991). However, premutation alleles have a propensity to change size intergenerationally and when transmitted by a female, can expand to form full mutations with >220 repeats (Fu et al. 1991; Heitz et al. 1991; Oberle et al. 1991). Full mutation alleles are almost always associated with hypermethylation of the CGG repeat and a surrounding CpG island (Bell et al. 1991; Oberle et al. 1991; Vincent et al. 1991; Sutcliffe et al. 1992). Hypermethylation initiates and/or stabilizes transcriptional silencing of the *FMR1* gene and the resulting deficit of FMR1 protein (FMRP) renders individuals at risk to express the fragile X syndrome phenotype (Pieretti et al. 1991; Devys et al. 1993; Siomi et al. 1993).

Full mutation alleles can be highly unstable in somatic cells (Fu et al. 1991; Snow et al. 1993). As a result, most full mutation carriers possess a high degree of repeat length mosaicism, which is manifest on Southern blot analysis as a complex smeared and poly-banded mutation pattern. These mutation patterns are typically well conserved among multiple tissues within individuals (Sutherland et al. 1991; Devys et al. 1992; Wohrle et al. 1992; Wohrle et al. 1993; de Graaff et al. 1995; Wohrle et al. 1995; Moutou et al. 1997; Reyniers et al. 1999; Tassone et al. 1999), and can be conserved in monozygous twins (Devys et al. 1992; Kruyer et al. 1994; Antinolo et al. 1996). Furthermore, analysis of cultured fibroblasts derived form adult full mutation

carriers has demonstrated that methylated full mutation alleles are stable in these cells (Wohrle et al. 1993; Burman et al. 1999a). Together, this evidence suggests that somatic instability is normally limited to a period during embryogenesis and thereafter, the repeats are maintained with little or no variation in size (Devys et al. 1992). Somatic mutation patterns might then represent a snapshot of the mutational process frozen at the point the cells cease to be permissive to further changes in repeat number.

Although the process that generates somatic size mosaicism is likely to involve multiple consecutive changes in repeat number, the instant at which repeat sizes become fixed could reveal a degree of familial similarity if the process is more determinate than stochastic. To our knowledge, analysis of familial similarity in full mutation sibships has not been reported. However, similar mutation patterns have been observed in at least one full mutation brother pair and was theorized to indicate that other gene(s) may influence the pattern of expansion (Rousseau et al. 1994b). Furthermore, examination of sibships containing only offspring with a premutation has demonstrated familial clustering in which premutation size is significantly more similar within families than between families (Nolin et al. 1996). Thus, a so far unidentified familial factor(s) may influence CGG repeat dynamics.

In the present study, we assessed mutation pattern similarity in a population of 56 full mutation carriers to test the hypothesis that mutation patterns are more similar within sibships than between unrelated individuals. By examining full mutations, we sought to determine if a familial factor(s) may influence repeat dynamics in somatic

cells. Our data demonstrate that, within the patient population under study, full mutation patterns appear to be more similar within sibships than in unrelated patients.

MATERIALS & METHODS

DNA Isolation and Analysis

DNA was extracted from the peripheral blood leukocytes of 56 fragile X patients who were referred to the Oregon Health Sciences University DNA Diagnostic Laboratory for diagnostic testing. DNA isolation and Southern blot analysis were performed as described (Burman et al. 1999b).

Similarity Measurements and Statistical Analysis

Digital representations of Southern blot membranes were captured with a phosphorimaging system (Molecular Dynamics; phosphorimager SI) scanning at a resolution of 100 pixels/cm. The portion of a typical Southern blot that contained relevant data was usually restricted to a region ~ 12-15 cm in length. Thus each lane was composed of 1200-1500 rows of data points. Profile plots for each lane in the Southern blot images were generated with ImageQuant software (Molecular Dynamics) and the raw data was exported to Excel version 8.0 (Microsoft) for analysis. Because the profile plots are not smooth curves but rather a series of interconnected points, the area under a plot can be estimated by simply summing signal intensity values. For each sample pair used in this study, an Excel spreadsheet was used to calculate total area and area in common while excluding from analysis that portion of the data that fell below an

adjustable baseline. The weighted mean number of CGG repeats in each of the 56 sib samples was also calculated in Excel as follows: weighted mean lane position = Σ ((lane position)(signal intensity))/Σ (signal intensity). Signal intensity values were obtained with the baseline set at the level of the system background (normal alleles in samples from females were excluded from analysis). Weighted mean lane positions were converted to CGG repeat units by reference to the molecular weight markers. All statistical analyses, including ANOVA and *t*-tests, were performed with SPSS software version 6.1 (SPSS Inc.). Significance levels are presented without consideration of multiple comparisons.

RESULTS

To determine whether a familial component influences somatic instability, we studied 56 full mutation carriers by Southern blot analysis. The patients were grouped into 16 sibling pairs in 12 families (including 2 families with 3 affected sibs) and 15 pairs of randomly matched unrelated patients. Genomic DNA samples derived from peripheral blood were digested with *Pst*I and each pair was run in adjacent lanes for analysis. *Pst*I was utilized as the restriction endonuclease because digestion produces CGG repeat containing fragments that are significantly smaller than those obtained with the more common *Eco*RI/*Eag*I assay and thus yield much better resolution of the peaks and smears that comprise the mutation pattern. Figure 15 shows representative Southern blot results for seven patients within three sibships that are suggestive of familial mutation pattern similarity.

Numerical descriptions of full mutations are complicated by the often highly heterogeneous mix of allele sizes present within an individual. Consequently we sought a system in which comparison of mutation patterns could be made directly. One means of representing mutation patterns is the creation of profile plots from phosphorimager-derived Southern blot images (figure 16). Using the profile plots, we compared mutation pattern similarity by overlaping two plots and determining the amount of area that the plots shared in common. By measuring similarity as the proportion of the total area that was shared in common, the data was normalized for gel-to-gel variation in signal strength allowing intergel comparisons to be made. One variable that affects the measurement of area under the profile plots is the selection of the baseline level, below which area is excluded from analysis. When the baseline is set at the level of the system background, the area under the plots is derived entirely from the sample signals with no contribution from system noise (figure 17, panels with "1x baseline level"). To distinguish between specific similarities such as coincident peaks, and less specific similarities that occur near the level of the system background, the baseline was systematically varied by fractional multiples of the system background level (figure 17). Using this pattern overlap approach, we obtained similarity measures for each of the full mutation sample pairs at each baseline setting to test the hypothesis that mutation patterns, as measured by shared area, are more similar in the sib pairs than the unrelated pairs. Comparison of the two groups demonstrated that similarity was not significantly greater in the sib pairs until the baseline was raised to a level 1.5 times greater than the

system background (table II). Thereafter, the difference between the sibs and unrelated pairs increased in significance as the level of the baseline was raised.

As an alternative to the profile plot overlap strategy, we used the phosphorimager-derived data to compute the weighted mean number of CGG repeats (wMean-CGG) in each full mutation sample as presented in Material and Methods. In the population of 56 full mutation carries used in this study, the arithmetic mean of the wMean-CGG values was 575 repeats (range: 318-649 repeats). We used the wMean-CGG values of the sib samples to test the hypothesis that variation in the wMean-CGG number is greater between families than within families. Nolin and colleagues have previously used ANOVA to demonstrate highly significant clustering of CGG repeat size in sibships containing only offspring with a premutation (Nolin et al. 1996). We applied the ANOVA strategy to the 12 full mutation sibships and found that the wMean-CGG repeat sizes were significantly more similar within families than between families ($P = .004$). To verify the suitability of the randomly matched unrelated samples as controls in the previous comparison strategy, we applied the ANOVA analysis to the 15 unrelated pairs. As expected, we found no significant difference within and between groups ($P = .992$).

DISCUSSION

The molecular basis of size mosaicism in fragile X full mutation carriers is largely unknown. Because the mutational status of the repeat tract in oocytes of premutation females has yet to be described, it is unknown if the transition from

premutation to full mutation occurs pre- or post-zygotically. Thus size mosaicism may be the product of either repeat expansion from an inherited premutation or retraction from an inherited full mutation. Some experimental evidence indicates that full mutations can be tolerated in the oocytes of full mutation carriers (Malter et al. 1997) and that premutation alleles in individuals who possess both a full and premutation are better described as retraction events rather than remnants of an inherited premutation that has failed to expand in a subset of cells (Mingroni-Netto et al. 1996; Moutou et al. 1997). Together, this evidence favors a model in which an inherited full mutation undergoes repeat retraction in the cells of the early embryo to produce size mosaicism.

Most models proposed to explain repeat length variability are based on DNA replication errors involving polymerase slippage (Richards and Sutherland 1994) and the formation of secondary structures, such as hairpins, that stabilize slippage structures by minimizing the energy difference between the duplex and slippage states (McMurray 1995; Sinden 1999). When CTG repeats are replicated in *Escherichia coli*, the products of repeat contraction have shown a regularity in size that implies defined DNA structures may be involved in the contraction process (Kang et al. 1995). Some evidence suggests that defined DNA structures may likewise play a role in CGG repeat variability. In particular, a hotspot for deletions (de Graaff et al. 1995) has been described in a number of unrelated fragile X patients (de Graaff et al. 1995; Quan et al. 1995; Mannermaa et al. 1996; Mila et al. 1996; Schmucker et al. 1996) and in cloned *FMR1* sequence replicated in *E. coli* (Hirst and White 1998) and *Saccharomyces cerevisiae* (Kremer et al. 1991b). One explanation for the existence of this hotspot is

that the sequence in the vicinity may be involved in the formation of secondary structure(s) that mediates the deletion process (de Graaff et al. 1995). Because many of the deletions in the fragile X patients have unmethylated flanking sequence and occur in mosaic form in conjunction with full mutation alleles, the deletions must be the product of somatic events and may be a consequence of early embryonic repeat contraction extending into flanking sequences prior to *de novo* methylation. The identification of similar deletion breakpoints in multiple unrelated patients and other organisms implies that the process may involve conserved DNA structures.

In the current study, we have applied a strategy of overlapping plots that represent patterns of size mosaicism to determine if a familial component influences the generation of size mosaicism and if, by extension, somatic variability may occur in a non-random manner as suggested by the deletion hotspot. One advantage of the profile plot overlap method is that description and comparison of patterns can be made directly with little manipulation of the underlying data. This somewhat mechanical approach has shortcomings as well, including allowing specific similarities such as two overlapping peaks to be obscured by less intense but more numerous overlaps that occur as a consequence of most full mutation alleles falling within a common size range. To overcome this weakness, we have employed a comparison strategy that utilizes an adjustable baseline level. Comparison of the sib pairs and randomly matched unrelated pairs shows that shared area in both groups decreases as the baseline is raised, but in the unrelated group this decrease occurs at a faster rate than in the sib group. As a result, the difference in mean similarity between the groups grows larger and eventually

reaches marginal significance when the baseline is set to a level 1.5 times greater than the system background. Further increases in the base line result in an increasingly significant difference between the sib and unrelated pairs. It should be noted that we used a one-tailed *t*-test to assess the *a priori* hypothesis that similarity is greater in the sib pairs than in the unrelated pairs. Consequently the *P*-values we have obtained, when considered in conjunction with the relatively small sample size, are of somewhat questionable significance. Nonetheless, the inverse relation between baseline level and *P*-value does indicate a trend in which similarities among the sibling cohort are driven by overlapping peaks while similarities in the unrelated group are the product of low-intensity overlaps that are more readily obscured as the baseline is raised.

In a study addressing familial transmission of the fragile X repeat, Nolin and colleagues have evaluated familial clustering in sibships containing only offspring with a premutation (Nolin et al. 1996). In that study, ANOVA was used to test for statistical significance and demonstrated that repeat size in the offspring of both male and female carriers was significantly more similar within families than among families (Nolin et al. 1996). The analysis of familial clustering in premutation sibships was facilitated by the somatic stability of premutation alleles that results in an absence of mosaicism and allows a simple numerical description of repeat size. In an effort to numerically describe full mutations in a systematic manner, we have used signal intensity to weight repeat size and generate a weighted mean CGG repeat value. These weighted mean values have the advantage that all of the numerous alleles present in a Southern blot representation of a complex mosaic mutation contribute to calculation of the weighted

mean. In addition, weighted mean estimates are largely free of observer bias and the simplicity of the description (e.g. one value per full mutation carrier) facilitates statistical analysis. When the wMean-CGG repeat values were used in an ANOVA analysis, we found that significantly less variation occurs within families than between families.

Our results suggest that somatic mosaicism is generated in a non-random process that may be influenced by familial factors. The identity of these factors and the basis of the apparent similarity are not known. One possibility is that repeat variability "pathways" depend on defined DNA secondary structures which are in turn determined by the sequence composition of the repeat and adjacent genomic regions. This possibility is supported by the observation of ordered deletions of CTG repeats in *E. coli* and the fragile X deletion hotspot as mentioned previously. Support for a potential role of non-repeat DNA sequence in triplet repeat stability is provided by a study that showed intergenerational stability of the Machado-Joseph disease [MIM 109150] CAG repeat is influenced by a C/G polymorphism adjacent to the repeat tract (Igarashi et al. 1996). Some evidence indicates that familial factors may also influence the inheritance of the Huntington's disease (HD)[MIM 143100] triplet repeat. In a study of juvenile HD, Telenius and colleagues found a significant correlation between repeat length in sib pairs (Telenius et al. 1993). Potential familial influences were further suggested in a study of germline mutation at the HD locus, which identified a father and son that shared an unusual mutation spectrum (Leeflang et al. 1999).

Although it is not known if fragile X full mutation carriers inherit a full or premutation, mutation pattern similarity in siblings is more easily accounted for if both sibs inherit the same or similar allele size. In the case of post-zygotic expansion, both sibs might inherit the premutation allele size present in the soma and germline of their mother. Pre-zygotic expansion is more difficult to reconcile and would require either that the premutation to full mutation transition occurs similarly in two or more independent events in the mother's germline or that the expansion occurs in the primordial germ cells and as a consequence the same expansion allele is present in multiple oocytes. In either case, it is worth noting that the unrelated full mutation carriers used as controls in the pattern overlap strategy were randomly matched without regard to maternal premutation size. Thus the pattern overlap comparison we have made is one between sibs with identical maternal premutations and non-sibs with random maternal premutation sizes. Although it might have been instructive to pair the non-sibs according to maternal premutation size, we were unable to do so in this study due to limited maternal information for most of the isolated unrelated full mutation cases. Our results contradict, to a very limited extent, a report that found mutation mosaicism (coexistence of full and premutations in an individual) does not have a familial basis (Nolin et al. 1994). However, because mutation mosaicism was addressed as a binary variable (present or absent), the methodology utilized was quite different from that of the current study making comparison of the findings difficult.

In conclusion, our study indicates that patterns of full mutation mosaicism are more similar within sibships than in unrelated individuals. This finding suggests that

familial factors may influence somatic repeat variation and that careful examination of mosaic patterns could provide mechanistically relevant information.

ACKNOWLEDGMENTS

The authors wish to thank Drs. Mitch Turker, Peter Jacky and Charles Brenner for helpful discussions. We also thank Dr. Bob Wildin for helpful discussions and assistance with data analysis. In addition, we gratefully acknowledge the input of Dr. Will Bloch, who independently analyzed some of the samples presented in this study with a PCR-GeneScan assay. Although the PCR assay did not contribute to the analysis presented here, the GeneScan data provided the impetus for this study by first suggesting the possibility of biologically relevant allele sharing in fragile X siblings.This study was supported in part by a National Institutes of Health training grant T32 GM08617 (R.W.B.).

ELECTRONIC DATABASE INFORMATION

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/omim (for fragile X syndrome [309550], Machado-Joseph disease [109150], and Huntington disease [143100])

Figure 15 legend

ON BACK OF THIS PAGE

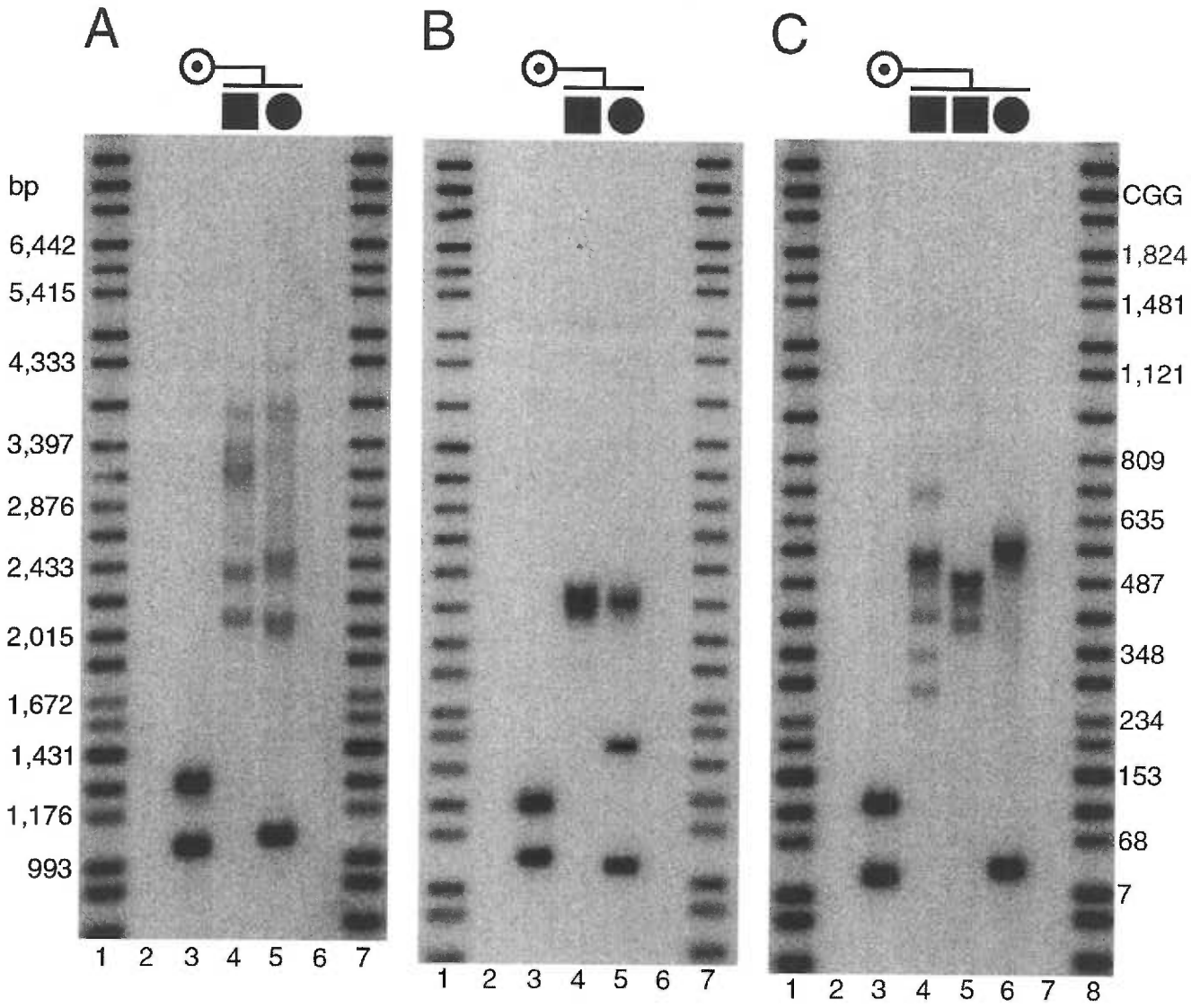

A
B
C
bp
6,442
5,415
4,333
3,397
2,876
2,433
2,015
1,672
1,431
1,176
993
CGG
1,824
1,481
1,121
809
635
487
348
234
153
68
7
1 2 3 4 5 6 7
1 2 3 4 5 6 7
1 2 3 4 5 6 7 8

Figure 16 legend

ON BACK OF THIS PAGE

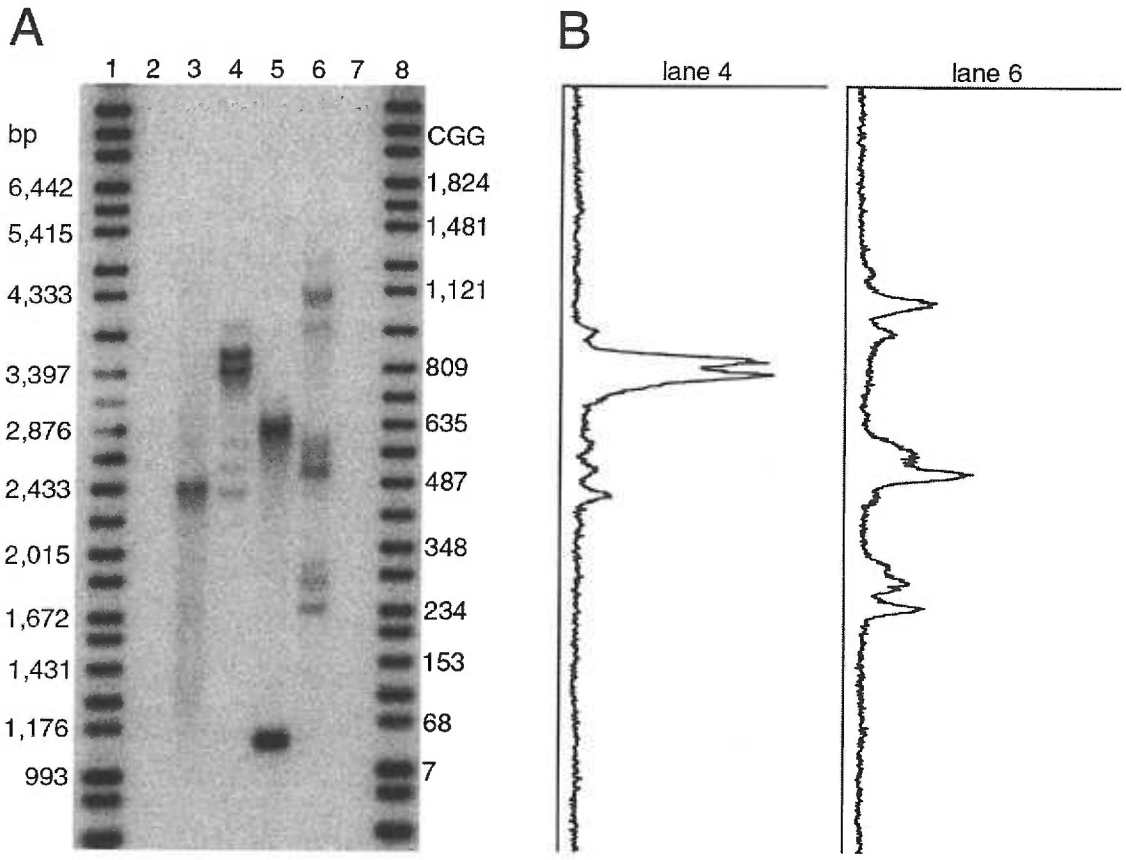

A
1 2 3 4 5 6 7 8
bp
6,442
5,415
4,333
3,397
2,876
2,433
2,015
1,672
1,431
1,176
993
CGG
1,824
1,481
1,121
809
635
487
348
234
153
68
7
B
lane 4
lane 6

Figure 17 legend

ON BACK OF THIS PAGE

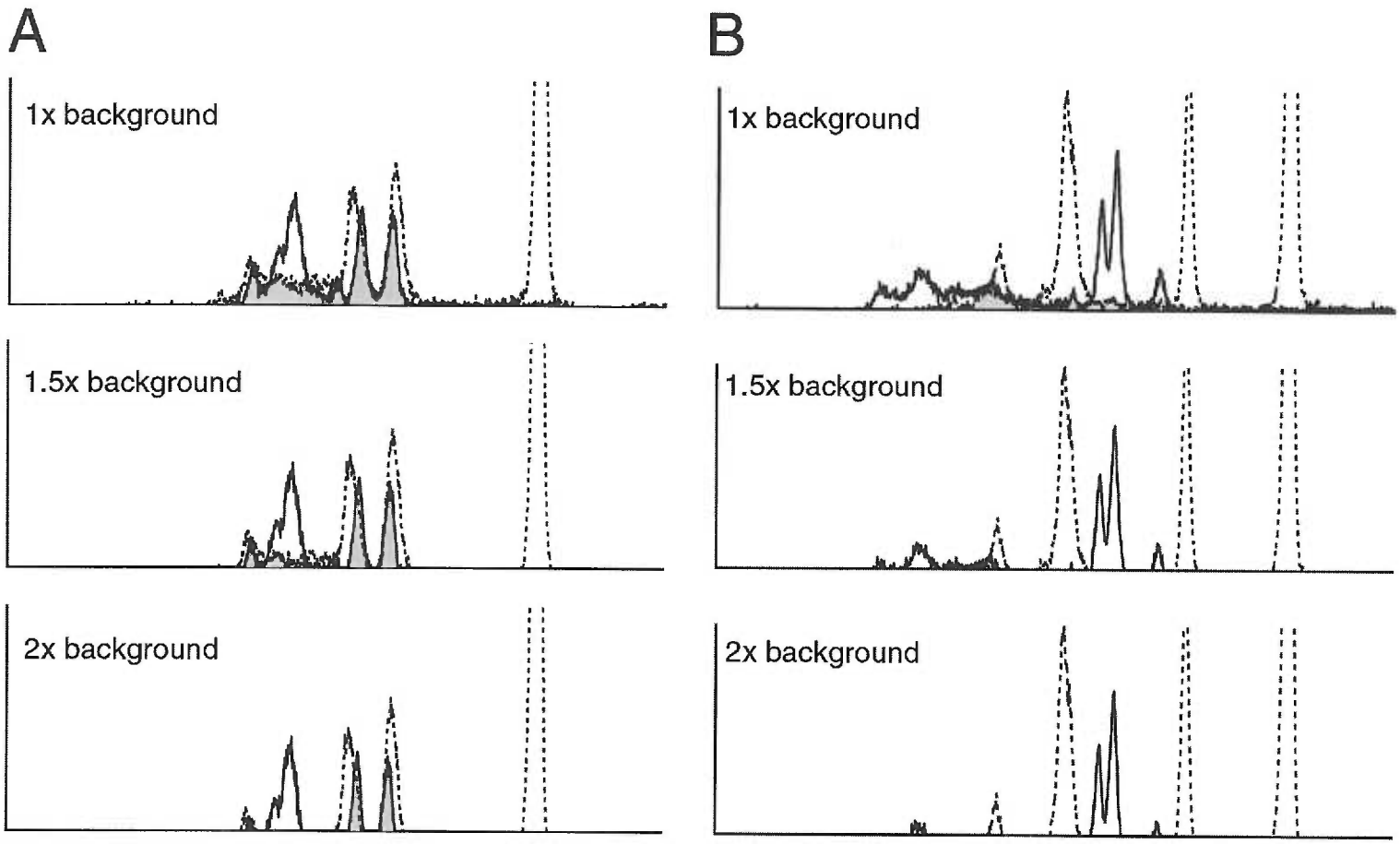

A
1x background
1.5x background
2x background
B
1x background
1.5x background
2x background

TABLE II. Full Mutation Similarity as Measured by Area in Common

Baseline Level	Mean Similarity (%)		*t*-TEST*P* (one-tailed)
	Sibling Pairs (n = 16)	Unrelated Pairs (n = 15)	
1x	30.0	22.4	.09
1.5x	20.3	11.2	.05
2x	15.2	6.2	.04
2.5x	12.3	3.6	.03
3x	10.0	1.5	.02

CHAPTER 5

CONCLUSION

Despite the accumulation of much empirical data describing the inheritance of the fragile X CGG repeat, little knowledge is available regarding the actual biological basis of repeat methylation and length variability. A number of fundamental questions remain to be answered, including whether the transition from premutation to full mutation occurs pre- or post-zygotically and why expansions in excess of ~200 repeats trigger DNA methylation. In addition, description of the mechanism(s) of repeat instability has been elusive, despite a plethora of models, each partially supported by experimental data collected in model organisms. Notwithstanding these deficiencies in knowledge, it is clear that the behavior of the fragile X CGG repeat can be quite varied in different contexts. For instance, it is well established that a strong parent of origin effect determines the probability of intergenerational expansion (Fu et al. 1991). Evidence also indicates that somatic repeat instability occur predominately in embryonic cells but not those of the adult (Devys et al. 1992; Wohrle et al. 1993; Burman et al. 1999a). The cellular and molecular variables that influence this contextual plasticity have not been described. The goal of this thesis has been an investigation of CGG repeat variability in somatic cells. I have studied variation in both repeat length and methylation using a combination of molecular, cellular, and family-study approaches. In the course of these studies I have generated a viable *in vitro* model system of repeat dynamics and have identified novel determinants of somatic variation both in this model system and *in vivo*.

To better understand the relation between repeat expansion and hypermethylation, I have studied DNA methylation patterns in a rare individual (MK)

who harbors a fully expanded, yet unmethylated, CGG repeat. Analysis of *FMR1* methylation was performed with a novel assay (as applied to the fragile X syndrome) that provided data on the methylation status of numerous CpG sites, including those located within the CGG repeat. I demonstrated that the full mutation in MK is apparently free of methylation both in the repeat and in ~1 kb of flanking sequence. This finding has contributed to the fragile X literature by defining the extent of hypomethylation in one patient who belongs to a rare class of hypomethylated full mutation carriers (Loesch et al. 1993; McConkie-Rosell et al. 1993; Hagerman et al. 1994; Merenstein et al. 1994; Rousseau et al. 1994b; Smeets et al. 1995; de Vries et al. 1996; Lachiewicz et al. 1996; Wang et al. 1996; Wohrle et al. 1998; Burman et al. 1999b; Taylor et al. 1999). Prior to this analysis, methylation of these hypomethylated full mutation alleles was described with conventional methylation-sensitive restriction enzymes that have the disadvantage of assessing the methylation status of one, or at most two, CpG dinucleotides. Since the fragile X promoter is located in a CpG island (Bell et al. 1991), an open question has been how well methylation of these restriction-enzyme sites reflects the overall methylation status of the entire CpG island in hypomethylated patients. Previous studies have demonstrated that in the case of a hypermethylated CpG island, methylation of the diagnostic *Eag*I site was representative of surrounding DNA (Hansen et al. 1992; Hornstra et al. 1993; Stoger et al. 1997). My analysis supports and extends the findings of these earlier studies by further validating the notion that hypomethylation of the *Eag*I site can also be representative of surrounding sites and that the methylation of individual molecules tends to be

coordinated with all CpG sites either largely hypo- or hyper-methylated (Stoger et al. 1997). The finding of a completely unmethylated CpG island also speaks to the origin of this unusual methylation variant. Because the CGG repeats are theorized to be the initial targets of *de novo* methylation (Bestor and Tycko 1996; Bender 1998; Chen et al. 1998), one tenable hypothesis regarding the origin of promoter hypomethylation was that the repeats were somehow methylated in the absence of methylation at flanking sequences. My results indicate that this possibility is not the explanation for the unmethylated full mutation in patient MK. It would be instructive in future studies to assess *FMR1* methylation in other hypomethylated full mutation carriers with the McrBC assay to determine if the same methylation pattern is observed in unrelated individuals. Because males carrying these unusual mutations tend to be "high-functioning" with phenotypic features in the mild to normal range of affectedness, some investigators have pursued the possibility of therapeutic demethylation of typical methylated full mutations (Chiurazzi et al. 1998). These demethylation experiments have shown that partial restoration of transcriptional activity, as measured by RT-PCR, can be achieved by treating cultured full mutation lymphoblasts with 5-azacytidine (Chiurazzi et al. 1998). An outstanding question that remains unanswered is the duration of reactivation that is achieved with a single treatment. If methylation is not completely stripped from the full mutation, as one might expect given the very high density of CpG sites, methylation may spread after treatment and again silence the *FMR1* gene. My results speak indirectly to the feasibility of therapeutic demethylation by showing that the favorable phenotypic outcome observed in patient MK is the result

of complete hypomethylation and not a mixed pattern of methylated and unmethylated CpGs. Finally, the demonstration of a completely unmethylated CpG island validates the use of expansion alleles derived from MK for analysis of the proposed relation between methylation and repeat stability discussed below.

To investigate the biological basis of MK's unmethylated expansion, I examined the methylation status of selected linked and unlinked repetitive element loci. This analysis was undertaken to learn if a defect in a fundamental methylation process such as *de novo* or maintenance methylation might explain the existence of MK's unusual methylation variant. Although my analysis did not ultimately determine the origin of the unmethylated expansion, I have tested a number of theories that for the first time begin to address the basis of these rare but potentially important methylation variants. One possible explanation for MK's unmethylated expansion is that methylation that may have once been present has been lost due to a failure of maintenance methylation. The term maintenance methylation refers to the process in which methylation patterns present on the DNA strands that serve as templates for replication are copied onto the newly synthesized DNA strands. Murine cells that carry a homozygous targeted disruption of the maintenance methyltransferase gene *dnmt1*, exhibit global hypomethylation of the genome (Li et al. 1992; Lei et al. 1996). I postulated that a region-specific failure of maintenance methylation could have rendered MK's full mutation free of methylation. The *FMR1* specificity of this hypothesis was suggested by the *Arabidopsis thaliana* mutant *ddm1* that displays a hypomethylation phenotype specific to repeated sequence (Vongs et al. 1993; Kakutani et al. 1996; Jeddeloh et al.

1998). One model of DDM1 function is that the protein mediates accessibility of the DNA to the maintenance methyltransferase (Jeddeloh et al. 1999). Thus regions of the genome that exist in condensed chromatin, such as fragile X full mutations (Luo et al. 1993; Eberhart and Warren 1996), might be susceptible to the loss of methylation patterns over time if a *trans* defect existed at a gene similar to *ddm1*. I assessed the efficacy of maintenance methylation in the *FMR1* region by examining methylation of specific *Alu* elements, which are highly methylated in somatic cells (Schmid 1991). I found *Alu* elements flanking MK's unmethylated expansion at a distance of ~ 5 kb were completely methylated in a pattern indistinguishable from controls. This result indicates that methylation can be successfully maintained near the unmethylated expansion and suggests that a *trans* defect that affects maintenance of methylation patterns in the *FMR1* region is unlikely to account for MK's unmethylated expansion.

Since perturbed methylation maintenance appears not to be the explanation for MK's unmethylated expansion, an alternative hypothesis is that the lack of methylation is caused by a *trans* defect in a gene that mediates *de novo* methylation of the expanded CGG repeat. Although the precise timing of expansion-mediated methylation has not been determined, expansion is believed to precede the methylation event (Devys et al. 1992; Malter et al. 1997), which may occur during the wave of genome-wide *de novo* methylation in early embryogenesis (Razin and Kafri 1994). One class of specific targets of *de novo* methylation are repetitive DNA elements such as *Alu*, *L1*, and retroviral elements (Yoder et al. 1997). Methylation of these elements is believed to be important for controlling transposition and maintaining genome integrity (Yoder et al.

1997). Consistent with this "genome protection" hypothesis, *B1* elements in the promoter region of the mouse *aprt* gene have been demonstrated to serve as specific targets for methylation (Yates et al. 1999). Depending on the parental origin, repetitive elements can be inherited in an unmethylated state (Yoder et al. 1997). These unmethylated elements are subject to *de novo* methylation during early development at the time of genome-wide *de novo* methylation. Although it is not understood how repetitive elements signal methylation, one hypothesis asserts that the formation of DNA secondary structures may serve as the signal (Bestor and Tycko 1996; Bender 1998). Methylation of the fragile X CGG repeat may also occur as the result of secondary structure as suggested by *in vitro* studies that have demonstrated certain types of CGG repeat structures are efficient substrates for *de novo* methylation (Chen et al. 1995; Chen et al. 1998). Thus the timing and specificity of *de novo* methylation of both repetitive elements and the *FMR1* CGG repeat may have common characteristics. I sought to determine if the unmethylated full mutation in MK was caused by a defect in *de novo* methylation that may be reflected in the methylation of repetitive element loci. I studied methylation at loci likely to have been inherited in the unmethylated state and subject to *de novo* methylation during MK's early development. My analysis demonstrated that methylation of three such loci had occurred in MK with a pattern indistinguishable from controls. These results suggest that if a *trans*-defect in *de novo* methylation is responsible for MK's unmethylated expansion, then methylation of the repetitive elements and the expanded CGG repeat is likely to be mediated by distinct pathways.

To further assess the possibility that a *trans*-defect in *de novo* methylation may be responsible for MK's unmethylated expansion, I attempted to complement the apparent deficiency by cell fusion. Cell fusion has been used by others to partially complement the methylation deficiency present in the cells of patients with the ICF syndrome (Schuffenhauer et al. 1995). I utilized the microcell mediated chromosome transfer technique to place an unmethylated expansion of ~350 repeats in a mouse EC cell line known to be capable of *de novo* methylating targets composed of repetitive elements (Turker et al. 1991; Yates et al. 1999). Despite propagation of this allele through ~24 population doublings in the EC cell background, the expansion remained unmethylated. The failure of these cells to methylate this allele could reflect a *cis*-variation on MK's chromosome that renders his expansion inherently resistant to *de novo* methylation. Another explanation is that mouse and human cells have an inherent difference in the way CGG repeat expansions are metabolized. In light of the ongoing efforts to create an animal model of CGG repeat expansion, these results indicate that previous assumptions regarding methylation of CGG repeats in the mouse should be re-examined (Bontekoe et al. 1997). In a future analysis, it may be instructive to test the ability of human embryonic stem cells to methylate MK's full mutation. If at all possible, this type of complementation experiment would be greatly strengthened by analysis of an unmethylated full mutation allele that is known to be capable of serving as a target for *de novo* methylation. One means of obtaining such an allele may be through demethylation of a typical full mutation. However, given the potential difficulty of producing a completely unmethylated *FMR1* CpG island, this approach may not be

successful. An alternate strategy may be to isolate clonal cell lines from extra-embryonic tissue such as chorionic villi that can contain full mutation alleles in an undermethylated state (Rousseau et al. 1991a; Sutherland et al. 1991; Sutcliffe et al. 1992). If a completely unmethylated expansion could be isolated in clonal form, these cells would serve as a convenient source of "methylatable" full mutation alleles.

One prominent barrier that has hindered detailed understanding of fragile X CGG repeat dynamics has been the absence of an animal model. Because a naturally occurring animal model of CGG repeat variability has not been described, three independent groups have created transgenic mice that carry human premutation alleles integrated at random sites in the murine genome. Unfortunately, none of these transgenic mice displayed either intergenerational or somatic instability despite carrying alleles of sufficient size to be unstable in humans (Bontekoe et al. 1997; Lavedan et al. 1997; Lavedan et al. 1998). In light of these difficulties, I sought an *in vitro* model of repeat dynamics that would allow analysis of variables that influence CGG repeat instability. I have studied repeat dynamics in both primary human fibroblasts and human-mouse somatic cell hybrids. Analysis in both cell types has the distinct advantage that repeat behavior is assessed within the native human chromosomal context. The cell hybrids have the further advantage that repeat behavior can be compared in different cellular backgrounds and the hybrid cells are not subject to replicative senescence as are primary human fibroblasts. I utilized human fibroblasts from typical full mutation carriers and the hypomethylated full mutation carrier MK to test the hypothesis that methylated fragile X expansions are inherently more stable than

unmethylated expansions. This hypothesis is not my own, but has rather been present in the fragile X literature for some time (Wohrle et al. 1996). Experimental evidence in support of this hypothesis has been incomplete and consists of the findings that methylated full mutation alleles are stable in cultured fibroblasts (Wohrle et al. 1993) and that the mutation pattern of unmethylated alleles in a heterogeneous mix of cells tends to be more smeared and diffuse than is normally observed in cells with methylated expansions (Wohrle et al. 1998; Glaser et al. 1999). I compared the stability of methylated and unmethylated expansions in primary human fibroblasts by isolating clonal cell lines and assessing stability by Southern blot analysis of DNA samples derived from the clones after propagation in culture for ~ 24 population doublings. My data confirmed a previous report of *in vitro* stability of methylated full mutations (Wohrle et al. 1993) and demonstrated for the first time that unmethylated alleles are unstable relative to their methylated counterparts in clonal fibroblast lines. This finding is important because it has provided the best evidence so far available that methylation status correlates with repeat stability. Furthermore, this data partially supports the hypothesis that somatic repeat instability may occur during the period of genome-wide undermethylation in early embryogenesis.

The Southern blot hybridization patterns observed with the unmethylated alleles were in most cases characterized by smears that seemed to extended into regions of low molecular weight and were thus suggestive of instability occurring primarily by repeat contraction. Although I hoped to study the specific types of repeat length changes that seemed to occur in the fibroblasts with an unmethylated expansion, replicative

senescence limited my ability to analyze these cells. To overcome this difficulty and gain a better understanding of the magnitude and direction of repeat length changes, I generated somatic cell hybrids in which both methylated and unmethylated full mutations were contained in a differentiated, and essentially imortal, mouse background. Analysis of the hybrid cells showed that both methylated and unmethylated expansions were highly unstable in the mouse background. Furthermore, both expansion and contraction events were observed in the cell hybrids, although contraction events significantly outnumbered instances of repeat expansion. Instability was a function of repeat length with large alleles more unstable than smaller alleles. These results have contributed to understanding fragile X repeat dynamics in various ways. First, in contrast to the situation in human fibroblasts, methylation is not sufficient to confer repeat stability in the mouse background. This finding suggests that the correlation between repeat stability and methylation observed in the human fibroblasts is either not causative or is mediated by cell-specific factors that may be absent in the differentiated hybrid cells. These results also demonstrate that the fragile X CGG repeats are potentially unstable in mouse cells, when located in the native human chromosomal context. This finding should have direct implications for ongoing efforts in other laboratories to create a mouse model of repeat expansion. Although I have shown that repeat expansion can potentially occur in a mouse background, my data indicates that instability in the hybrids occurs primarily by contraction and that a stability threshold exists at ~200-250 repeats. Thus expanded alleles are not well tolerated in differentiated cell hybrids and may not be tolerated in mouse cells *in vivo*.

At a minimum, these results suggest caution when trying to engineer large expansions in mice.

Interestingly, the stability threshold I have identified in the cell hybrids corresponds to the boundary in humans between somatically stable premutations and unstable full mutations (Fu et al. 1991). Although the mechanism and direction (expansion or contraction) of repeat length variability in human somatic cells is not known, the results I have obtained in the primary human fibroblasts and cell hybrids favor a model in which repeat contraction is the predominant type of instability in somatic cells. To the extent that cell hybrids can serve as a general model of repeat instability in mitotically dividing cells, it may be worthwhile to use the differentiated hybrids in future studies that seek to describe the molecular mechanism(s) of somatic instability.

In addition to the influence of repeat length on instability, I demonstrated that another potential determinant of repeat behavior is cellular differentiation. I found that large methylated expansions of sufficient size to be highly unstable in the differentiated cell hybrids were surprisingly stable in undifferentiated hybrids produced with EC cells. Significantly, instability in these EC hybrids could by induced by spontaneous or retinoic acid-induced differentiation. The unexpected role of cellular differentiation in repeat stability is another indication that repeat behavior can be critically influenced by cellular context. Because the expanded repeats behave differently in differentiated and undifferentiated cells with the same genetic background, I believe that comparison of these hybrid cell types offers a unique opportunity to further define variables that

influence CGG repeat dynamics. In future studies it will be important to more thoroughly compare the behavior of multiple allele sizes in the undifferentiated hybrids to better understand the limits of repeat stability in these cells. Although I can only speculate about the type of changes that accompany differentiation and induce instability, one logical candidate is a change in the direction of DNA replication through the repeat. Analysis of repeat dynamics in *E. coli* and *S. cerevisiae* by other groups has shown that the direction of replication is an important determinant of instability (Shimizu et al. 1996; Hirst and White 1998; White et al. 1999; Balakumaran et al. 2000). If replication direction has a similar influence on repeat stability in cell hybrids, I would predict that the direction of replication switches upon differentiation. Future studies to determine the direction of replication in both hybrid cell types might reveal a correlation between stability and direction. Although this type of analysis would not necessarily indicate a causative relation, when considered in conjunction with data previously obtained in *E. coli* and *S. cerevisiae*, this experimental evidence would at least determine whether replication polarity is a plausible determinant of stability in mammals.

In an experimental approach quite distinct from the cell hybrid model described above, I retrospectively analyzed mutation patterns in 56 fragile X patients to determine if a familial factor(s) may influence somatic repeat length variability. This study was prompted by the work of Nolin and colleagues who showed that in sibships containing only premutation carriers, repeat size was significantly more similar within families than between families (Nolin et al. 1996). This result suggested that some type of

familial factor influences the degree of expansion (premutation to larger premutation) between generations (Nolin et al. 1996). Because the study was deliberately restricted to sibships with only premutation carriers, it has been an open question whether or not a similar effect operates when intergenerational transmission results in siblings who possess full mutations.

One prominent limitation that has hindered analysis of full mutations is the complexity of the mutation pattern. When visualized by Southern blot analysis, most full mutations display a hybridization pattern composed of multiple bands and smears that can be spread over a wide range of repeat sizes. Numerical description is thus complicated by the large number of alleles that contribute, in varying degrees, to the mutation pattern. To test mutation pattern similarity in siblings, I devised two novel comparison strategies that were designed to overcome these difficulties. One method was based on overlapping phosphorimager derived profile plot representations of mutation patterns to determine the amount of area two plots share in common. I then used area in common as an estimate of similarity. As a whole, the data I obtained with this strategy was somewhat inconclusive but did suggest a trend in which sibling samples appeared to share peaks in common to a greater extent than unrelated samples.

The second method I used to examine pattern similarity was based on calculation of a weighted mean CGG repeat value. This value was essentially a measure of the mean allele size present in a full mutation sample with the contribution of each allele weighted by its relative abundance compared to the other alleles in the sample. By utilizing raw data captured directly with a phosphorimager, the weighted means were

designed to be free of observer bias and an accurate representation of all the alleles that comprise a particular mutation pattern. I used the weighted means in an ANOVA analysis to determine if these values were clustered in families as was previously observed in premutation sibships (Nolin et al. 1996). My data demonstrated that weighted mean CGG repeat values were significantly more similar within families than between families ($P = .004$). This result suggests that some type of familial factor influences the pattern of mosaicism observed in fragile X siblings. The study did not address the identity of the factor(s) responsible for familial similarity. One possibility is that the size of the inherited allele determines the pattern of mosaicism in the offspring. Pattern similarity in siblings would then imply that sibs inherit alleles of the same or similar size. In the case of a pre-zygotic premutation to full mutation expansion event, co-inheritance of the same allele would be most easily explained by expansion in the maternal primordial germ cells that results in multiple oocytes with the same expansion size. It is also possible that sequence variation in *cis* or genes in *trans* influence the "pathway" of repeat instability. Siblings could reasonably be expected to share in common *cis* variations and some alleles at other loci that may influence repeat variability in *trans*. As a consequence, full mutation patterns in siblings could be more similar than those of unrelated full mutation carriers. Of course, the postulated role of inherited allele size and *cis/trans* effects are not exclusive. Further analysis of mutation patterns in families is somewhat unlikely to distinguish between these possibilities. An alternate avenue of investigation may instead be analysis of instability in cultured cells. Since I have demonstrated that instability of full mutation alleles can be triggered by

fusion with a differentiated mouse cell, it may be possible to compare the instability of alleles with similar size from different sources. For instance, clonal human fibroblast lines that each contain ~600 repeats could be isolated from two unrelated fragile X patients. Fusion of these fibroblasts with mouse cells would trigger instability and the direction and magnitude of change could be monitored in a large number of clones derived from each fusion. If "familial" factors influence instability then one might expect to observe different spectrums of instability in the hybrids derived from the two unrelated patients. This type of analysis could also be carried out with cells derived from siblings and half-siblings to determine if the spectrum of instability is influenced by the degree of relatedness.

Analysis of CGG repeat dynamics in model systems such as *E. coli* and *S. cerevisiae*, and now cell hybrids has clearly contributed to understanding determinants that can influence CGG repeat stability. In addition, studies that have described the inheritance of the fragile X mutation have defined additional variables such as sex of the transmitting parent, size of the transmitted allele, number of uninterrupted CGG repeats, and other unidentified familial factors that determine the risk, and possibly extent, of expansion in humans. Perhaps the one phrase that best describes the behavior of the fragile X CGG repeat is contextual variation. Even within humans, stability of the CGG repeat is quite variable as demonstrated by varying degrees of stability in the male and female germlines, in the cells of the early embryo and adult, and when methylated and unmethylated. Extending analysis of CGG repeat dynamics to other organisms has demonstrated further variability between human fibroblasts and mouse-human hybrids,

and differentiated and undifferentiated cell hybrids. In the absence of an animal model, a complete description of CGG repeat dynamics will depend on a marriage of human and model system data to obtain a unified understanding of CGG repeat behavior at all stages of human development including gametogenesis and early embryogenesis.

REFERENCES

Abitbol M, Menini C, Delezoide AL, Rhyner T, Vekemans M, Mallet J (1993) Nucleus basalis magnocellularis and hippocampus are the major sites of FMR-1 expression in the human fetal brain. Nature Genetics 4:147-53

Aladjem MI, Wahl GM (1997) Mapping replication fork direction by leading strand analysis. Methods 13:281-92

Allingham-Hawkins DJ, Babul-Hirji R, Chitayat D, Holden JJ, Yang KT, Lee C, Hudson R, et al (1999) Fragile X premutation is a significant risk factor for premature ovarian failure: the International Collaborative POF in Fragile X study- -preliminary data. American Journal of Medical Genetics 83:322-5

Antinolo G, Borrego S, Cabeza JC, Sanchez R, Sanchez J, Sanchez B (1996) Reverse mutation in fragile X syndrome. American Journal of Human Genetics 58:237-9

Armour JA, Harris PC, Jeffreys AJ (1993) Allelic diversity at minisatellite MS205 (D16S309): evidence for polarized variability. Human Molecular Genetics 2:1137-45

Ashley CT, Sutcliffe JS, Kunst CB, Leiner HA, Eichler EE, Nelson DL, Warren ST (1993a) Human and murine FMR-1: alternative splicing and translational initiation downstream of the CGG-repeat. Nature Genetics 4:244-51

Ashley CT, Jr., Warren ST (1995) Trinucleotide repeat expansion and human disease. Annual Review of Genetics 29:703-28

Ashley CT, Jr., Wilkinson KD, Reines D, Warren ST (1993b) FMR1 protein: conserved RNP family domains and selective RNA binding. Science 262:563-6

Ashley-Koch AE, Robinson H, Glicksman AE, Nolin SL, Schwartz CE, Brown WT, Turner G, et al (1998) Examination of factors associated with instability of the FMR1 CGG repeat. American Journal of Human Genetics 63:776-85

Atkins L, Taft PD, Kalal KP (1962) Asynchronous DNA synthesis of sex chromatin in human interphase nuclei. Journal of Cell Biology 15:390-393

Bakker CE, Consortium D-BFX (1994) Fmr1 knockout mice: a model to study fragile X mental retardation. Cell 78:23-33

Balakumaran BS, Freudenreich CH, Zakian VA (2000) CGG/CCG repeats exhibit orientation-dependent instability and orientation-independent fragility in saccharomyces cerevisiae. Human Molecular Genetics 9:93-100

Bell MV, Hirst MC, Nakahori Y, MacKinnon RN, Roche A, Flint TJ, Jacobs PA, et al (1991) Physical mapping across the fragile X: hypermethylation and clinical expression of the fragile X syndrome. Cell 64:861-6

Bender J (1998) Cytosine methylation of repeated sequences in eukaryotes: the role of DNA pairing. Trends in Biochemical Sciences 23:252-6

Bennetto MA, Pennington BF (1996) The Neuropsychology of Fragile X Syndrome. In: Hagerman RJ, Cronister A (eds) Fragile X Syndrome: Diagnosis, Treatment, and Research. The John Hopkins University Press, Baltimore, pp 210-248

Bestor TH, Tycko B (1996) Creation of genomic methylation patterns. Nature Genetics 12:363-7

Bingham PM, Scott MO, Wang S, McPhaul MJ, Wilson EM, Garbern JY, Merry DE, et al (1995) Stability of an expanded trinucleotide repeat in the androgen receptor gene in transgenic mice. Nature Genetics 9:191-6

Bontekoe CJ, de Graaff E, Nieuwenhuizen IM, Willemsen R, Oostra BA (1997) FMR1 premutation allele (CGG)81 is stable in mice. European Journal of Human Genetics 5:293-8

Burman RW, Popovich BW, Jacky PB, Turker MS (1999a) Fully expanded FMR1 CGG repeats exhibit a length- and differentiation-dependent instability in cell hybrids that is independent of DNA methylation. Human Molecular Genetics 8:2293-2302

Burman RW, Yates PA, Green LD, Jacky PB, Turker MS, Popovich BW (1999b) Hypomethylation of an expanded FMR1 allele is not associated with a global DNA methylation defect. American Journal of Human Genetics 65:1375-1386

Burright EN, Clark HB, Servadio A, Matilla T, Feddersen RM, Yunis WS, Duvick LA, et al (1995) SCA1 transgenic mice: a model for neurodegeneration caused by an expanded CAG trinucleotide repeat. Cell 82:937-48

Cantu JM, Scaglia HE, Medina M, Gonzalez-Diddi M, Morato T, Moreno ME, Perez-Palacios G (1976) Inherited congenital normofunctional testicular hyperplasia and mental deficiency. Human Genetics 33:23-33

Chen X, Mariappan SV, Catasti P, Ratliff R, Moyzis RK, Laayoun A, Smith SS, et al (1995) Hairpins are formed by the single DNA strands of the fragile X triplet

repeats: structure and biological implications. Proceedings of the National Academy of Sciences of the United States of America 92:5199-203

Chen X, Mariappan SV, Moyzis RK, Bradbury EM, Gupta G (1998) Hairpin induced slippage and hyper-methylation of the fragile X DNA triplets. Journal of Biomolecular Structure & Dynamics 15:745-56

Chiurazzi P, Macpherson J, Sherman S, Neri G (1996) Significance of linkage disequilibrium between the fragile X locus and its flanking markers. American Journal of Medical Genetics 64:203-8

Chiurazzi P, Pomponi MG, Willemsen R, Oostra BA, Neri G (1998) In vitro reactivation of the FMR1 gene involved in fragile X syndrome. Human Molecular Genetics 7:109-13

Coffee B, Zhang F, Warren ST, Reines D (1999) Acetylated histones are associated with FMR1 in normal but not fragile X-syndrome cells. Nature Genetics 22:98-101

Comery TA, Harris JB, Willems PJ, Oostra BA, Irwin SA, Weiler IJ, Greenough WT (1997) Abnormal dendritic spines in fragile X knockout mice: maturation and pruning deficits. Proceedings of the National Academy of Sciences of the United States of America 94:5401-4

Corbin F, Bouillon M, Fortin A, Morin S, Rousseau F, Khandjian EW (1997) The fragile X mental retardation protein is associated with poly(A)+ mRNA in actively translating polyribosomes. Human Molecular Genetics 6:1465-72

Coverley D, Laskey RA (1994) Regulation of eukaryotic DNA replication. Annual Review of Biochemistry 63:745-76

Cronister A, Hagerman RJ, Wittenberger M, Amiri K (1991a) Mental impairment in cytogenetically positive fragile X females. American Journal of Medical Genetics 38:503-4

Cronister A, Schreiner R, Wittenberger M, Amiri K, Harris K, Hagerman RJ (1991b) Heterozygous fragile X female: historical, physical, cognitive, and cytogenetic features. American Journal of Medical Genetics 38:269-74

Darlow JM, Leach DR (1998) Secondary structures in d(CGG) and d(CCG) repeat tracts. Journal of Molecular Biology 275:3-16

De Boulle K, Verkerk AJ, Reyniers E, Vits L, Hendrickx J, Van Roy B, Van den Bos F, et al (1993) A point mutation in the FMR-1 gene associated with fragile X mental retardation. Nature Genetics 3:31-5

de Graaff E, Willemsen R, Zhong N, de Die-Smulders CE, Brown WT, Freling G, Oostra B (1995) Instability of the CGG repeat and expression of the FMR1 protein in a male fragile X patient with a lung tumor. American Journal of Human Genetics 57:609-18

de Vries BB, Halley DJ, Oostra BA, Niermeijer MF (1998) The fragile X syndrome. Journal of Medical Genetics 35:579-89

de Vries BB, Jansen CC, Duits AA, Verheij C, Willemsen R, van Hemel JO, van den Ouweland AM, et al (1996) Variable FMR1 gene methylation of large

expansions leads to variable phenotype in three males from one fragile X family. Journal of Medical Genetics 33:1007-10

de Vries BB, van den Ouweland AM, Mohkamsing S, Duivenvoorden HJ, Mol E, Gelsema K, van Rijn M, et al (1997) Screening and diagnosis for the fragile X syndrome among the mentally retarded: an epidemiological and psychological survey. Collaborative Fragile X Study Group. American Journal of Human Genetics 61:660-7

Devys D, Biancalana V, Rousseau F, Boue J, Mandel JL, Oberle I (1992) Analysis of full fragile X mutations in fetal tissues and monozygotic twins indicate that abnormal methylation and somatic heterogeneity are established early in development. American Journal of Medical Genetics 43:208-16

Devys D, Lutz Y, Rouyer N, Bellocq JP, Mandel JL (1993) The FMR-1 protein is cytoplasmic, most abundant in neurons and appears normal in carriers of a fragile X premutation. Nature Genetics 4:335-40

Dietrich A, Kioschis P, Monaco AP, Gross B, Korn B, Williams SV, Sheer D, et al (1991) Molecular cloning and analysis of the fragile X region in man. Nucleic Acids Research 19:2567-72

Drouin R, Angers M, Dallaire N, Rose TM, Khandjian W, Rousseau F (1997) Structural and functional characterization of the human FMR1 promoter reveals similarities with the hnRNP-A2 promoter region. Human Molecular Genetics 6:2051-60

Eberhart DE, Malter HE, Feng Y, Warren ST (1996) The fragile X mental retardation protein is a ribonucleoprotein containing both nuclear localization and nuclear export signals. Human Molecular Genetics 5:1083-91

Eberhart DE, Warren ST (1996) Nuclease sensitivity of permeabilized cells confirms altered chromatin formation at the fragile X locus. Somatic Cell & Molecular Genetics 22:435-41

Eichler EE, Holden JJ, Popovich BW, Reiss AL, Snow K, Thibodeau SN, Richards CS, et al (1994) Length of uninterrupted CGG repeats determines instability in the FMR1 gene. Nature Genetics 8:88-94

Eichler EE, Kunst CB, Lugenbeel KA, Ryder OA, Davison D, Warren ST, Nelson DL (1995) Evolution of the cryptic FMR1 CGG repeat. Nature Genetics 11:301-8

Eichler EE, Richards S, Gibbs RA, Nelson DL (1993) Fine structure of the human FMR1 gene [published erratum appears in Human Molecular Genetics 1994 Apr;3(4):684-5]. Human Molecular Genetics 2:1147-53

Escalante JA, Grunspun H, Frota-Pessoa O (1971) Severe sex-linked mental retardation. Journal Genetica Human 19:137-140

Feng Y, Absher D, Eberhart DE, Brown V, Malter HE, Warren ST (1997a) FMRP associates with polyribosomes as an mRNP, and the I304N mutation of severe fragile X syndrome abolishes this association. Molecular Cell 1:109-18

Feng Y, Gutekunst CA, Eberhart DE, Yi H, Warren ST, Hersch SM (1997b) Fragile X mental retardation protein: nucleocytoplasmic shuttling and association with somatodendritic ribosomes. Journal of Neuroscience 17:1539-47

Feng Y, Zhang F, Lokey LK, Chastain JL, Lakkis L, Eberhart D, Warren ST (1995) Translational suppression by trinucleotide repeat expansion at FMR1. Science 268:731-4

Fisch GS, Shapiro LR, Simensen R, Schwartz CE, Fryns JP, Borghgraef M, Curfs LM, et al (1992) Longitudinal changes in IQ among fragile X males: clinical evidence of more than one mutation? American Journal of Medical Genetics 43:28-34

Fournier RE (1981) A general high-efficiency procedure for production of microcell hybrids. Proceedings of the National Academy of Sciences of the United States of America 78:6349-53

Freudenreich CH, Kantrow SM, Zakian VA (1998) Expansion and length-dependent fragility of CTG repeats in yeast. Science 279:853-6

Freudenreich CH, Stavenhagen JB, Zakian VA (1997) Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. Molecular & Cellular Biology 17:2090-8

Fu YH, Kuhl DP, Pizzuti A, Pieretti M, Sutcliffe JS, Richards S, Verkerk AJ, et al (1991) Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. Cell 67:1047-58

Gacy AM, Goellner G, Juranic N, Macura S, McMurray CT (1995) Trinucleotide repeats that expand in human disease form hairpin structures in vitro. Cell 81:533-40

Gacy AM, McMurray CT (1998) Influence of hairpins on template reannealing at trinucleotide repeat duplexes: a model for slipped DNA. Biochemistry 37:9426-34

Glaser D, Wohrle D, Salat U, Vogel W, Steinbach P (1999) Mitotic behavior of expanded CGG repeats studied on cultured cells: further evidence for methylation-mediated triplet repeat stability in fragile X syndrome. American Journal of Medical Genetics 84:226-8

Glover TW (1981) FUdR induction of the X chromosome fragile site: evidence for the mechanism of folic acid and thymidine inhibition. American Journal of Human Genetics 33:234-42

Godde JS, Kass SU, Hirst MC, Wolffe AP (1996) Nucleosome assembly on methylated CGG triplet repeats in the fragile X mental retardation gene 1 promoter. Journal of Biological Chemistry 271:24325-8

Goldberg YP, Kalchman MA, Metzler M, Nasir J, Zeisler J, Graham R, Koide HB, et al (1996) Absence of disease phenotype and intergenerational stability of the CAG repeat in transgenic mice expressing the human Huntington disease transcript. Human Molecular Genetics 5:177-85

Goldman MA, Holmquist GP, Gray MC, Caston LA, Nag A (1984) Replication timing of genes and middle repetitive sequences. Science 224:686-92

Goodfellow PN, Davies KE, Ropers HH (1985) Report of the Committee on the Genetic Constitution of the X and Y Chromosomes. Cytogenetics & Cell Genetics 40:296-352

Gordenin DA, Kunkel TA, Resnick MA (1997) Repeat expansion--all in a flap? [news]. Nature Genetics 16:116-8

Gourdon G, Radvanyi F, Lia AS, Duros C, Blanche M, Abitbol M, Junien C, et al (1997) Moderate intergenerational and somatic instability of a 55-CTG repeat in transgenic mice. Nature Genetics 15:190-2

Gustavson KH, Blomquist HK, Holmgren G (1986) Prevalence of the fragile-X syndrome in mentally retarded boys in a Swedish county. American Journal of Medical Genetics 23:581-7

Hagerman RJ (1996) Physical and Behavioral Phenotype. In: Hagerman RJ, Cronister A (eds) Fragile X Syndrome: Diagnosis, Treatment, and Research. The John Hopkins University Press, Baltimore, pp 3-87

Hagerman RJ, Hull CE, Safanda JF, Carpenter I, Staley LW, O'Connor RA, Seydel C, et al (1994) High functioning fragile X males: demonstration of an unmethylated fully expanded FMR-1 mutation associated with protein expression. American Journal of Medical Genetics 51:298-308

Hagerman RJ, Jackson C, Amiri K, Silverman AC, O'Connor R, Sobesky W (1992) Girls with fragile X syndrome: physical and neurocognitive status and outcome. Pediatrics 89:395-400

Hammond LS, Macias MM, Tarleton JC, Shashidhar Pai G (1997) Fragile X syndrome and deletions in FMR1: new case and review of the literature. American Journal of Medical Genetics 72:430-4

Hansen RS, Canfield TK, Fjeld AD, Mumm S, Laird CD, Gartler SM (1997) A variable domain of delayed replication in FRAXA fragile X chromosomes: X inactivation-like spread of late replication. Proceedings of the National Academy of Sciences of the United States of America 94:4587-92

Hansen RS, Canfield TK, Lamb MM, Gartler SM, Laird CD (1993) Association of fragile X syndrome with delayed replication of the FMR1 gene. Cell 73:1403-9

Hansen RS, Gartler SM, Scott CR, Chen SH, Laird CD (1992) Methylation analysis of CGG sites in the CpG island of the human FMR1 gene. Human Molecular Genetics 1:571-8

Hatton KS, Dhar V, Brown EH, Iqbal MA, Stuart S, Didamo VT, Schildkraut CL (1988) Replication program of active and inactive multigene families in mammalian cells. Molecular & Cellular Biology 8:2149-58

Heitz D, Rousseau F, Devys D, Saccone S, Abderrahim H, Le Paslier D, Cohen D, et al (1991) Isolation of sequences that span the fragile X and identification of a fragile X-related CpG island. Science 251:1236-9

Helderman-van den Enden AT, Maaswinkel-Mooij PD, Hoogendoorn E, Willemsen R, Maat-Kievit JA, Losekoot M, Oostra BA (1999) Monozygotic twin brothers with the fragile X syndrome: different CGG repeats and different mental capacities. Journal of Medical Genetics 36:253-7

Hellmann-Blumberg U, Hintz MF, Gatewood JM, Schmid CW (1993) Developmental differences in methylation of human Alu repeats. Molecular & Cellular Biology 13:4523-30

Hinds HL, Ashley CT, Sutcliffe JS, Nelson DL, Warren ST, Housman DE, Schalling M (1993) Tissue specific expression of FMR-1 provides evidence for a functional role in fragile X syndrome. Nature Genetics 3:36-43

Hinton VJ, Brown WT, Wisniewski K, Rudelli RD (1991) Analysis of neocortex in three males with the fragile X syndrome. American Journal of Medical Genetics 41:289-94

Hirst MC, Grewal PK, Davies KE (1994) Precursor arrays for triplet repeat expansion at the fragile X locus. Human Molecular Genetics 3:1553-60

Hirst MC, Rack K, Nakahori Y, Roche A, Bell MV, Flynn G, Christadoulou Z, et al (1991a) A YAC contig across the fragile X site defines the region of fragility. Nucleic Acids Research 19:3283-8

Hirst MC, Roche A, Flint TJ, MacKinnon RN, Bassett JH, Nakahori Y, Watson JE, et al (1991b) Linear order of new and established DNA markers around the fragile site at Xq27.3. Genomics 10:243-9

Hirst MC, White PJ (1998) Cloned human FMR1 trinucleotide repeats exhibit a length- and orientation-dependent instability suggestive of in vivo lagging strand secondary structure. Nucleic Acids Research 26:2353-8

Hornstra IK, Nelson DL, Warren ST, Yang TP (1993) High resolution methylation analysis of the FMR1 gene trinucleotide repeat region in fragile X syndrome. Human Molecular Genetics 2:1659-65

Hwu WL, Lee YM, Lee SC, Wang TR (1993) In vitro DNA methylation inhibits FMR-1 promoter. Biochemical & Biophysical Research Communications 193:324-9

Igarashi S, Takiyama Y, Cancel G, Rogaeva EA, Sasaki H, Wakisaka A, Zhou YX, et al (1996) Intergenerational instability of the CAG repeat of the gene for Machado-Joseph disease (MJD1) is affected by the genotype of the normal chromosome: implications for the molecular mechanisms of the instability of the CAG repeat. Human Molecular Genetics 5:923-32

Ikeda H, Yamaguchi M, Sugai S, Aze Y, Narumiya S, Kakizuka A (1996) Expanded polyglutamine in the Machado-Joseph disease protein induces cell death in vitro and in vivo. Nature Genetics 13:196-202

Jacky P (1996) Cytogenetics. In: Hagerman RJ, Cronister A (eds) Fragile X Syndrome: Diagnosis, Treatment, and Research. The John Hopkins University Press, Baltimore, pp 114-164

Jaenisch R (1997) DNA methylation and imprinting: why bother? [see comments]. Trends in Genetics 13:323-9

Jaworski A, Rosche WA, Gellibolian R, Kang S, Shimizu M, Bowater RP, Sinden RR, et al (1995) Mismatch repair in Escherichia coli enhances instability of (CTG)n triplet repeats from human hereditary diseases. Proceedings of the National Academy of Sciences of the United States of America 92:11019-23

Jeddeloh JA, Bender J, Richards EJ (1998) The DNA methylation locus DDM1 is required for maintenance of gene silencing in Arabidopsis. Genes & Development 12:1714-25

Jeddeloh JA, Stokes TL, Richards EJ (1999) Maintenance of genomic methylation requires a SWI2/SNF2-like protein. Nature Genetics 22:94-7

Johannisson R, Rehder H, Wendt V, Schwinger E (1987) Spermatogenesis in two patients with the fragile X syndrome. I. Histology: light and electron microscopy. Human Genetics 76:141-7

Kahkonen M, Alitalo T, Airaksinen E, Matilainen R, Launiala K, Autio S, Leisti J (1987) Prevalence of the fragile X syndrome in four birth cohorts of children of school age. Human Genetics 77:85-7

Kakutani T, Jeddeloh JA, Flowers SK, Munakata K, Richards EJ (1996) Developmental abnormalities and epimutations associated with DNA hypomethylation mutations. Proceedings of the National Academy of Sciences of the United States of America 93:12406-11

Kang S, Jaworski A, Ohshima K, Wells RD (1995) Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in E. coli. Nature Genetics 10:213-8

Kaytor MD, Burright EN, Duvick LA, Zoghbi HY, Orr HT (1997) Increased trinucleotide repeat instability with advanced maternal age. Human Molecular Genetics 6:2135-9

Khandjian EW, Fortin A, Thibodeau A, Tremblay S, Cote F, Devys D, Mandel JL, et al (1995) A heterogeneous set of FMR1 proteins is widely distributed in mouse tissues and is modulated in cell culture. Human Molecular Genetics 4:783-9

Kochanek S, Renz D, Doerfler W (1993) DNA methylation in the Alu sequences of diploid and haploid primary human cells. EMBO Journal 12:1141-51

Kremer EJ, Pritchard M, Lynch M, Yu S, Holman K, Baker E, Warren ST, et al (1991a) Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. Science 252:1711-4

Kremer EJ, Yu S, Pritchard M, Nagaraja R, Heitz D, Lynch M, Baker E, et al (1991b) Isolation of a human DNA sequence which spans the fragile X. American Journal of Human Genetics 49:656-61

Kruyer H, Mila M, Glover G, Carbonell P, Ballesta F, Estivill X (1994) Fragile X syndrome and the (CGG)n mutation: two families with discordant MZ twins. American Journal of Human Genetics 54:437-42

Kunst CB, Warren ST (1994) Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. Cell 77:853-61

La Spada AR (1997) Trinucleotide repeat instability: genetic features and molecular mechanisms. Brain Pathology 7:943-63

La Spada AR, Peterson KR, Meadows SA, McClain ME, Jeng G, Chmelar RS, Haugen HA, et al (1998) Androgen receptor YAC transgenic mice carrying CAG 45 alleles show trinucleotide repeat instability. Human Molecular Genetics 7:959-67

Lachiewicz AM, Spiridigliozzi GA, McConkie-Rosell A, Burgess D, Feng Y, Warren ST, Tarleton J (1996) A fragile X male with a broad smear on Southern blot analysis representing 100-500 CGG repeats and no methylation at the EagI site of the FMR-1 gene. American Journal of Medical Genetics 64:278-82

Lavedan C, Grabczyk E, Usdin K, Nussbaum RL (1998) Long Uninterrupted Cgg Repeats Within the First Exon Of the Human Fmr1 Gene Are Not Intrinsically Unstable In Transgenic Mice. Genomics 50:229-240

Lavedan CN, Garrett L, Nussbaum RL (1997) Trinucleotide repeats (CGG)22TGG(CGG)43TGG(CGG)21 from the fragile X gene remain stable in transgenic mice. Human Genetics 100:407-14

Leeflang EP, Tavare S, Marjoram P, Neal CO, Srinidhi J, MacFarlane H, MacDonald ME, et al (1999) Analysis of germline mutation spectra at the Huntington's disease locus supports a mitotic mutation mechanism. Human Molecular Genetics 8:173-83

Lei H, Oh SP, Okano M, Juttermann R, Goss KA, Jaenisch R, Li E (1996) De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development 122:3195-205

Li E, Bestor TH, Jaenisch R (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 69:915-26

Liu WM, Schmid CW (1993) Proposed roles for DNA methylation in Alu transcriptional repression and mutational inactivation. Nucleic Acids Research 21:1351-9

Loesch DZ, Huggins R, Hay DA, Gedeon AK, Mulley JC, Sutherland GR (1993) Genotype-phenotype relationships in fragile X syndrome: a family study. American Journal of Human Genetics 53:1064-73

Lubs HA (1969) A marker X chromosome. American Journal of Human Genetics 21:231-44

Lugenbeel KA, Peier AM, Carson NL, Chudley AE, Nelson DL (1995) Intragenic loss of function mutations demonstrate the primary role of FMR1 in fragile X syndrome. Nature Genetics 10:483-5

Luo S, Robinson JC, Reiss AL, Migeon BR (1993) DNA methylation of the fragile X locus in somatic and germ cells during fetal development: relevance to the fragile X syndrome and X inactivation. Somatic Cell & Molecular Genetics 19:393-404

Macpherson J, Murray A, Webb J, Jacobs P (1999) Fragile X syndrome: of POF and premutations. Journal of Medical Genetics 36:171-2

Malter HE, Iber JC, Willemsen R, de Graaff E, Tarleton JC, Leisti J, Warren ST, et al (1997) Characterization of the full fragile X syndrome mutation in fetal gametes. Nature Genetics 15:165-9

Mangiarini L, Sathasivam K, Mahal A, Mott R, Seller M, Bates GP (1997) Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. Nature Genetics 15:197-200

Mannermaa A, Pulkkinen L, Kajanoja E, Ryynanen M, Saarikoski S (1996) Deletion in the FMR1 gene in a fragile-X male. American Journal of Medical Genetics 64:293-5

Marozzi A, Vegetti W, Manfredini E, Tibiletti MG, Testa G, Crosignani PG, Ginelli E, et al (2000) Association between idiopathic premature ovarian failure and fragile X premutation. Human Reproduction 15:197-202

Martin JP, Bell J (1943) A pedigree of mental defect showing sex-linkage. Journal of Neurological Psychiatry 6:154-157

Maurer DJ, O'Callaghan BL, Livingston DM (1996) Orientation dependence of trinucleotide CAG repeat instability in Saccharomyces cerevisiae. Molecular & Cellular Biology 16:6617-22

McConkie-Rosell A, Lachiewicz AM, Spiridigliozzi GA, Tarleton J, Schoenwald S, Phelan MC, Goonewardena P, et al (1993) Evidence that methylation of the FMR-I locus is responsible for variable phenotypic expression of the fragile X syndrome. American Journal of Human Genetics 53:800-9

McMurray CT (1995) Mechanisms of DNA expansion. Chromosoma 104:2-13

Meijer H, de Graaff E, Merckx DM, Jongbloed RJ, de Die-Smulders CE, Engelen JJ, Fryns JP, et al (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. Human Molecular Genetics 3:615-20

Merenstein SA, Shyu V, Sobesky WE, Staley L, Berry-Kravis E, Nelson DL, Lugenbeel KA, et al (1994) Fragile X syndrome in a normal IQ male with learning and emotional problems. Journal of the American Academy of Child & Adolescent Psychiatry 33:1316-21

Merenstein SA, Sobesky WE, Taylor AK, Riddle JE, Tran HX, Hagerman RJ (1996) Molecular-clinical correlations in males with an expanded FMR1 mutation. American Journal of Medical Genetics 64:388-94

Mila M, Castellvi-Bel S, Sanchez A, Lazaro C, Villa M, Estivill X (1996) Mosaicism for the fragile X syndrome full mutation and deletions within the CGG repeat of the FMR1 gene. Journal of Medical Genetics 33:338-40

Mingroni-Netto RC, Haddad LA, Vianna-Morgante AM (1996) The number of CGG repeats of the FMR1 locus in premutated and fully mutated heterozygotes and their offspring: implications for the origin of mosaicism. American Journal of Medical Genetics 64:270-3

Miniou P, Bourc'his D, Molina Gomes D, Jeanpierre M, Viegas-Pequignot E (1997a) Undermethylation of Alu sequences in ICF syndrome: molecular and in situ analysis. Cytogenetics & Cell Genetics 77:308-13

Miniou P, Jeanpierre M, Bourc'his D, Coutinho Barbosa AC, Blanquet V, Viegas-Pequignot E (1997b) alpha-satellite DNA methylation in normal individuals and in ICF patients: heterogeneous methylation of constitutive heterochromatin in adult and fetal tissues. Human Genetics 99:738-45

Monckton DG, Coolbaugh MI, Ashizawa KT, Siciliano MJ, Caskey CT (1997) Hypermutable myotonic dystrophy CTG repeats in transgenic mice [see comments]. Nature Genetics 15:193-6

Morton JE, Bundey S, Webb TP, MacDonald F, Rindl PM, Bullock S (1997) Fragile X syndrome is less common than previously estimated. Journal of Medical Genetics 34:1-5

Moutou C, Vincent MC, Biancalana V, Mandel JL (1997) Transition from premutation to full mutation in fragile X syndrome is likely to be prezygotic. Human Molecular Genetics 6:971-9

Mummaneni P, Bishop PL, Turker MS (1993) A cis-acting element accounts for a conserved methylation pattern upstream of the mouse adenine phosphoribosyltransferase gene. Journal of Biological Chemistry 268:552-8

Mummaneni P, Walker KA, Bishop PL, Turker MS (1995) Epigenetic gene inactivation induced by a cis-acting methylation center. Journal of Biological Chemistry 270:788-92

Ng HH, Bird A (1999) DNA methylation and chromatin modification. Current Opinion in Genetics and Development 9:158-63

Nolin SL, Glicksman A, Houck GE, Jr., Brown WT, Dobkin CS (1994) Mosaicism in fragile X affected males. American Journal of Medical Genetics 51:509-12

Nolin SL, Lewis FA, 3rd, Ye LL, Houck GE, Jr., Glicksman AE, Limprasert P, Li SY, et al (1996) Familial transmission of the FMR1 CGG repeat. American Journal of Human Genetics 59:1252-61

Oberle I, Heilig R, Moisan JP, Kloepfer C, Mattei GM, Mattei JF, Boue J, et al (1986) Genetic analysis of the fragile-X mental retardation syndrome with two flanking

polymorphic DNA markers. Proceedings of the National Academy of Sciences of the United States of America 83:1016-20

Oberle I, Rousseau F, Heitz D, Kretz C, Devys D, Hanauer A, Boue J, et al (1991) Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. Science 252:1097-102

Parniewski P, Bacolla A, Jaworski A, Wells RD (1999) Nucleotide excision repair affects the stability of long transcribed (CTG*CAG) tracts in an orientation-dependent manner in Escherichia coli. Nucleic Acids Research 27:616-23

Pembrey ME, Winter RM, Davies KE (1985) A premutation that generates a defect at crossing over explains the inheritance of fragile X mental retardation. American Journal of Medical Genetics 21:709-17

Pieretti M, Zhang FP, Fu YH, Warren ST, Oostra BA, Caskey CT, Nelson DL (1991) Absence of expression of the FMR-1 gene in fragile X syndrome. Cell 66:817-22

Price DK, Zhang F, Ashley CT, Jr., Warren ST (1996) The chicken FMR1 gene is highly conserved with a CCT 5'-untranslated repeat and encodes an RNA-binding protein. Genomics 31:3-12

Quan F, Grompe M, Jakobs P, Popovich BW (1995) Spontaneous deletion in the FMR1 gene in a patient with fragile X syndrome and cherubism. Human Molecular Genetics 4:1681-4

Razin A (1998) CpG methylation, chromatin structure and gene silencing-a three-way connection. EMBO Journal 17:4905-8

Razin A, Kafri T (1994) DNA methylation from embryo to adult. Progress in Nucleic Acid Research & Molecular Biology 48:53-81

Reddy PS, Housman DE (1997) The complex pathology of trinucleotide repeats. Current Opinion in Cell Biology 9:364-72

Reiss AL, Abrams MT, Greenlaw R, Freund L, Denckla MB (1995) Neurodevelopmental effects of the FMR-1 full mutation in humans. Nature Medicine 1:159-67

Reiss AL, Aylward E, Freund LS, Joshi PK, Bryan RN (1991a) Neuroanatomy of fragile X syndrome: the posterior fossa. Annals of Neurology 29:26-32

Reiss AL, Freund L, Tseng JE, Joshi PK (1991b) Neuroanatomy in fragile X females: the posterior fossa. American Journal of Human Genetics 49:279-88

Reiss AL, Lee J, Freund L (1994) Neuroanatomy of fragile X syndrome: the temporal lobe. Neurology 44:1317-24

Reyniers E, Martin JJ, Cras P, Van Marck E, Handig I, Jorens HZ, Oostra BA, et al (1999) Postmortem examination of two fragile X brothers with an FMR1 full mutation. American Journal of Medical Genetics 84:245-9

Reyniers E, Vits L, De Boulle K, Van Roy B, Van Velzen D, de Graaff E, Verkerk AJ, et al (1993) The full mutation in the FMR-1 gene of male fragile X patients is absent in their sperm. Nature Genetics 4:143-6

Richards BW, Sylvester PE, Brooker C (1981) Fragile X-linked mental retardation: the Martin-Bell syndrome. Journal of Mental Deficiency Research 25:253-6

Richards RI, Sutherland GR (1994) Simple repeat DNA is not replicated simply [news]. Nature Genetics 6:114-6

Rousseau F, Heitz D, Biancalana V, Blumenfeld S, Kretz C, Boue J, Tommerup N, et al (1991a) Direct diagnosis by DNA analysis of the fragile X syndrome of mental retardation. New England Journal of Medicine 325:1673-81

Rousseau F, Heitz D, Tarleton J, MacPherson J, Malmgren H, Dahl N, Barnicoat A, et al (1994a) A multicenter study on genotype-phenotype correlations in the fragile X syndrome, using direct diagnosis with probe StB12.3: the first 2,253 cases. American Journal of Human Genetics 55:225-37

Rousseau F, Robb LJ, Rouillard P, Der Kaloustian VM (1994b) No mental retardation in a man with 40% abnormal methylation at the FMR-1 locus and transmission of sperm cell mutations as premutations. Human Molecular Genetics 3:927-30

Rousseau F, Rouillard P, Morel ML, Khandjian EW, Morgan K (1995) Prevalence of carriers of premutation-size alleles of the FMRI gene--and implications for the population genetics of the fragile X syndrome. American Journal of Human Genetics 57:1006-18

Rousseau F, Vincent A, Rivella S, Heitz D, Triboli C, Maestrini E, Warren ST, et al (1991b) Four chromosomal breakpoints and four new probes mark out a 10-cM region encompassing the fragile-X locus (FRAXA). American Journal of Human Genetics 48:108-16

Rudelli RD, Brown WT, Wisniewski K, Jenkins EC, Laure-Kamionowska M, Connell F, Wisniewski HM (1985) Adult fragile X syndrome. Clinico-neuropathologic findings. Acta Neuropathologica 67:289-95

Samadashwily GM, Raca G, Mirkin SM (1997) Trinucleotide repeats affect DNA replication in vivo. Nature Genetics 17:298-304

Sarkar PS, Chang HC, Boudi FB, Reddy S (1998) CTG repeats show bimodal amplification in E. coli. Cell 95:531-40

Sato T, Oyake M, Nakamura K, Nakao K, Fukusima Y, Onodera O, Igarashi S, et al (1999) Transgenic mice harboring a full-length human mutant DRPLA gene exhibit age-dependent intergenerational and somatic instabilities of CAG repeats comparable with those in DRPLA patients. Human Molecular Genetics 8:99-106

Schmid CW (1991) Human Alu subfamilies and their methylation revealed by blot hybridization. Nucleic Acids Research 19:5613-7

Schmid CW (1996) Alu: structure, origin, evolution, significance and function of one-tenth of human DNA. Progress in Nucleic Acid Research & Molecular Biology 53:283-319

Schmucker B, Ballhausen WG, Pfeiffer RA (1996) Mosaicism of a microdeletion of 486 bp involving the CGG repeat of the FMR1 gene due to misalignment of GTT tandem repeats at chi-like elements flanking both breakpoints and a full mutation. Human Genetics 98:409-14

Schuffenhauer S, Bartsch O, Stumm M, Buchholz T, Petropoulou T, Kraft S, Belohradsky B, et al (1995) DNA, FISH and complementation studies in ICF

syndrome: DNA hypomethylation of repetitive and single copy loci and evidence for a trans acting factor. Human Genetics 96:562-71

Schweitzer JK, Livingston DM (1997) Destabilization of CAG trinucleotide repeat tracts by mismatch repair mutations in yeast. Human Molecular Genetics 6:349-55

Schweitzer JK, Livingston DM (1998) Expansions of CAG repeat tracts are frequent in a yeast mutant defective in Okazaki fragment maturation. Human Molecular Genetics 7:69-74

Schwemmle S (1999) In vivo footprinting analysis of the FMR1 gene: proposals concerning gene regulation in high-functioning males. American Journal of Medical Genetics 84:266-7

Schwemmle S, de Graaff E, Deissler H, Glaser D, Wohrle D, Kennerknecht I, Just W, et al (1997) Characterization of FMR1 promoter elements by in vivo-footprinting analysis. American Journal of Human Genetics 60:1354-62

Sherman S (1996) Epidemiology. In: Hagerman RJ, Cronister A (eds) Fragile X Syndrome: Diagnosis, Treatment, and Research. The John Hopkins University Press, Baltimore, pp 165-192

Sherman SL (1995) The High Prevalence of Fragile X Premutation Carrier Females: Is This Frequency Unique to the French Canadian Population? American Journal of Human Genetics 57:991-993

Sherman SL, Jacobs PA, Morton NE, Froster-Iskenius U, Howard-Peebles PN, Nielsen KB, Partington MW, et al (1985) Further segregation analysis of the fragile X

syndrome with special reference to transmitting males. Human Genetics 69:289-99

Sherman SL, Morton NE, Jacobs PA, Turner G (1984) The marker (X) syndrome: a cytogenetic and genetic analysis. Annals of Human Genetics 48:21-37

Shimizu M, Gellibolian R, Oostra BA, Wells RD (1996) Cloning, characterization and properties of plasmids containing CGG triplet repeats from the FMR-1 gene. Journal of Molecular Biology 258:614-26

Sinden RR (1999) Biological implications of the DNA structures associated with disease-causing triplet repeats. American Journal of Human Genetics 64:346-53

Siomi H, Choi M, Siomi MC, Nussbaum RL, Dreyfuss G (1994) Essential role for KH domains in RNA binding: impaired RNA binding by a mutation in the KH domain of FMR1 that causes fragile X syndrome. Cell 77:33-9

Siomi H, Siomi MC, Nussbaum RL, Dreyfuss G (1993) The protein product of the fragile X gene, FMR1, has characteristics of an RNA-binding protein. Cell 74:291-8

Siomi MC, Zhang Y, Siomi H, Dreyfuss G (1996) Specific sequences in the fragile X syndrome protein FMR1 and the FXR proteins mediate their binding to 60S ribosomal subunits and the interactions among them. Molecular & Cellular Biology 16:3825-32

Sittler A, Devys D, Weber C, Mandel JL (1996) Alternative splicing of exon 14 determines nuclear or cytoplasmic localisation of fmr1 protein isoforms. Human Molecular Genetics 5:95-102

Slegtenhorst-Eegdeman KE, de Rooij DG, Verhoef-Post M, van de Kant HJ, Bakker CE, Oostra BA, Grootegoed JA, et al (1998) Macroorchidism in FMR1 knockout mice is caused by increased Sertoli cell proliferation during testicular development. Endocrinology 139:156-62

Smeets HJ, Smits AP, Verheij CE, Theelen JP, Willemsen R, van de Burgt I, Hoogeveen AT, et al (1995) Normal phenotype in two brothers with a full FMR1 mutation. Human Molecular Genetics 4:2103-8

Smit AF (1996) The origin of interspersed repeats in the human genome. Current Opinion in Genetics & Development 6:743-8

Snow K, Doud LK, Hagerman R, Pergolizzi RG, Erster SH, Thibodeau SN (1993) Analysis of a CGG sequence at the FMR-1 locus in fragile X families and in the general population. American Journal of Human Genetics 53:1217-28

Snow K, Tester DJ, Kruckeberg KE, Schaid DJ, Thibodeau SN (1994) Sequence analysis of the fragile X trinucleotide repeat: implications for the origin of the fragile X mutation. Human Molecular Genetics 3:1543-51

Spiro C, Pelletier R, Rolfsmeier ML, Dixon MJ, Lahue RS, Gupta G, Park MS, et al (1999) Inhibition of FEN-1 processing by DNA secondary structure at trinucleotide repeats. Molecular Cell 4:1079-85

Spring NV, Monckton DG, Krahe R, Siciliano MJ (1998) Genomic instability in somatic cell hybrids containing the human myotonic dystrophy (CTG)n repeat. American Journal of Human Genetics 63 (supplement):A387

Stewart CL, Stuhlmann H, Jahner D, Jaenisch R (1982) De novo methylation, expression, and infectivity of retroviral genomes introduced into embryonal carcinoma cells. Proc Natl Acad Sci U S A 79:4098-102

Stewart FJ, Raleigh EA (1998) Dependence of McrBC cleavage on distance between recognition elements. Biological Chemistry 379:611-6

Stoger R, Kajimura TM, Brown WT, Laird CD (1997) Epigenetic variation illustrated by DNA methylation patterns of the fragile-X gene FMR1. Human Molecular Genetics 6:1791-801

Strauss WM (1998) Unit 6.3: Using DNA Fragments as Probes. In: Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K et al (eds) Current Protocols in Molecular Biology. Vol. 1. John Wiley & Sons, Inc., pp 6.3.1-6.3.6

Sutcliffe JS, Nelson DL, Zhang F, Pieretti M, Caskey CT, Saxe D, Warren ST (1992) DNA methylation represses FMR-1 transcription in fragile X syndrome. Human Molecular Genetics 1:397-400

Sutherland GR (1977) Fragile sites on human chromosomes: demonstration of their dependence on the type of tissue culture medium. Science 197:265-6

Sutherland GR (1979) Heritable fragile sites on human chromosomes I. Factors affecting expression in lymphocyte culture. American Journal of Human Genetics 31:125-35

Sutherland GR, Gedeon A, Kornman L, Donnelly A, Byard RW, Mulley JC, Kremer E, et al (1991) Prenatal diagnosis of fragile X syndrome by direct detection of the unstable DNA sequence. New England Journal of Medicine 325:1720-2

Sutherland GR, Hecht F, Mulley JC, Glover TW, Hecht BK (1985) Biochemistry of Fragile Site Expression. In: Sutherland GR (ed) Fragile Sites on Human Chromosomes. Oxford University Press, Oxford, pp 80-94

Suthers GK, Hyland VJ, Callen DF, Oberle I, Rocchi M, Thomas NS, Morris CP, et al (1990) Physical mapping of new DNA probes near the fragile X mutation (FRAXA) by using a panel of cell lines. American Journal of Human Genetics 47:187-95

Suthers GK, Mulley JC, Voelckel MA, Dahl N, Vaisanen ML, Steinbach P, Glass IA, et al (1991) Linkage homogeneity near the fragile X locus in normal and fragile X families. Genomics 10:576-82

Syrrou M, Georgiou I, Patsalis PC, Bouba I, Adonakis G, Pagoulatos GN (1999) Fragile X premutations and (TA)n estrogen receptor polymorphism in women with ovarian dysfunction. American Journal of Medical Genetics 84:306-8

Tassone F, Hagerman RJ, Gane LW, Taylor AK (1999) Strong similarities of the FMR1 mutation in multiple tissues: postmortem studies of a male with a full mutation and a male carrier of a premutation. American Journal of Medical Genetics 84:240-4

Taylor AK, Tassone F, Dyer PN, Hersch SM, Harris JB, Greenough WT, Hagerman RJ (1999) Tissue heterogeneity of the FMR1 mutation in a high-functioning male with fragile X syndrome. American Journal of Medical Genetics 84:233-9

Taylor JH (1960) Asynchronous duplication of chromosomes in cultured cells of Chinese hamster. Journal of Biophysical and Biochemical Cytology 7:455-463

Telenius H, Kremer HP, Theilmann J, Andrew SE, Almqvist E, Anvret M, Greenberg C, et al (1993) Molecular analysis of juvenile Huntington disease: the major influence on (CAG)n repeat length is the sex of the affected parent. Human Molecular Genetics 2:1535-40

Timchenko LT, Caskey CT (1999) Triplet repeat disorders: discussion of molecular mechanisms. Cellular and Molecular Life Sciences 55:1432-47

Tommerup N, Poulsen H, Brondum-Nielsen K (1981) 5-Fluoro-2'-deoxyuridine induction of the fragile site on Xq28 associated with X linked mental retardation. Journal of Medical Genetics 18:374-6

Torchia BS, Call LM, Migeon BR (1994) DNA replication analysis of FMR1, XIST, and factor 8C loci by FISH shows nontranscribed X-linked genes replicate late. American Journal of Human Genetics 55:96-104

Turker MS, Mummaneni P, Bishop PL (1991) Region- and cell type-specific de novo DNA methylation in cultured mammalian cells. Somatic Cell & Molecular Genetics 17:151-7

Turker MS, Stambrook PJ, Tischfield JA, Smith AC, Martin GM (1989a) Allelic variation linked to adenine phosphoribosyltransferase locus in mouse teratocarcinoma cell line and feral-derived mouse strains. Somatic Cell & Molecular Genetics 15:159-66

Turker MS, Swisshelm K, Smith AC, Martin GM (1989b) A partial methylation profile for a CpG site is stably maintained in mammalian tissues and cultured cell lines. Journal of Biological Chemistry 264:11632-6

Turner G, Robinson H, Laing S, Purvis-Smith S (1986) Preventive screening for the fragile X syndrome. New England Journal of Medicine 315:607-9

Turner G, Till R, Daniel A (1978) Marker X chromosomes, mental retardation and macro-orchidism. New England Journal of Medicine 299:1472

Turner G, Webb T, Wake S, Robinson H (1996) Prevalence of fragile X syndrome. American Journal of Medical Genetics 64:196-7

Uzielli ML, Guarducci S, Lapi E, Cecconi A, Ricci U, Ricotti G, Biondi C, et al (1999) Premature ovarian failure (POF) and fragile X premutation females: from POF to to fragile X carrier identification, from fragile X carrier diagnosis to POF association data. American Journal of Medical Genetics 84:300-3

Verheij C, Bakker CE, de Graaff E, Keulemans J, Willemsen R, Verkerk AJ, Galjaard H, et al (1993) Characterization and localization of the FMR-1 gene product associated with fragile X syndrome. Nature 363:722-4

Verheij C, de Graaff E, Bakker CE, Willemsen R, Willems PJ, Meijer N, Galjaard H, et al (1995) Characterization of FMR1 proteins isolated from different tissues. Human Molecular Genetics 4:895-901

Verkerk AJ, de Graaff E, De Boulle K, Eichler EE, Konecki DS, Reyniers E, Manca A, et al (1993) Alternative splicing in the fragile X gene FMR1. Human Molecular Genetics 2:399-404

Verkerk AJ, Pieretti M, Sutcliffe JS, Fu Y-H, Kuhl DP, Pizzuti A, Reiner O, et al (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident

with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell 65:905-14

Vianna-Morgante AM (1999) Twinning and premature ovarian failure in premutation fragile X carriers. American Journal of Medical Genetics 83:326

Vincent A, Heitz D, Petit C, Kretz C, Oberle I, Mandel JL (1991) Abnormal pattern detected in fragile-X patients by pulsed-field gel electrophoresis. Nature 349:624-6

Vongs A, Kakutani T, Martienssen RA, Richards EJ (1993) Arabidopsis thaliana DNA methylation mutants. Science 260:1926-8

Wang YC, Lin ML, Lin SJ, Li YC, Li SY (1997) Novel point mutation within intron 10 of FMR-1 gene causing fragile X syndrome. Human Mutation 10:393-9

Wang YH, Griffith J (1996) Methylation of expanded CCG triplet repeat DNA from fragile X syndrome patients enhances nucleosome exclusion. Journal of Biological Chemistry 271:22937-40

Wang Z, Taylor AK, Bridge JA (1996) FMR1 fully expanded mutation with minimal methylation in a high functioning fragile X male. Journal of Medical Genetics 33:376-8

Warren ST, Davidson RL (1984) Expression of fragile X chromosome in human-rodent somatic cell hybrids. Somatic Cell & Molecular Genetics 10:409-13

Warren ST, Knight SJ, Peters JF, Stayton CL, Consalez GG, Zhang FP (1990) Isolation of the human chromosomal band Xq28 within somatic cell hybrids by fragile X

site breakage. Proceedings of the National Academy of Sciences of the United States of America 87:3856-60

Webb TP, Bundey S, Thake A, Todd J (1986) The frequency of the fragile X chromosome among schoolchildren in Coventry. Journal of Medical Genetics 23:396-9

Wheeler VC, Auerbach W, White JK, Srinidhi J, Auerbach A, Ryan A, Duyao MP, et al (1999) Length-dependent gametic CAG repeat instability in the Huntington's disease knock-in mouse. Human Molecular Genetics 8:115-22

White PJ, Borts RH, Hirst MC (1999) Stability of the human fragile X (CGG)(n) triplet repeat array in Saccharomyces cerevisiae deficient in aspects of DNA metabolism. Molecular & Cellular Biology 19:5675-84

Wierdl M, Dominska M, Petes TD (1997) Microsatellite instability in yeast: dependence on the length of the microsatellite. Genetics 146:769-79

Willemsen R, Bontekoe C, Tamanini F, Galjaard H, Hoogeveen A, Oostra B (1996) Association of FMRP with ribosomal precursor particles in the nucleolus. Biochemical & Biophysical Research Communications 225:27-33

Wohrle D, Hennig I, Vogel W, Steinbach P (1993) Mitotic stability of fragile X mutations in differentiated cells indicates early post-conceptional trinucleotide repeat expansion. Nature Genetics 4:140-2

Wohrle D, Hirst MC, Kennerknecht I, Davies KE, Steinbach P (1992) Genotype mosaicism in fragile X fetal tissues. Human Genetics 89:114-6

Wohrle D, Kennerknecht I, Wolf M, Enders H, Schwemmle S, Steinbach P (1995) Heterogeneity of DM kinase repeat expansion in different fetal tissues and further expansion during cell proliferation in vitro: evidence for a casual involvement of methyl-directed DNA mismatch repair in triplet repeat stability. Human Molecular Genetics 4:1147-53

Wohrle D, Salat U, Glaser D, Mucke J, Meisel-Stosiek M, Schindler D, Vogel W, et al (1998) Unusual mutations in high functioning fragile X males: apparent instability of expanded unmethylated CGG repeats. Journal of Medical Genetics 35:103-11

Wohrle D, Schwemmle S, Steinbach P (1996) DNA methylation and triplet repeat stability: new proposals addressing actual questions on the CGG repeat of fragile X syndrome. American Journal of Medical Genetics 64:266-7

Yates PA, Burman RW, Mummaneni P, Krussel S, Turker MS (1999) Tandem B1 elements located in a mouse methylation center provide a target for de novo DNA methylation. Journal of Biological Chemistry 274:36357-61

Yoder JA, Walsh CP, Bestor TH (1997) Cytosine methylation and the ecology of intragenomic parasites. Trends in Genetics 13:335-40

Yu S, Pritchard M, Kremer E, Lynch M, Nancarrow J, Baker E, Holman K, et al (1991) Fragile X genotype characterized by an unstable region of DNA. Science 252:1179-81

Zhong N, Yang W, Dobkin C, Brown WT (1995) Fragile X gene instability: anchoring AGGs and linked microsatellites. American Journal of Human Genetics 57:351-61