

**UC Davis**

**UC Davis Electronic Theses and Dissertations**

**Title**

Genetic factors associated with CGG repeat instability in FMR1 Premutation Carriers with Various Degrees of Mosaicism

**Permalink**

<https://escholarship.org/uc/item/8nd623tt>

**Author**

Hwang, Ye Hyun

**Publication Date**

2022

Peer reviewed|Thesis/dissertation

**Genetic Factors Associated with CGG Repeat Instability in *FMRI* Premutation Carriers  
with Various Degrees of Mosaicism.**

**By**

**YE HYUN HWANG  
THESIS**

**Submitted in partial satisfaction of the requirements for the degree of**

**MASTER OF SCIENCE**

**in**

**Integrative Genetics and Genomics**

**in the**

**OFFICE OF GRADUATE STUDIES**

**of the**

**UNIVERSITY OF CALIFORNIA**

**DAVIS**

**Approved:**

---

**Flora Tassone, Chair**

---

**Susan Rivera**

---

**Jacqueline Barlow**

**Committee in Charge**

**2022**

# Table of Contents

I. Abstract .....	iii
II. Objective .....	iv
III. Introduction .....	1
Fragile X Syndrome and <i>FMR1</i> Associated Disorders .....	1
<i>FMR1</i> Alleles and Gene Expressions .....	2
<i>FMR1</i> Gene Structure .....	3
FMRP function .....	5
CGG Repeat Instability .....	6
Mosaicism of the <i>FMR1</i> gene .....	8
Mosaicism in <i>FMR1</i> Premutation Carriers .....	8
Potential Role of DNA Repair Proteins in allelic instability .....	10
IV. Methods .....	12
Isolation of Genomic DNA .....	12
Measurement of DNA Concentration .....	12
CGG Allele Sizing by PCR Analysis .....	13
Capillary electrophoresis (CE) .....	14
Characterization of AGG Interruptions .....	14
HpaII Digestion .....	15
Southern Blot Analysis .....	16
SNP Genotyping Analysis .....	17
V. Published Manuscripts .....	18
i. Manuscript: <i>Repeat Instability in the Fragile X-Related Disorders: Lessons from a Mouse Model</i> .....	19
ii. Manuscript: <i>Both cis and trans-acting genetic factors drive somatic instability in female carriers of the FMR1 premutation</i> .....	34
VI. <i>Factors Associated with CGG Repeat Instability and Mosaicism in FMR1 Premutation Males</i> .....	46
VI. Discussion .....	59
VII. Conclusion .....	60
VIII. Bibliography .....	62

## I. Abstract

Allele instability in trinucleotide repeat disorders has been associated with many different physical and psychological conditions, including inherited forms of autism and neurodegenerative disorders. This form of instability is observed in Fragile X Syndrome, a trinucleotide disorder, in which a CGG repeat located in the 5'UTR of the fragile X messenger ribonucleoprotein 1 (*FMR1*) gene is greater than 200 CGG repeats. This leads to methylation of the gene, transcriptional silencing and consequent absence or reduction of the encoded protein, FMRP. In *FMR1* premutation (PM) carriers, who have an allele containing between 55 – 200 CGG repeat length, a CGG repeat expansion during transmission to the offspring, leads to Fragile X Syndrome. Although allele instability has been observed mainly in full mutation alleles (>200 CGG repeats), it has been observed throughout the CGG range, and leading to somatic mosaicism.

However, it is unclear what molecular factors are associated with the risk of *FMR1* CGG repeat expansion and if instability correlates with clinical conditions throughout the lifespan of individuals carrying an expanded allele. Furthermore, it is unknown if these factors confer difference in the degrees of risk based on the individual's biological sex, or age, or if instability or changes may occur over time within individuals.

In this study, we investigated 426 PM female and 454 PM male carriers. Within these two cohorts, we observed that CGG repeat size correlates with *FMR1* mRNA levels, and with the number of AGG interruptions, confirming previous reports. Interestingly, a lower number of AGG interruptions and higher CGG repeat size increases the risk of expansion from mother to offspring during transmission, relevant to the new observation here reported. When studying CGG instability over time, we found that eight PM females (n=24) underwent allele expansion as they aged, with three individuals displaying an increase of three or greater repeats in CGG repeat number. Likewise, in the PM male group (n=50), 19 individuals showed an increased CGG repeat number over time, with two undergoing CGG allele size decrease. We also found that the expanded unmethylated regions, significantly correlated with CGG repeat size and AGG interruptions in both females and males.

Thus, CGG repeat size and AGG interruptions are significantly correlated with somatic instability, regardless of gender, although differences in allele expansion patterns between PM males and females exist. Trans molecular factors such as DNA repair-associated genes are also capable of affecting risk of expansion. Indeed, our preliminary observations found a significant correlation between allele in two genes, *MSH3* and *FAN1*, both of which play a role in DNA repair.

These findings carry the implications that molecular measures could be used to determine individuals with higher likelihood of allele instability, which may influence the phenotypic expression. Overall, this study can help future diagnostics in determining which PM individuals, both males and females, are more likely to experience allele expansion and be at risk of developing Fragile X associated conditions.

## II. Objective

The objective of this research was to investigate and identify molecular factors that potentially contribute to *FMR1* CGG repeat instability and somatic mosaicism (presence of more than one allele in an individual's somatic cell population) in PM carriers. In this study I investigated allele instability in both PM males and females, and characterized the molecular measures at the *FMR1* locus, including CGG repeats, AGG interruptions, methylation (methylation status and AR in females), *FMR1* mRNA expression levels and a measure of allele instability. In addition, I investigated which genetic factors may affect allele expansion of individuals with differing degrees of instability and investigated how mosaic patterns changes over time in individuals displaying mosaicism.

Ultimately, the goal would be to determine if PM instability contributes to the premutation related phenotypes observed in female and male *FMR1* PM carriers.

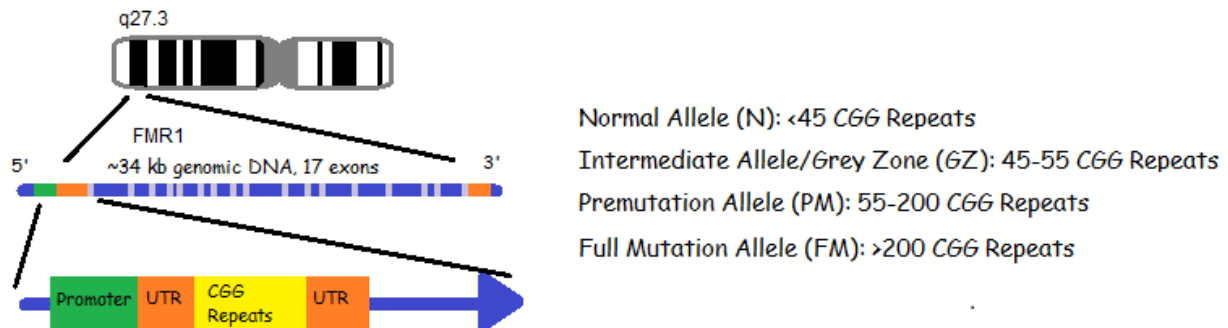
We accomplished these objectives through the following **Specific Aims**:

- 1.) To characterize the instability profile and features of the *FMR1* allele in both male and female premutation carriers
- 2.) To investigate the degree of instability in PM carriers and correlate it to the CGG repeat length and to the presence of AGG interruptions.
- 3.) To investigate how mosaicism patterns and degree of expansion derived from allelic instability differ between PM males and females.
- 4.) To investigate the role of Activation Ratio (AR) in allele instability in PM females
- 5.) To investigate the correlations between allele instability and trans molecular factors to determine if they constitute potential risk factors of PM allele instability.
- 6.) To characterize the allele expansion profile in the same individual over a period of years to determine if and how mosaicism patterns may change over time.

### III. Introduction

#### Fragile X Syndrome and Associated Disorders

Fragile X Syndrome (FXS) is the most common form of intellectual disabilities and the most common inherited cause of autism disorder. This condition is preceded by an expansion of a CGG trinucleotide repeats in the 5' untranslated (UTR) region of the *FMR1* (*Fragile X messenger ribonucleoprotein 1*) gene, which maps in the Xq27.3 region of the X chromosome [Zhou et al, 2016] (Figure 1). FXS is caused by an expansion of greater than 200 CGG repeats in the promoter region, causing the *FMR1* gene to be epigenetically silenced [Hansen et al, 1992]. The loss of function, in turn, leads to loss of FMRP (Fragile X messenger ribonucleoprotein), a protein involved in pathways important to neurological development [Bagni et al, 2012]. This loss of FMRP leads to developmental delays that can be observed in early childhood years [Bailey et al, 2001]. Even lowered amounts of FMRP in expanded unsilenced *FMR1* can result in lowered IQ scores in these individuals [Roth et al, 2021].



**Figure 1: *FMR1* structure**

The region in which the *FMR1* gene is located within the X chromosome is shown, along with the basic structure of the 5' UTR region of *FMR1*. The length of the CGG repeats determine the categorization of the *FMR1* allele, in which full mutation and premutation alleles lead to FXS and *FMR1* associated disorders

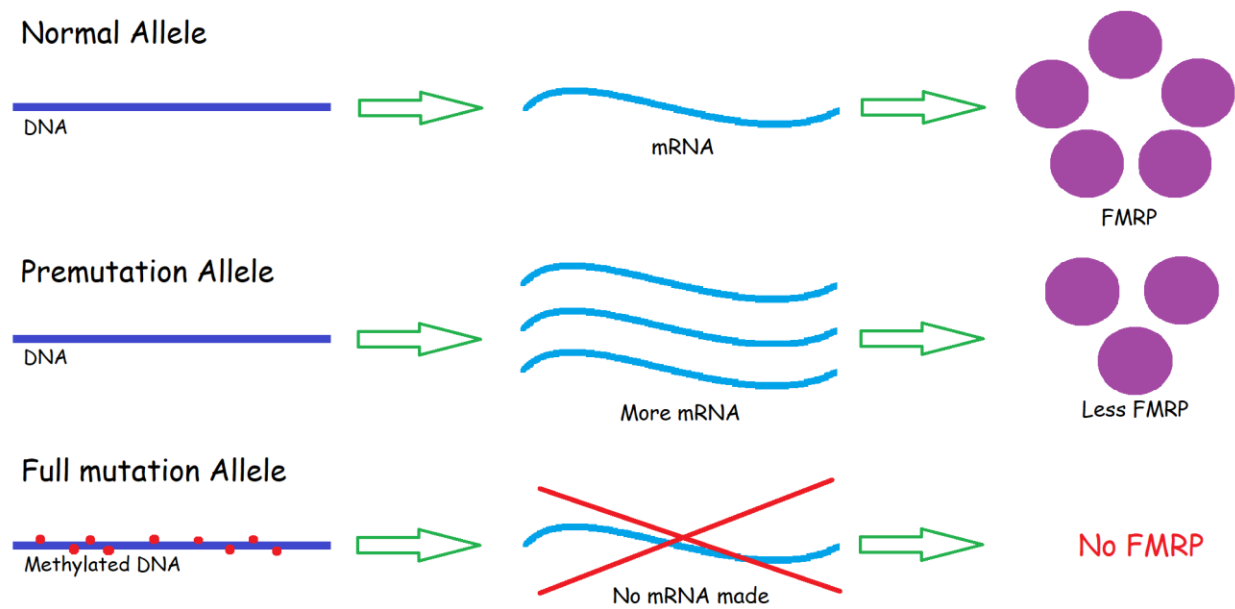
Other disorders related to the *FMR1* gene, so named *FMR1* associated disorders or conditions, include FXPOI (Fragile X-associated primary ovarian insufficiency), a condition of lowered fertility in females, FXTAS (Fragile X-associated tremor/ataxia syndrome), and FXAND (Fragile X-associated neuropsychiatric disorders). The FXPOI prevalence is ~20% in female PM carriers [Allen et al, 2020] compared to ~1% in the general population; FXTAS is observed in about 40% in PM males [Jacquemont et al, 2004] and 6-15% in female carriers [Allen et al, 2020], and FXAND is present in ~50% of individuals with the PM allele [Hagerman et al, 2018].

All these conditions result from the presence of a premutation allele (CGG repeats between 50-200 repeats), when the *FMR1* gene is not silenced and resulting in an increased expression of *FMR1* mRNA and decreased FMRP expression due to a deficit in translation efficiency of the expanded CGG containing RNA. The characteristic features observed in FXTAS, including tremors and ataxia are likely due to *FMR1* RNA “toxicity” from the mRNA overload present in PM carriers [Hagerman and Hagerman, 2015; Tassone et al, 2000].

### FMR1 Alleles and gene expression

According to the American Society of Medical Genetics, *FMR1* can be categorized based on CGG repeat numbers as normal (<45 CGG repeats), grey zone/intermediate (45-54 CGG repeats, GZ), premutation (55-200 CGG repeats, PM), and full mutation (>200 CGG repeats, FM) [Maddalena et al, 2001]. Among these alleles, from a meta-analysis study, the frequency of the FM allele is around 1:6,000 in males and 1:8300 in females, and the prevalence of the PM allele is 1:730 in males and 1:164 in females [Hunter et al, 2014].

The *FMR1* gene encodes for FMRP (Fragile X mental retardation protein), an RNA binding protein which plays a key role in cognitive development, learning and memory. The CGG repeat length affects the expression and production of both *FMR1* mRNA and FMRP, as illustrated in Figure 2.



**Figure 2: *FMR1* gene expression depends on CGG repeat number**

This figure shows how the progression from transcription and translation differs based on the size of the *FMR1* allele. Compared to the normal allele, the PM allele produces more mRNA but less FMRP. FM alleles are methylated (red dots on the DNA), preventing the mRNA from being transcribed, which leads to no FMRP production and ultimately to FXS.

In PM cases, *FMR1* mRNA levels are elevated from two to tenfold compared to normal levels, and yet less FMRP is produced, possibly because expanded *FMR1* mRNA may result in impaired translation, due to deficit in translation efficiency [Primerano et al, 2002]. PM allele also leads to clinical involvements, including FXPOI and FXTAS through mRNA overexpression and toxicity as stated before. This mechanism contrasts with the fully methylated *FMR1* FM gene unable to produce mRNA and FMRP, which lack thereof, leads to FXS phenotype [Sutcliffe et al, 1992].

This gene is expressed through many different mRNA isoforms (alternative splicing variants), and the expression level of each type of isoform varies based on CGG repeat number.

Indeed, the *FMR1* gene, when being transcribed, undergoes alternative splicing, which alters its function in the cell. The splicing occurs mostly with inclusions/exclusions of Exons 12 and 14, and alternative splice acceptors in Exons 15 and 17 [Pretto et al, 2015]. The most abundant types of *FMR1* isoforms in both normal and PM individuals are isoforms that were missing Exon 12, and these isoforms are displayed evenly across different tissues of the same individual. *Iso7* and *Iso17* are the most expressed isoforms in both groups, though the PM individuals expressed these at a higher level than their normal allele counterparts [Pretto et al, 2015].

Furthermore, PM individuals display higher amounts of truncated isoforms as well, particularly *isoform 10* and *isoform 10b*, which both lack the C-terminal site [Tseng et al, 2017]. The lack of a C-terminal site in both *Iso10* and *Iso10b* implies that the PM versions of FMRP may be impaired in its ability to export the needed mRNA to its proper location. Furthermore, *Iso10* and *Iso10b* retain Exon 15 in their sequence, but also contain a frameshift, resulting in the deletion of the RGG box, which in turn can affect the binding to MAP1B (microtubule-associated protein 1B), a gene involved in nervous system development and adult brain physiology [Lu et al, 2004].

The altered transcriptional signature observed in premutation, could play a role in the clinical features of FXS and Fragile X related conditions, such as behavioral problems, developmental abnormalities, and neurodegeneration.

### *FMR1* Gene Structure

The *FMR1* gene consists of 38 kb of genomic DNA and 17 exons total [Eichler et al, 1993]. The origin of replication is located about 300-450 nucleotides upstream of the transcriptional initiation site. The CGG repeat is located within the 5'UTR of the gene, and the *FMR1* promoter and the initiation site for RNA transcription are located immediately upstream of the CGG repeat. The promoter lacks a TATA box but has many binding motifs for AP2 and multiple Spl sites (transcription factor domains), in a region about 1kb upstream of the transcription initiation site. This gene also possesses distinct boundaries of methylated and unmethylated sites that are lost in FXS individuals, with heavily methylated *FMR1* promoter region. [Naumann et al, 2009]

The CGG repeat element can be interspersed with AGG interruptions, which are interruption sequences located within the CGG repeats (Figure 3) and shown to contribute to allele stability of the *FMR1* gene. They are often found in the 5' region of the CGG repeat region and interspersed within the region. AGG interruptions at the 5' end of this region tends to have a periodicity of around 9-11 repeats long [Zhao et al, 2016].



**Figure 3: AGG interruptions within the CGG repeat**

This image depicts AGG interruptions, emphasized in yellow, within the CGG repeat region located in the 5' UTR of *FMR1*. The most common range of AGG interruptions in an individual range between 0 – 2 AGG interruptions in males and 0 – 3 in females depending on the length [Yrigollen et al, 2013].



Nonetheless, by interrupting the CGG repeat, AGGs reduce the risk of CGG repeat expansion, a risk faced particularly during DNA replication due to the CGG repeat region experiencing strand slippage followed by repriming of the slipped region which leads to the incorporation of extra CGG repeats [Zhao et al, 2016]. The number of AGG interruptions also correlates with the CGG repeat length, though this effect is more noticeable when comparing CGG repeat number of individuals with no AGG and those with AGG interruptions, whereas the CGG difference between individuals with 1 vs 2 or more AGG repeats are more subtle [Yrigollen et al, 2011]. Alleles with larger numbers of CGG repeats tend to have fewer AGG interruptions overall [Zhao et al, 2016; Yrigollen et al, 2011].

The continuous chain of CGG repeats in the *FMR1* gene can form secondary structures such as tetraplexes and hairpins, and the transcribed RNA are also known to create hairpin structures, with AGG interruptions creating bulges and loops within the DNA [Yrigollen et al, 2011]. As CGG repeats are known to form hairpin structures and guanines in hairpin structures can be prone to oxidative damage, the resulting CGG hairpin structures could block DNA synthesis and cause replication fork stalling and collapse [Zhao et al, 2016]. Furthermore, the fragility of the single-stranded R-loops formed by the CGG repeats puts them at risk of damage, which can predispose the allele toward risk of triggering expansions and secondary structures, and thus increase risk of repeat instability.

As CGG repeats are prone to expansion when passed from parent to child, particularly through maternal transmission, there is interest in whether the AGG interruptions serve as protection from expansion as the allele is inherited. Normal alleles most often possess 2 AGG interruptions (rarely more than 2, or 1 or none), whereas premutation alleles tend to possess 0-2 alleles with larger premutation alleles having less AGG interruption, but more often none. This observation can be explained by the loss of AGG interruptions, a relatively common event which involves a mechanism that leads to the loss of AGG in the 3' end while keeping the over CGG repeat element [Eichler et al, 1995]. This potentially causes a premutation allele inherited from a parent to expand into a full mutation allele in the offspring [Yrigollen et al, 2012].

Different cases investigated the CGG expansion mechanism. One study [Yrigollen et al, 2012] observed a correlation between the mother's AGG profile and the presence of a FM expansion (>200 CGG repeats) in their offspring. The results indicated that the total CGG length highly correlated with intergenerational FM allele transmission, whereas high AGG interruptions correlated with fewer FM allele transmission from mother to offspring. Alleles with the longest uninterrupted CGG repeat region had the highest risk for repeat instability and alleles with long CGG repeats inherited from mothers with higher maternal age were at a higher risk of expansion [Yrigollen et al, 2014], suggesting that both repeat length and maternal age affect intergenerational CGG repeat allele stability.

Furthermore, locations of CGG expansion sites are often in the 3' end of the CGG repeats, particularly with longer CGG repeats [Yrigollen et al, 2013], which would be expected as the loss of AGG interruptions mainly occurs in the 3' end. Without an AGG buffer to lower

expansion risk, the allele undergoes a slow progression of transitioning from a several hundred repeats when transmitted from parent to child. Observations of this phenomenon can be seen in families with PM carriers and no histories of FXS [Yrigollen et al, 2013].

This inference of AGG interruptions as protective against expansion is supported by another observation which found that populations with high interspersed AGG interruptions' number had lower estimated expansion rates than those with less interspersed AGG interruptions or less AGG interruptions, which supports the idea that the presence and the number of AGG interruptions affect risk of CGG allele expansion [Nolin et al, 2013]. Furthermore, the study found that AGG interruptions had a positive effect in the prevention of instability and changes in CGG repeat number, which reinforces the belief that loss of AGG interruptions, though a rare event, increases risk of *FMR1* expansion in offspring, and possibly in somatic mutations as well.

### FMRP function

As previously mentioned, the FMRP protein is known to play a key role in neurodevelopment. Structural features in FMRP contribute to its function in the cell. The full length FMRP, referred to as *Iso1*, contains two KH domains, an RGG box, and phosphorylation and methylation sites within Exon 15 [Pretto et al, 2015]. FMRP also contains a functional nuclear localization signal in the N-terminus and a nuclear export signal in the C-terminus [Kaytor et al, 2001], which aid in its function of RNA transport between nucleus and cytoplasm.

Through studies involving these interactions, FXS symptoms have been concluded to be correlated with FMRP deficit. As the *FMR1* gene is responsible for the production of FMRP, *FMR1* methylation and expansion can both cause FMRP deficit, leading to Fragile X-associated conditions. Males have a more severe phenotype compared to their female counterparts because they only possess one X chromosome. On the other hand, females undergo X-inactivation, a process in which one of the X-chromosomes is switched off, which can lead to a wide range of severity of FXS related conditions due to differences in AR depending on the individual [Rajaratnam et al, 2017]. This is speculated to lead to gender bias in overall *FMR1* expression and severity of symptoms. Thus, differences in gender lead to visible differences in phenotypic expression, as females tend to have a second chromosome carrying a normal allele that mitigates the negative effects of the mutated expanded one.

As stated before, *FMR1* alleles have a wide range of CGG repeat numbers, from less than 45 to over 200 repeats, and the number of repeats affects the levels of FMRP in an individual. PM alleles do not cause FXS, but still result in other FX related disorders, and these conditions can also be affected by gender bias. Because of the lower number of CGG repeats compared to FM alleles, PM alleles are mostly unmethylated and thus, still encoding for FMRP, though not to the extent of the wild type, leading to FMRP deficiency. FM alleles on the other hand are methylated, and the production of FMRP is prevented in the individual, leading to the FX-related phenotype. FMRP demonstrates RNA binding activity, and its N-terminal also binds poly(G). It can bind to many RNAs in the brain, including its own, suggesting that FMRP functions as a repressor of mRNA translation of protein in the brain [Kaytor et al, 2001]. This implication is further supported by a study which found that FMRP levels are related to the level of total

development, but not the rate of development over time [Nimchinsky et al, 2001], implying that FMRP is involved in some but not all mechanisms of brain development.

FMRP is distributed mostly in the cytoplasm, where it associates with polyribosome transcribing proteins, targeting therefore mRNAs transcribed in ribosomes [Kaytor et al, 2001]. Studied cases of FXS due to point mutations showed that defective FMRP could not dimerize properly, and is unable to regulate mRNA, suggesting that FMRP binds a subset of mRNAs localized in the dendrite, and thus critical for proper synaptic function [Kaytor et al, 2001]. Specifically, dendrite spine density and development were shown to be correlated with *FMR1* genotypes in one study with knock-out mice [Nimchinsky et al, 2001], as the spine length was increased in *FMR1* knock-out mice compared to controls during their 1st week. Afterwards, dendrite spines were elevated selectively in second-order dendrites, which could be a possible trend that continued into adulthood and contributed to FXS related symptoms in adults. However, when observed in vitro studies, FMRP itself does not cause abnormal phenotype in dendrites, which makes it likely that it coordinates with other activities, most likely with synaptic activity [Nimchinsky et al, 2001]. Regardless, the lack of FMRP due to methylated *FMR1* induce a similar effect as the point mutation version of *FMR1*, which leads to FXS. Indeed, abnormal spine morphology has also been documented in humans [Taylor et al, 1999; Irwin et al, 2000; Hinton et al, 1991].

Finally, FMRP expression can be present and be variable even in cases of FXS, which affects development function and growth in males. This is in part due to the fact that *FMR1* FM alleles, in many cases, are not completely fully methylated, leading to mosaicism (for methylation or size) which degree associates with the severity of the clinical symptoms [Pretto et al. 2014 (2); Meng et al, 2021]

### CGG Repeat Instability

The general definition of instability is the condition in which a gene does present with a spectrum of larger or smaller alleles than that tend to change in size. In the case of the *FMR1* gene, this is due to expansions and contractions that occur in the trinucleotide repeat region, which results in instability of repeat size [Zhao et al, 2019]. Because of the repetitive nature of the CGG trinucleotide repeat region, some unusual structures can be formed including stable hairpins, G-quadruplexes, Z-DNA, and persistent R-loops, which are RNA-DNA hybrids that form during transcription [Usdin et al, 2015]. The formation of these structures predisposes this area to DNA damage [Zhao et al, 2016]. To counter this damage, DNA mismatch repair (MMR) proteins are utilized, but these proteins can potentially increase the repeat number while repairing detected mismatches from the aforementioned damage [Zhao et al, 2016].

When the DNA is repaired, the CGG repeat is prone to strand slippage, which can result in extra repeats incorporated into the complementary strand. As a result, the DNA mismatch repair procedure affect *FMR1* CGG instability, and expansions result from the errors generated while DNA damage repairing is attempted [Zhao et al, 2016]. Strand slippage is an event that can occur during both DNA repair and replication [Usdin et al, 2015], though in this case, this work will focus on the DNA repair aspect, as this mechanism more closely relates with somatic mosaicism, which will be elaborated later in this study.

In the process of DNA repair, MutS $\beta$ , a complex involved in DNA mismatch repair (MMR), generates both expansions and contractions of CGG repeats in a mouse model of the FX associated disorders, which makes it likely that a similar mechanism can cause expansions of CGG in humans. The complex is composed of proteins MSH2 and MSH3, which together are involved in MMR. It has been shown that MSH2 is essential to expansion in mouse models [Lokanga et al, 2014]. Because MutS $\beta$  is required in at least 98% of expansions known, SNP mutations in either component are linked to an increased risk of expansion, including the one observed in FX related conditions. [Zhao et al, 2016] Repeats also impede replication, which can be another issue when DNA is repaired [Zhou et al, 2016].

MutS $\alpha$ , another complex, is also involved in expansion and hypothesized to protect against large contractions, which is when repeats are decreased in a gene, and this can be observed when observed with an absence of MutS $\beta$  [Zhao et al, 2016]. However, it is not essential to expansion generation. On the other hand, Pol $\beta$ , another DNA polymerase, is not involved in MMR but is involved in base excision repair (BER). BER is the major pathway to oxidative damage repair, a type of degradation that increases expansion risk. Its dysfunction is correlated with increased expansion risk, since BER is used to repair DNA oxidative damage [Zhao et al, 2016]. This has been supported by an FX mouse model study focused on Pol $\beta$  involvement in BER [Zhao et al, 2016].

Expansions and contractions are both involved with DNA repair of the trinucleotide repeat region, but expansion is much more common than contraction in the *FMR1* gene [Zhao et al, 2016]. Thus, repeat instability in this study mostly focuses on and refers to somatic allelic instability due to the CGG expansion, which is likely more relevant to increased risk of FX-related disorders, as stated in previous sections. Expansion types fall into two categories: replication based, and recombination based; FX related disorders display characteristics of both and linked with CGG repeat length.

Finally, an interesting feature about CGG instability is that the tissue source of the DNA sample affects the amount of observed instability, a phenomenon also noted in other expansion-based diseases, in which some cell types are more prone to expansion in the same gene than other types [Usdin et al, 2015]. As methylation is known to affect allele instability [Zhou et al, 2016], any difference between tissue methylation implies that there are potentially different mechanisms, tissue specific, that affect allele instability between different tissues of the same individual. This speculation is supported by observing samples from blood and samples from fibroblast cells, which showed a recognizable difference in methylation and instability. Blood tended to show more CGG instability than fibroblasts, implying that brain CGG instability cannot be determined for certainty by looking at other tissues. This variation of CGG instability observed within a tissue and within different tissues of any given individual leads to the phenomenon of intra-tissue and inter-tissue somatic mosaicism, respectively.

### Mosaicism of the *FMR1* gene

Mosaicism refers to a population of cells in an individual which display more than one allele size. Mosaicism, in the case of *FMR1*, is caused by the CGG repeat instability, which leads to two types of mosaicism: somatic and inherited (also known as germline). Somatic mosaicism is when somatic cells display two or more genotypes in an individual due to expansion within the individual. It can be either intra-tissue and/or inter-tissue mosaicism. This contrasts with inherited mosaicism, or germline mosaicism, which is inherited from a parent during transmission in the next generation, and due either to gamete or to parental mosaicism. [Freed et al, 2014]

Of the two basic types, somatic and germline, this paper focuses on somatic mosaicism, which is more of interest in this study because this type of mosaicism is observed in somatic cells in tissues of FX PM carriers, including brain and blood [Lokanga et al, 2013; Hwang and Hayward et al, 2022]. The expression of somatic mosaicism is not inherited through the germline, as it occurs during development after the zygote is formed, thus is not present in the gametes that formed the zygote.

Somatic mutations do not necessarily appear in a short amount of time, as it can be a gradual process. In the case of *FMR1*, somatic mosaicism occurs early on, during embryonic development [Helderman-van den Enden et al, 1999], during which large active alleles appear to be at a selective disadvantage, as they are correlated with cell death. Hypermethylation of *FMR1* in FM individuals have been associated with somatic stability in FM CGG repeats [Wöhrle et al, 1993]. As a result, the methylated FM alleles are favored in the individual, though remaining active FM alleles can contribute to mosaicism in the individual.

However, somatic mosaicism can be “inherited” by cells that carry the lineage of the original mutant mosaic cell, and this can, in turn, affect the phenotype of the individual [Freed et al, 2014]. Mosaic variation occurring during development can result in diseases [Freed et al, 2014] in many different tissues. This phenomenon is also linked to many medical conditions, such as cancer, neurodegenerative disease, and neurodevelopmental disorders as in the case of FXS and associated disorders.

Furthermore, the different degree of allelic instability displayed by different tissues, correspond to the various patterns of somatic mosaicism we observed in our study. It has been shown that somatic stability correlates with DNA methylation in fibroblasts but not in somatic cell hybrids [Freed et al, 2014]. Unmethylated alleles are also at greater risk for mitotic instability, which in turn affect mosaicism expression in different tissues [Pretto et al, 2014].

### Mosaicism in *FMR1* Premutation Carriers

Prior studies have mainly focused on mosaicism detected in FM individuals and compared mosaic individuals with non-mosaic individuals which have shown that FM mosaics had different amounts of FXS penetrance and resulting in a less severe phenotype, including higher adaptive skills development than the non-mosaic individuals [Zhao et al, 2016; Pretto et al. 2014]. This is likely due to the presence of unmethylated alleles, transcribed and thus translated, resulting in

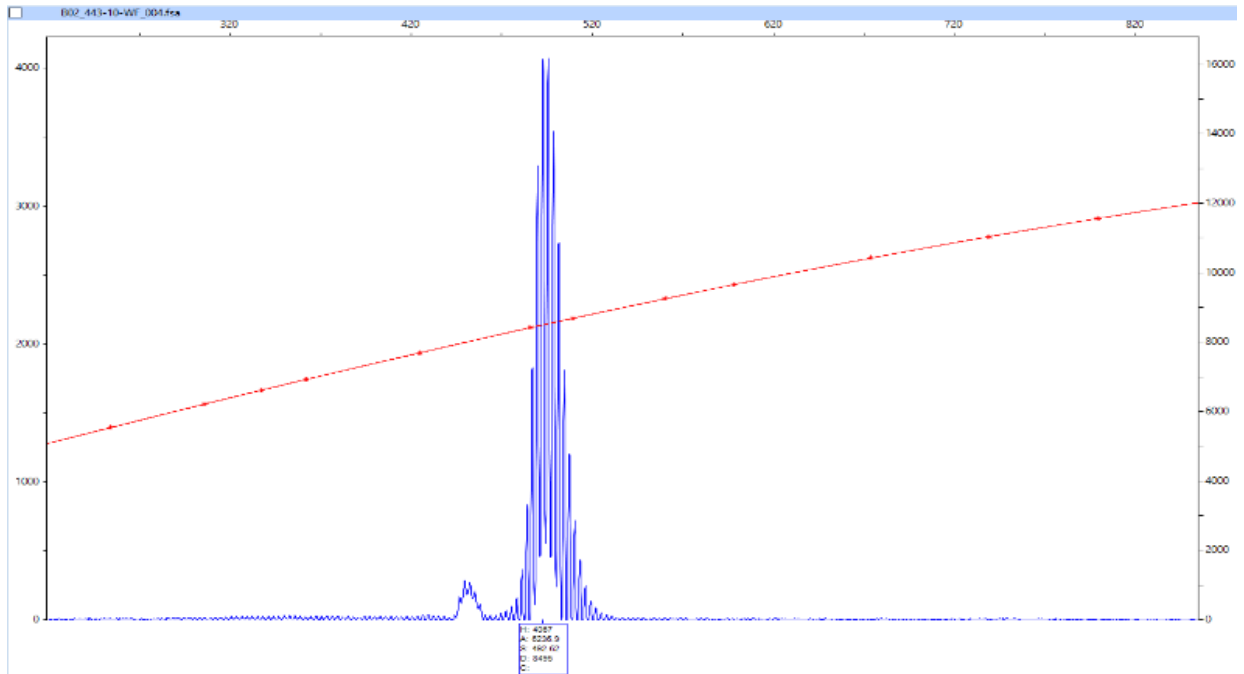
the expression of variable amount of FMRP. Phenotypic differences were also observed in FM mosaic females compared to FM non-mosaic females, as demonstrated in a study which showed less severe phenotypes [Mailick et al, 2018]. Thus, a distinction is observable between mosaic and non-mosaic groups, though what causes these groups to differ in mosaicism, and ultimately in the clinical outcome, remains unclear.

Somatic mosaicism in PM individuals, which is the focus of this work, is of particular interest because not all PM individuals display mosaicism, and we are the first to document and demonstrate this phenomenon. It is unclear what is the direct biological effect or role on the individuals' phenotypes who possess it.

There are observed gender differences in mosaicism of *FMR1* instability, due to differing numbers of X chromosomes. When studying females, *FMR1* PM instability is of interest because female cells carry two copies of the *FMR1* gene, yet there are females who display variable amounts of expansion in their PM allele. A normal female displays two peaks, correspondent to the two *FMR1* alleles, one for each X chromosome, in the PCR profile visualized by capillary electrophoresis (CE). Sometimes, females can display more than two peaks in their CE graph, which is attributed to allele instability. This instability is also what causes the shoulder effects, a distinct cluster of peaks often located to the right of the main PM allele peak. It has been observed in mice and humans that in the event of X-inactivation in PM females, repeat instability is confined to *FMR1* alleles located on the active X-chromosome [Zhao et al, 2019]. Human females with a higher activation ratio (AR), and thus, with a higher percentage of normal allele on the active X-chromosomes, have been observed to less likely display mosaicism [Zhao et al, 2019].

The number of CGG repeats also affect the type of conditions the PM female individual is likely to be affected from. For instance, individuals with FXTAS are more likely to be females at the higher end of PM CGG repeat range, and those with FXPOI are more likely to be in the middle CGG repeat range. Oddly, in a newer study, female who display PM/FM mosaicism overall have less symptoms than those with PM non-mosaic and PM only mosaic. In contrast, the PM-only mosaic individuals have more severe symptoms compared to the other two groups, which may imply that the PM/FM mosaicism is correlated with a symptom-protective effect, whereas the presence of only PM expansion has a negative effect on the health of female premutation carriers [Mailick et al, 2018].

On the other hand, males only possess a single *FMR1* allele on their only X-chromosome and display a different type of mosaic pattern. Generally, one peak is usually observed in a PCR profile. However, in a mosaic male, two or more alleles would be observed. This mosaicism is caused by CGG repeat instability in the individual. The more common pattern of allelic instability observed in PM males is represented by wide broad CE peaks, demonstrating the presence of a series of alleles, smaller and larger in size from the main one (Figure 4). To note, PM males do not possess FM alleles in their sperm, and do not transmit FM alleles to their offspring [Zhou et al, 2016].



**Figure 4: PM Male Expansion Pattern**

Example of a typical expansion pattern in PM males. Note the broadened alleles to the left and right of the main allele in the center of the alleles.

Observation in a study on somatic mosaicism in PM female and male mice implies that X-inactivation is a large component in the gender differences of *FMR1* mosaicism. Females have twice the number of X chromosomes, in which one is randomly deactivated, mitigating its overall effect on females compared to males. This may be supported by the observation that expansion occurs more extensively in males than females despite similar expanded allele sizes, and that expansion is only observed thus far on the active X chromosome. Furthermore, somatic expansion in PM carriers appears to increase with age, particularly in females, as we reported in this study.

Despite this, X-inactivation does not completely account for gender differences in somatic expansion since PM male mice had noticeably higher somatic instability (about 4.7 times more) than females in certain tissues than what would be the expected number (females having 50% the amount of somatic expansion compared to males). It has also been found that gender differences in somatic expression are not due to gender hormonal differences [Lokanga et al, 2014]. When comparing mice that had reproductive organs removed to those without removed organs, there was no difference in the levels of expansion, though there was a visible sex related difference.

#### Potential Role of DNA Repair Proteins in allelic instability

Allelic variations of genes involved in DNA repair affect DNA repair pathways and are known to be involved with repeat expansion. Loss-of-function variant SNPs (single nucleotide variation) in DNA repair genes have been correlated in trinucleotide repeat-based disorders, including ataxia [Bettencourt, C. et al, 2016] and with the age at onset of Huntington's Disease, a CAG

trinucleotide repeat disorder-based motor disorders, and of Myotonic Dystrophy 1, a CTG trinucleotide repeat disorder [Bettencourt, C. et al, 2016; Flower et al, 2019].

Thus, it is not much of a stretch to hypothesize that a similar relationship between allelic variants in DNA repair gene and CGG expansion in the *FMR1* gene may also exist. One such gene, *MSH3*, is known to modify expansion in HD and other trinucleotide repeat disorders [Moss et al, 2017]. Likewise, a similar *FMR1* CGG expansion reduction effect was observed when *MSH3* was lost in mutant male mice [Zhao et al, 2015], which could hint at a potential relationship between *MSH3* and CGG expansion. It is also known that mutations in other MMR relevant SNP variants lead to decreased incidents of CGG expansion or to the lack of CGG expansion [Zhao et al, 2018; Hwang and Hayward et al, 2022], suggesting that they affect mosaicism and DNA repair in *FMR1* CGG repeats. However, it is unclear whether this also correlates with somatic mosaicism specifically.

We have stated earlier in this paper that the MutSB complex and the BER pathway are involved in CGG expansion. Thus, we hypothesize that SNPs within these genes would affect CGG expansion in *FMR1* and contribute to PM mosaicism in an individual. Although there have been prior studies of some SNPs involved in DNA repair that were found to correlate with other expansion related disorders [Ciosi et al, 2019], the relationship of many of these SNPs with *FMR1* CGG repeat expansion have not yet been investigated. As consequence, any specific SNPs and genes of interest which have been know investigation with CGG repeat size expansion as well.

In this study, we investigate whether any such SNPs contribute to instability of the *FMR1* repeat expansions in PM females, and if so, to what degree any significant SNPs contribute to CGG allele expansion. The findings and results are elaborated further in a later section of this study.



## IV. Methods

Molecular measures included in this study were CGG repeat number, AGG interruptions, methylation status, *FMR1* mRNA expression levels, allelic instability, and genotyping. Molecular measures were obtained from DNA/RNA isolated from whole blood derived from PM participants, both males and females.

### Isolation of genomic DNA

Blood samples were obtained from the UC Davis Medical Center, and most of the samples used in this study were derived from peripheral blood leukocytes. Standard protocol used for DNA extraction was modified from the standard DNA purification from whole blood DNA protocol from Qiagen, specifically the Gentra Puregene Blood Kit.

9ml of RBC Lysis Solution was dispensed into a 15 ml centrifuge tube. 3 ml of whole blood was added into the tube and mixed by inverting 10 times. The tube was then incubated for 10 min at room temperature (15–25°C) and inverted at least once during the incubation. Afterwards, tube was vortexed and centrifuged for 10 min at 2000 x g (4000 rpm) to pellet the white blood cells and the resulting supernatant was discarded through pouring and pipetting, leaving a small amount of residual liquid and the white blood cell pellet. The tube was then vortexed vigorously to resuspend the pellet in the residual liquid to facilitate cell lysis and DNA extraction. 3 ml Cell Lysis Solution was added, and the mixture was pipetted up and down to lyse the cells, then vortexed vigorously for 10 s (seconds) before sitting at room temperature. After overnight incubation at room temperature, 15 µl of RNase A Solution was added, and the solution was mixed by inverting 25 times, then incubated for 30 min at 37°C and incubated for 3 min on ice to quickly cool the sample. 1 ml of Protein Precipitation Solution was added; the tube was vortexed vigorously for 20 s at high speed, then centrifuged for 10 min at 2000 x g (4000 rpm). 3 ml of Isopropanol was poured into a clean 15 ml tube and the supernatant from the previous step was added by pouring carefully into the 15 ml tube. The mixture was inverted gently 50 times until the DNA was visible as threads or a clump. The supernatant was discarded, and the tube drained by inverting on a clean piece of absorbent paper, taking care that the pellet remains in the tube. 3 ml of 70% Ethanol was added into the tube and the tube was inverted several times to wash the DNA pellet, and the tube was centrifuged 1 min at 2000 x g (4000 rpm). After centrifugation, the supernatant was discarded, and the pellet was air-dried for 5–10 min. 400ul DNA Hydration Solution was added and the solution was vortexed for 5 s at medium speed to mix. The tube was then incubated at room temperature overnight to dissolve the DNA. After the DNA was dissolved, the concentration of the DNA sample was measured using the Nanodrop or Qubit, and the DNA was stored at -20°C.

### Measurements of DNA Concentration

DNA concentrations of genomic DNA were determined by Nanodrop or Qubit. When using Nanodrop measurement (Thermofisher), the NanoDrop 2000 spectrophotometer was used and the sample was analyzed with the NanoDrop 2000 software. The measurement was saved by selecting Add to report to save the data onto a workbook in the software.

2  $\mu\text{L}$  of DNA hydration solution was then used as the blanking solution and placed onto the bottom pedestal of the spectrophotometer. The blank was then wiped away and 1  $\mu\text{L}$  of sample was added onto the pedestal and measured by clicking measure in the software. After each measurement, the sample was wiped with a dry wipe and the next sample was added and measured as before until all samples were measured. When following the Qubit protocol (Invitrogen), 2 assay tubes were prepared for DNA broad range standards, along with 1 assay tube per sample to be measured. The Qubit working solution was made by diluting the Qubit reagent in a ratio of 1:200 in Qubit buffer. 200  $\mu\text{L}$  of working solution was prepared for each standard and sample. The assay tubes were then prepared so that 10  $\mu\text{L}$  of the 2 Qubit standards were added into 190  $\mu\text{L}$  of Qubit mix and 1  $\mu\text{L}$  of sample was added into 199  $\mu\text{L}$  of Qubit mix. Once all samples and standards were added into the assay tubes, the tubes were vortexed for 2-3 seconds and incubated for 2 minutes at room temperature. After incubation, the Qubit Fluorometer was calibrated with the 2 standards, and the sample tubes were added into the Qubit fluorometer to measure the concentration of each sample.

After concentration measurement by NanoDrop or Qubit, the DNA stock samples were finally diluted with DNA hydration solution in 100 ng/ $\mu\text{L}$  aliquots based on calculations using the DNA stock measurements determined by the Nanodrop or the Qubit.

### CGG Allele Sizing by PCR Analysis

*FMR1* specific PCR was used to determine the number of CGG repeats within the *FMR1* allele.

PCR cycles protocol were as follows:

1 hold (98°C, 5 min), 25 cycles (97°C for 35 seconds, 62°C for 35 seconds, and 72°C for 4 minutes), and a second hold (72°C for 10 minutes) and hold at 4°C. The PCR tubes were taken out and labelled accordingly to the DNA samples to be used. Then, the reagents, which consist of GC buffer, *FMR1* primer, Diluent, GC enzyme, and H<sub>2</sub>O, were all taken out of -20°C, apart from the GC enzyme to prevent enzyme degradation. The taken reagents were put on ice and taken to an isolated PCR room to create the MM (Master Mix).

GC buffer, *FMR1* primers, Diluent, GC enzyme, and H<sub>2</sub>O were added in a ratio of (11.45 $\mu\text{L}$ :0.5 $\mu\text{L}$ :1 $\mu\text{L}$ :0.05 $\mu\text{L}$ :1 $\mu\text{L}$ ), respectively. The GC enzyme was taken out of -20°C only at time of use. All reagents except the GC enzyme were vortexed and spun prior to being added, and the enzyme was only spun briefly to prevent damage to the enzyme. The total amount of MM was calculated to ensure there was at least enough for one extra PCR sample after aliquoting the total DNA samples. 14 $\mu\text{L}$  of MM and 2 $\mu\text{L}$  of the appropriate DNA were aliquoted into each PCR tube.

PCR tubes were vortexed and spun, then placed in the PCR machine. At the end of the PCR cycles, PCR products were removed from the PCR machine tray, and the samples were then stored at -20°C.

### Capillary electrophoresis (CE)

CE plates were prepared as follows:

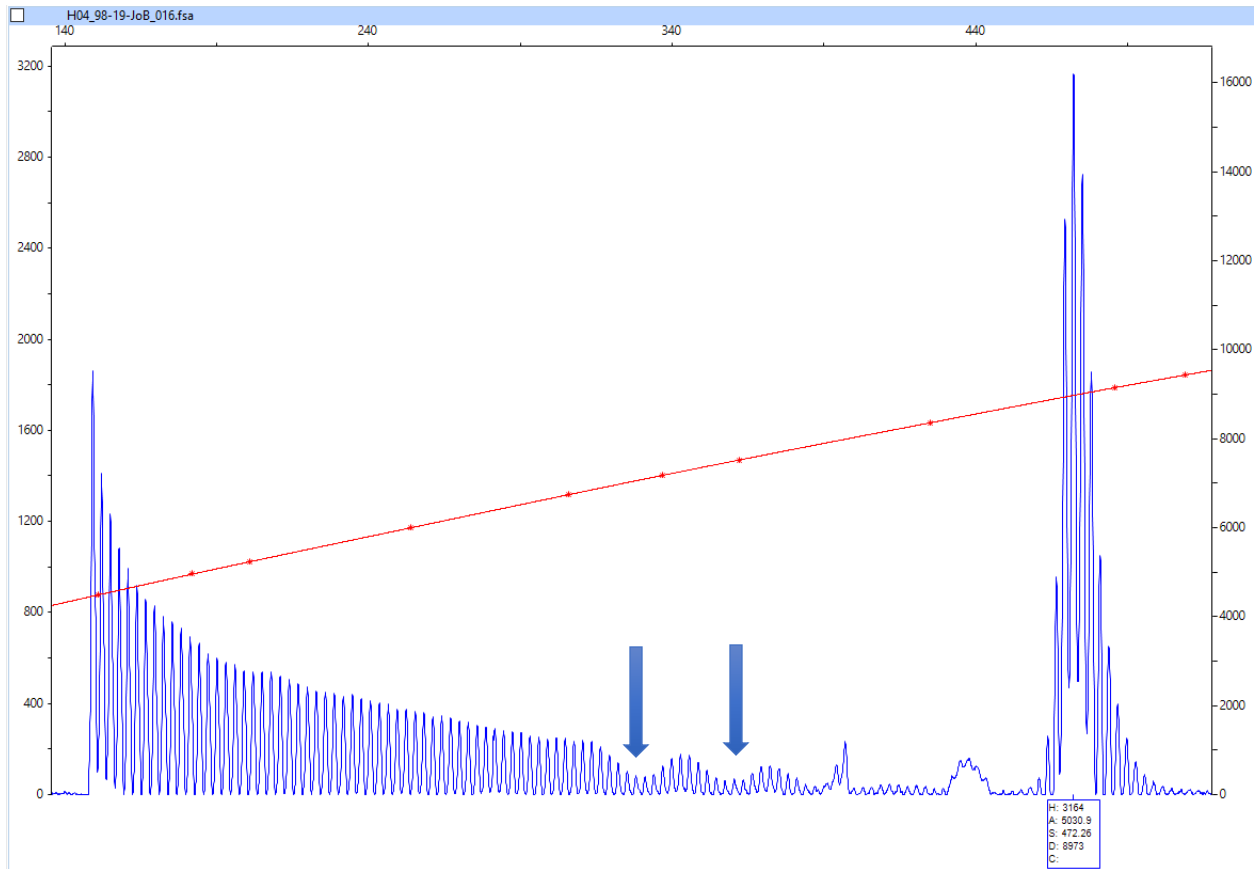
PCR mix, ROX ladder and formamide in a ratio of (2 $\mu$ L:11 $\mu$ L:2 $\mu$ L) were added into a mixture in a 96-well plate format. The ROX ladder and formamide were mixed in the PCR room to prevent contamination, and the mixture was aliquoted into CE plated wells. The wells were analyzed every 2 columns, with each column containing 8 wells. Also, the PCR products and ROX/formamides mixture were added according to a CE plate chart to prevent sample mix-ups after analysis. Extra formamide or nuclease-free H<sub>2</sub>O was added to fill in any empty wells. Once all samples were added, the plate was tightly sealed with adhesive aluminum and flattened to make sure no loose areas, in which the mixture could leak out, was present. Then the sealed plate was vortexed and spun briefly in a large centrifuge. The PCR plate was placed in the PCR machine and the CE prep protocol was activated. Once the PCR was done, the plate was stored on ice and raw data was generated using GeneMapper V4 software. The raw generated data was analyzed using Peak Scanner, a software that determines CGG peak size. ROX Ladder size standard (79, 90, 105, 131, 151, 182, 201, 254, 306, 337, 362, 425, 486, 509, 560, 598, 674, 739, 799, 902, and 1007) and Asuragen FX Analysis Method was used to determine CGG repeat sizes. From generated CE plots, the tallest peaks were chosen from the sample and used to calculate the number of CGG repeats for each allele shown.

### Characterization of AGG Interruptions

To determine the number of AGG repeats, a protocol similar to the CGG sizing protocol was used. In this case, however, a CGG primer was used alongside the other two CGG flanking *FMR1* specific primers was used (tri-primer PCR), and the amount of diluent was halved to accommodate the new reagent, to create a mixture with the ratio (11.45 $\mu$ L GC buffer, 0.5 $\mu$ L *FMR1* primer, 0.5 $\mu$ L Diluent, 0.5 $\mu$ L CGG Primer, 0.05 $\mu$ L GC Enzyme, 1 $\mu$ L H<sub>2</sub>O), with a total of 14  $\mu$ L per reaction. 14 $\mu$ L of MM and 2 $\mu$ L of the appropriate DNA were aliquoted into each PCR tube. Specific details of this method are as described in [Yrigollen et al, 2012].

The PCR protocol was as follows:

Initial hold of 95C, followed by two sets of 10 cycles (97C for 35 sec, 62C for 35 sec, and 68C for 4 minutes) and 20 cycles (97C for 35 sec, 62C for 35 sec, and 68C for 4 min + 20 s/cycle). After running PCR cycles, PCR products were removed from the PCR machine tray, and stored at -20°C. Raw data for PCR products were then generated using CE with GeneMapper V4 and analyzed with Peak Scanner. The number of AGG interruptions was determined by counting the number of depressions within the general pattern of the peak before the peak of the allele itself as shown in Figure 5.

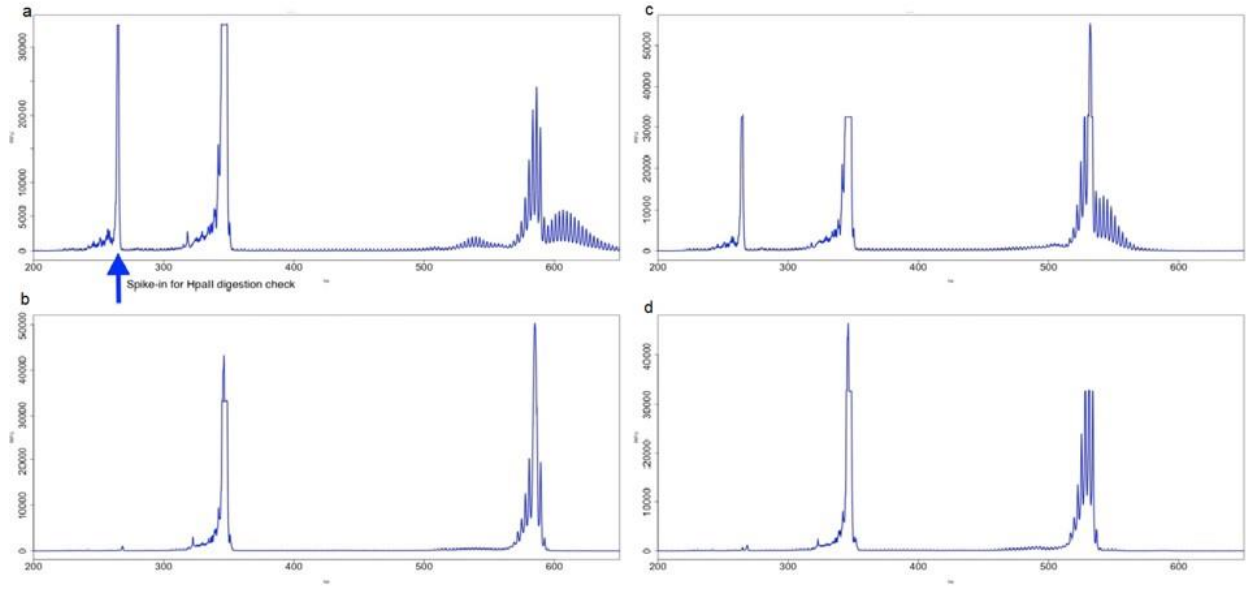


**Figure 5: CE Plots showing the presence of 2 visible AGG interruptions**

Example of CE plot displaying the number of depressions within the CGG peaks which correspond to the presence of the AGG interruption. In this case, the sample has 2 visible AGG interruptions (pointed by the arrows).

### HpaII Digestion

gDNA was digested with HpaII restriction enzyme, which was used to determine which alleles were unmethylated, as HpaII is a methylation restriction enzyme that cut only unmethylated DNA sequence (5' C<sup>^</sup>CGG 3', 5' GGC<sup>^</sup>C 3', Biolabs). HpaII was added to 2  $\mu$ L of 100ng/uL DNA aliquots prior to running PCR. PCR products were visualized through CE, showing which alleles of the CGG repeat expansions in the samples were methylated or not in the CE plots through comparison with the HpaII-undigested samples. Two case examples are showed in Figure 6, which demonstrate that the expanded unstable alleles are unmethylated, as they presence disappear when digested with HpaII.



**Figure 6: HpaII Digestion and Unmethylated Regions**

The blue arrow points at an unmethylated control which is digested upon treatment with HpaII. Figures 6a and 6c depict two samples not treated with HpaII before PCR. Figures 6b and 6d are the same samples as 6a and 6c respectively, but after HpaII treatment prior to PCR. The control is digested along with the unmethylated alleles.

### Southern Blot Analysis

Genomic DNA was isolated from peripheral blood leukocytes (3 to 5 ml of whole blood) using standard methods (Puregene Kit; Gentra Inc., Minneapolis, MN). For Southern blot analysis, 5 to 10 g of isolated DNA was digested with EcoRI and NruI. Digested samples were separated on a 0.8% agarose/Tris acetate EDTA (TAE) gel, followed by partial depurination with HCl (0.4 N) for 15 minutes and denaturation in NaOH (0.5 N) for 30 minutes. DNA was transferred in 10x standard saline citrate (SSC) to a charged nylon membrane (Roche Diagnostics, Basel, Switzerland) using a vacuum transfer apparatus (Vacuum Blotter 785; Bio-Rad, Hercules, CA). A 1-kb DNA ladder (Invitrogen, Carlsbad, CA) was used as a size standard. The membranes were cross-linked (UV Cross linker; Fisher Scientific, Pittsburgh, PA) and were hybridized overnight at 42°C in roller bottles (Isotemp, Fisher Scientific) in Dig Easy Hybridization Buffer (Roche Diagnostics) with the *FMR1* genomic probe StB12.3, labeled with Dig-11-dUTP by PCR (PCR Dig Synthesis Kit; Roche Diagnostics). After denaturation (boiling for 15 minutes), the probe was blocked with Cot1 DNA (Invitrogen) for 20 minutes at 65°C. Filters were washed twice for 5 minutes in 2 SSC/1% SDS and twice for 15 minutes in 1x SSC/0.1% SDS at 65°C. Filter blocking and *FMR1* gene detection were performed using blocking solution and detection buffer according to the manufacture (Roche Diagnostics). Filters were exposed to X-ray film (Super RX; Fuji Medical X-Ray Film, Bedfordshire, UK) for 2 hours.

Genomic DNA was also amplified by PCR with primers c and f32 using the osmolite betaine according to Saluto et al.<sup>30</sup> PCR fragments were visualized on a 2% agarose gel, ethidium bromide stained. For a correct sizing, PCR products were separated on 6% denaturing polyacrylamide gels, followed by electroblot transfer (TE62; Amersham-Pharmacia Biotech, Pittsburgh, PA) to a nylon membrane (Roche Diagnostics) at 4 volts for 2 hours. Membranes

were then cross-linked (UV Cross linker; Fisher Scientific). Dig-labeled DNA molecular weight Marker V (Roche Diagnostics) and a known size marker were used as size standards. Filters were hybridized overnight at 42°C in roller bottles (Isotemp; Fisher Scientific) in Dig Easy Hybridization Buffer (Roche Diagnostics) with a Dig-end-labeled oligonucleotide probe [(CGG)<sub>10</sub>] and Dig-labeled pBR322 DNA. Filters were washed at room temperature, twice for 5 minutes with 2x SSC/0.1% SDS (100 ml) and twice for 7 minutes in a larger volume (400 ml) with the same washing solution, followed by two washes of 25 minutes, each in 0.5x SSC/0.1 SDS at 45°C. Detection of the *FMR1* PCR products was performed according to the manufacturer (Roche Diagnostics). Filters were exposed to X-ray film (Super RX; Fuji Medical X-Ray Film) at room temperature for approximately 30 minutes. Analysis of the repeat number for both Southern blot and PCR used an Alpha Innotech FluorChem 8800 Image Detection System (Alpha Innotech, San Leandro, CA).

### SNP Genotyping



For genotyping 2.4 µL of gDNA (50 µg/µL), the samples were aliquoted into 96-well plates or 384-well plates and 2.6 µL of TaqMan reagent were added into each well. The plates were then sealed, vortexed, and spun. Plates were then placed into the StepOne System (Applied Biosystems) and analyzed using the StepOne program, through Advanced Setup, which was set up to detect the target SNPs. The TaqMan genotyping assay include a QSY quencher probe attached to 2 different signals, VIC, and FAM, which detect two allele types of the target gene within an individual. The signal was detected for each cycle as the product gets amplified.

The generated raw data was saved and uploaded into the Genotyping qPCR program, in which the analyzed samples were placed into 3 different categories: homozygous for one allele, homozygous for the other allele, and heterozygous for both. The data was then sorted into a larger master table and compared with the data of known genotyped sample controls.

## V. Published manuscripts

Perspective

## Repeat Instability in the Fragile X-Related Disorders: Lessons from a Mouse Model

Xiaonan Zhao <sup>1,†</sup> , Inbal Gazy <sup>1,†</sup>, Bruce Hayward <sup>1</sup>, Elizabeth Pintado <sup>2</sup>, Ye Hyun Hwang <sup>3</sup>, Flora Tassone <sup>3</sup>  and Karen Usdin <sup>1,\*</sup>

<sup>1</sup> Section on Gene Structure and Disease, Laboratory of Cell and Molecular Biology, National Institute of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA; xiaonan.zhao@nih.gov (X.Z.); inbal.gazy@nih.gov (I.G.); bruce.hayward@nih.gov (B.H.)

<sup>2</sup> Department of Medical Biochemistry and Molecular Biology, School of Medicine, University Hospital Virgen Macarena, University of Seville, 41009 Seville, Spain; elizabet@us.es

<sup>3</sup> Department of Biochemistry and Molecular Medicine and MIND Institute, UC Davis Medical Center, Sacramento, CA 95817, USA; yehhwang@ucdavis.edu (Y.H.H.); ftassone@ucdavis.edu (F.T.)

\* Correspondence: ku@helix.nih.gov; Tel.: +1-301-496-2189

† These authors contributed equally to this work and thus should be considered co-first authors.

Received: 12 January 2019; Accepted: 27 February 2019; Published: 1 March 2019



**Abstract:** The fragile X-related disorders (FXDs) are a group of clinical conditions that result primarily from an unusual mutation, the expansion of a CGG-repeat tract in exon 1 of the *FMR1* gene. Mouse models are proving useful for understanding many aspects of disease pathology in these disorders. There is also reason to think that such models may be useful for understanding the molecular basis of the unusual mutation responsible for these disorders. This review will discuss what has been learnt to date about mechanisms of repeat instability from a knock-in FXD mouse model and what the implications of these findings may be for humans carrying expansion-prone *FMR1* alleles.

**Keywords:** CGG Repeat Expansion Disease; DNA instability; expansion; contraction; mismatch repair (MMR); base excision repair (BER); transcription coupled repair (TCR); double-strand break repair (DSBR); Non-homologous end-joining (NHEJ); mosaicism

### 1. Introduction

The fragile X-related disorders (FXDs) are X-linked disorders that include a form of ovarian dysfunction known as fragile X-associated primary ovarian insufficiency (FXPOI; MIM# 311360), a neurodegenerative condition, fragile X-associated tremor/ataxia syndrome (FXTAS; MIM# 300623) and fragile X syndrome (FXS; MIM# 300624), a major cause of intellectual disability and autism [1]. FXPOI and FXTAS are seen in carriers of *FMR1* premutation (PM) alleles, alleles that have a tandem array of 55–200 CGG repeats in exon 1. Most cases of FXS are seen in carriers of *FMR1* full mutation (FM) alleles, alleles that have >200 repeats, with a minority of individuals having deletions or point mutations that affect the levels or functionality of FMRP, the *FMR1* gene product, an important regulator of translation in the brain. The difference in the clinical consequences of the inheritance of a PM versus a FM allele results from the paradoxical effects of the repeat on *FMR1* expression. Most FM alleles are epigenetically silenced, resulting in the absence of FMRP. In contrast PM alleles are transcriptionally active and can have transcript levels anywhere between 2 and 8 times the levels of normal alleles [2]. Both FM and PM carriers show wide variability in their clinical presentation and both FXTAS and FXPOI show incomplete penetrance suggesting the contribution of other genetic factors to disease severity.

The CGG-repeat tract is unstable and is prone to expand and contract in a manner dependent on repeat number and the number of AGG interruptions present at the 5' end of the tract [3,4]. Instability



occurring in somatic cells can lead to repeat size mosaicism. In fact, it has been estimated that >40% of individuals with alleles >55 repeats are mosaic [5]. Mosaicism for both PM and FM alleles results in some individuals showing symptoms characteristic of both PM and FM alleles [6,7]. The mechanism(s) responsible for the repeat instability is largely unknown. While some instability has been reported in various human cells in culture [8–10], most studies of the instability mechanisms have used mouse models [11–14]. Mouse models offer a number of clear advantages over some of these cell-based systems, including the high frequency of both expansions and contractions and the ability to examine instability in different biologically relevant organs at various stages of development. In addition, since the size of the original allele can be readily established based on the allele size at birth, alleles arising from expansion can be clearly distinguished from those arising by contraction. Of the various mouse models for the FXDs that have been generated, most work on repeat instability has made use of a model in which the short CGG-repeat tract present in the endogenous mouse *FMRI* gene was replaced with ~130 CGG-repeats [14]. In this review we will address what we have learnt to date about repeat instability from this mouse model, as well as from other model systems and other related Repeat Expansion Disorders. We will also discuss some of the implications of this information for diagnosis and disease risk assessment in humans.

## 2. Instability in Humans and Mice May Share a Common Molecular Basis

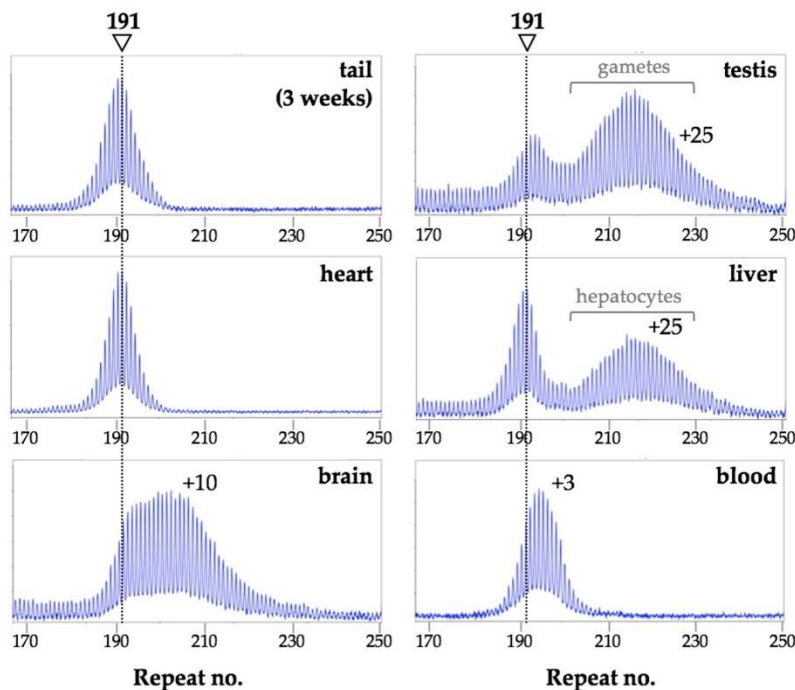
There are a number of reasons to think that instability in mice and humans share common mechanisms. Firstly, while both expansions and contractions are seen, in both species expansions predominate over contractions, at least in the PM range. Secondly, in both humans and mice expansion events require transcription or the presence of the PM allele in a region of open chromatin [15,16]. In addition, both mice and humans show a maternal age effect for expansion risk [3,17], suggesting that expansion occurs in the oocyte in both species. While some postnatal oogenesis has been observed in mice whose existing oocytes have been ablated [18], the contribution of a dividing oocyte stem cell population to postnatal oogenesis and the pool of viable oocytes that can be fertilized in normal mammals is still controversial. Thus, since oocytes do not divide, a maternal age effect for expansions is generally considered to reflect events that occur in the absence of cell division. This suggests that the underlying mechanism in both species involves aberrant DNA repair and/or recombination, rather than a problem with chromosomal replication. Finally, many of the same genetic factors that affect expansion risk in the FXD mouse are known to modulate expansion risk in other human Repeat Expansion Diseases [19–21]. Since current evidence supports a common mutational mechanism for all of these diseases, this suggests that the FXD mouse may accurately recapitulate at least some aspects of repeat expansion in the FXDs.

Whilst most intergenerational expansions in mice are relatively small, large expansions characterize the PM to FM transition on maternal transmission in humans. However, there is no clear evidence that the PM to FM transition occurs in a single step in women and, in principle, it is possible to generate an equivalent increase in repeat number by a series of small expansions occurring over the decades that the human oocyte spends in dictyate arrest prior to fertilization [17]. Furthermore, while small expansions predominate in mice, larger intergenerational expansions are seen albeit at a lower frequency [17]. These larger expansions are sensitive to the same genetic factors that affect small expansions and thus likely arise from the same basic mechanism [22,23]. While much less is known about contractions, work in both mice and humans suggest that the underlying mechanism is likely different from the mechanism that gives rise to expansions [4,24,25].

Although available evidence suggests that mice may be useful for understanding the instability of human PM alleles, mouse *Fmr1* alleles with repeat numbers in the FM range do not undergo repeat-mediated epigenetic silencing as do FM alleles in humans. However, there is reason to think that X chromosome inactivation in female mice can provide a window into the factors associated with the instability of silenced alleles [16].

### 3. Different Cell Types Show Different Propensities to Expand in Mice

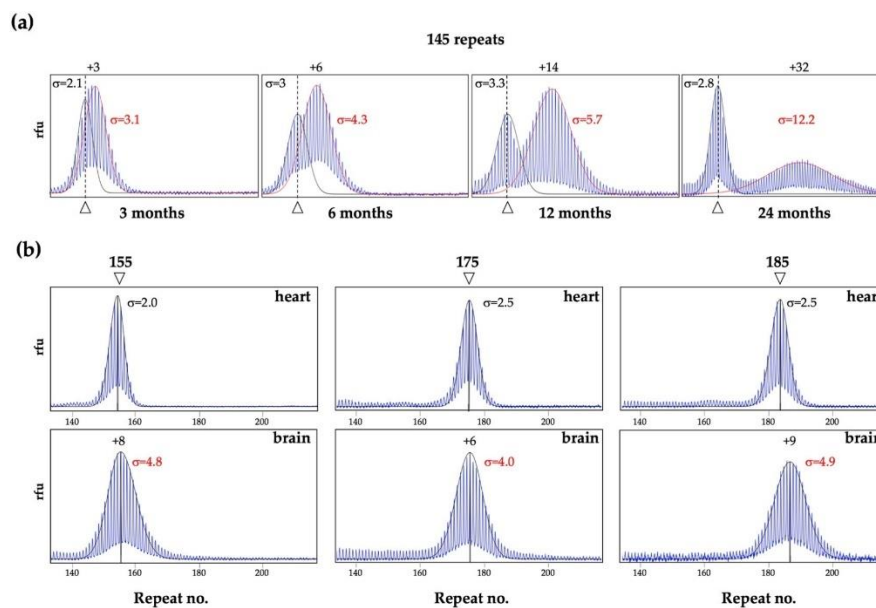
As can be seen in Figure 1, in the FXD mice some organs, like heart, show very little post-natal expansion, as evidenced by the fact that the repeat PCR profile of this organ does not change with age and remains indistinguishable from the profile seen in the tail DNA taken at 3 weeks of age [26]. In contrast, most other organs show some expansion as evidenced by the presence of alleles larger than the heart allele (Figure 1) [17,26]. In organs like testes and liver, expansions are apparent as a shift from a unimodal repeat PCR profile as seen in the tail DNA, to one that is more bimodal (Figures 1 and 2a). In young animals that have not yet accumulated many expansions, the second peak can appear more like a “shoulder” rather than a distinct peak, as in the example of the liver of a 3-month old mouse shown in Figure 2a. However, over time, as expansions continue to accumulate, expanding alleles diverge further from the original allele resulting in 2 clearly distinct allele peaks (Figures 1 and 2a). This bimodal peak distribution reflects the fact that, in some organs some cells are expansion-prone whilst others are not.



**Figure 1.** Different mouse organs show different propensities to expand. Repeat PCR was carried out on DNA extracted from different organs of a male mouse with 191 inherited repeats and analyzed as previously described [26]. The tail DNA sample was taken at 3 weeks, the remaining samples at 8 months of age. The arrowhead and dotted lines indicate the repeat number in the original inherited allele as assessed from tail DNA taken at 3 weeks of age. The numbers within each panel indicate the number of repeats added.

For example, in testes expansion is confined to the spermatogonia or primary spermatocytes [17], with the shorter alleles in the testes profile in Figure 1 corresponding to alleles in the somatic cells and the longer alleles corresponding to alleles in the gametes. Similarly, in the liver, expansion is confined to hepatocytes [27], while in the brain expansion is more extensive in the striatum and basolateral amygdala than in the medial prefrontal cortex [26]. Unlike PM alleles in testes and liver, PM alleles in

blood show relatively little expansion and what little expansion is seen has a unimodal distribution in males (Figure 1). This may reflect the fact that all white blood cells are equally prone to expansion. Why some cells are expansion-prone and others are not is not fully understood. It does not seem to be simply related to the amount of *Fmr1* transcription since tissues with similar levels of *Fmr1* mRNA can show very different propensities to expand [26]. Rather it may be related to the balance between the levels of expression of proteins involved in the generation of expansions and those involved in the pathway(s) that promotes contraction or error-free repair [26,28]. Computer simulations suggest that the expanded allele profile, even that seen in very expansion-prone tissue, is consistent with the addition of 1–2 repeats with each expansion event [29]. However, as will be discussed below, larger expansions and contractions also occur (Section 7).



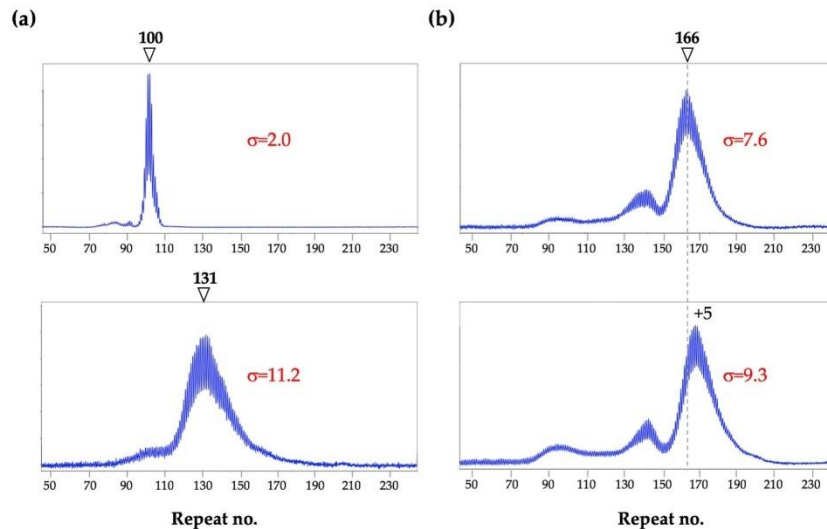
**Figure 2.** Change in the PM repeat PCR profiles and  $\sigma$  with age and extent of expansion. The PCR profiles and  $\sigma$ s were generated from male mice as previously described [26,27]. (a) Liver PCR profiles of mice of different ages, all with an original allele of ~145 repeats. The arrowhead on the bottom of each panel and the dotted lines indicate the original inherited allele as assessed from tail DNA taken at 3 weeks of age. The numbers above the panels indicate the number of repeats added to the expanded allele. The  $\sigma$  of the stable allele and the expanded allele are shown in black and red, respectively. The examples shown in this panel are derived from previously published work [27]; (b) PCR profiles and corresponding  $\sigma$  for alleles in hearts and brains of 1-year old (155 repeats) and 6-month old (175 and 185 repeats) mice. The number associated with each arrowhead represents the number of repeats in the indicated allele. For the hearts, this number corresponds to the original inherited allele. For the brains, the repeat size reflects a gain of 6–9 repeats from the original allele.

As alleles expand, their PCR profile widens as differences in the timing and size of expansions in different cells transforms the original discrete allele into a more heterogeneous mixture of allele sizes (Figure 2a). Thus, the broadness of the allele peak, as reflected in the standard deviation ( $\sigma$ ), can be a sensitive metric of expansion [27,29]. In our experience, stable alleles, like those in heart, show a  $\sigma$  of  $\leq 2.5$ , for a wide range of repeat sizes and mouse ages as illustrated in Figure 2b. In contrast, the  $\sigma$  of alleles in expansion-prone cells is  $>2.5$  and increases as expansion increases as the animals age (Figure 2a,b). Overlapping allele peaks can result in an overestimation of the  $\sigma$  of the smaller allele and

an underestimation of the larger one (Figure 2a). However, in unimodal PCR profiles or profiles with distinct peaks, an allele that has expanded has a larger  $\sigma$  than a stable allele of the same size (compare brain to heart in Figure 2b).

It should be noted that while the repeat PCR profiles for alleles in the heart of WT mice shows no evidence of post-natal expansion, they differ from the heart profiles in mice with mutations that abolish expansions completely [30]. Specifically, the profile in WT animals has a normal distribution while the profile in mutant mice is not only sharper, but is also left-skewed [30]. This left skew likely reflects PCR stutter products, while the normal distribution of the WT heart profiles likely reflects some expansion that occurs in these animals during early embryonic development.

*Implications for humans:* Analysis of the CGG repeat in the human *FMRI* gene is routinely performed using blood where, as in mice, a unimodal PCR profile is commonly seen in males. However, by analogy with what is seen in mice, a unimodal PCR profile may not mean that the allele is stable. As with mice, the  $\sigma$  of an allele profile likely reflects the extent of somatic expansion. This would be predicted to vary with total repeat number, the number of AGG interruptions and the effect of different genetic and/or environmental modifiers of expansion risk. As with mice (Figure 2a), age may also be a factor, at least for very unstable alleles. Figure 3a shows examples of unimodal repeat PCR profiles characteristic of stable (top panel) and unstable (bottom panel) alleles.



**Figure 3.** Blood repeat PCR profiles for 3 human male PM carriers. Repeat PCR analysis was carried out as described previously [31]. The number associated with each arrowhead represents the number of repeats in the indicated allele. (a) Profiles of two individuals showing a unimodal profile with a sharp peak (top) and a broad peak (bottom); (b) Profiles generated from the same individual using samples taken 4 years apart. The dotted line in these panels indicates the size of the major allele at the earlier timepoint. A shift corresponding to the gain of 5 CGG repeats is seen in the later sample. In each case the  $\sigma$  values are for the major allele peak.

As illustrated in Figure 3b, multiple peaks are also sometimes seen in males. These alleles may arise by contraction of larger alleles or from large expansions. As can be seen in the 2 different blood samples from the same individual taken 4 years apart (Figure 3b), these peaks can be broad, indicative of subsequent expansions. Notably, in this individual, both the repeat number and the  $\sigma$  were larger in the second sample taken 4 years later, consistent with an age effect.

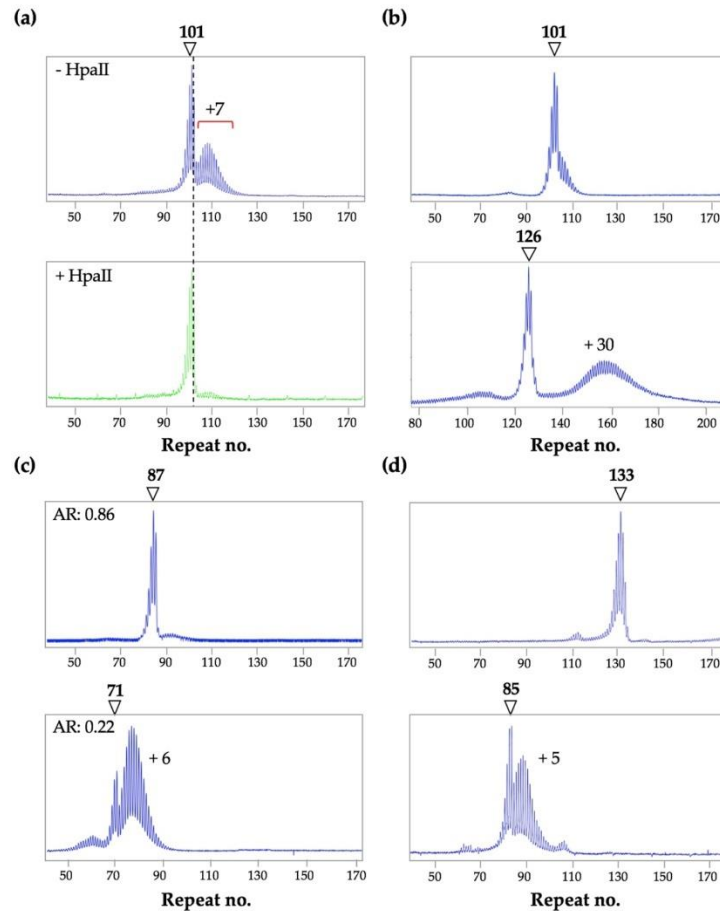
PM alleles are relatively stable in mouse blood compared to alleles from other cells. Thus, it is possible that any evidence of expansion in human blood, reflects the presence of even more extensive expansion in organs like brain or gonads. Different allele profiles in different tissues have been reported in some PM carriers [32–39] and 2 different studies support the idea that some men have larger expansions with broader  $\sigma$ s in sperm relative to blood [40,41]. Differences between allele sizes in blood and in brain or gonads in expansion-prone individuals could lead to an underestimation of the intergenerational expansion risk. It may be that it also contributes to the apparent variability in the penetrance of FXTAS and FXPOI.

#### 4. Expansions in Females Only Occurs on The Active X Chromosome

While in male mice a bimodal repeat PCR profile is seen in organs that have cell types with different expansion rates, most female mice show a bimodal PCR profile for the PM allele in all expansion-prone tissues [16]. This is a consequence of the fact that expansions in females are confined to alleles on the active X chromosome [16]. Thus, even in expansion-prone cell types, expansion only occurs in ~50% of cells in females with normal X chromosome inactivation (XCI).

*Implications for humans:* A bimodal PCR profile for the PM allele is also sometimes seen in human females (Figure 4). The larger alleles are lost from the repeat PCR profile when the DNA is digested prior to PCR with a methylation-sensitive enzyme that has one or more cleavage sites within the PCR amplicon. Such an enzyme cuts the *FMRI* allele on the active X chromosome, leaving the allele on the inactive X as the only template for PCR. Thus, in the example shown in Figure 4a, alleles on the active X have gained ~7 repeats relative to alleles on inactive X chromosomes (bottom panel). Notably, unlike the roughly normal distribution seen in the repeat PCR profiles of expanded alleles, the shape of the repeat PCR profile for alleles on the inactive X is asymmetric and closely resembles the shape of the PCR profile seen in mice with mutations that completely block expansions [30,42]. Even among women with a bimodal allele distribution, differences in the extent of instability can result in dramatic differences in their PCR profiles. Figure 4b illustrates the 2 extremes of the possible bimodal PCR profiles, with the woman in the top panel showing a very low level of somatic instability and the woman in the bottom panel showing unusually high levels.

However, not all women show a bimodal profile for the PM allele. For example, as shown in the upper panel of Figure 4c, women with a high activation ratio (AR), that is, a high proportion of cells in which the active X chromosome carries the normal allele, would show a single sharp peak for the PM allele, with little, if any, evidence of a second peak, since no expansion would take place on the inactive X. In contrast, a woman with a low AR, would be more likely to have a profile with more expanded alleles than stable ones as shown in the bottom panel of Figure 4c. An even lower AR might result in unimodal PM profile with a large  $\sigma$ , as reported for a woman with an AR of 0.06 [43]. A single, sharp PM allele profile can also be seen even in the absence of skewed XCI (Figure 4d). Such alleles may result from the presence of AGG interruptions that reduce the expansion frequency, a genetic background that is not prone to expansion or, potentially, to an effect of age, with very young females being more likely to show such a profile. In the cases shown in Figure 4d, both women were of similar age, showed no skewing of XCI and had no AGG interruptions. Thus, the sharp and asymmetric profile seen for the 133 repeat allele may reflect genetic or environmental factors that reduce expansion risk as in mice with mutations that block expansions [24,30,42].



**Figure 4.** PM Repeat PCR profiles from the blood of human female PM carriers. The arrow in each instance indicates the stable allele with the indicated number of repeats. (a) Profiles for a female PM carrier without (top panel) and with (bottom panel) HpaII pre-digestion were generated as previously described [16]. The alleles on both the active and inactive X are shown in blue in the top panel and the alleles on the inactive X in green in the bottom panel; (b) Examples of very different bimodal PCR profiles; (c) Profiles for 2 women with different activation ratios (ARs); (d) Profiles of two females of similar ages and ARs, both with no AGG interruptions showing very different levels of somatic expansion.

### 5. Expansion in the Male and Female Germline

More expansions are seen in the oocytes of older female mice than younger ones [17]. This is consistent with some expansions occurring postnatally in non-dividing oocytes. In contrast, germ line expansions in male mice occurs in the spermatogonial stem cells (SSCs), cells that undergo multiple rounds of cell division [17]. Furthermore, male mice show a higher frequency of germ line expansions than females [17]. Studies of mice with mutations in different genes shows that the same genetic factors affect expansions in males and females and the same genetic factors also account for large and small expansions [24,25,28,30,42,44]. This would be consistent with single mechanism being responsible for all expansions in both males and females.

*Implications for humans:* Almost all expansions from a PM to a FM allele in humans occurs on maternal transmission. However, men with PM alleles transmit more expansions at least for smaller PM alleles [4,45] consistent with what is observed in mice. The fact that expansions in male and female mice share a dependence on a common set of genes would argue against a female-specific mechanism for the generation of FM alleles in humans. The FX repeat forms unusual intrastrand secondary structures [46–50], that are thought to make the repeat tract difficult to replicate [49,51,52]. This may result in pressure for larger alleles in dividing gametes to contract over time, as is seen in FX embryonic stem cells [53]. Germ cells in a 30-year old man will have undergone ~400 divisions, compared to 31 in a woman the same age [54,55]. However, by analogy with mice, most expansions in the female germline likely occur during post-natal life, well after cell division is complete. Thus, expanded alleles in female gametes face little pressure to contract, whilst male gametes are under continuous pressure to do so. This might explain why FMRP was detected in primordial germ cells of a 17-week old male FM fetus but not in those of a 13-week old fetus [56] and why older FM males only have PM alleles in their sperm [57,58]. It may also provide an explanation for why male PM carriers do not generally transmit FM alleles to their children.

## 6. Genetic and Environmental Factors Affecting Instability

A number of genetic and environmental factors have been shown to impact expansion risk in mice. For example, an exogenous source of oxidative stress increases expansions [59]. Mutations in different DNA repair genes also affects the extent of expansion, with some mutations reducing expansions [24,25,28,30,42,44] and others increasing them [27,30,44,60]. For example, mutations in mismatch repair (MMR) proteins, including MSH2, MSH3, MSH6 and MLH3, either eliminate expansions altogether [24,30,42] or severely reduce their incidence [25]. Similarly, a single hypomorphic allele of Pol $\beta$ , a DNA polymerase that plays an essential role in base excision repair (BER), is sufficient to significantly reduce the expansion frequency [28]. Thus, proteins from multiple DNA repair pathways that normally work to prevent mutations, interact in such a way so as to actually cause the repeat expansion mutation. Work in vitro has shown that the FX repeats form unusual DNA structures including hairpins that have a mixture of Watson-Crick and non-Watson-Crick base pairs [46–50,61,62]. Current thinking is that these structures are the substrates upon which this process acts but the sequence of events and all the factors involved are still not fully understood (see [23] for recent review).

Mutations in other proteins, including two 5'-3' exonucleases, EXO1 and FAN1, lead to an increase in expansions, suggesting that these proteins are protective [30,60]. Loss of EXO1 affects expansions in the germ line and in the small intestine but not in the brain [30]. In contrast, loss of FAN1 affects expansion in multiple organs including brain but does not affect the germ line expansion frequency [60]. ERCC6/CSB plays a paradoxical role in repeat expansion playing a minor role in promoting expansions in some instances [44] and protecting against them in others [63]. This paradoxical effect may reflect this protein's ability to participate in multiple DNA repair pathways. Recently, DNA ligase IV (LIG4) has also been shown to protect against expansion [27]. Since LIG4 is essential for non-homologous end-joining (NHEJ), a form of double strand break (DSB) repair, it suggests that the expansion process competes with the NHEJ pathway for a common substrate. This supports the idea that expansion proceeds through a DSB intermediate, perhaps one generated by MutL $\gamma$  [27]. A simple gap-filling model for the generation of expansions from a staggered DNA DSB arising during transcription or DNA repair has been suggested [27].

Very little is known about the factors that promote contractions. In the mouse model, factors that abolish expansions do not necessarily reduce contractions [24,25,28,30,42]. In fact, the frequency of contractions usually increases when the expansion frequency drops. This suggests that some, if not all, contractions occur via a mechanism that differs from the expansion mechanism and that when expansions are blocked, a process or processes that favors contractions predominates. This would be consistent with the observation that AGG interruptions, which are an important modifier of expansion risk, do not affect the contraction frequency [4]. Moreover, while transcription or open chromatin

is required for expansion, contraction of methylated alleles can be seen in both mouse and humans, particularly in rapidly dividing cells such as those in the early embryo or in cells grown at low cell densities in vitro [10,17,53]. Contractions under these circumstances may reflect the difficulty the cell has in replicating the FX repeats [49,51,52], thus favoring cells in which repeats have been lost. Tandemly repeated sequences often contract via strand-slippage during replication in a variety of organisms [64–66]. Such a mechanism could explain the observed loss of AGG interruptions sometimes associated with contraction [4], if slippage occurred upstream of the interruption with re-priming of DNA synthesis occurring downstream of the interruption. Strand-slippage by the FX repeat also has the potential to generate point mutations within the repeat by frameshifting or limited intra-strand template switching with priming within the hairpin [67–69].

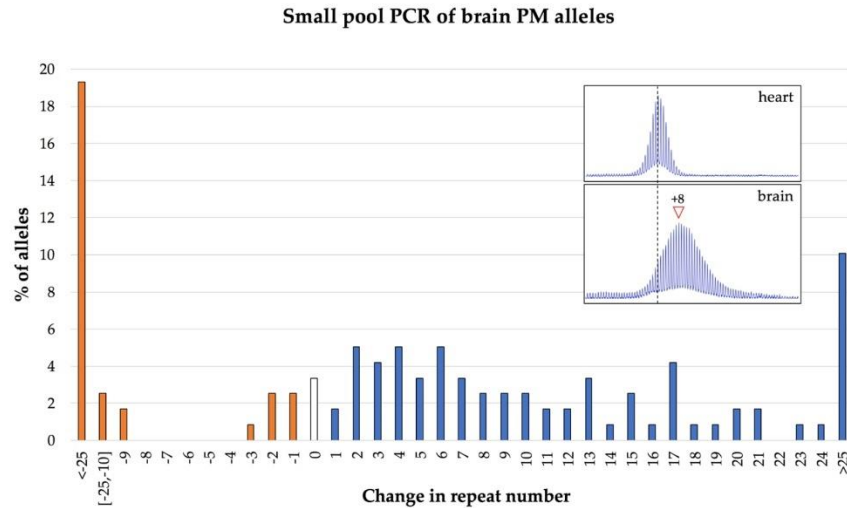
*Implications for humans:* Genetic factors that impact expansion may contribute to the increased expansion risk seen for PM carriers with a family history of FXS relative to carriers of similar PM alleles in the general population [70]. It may also account for why some individuals show more somatic expansion than others (Figures 3 and 4). Interestingly, single nucleotide polymorphisms (SNPs) in genes including *ERCC6/CSB* [71], *MSH3* [21,72] and *FAN1* [19,20] are thought to modify disease risk in other Repeat Expansion Diseases via their effect on somatic expansion. A SNP in *MLH1*, the binding partner of *MLH3* in the complex MutL $\gamma$ , has also been shown to have a similar effect [19]. Since most factors that affect expansion in somatic cells also affect expansion in germ cells in mice, it is likely that similar genetic factors would also impact the risk of intergenerational expansion. However, as illustrated by the differences between the effects of *FAN1* and *EXO1* mutations on expansion in different tissues in mice, some genetic factors may be more important modulators of expansion in some cells than in others and thus, may affect expansion differently in different organs. For example, a polymorphism in *FAN1* may result in increased expansion in brain but not necessarily in oocytes, while *EXO1* polymorphisms may result in increased expansion in gametes, but not liver or other somatic tissue. Thus, a thorough understanding of expansion predisposition may require testing of multiple tissues.

## 7. The Frequency of Large Contractions and Expansions can be Underestimated

Analysis of repeat length and somatic instability is routinely performed on bulk genomic DNA. Such analysis on mice tissue indicates that somatic instability mostly involves the gain of relatively small number of repeats (Figure 2). However, large expansions and contractions can be seen in intergenerational transmission in mice [17]. They can also be seen if they occur during early embryonic development when they represent a significant fraction of the alleles in the population. These observations indicate that large expansions and contractions can occur not only in humans but also in mice. However, similar events occurring postnatally are difficult to detect using PCR on bulk genomic DNA since the resultant alleles vary considerably in the number of repeats gained or lost. Thus, each of these alleles represents a very small proportion of alleles in the population and likely will not be detected in standard PCR analysis. For example, as seen in Figure 5, when bulk DNA from the brain of a 1-year old mouse is analyzed, the PCR profile is consistent with most changes in repeat number involving the gain of a small number of repeats. However, PCR on single genome equivalents from the same brain sample shows that almost 30% of alleles have lost or gained more than 25 repeats. Thus, larger expansions and contractions actually occur relatively frequently and may ultimately reflect a relatively large fraction of the total alleles in the population.

*Implications for humans:* This combination of expansion and contraction can result in individuals being highly mosaic for a variety of different alleles. The fact that larger expansions and contractions that occur later in development are difficult to detect in mice, raises the possibility that some humans may be even more mosaic than analysis of their bulk DNA suggests. Thus, careful analysis of the distribution of allele sizes in carriers might be needed to properly assess disease risk.





**Figure 5.** Distribution of the changes in the repeat number seen in the brain of a 1-year old mouse with an inherited allele of 162 repeats. The repeat number of individual alleles was determined by small pool PCR of single genome equivalents as described previously [30]. The data shown represent 119 individual PCR reactions. The inset panels show the bulk PCR profiles for the heart and brain of the same animal.

## 8. Concluding Remarks

The use of a mouse model allows the dynamics of repeat instability in the *FMR1* gene to be explored over time in multiple tissues. This has resulted in a number of observations that may be relevant to repeat instability in humans. For example, we have learnt that some cell types show more expansion than others [16,17,24,26]. In particular, expansions are more extensive in brain and gonads than in blood. This may mean that, in some people, the size of the repeat or the extent of somatic expansion seen in blood may not reflect what is present in disease-relevant cells. Examination of the repeat PCR profiles from mice over time has also shown that the  $\sigma$  of the allele peak provides a sensitive measure of the extent of somatic instability. Specifically, stable alleles have a low  $\sigma$  ( $<2.5$ ), while expanded alleles have a larger  $\sigma$ . A wide range of  $\sigma$  values can be seen for human PM alleles in blood that, by analogy with FXD mice, likely reflects a wide variation in the extent of somatic expansion in different people.

A number of genetic factors that promote or protect against expansions in the FXD mouse model have been identified [24,25,27,28,30,42,44,60,63,73,74]. Some of these factors have been implicated in expansion in humans with other Repeat Expansion Diseases [20,21,75–77], suggesting that they may be relevant for human carriers of unstable *FMR1* alleles as well. If so, the prediction would be that polymorphisms in these factors would be modifiers of both germ line and somatic expansion risk in FX families. As such, people who have elevated activities of protective factors or reduced activities of factors that promote expansion may show little, if any, somatic expansion. In contrast, those with elevated activities of expansion-promoting factors or reduced activities of protective factors may show more somatic expansion. Evidence from other Repeat Expansion Diseases suggests that somatic expansion contributes to differences in the age at onset and disease severity [20,21,75–77]. In a recent study, women who showed PM allele mosaicism reported more severe symptoms than women who were not mosaic [39], suggesting that somatic instability may exacerbate PM symptoms. However, additional studies are needed to fully understand the contribution of somatic expansion to disease pathology in PM carriers. In any event, an increased propensity for somatic expansion likely indicates

an increased propensity for germ-line expansion and thus an increased risk of intergenerational transmission of larger alleles. Additional work is also needed to assess whether the same genetic factors that affect expansion risk in mice also modulate the risk of somatic and intergenerational expansion of human FXD alleles.

**Author Contributions:** X.Z, I.G., B.H., E.P., F.T. and K.U. compiled the data and wrote the paper.

**Acknowledgments:** The work described in this manuscript was funded by a grant from the Intramural Program of the NIDDK to KU (DK057808) which includes funds for covering the costs to publish in open access.

**Conflicts of Interest:** The authors declare no conflict of interest. The funding sponsors had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript and in the decision to publish the results.

## References

- Lozano, R.; Rosero, C.A.; Hagerman, R.J. Fragile X spectrum disorders. *Intractable Rare Dis. Res.* **2014**, *3*, 134–146. [[CrossRef](#)] [[PubMed](#)]
- Tassone, F.; Hagerman, R.J.; Taylor, A.K.; Gane, L.W.; Godfrey, T.E.; Hagerman, P.J. Elevated levels of *FMRI* mRNA in carrier males: A new mechanism of involvement in the fragile-X syndrome. *Am. J. Hum. Genet.* **2000**, *66*, 6–15. [[CrossRef](#)] [[PubMed](#)]
- Yrigollen, C.M.; Martorell, L.; Durbin-Johnson, B.; Naudo, M.; Genoves, J.; Murgia, A.; Polli, R.; Zhou, L.; Barbouth, D.; Rupchock, A.; et al. AGG interruptions and maternal age affect *FMRI* CGG repeat allele stability during transmission. *J. Neurodev. Disord.* **2014**, *6*, 24. [[CrossRef](#)] [[PubMed](#)]
- Nolin, S.L.; Glicksman, A.; Ersalesi, N.; Dobkin, C.; Brown, W.T.; Cao, R.; Blatt, E.; Sah, S.; Latham, G.J.; Hadd, A.G. Fragile X full mutation expansions are inhibited by one or more AGG interruptions in premutation carriers. *Genet. Med.* **2015**, *17*, 358–364. [[CrossRef](#)] [[PubMed](#)]
- Nolin, S.L.; Glicksman, A.; Houck, G.E., Jr.; Brown, W.T.; Dobkin, C.S. Mosaicism in fragile X affected males. *Am. J. Med. Genet.* **1994**, *51*, 509–512. [[CrossRef](#)] [[PubMed](#)]
- Basuta, K.; Schneider, A.; Gane, L.; Polussa, J.; Woodruff, B.; Pretto, D.; Hagerman, R.; Tassone, F. High functioning male with fragile X syndrome and fragile X-associated tremor/ataxia syndrome. *Am. J. Med. Genet. A* **2015**, *167*, 2154–2161. [[CrossRef](#)] [[PubMed](#)]
- Hwang, Y.T.; Aliaga, S.M.; Arpone, M.; Francis, D.; Li, X.; Chong, B.; Slater, H.R.; Rogers, C.; Bretherton, L.; Hunter, M.; et al. Partially methylated alleles, microdeletion, and tissue mosaicism in a fragile X male with tremor and ataxia at 30 years of age: A case report. *Am. J. Med. Genet. A* **2016**, *170*, 3327–3332. [[CrossRef](#)] [[PubMed](#)]
- Wohrle, D.; Salat, U.; Hameister, H.; Vogel, W.; Steinbach, P. Demethylation, reactivation, and destabilization of human fragile X full-mutation alleles in mouse embryocarcinoma cells. *Am. J. Hum. Genet.* **2001**, *69*, 504–515. [[CrossRef](#)] [[PubMed](#)]
- Gerhardt, J.; Zaninovic, N.; Zhan, Q.; Madireddy, A.; Nolin, S.L.; Ersalesi, N.; Yan, Z.; Rosenwaks, Z.; Schildkraut, C.L. Cis-acting DNA sequence at a replication origin promotes repeat expansion to fragile X full mutation. *J. Cell Biol.* **2014**, *206*, 599–607. [[CrossRef](#)] [[PubMed](#)]
- Brykczynska, U.; Pecho-Vrieseling, E.; Thiemeyer, A.; Klein, J.; Fruh, I.; Doll, T.; Manneville, C.; Fuchs, S.; Iazeolla, M.; Beibel, M.; et al. CGG repeat-Induced *FMRI* silencing depends on the expansion size in human iPSCs and neurons carrying unmethylated full mutations. *Stem Cell Rep.* **2016**, *7*, 1059–1071. [[CrossRef](#)] [[PubMed](#)]
- Bontekoe, C.J.; de Graaff, E.; Nieuwenhuizen, I.M.; Willemsen, R.; Oostra, B.A. *FMRI* premutation allele (CGG)81 is stable in mice. *Eur. J. Hum. Genet.* **1997**, *5*, 293–298. [[PubMed](#)]
- Lavedan, C.; Grabczyk, E.; Usdin, K.; Nussbaum, R.L. Long uninterrupted CGG repeats within the first exon of the human *FMRI* gene are not intrinsically unstable in transgenic mice. *Genomics* **1998**, *50*, 229–240. [[CrossRef](#)] [[PubMed](#)]
- Willemsen, R.; Hoogeveen-Westerveld, M.; Reis, S.; Holstege, J.; Severijnen, L.A.; Nieuwenhuizen, I.M.; Schrier, M.; van Unen, L.; Tassone, F.; Hoogeveen, A.T.; et al. The *FMRI* CGG repeat mouse displays ubiquitin-positive intranuclear neuronal inclusions; implications for the cerebellar tremor/ataxia syndrome. *Hum. Mol. Genet.* **2003**, *12*, 949–959. [[CrossRef](#)] [[PubMed](#)]

14. Entezam, A.; Biacsi, R.; Orrison, B.; Saha, T.; Hoffman, G.E.; Grabczyk, E.; Nussbaum, R.L.; Usdin, K. Regional FMRP deficits and large repeat expansions into the full mutation range in a new fragile X premutation mouse model. *Gene* **2007**, *395*, 125–134. [[CrossRef](#)] [[PubMed](#)]
15. Glaser, D.; Wohrle, D.; Salat, U.; Vogel, W.; Steinbach, P. Mitotic behavior of expanded CGG repeats studied on cultured cells: further evidence for methylation-mediated triplet repeat stability in fragile X syndrome. *Am. J. Med. Genet.* **1999**, *84*, 226–228. [[CrossRef](#)]
16. Lokanga, R.; Zhao, X.N.; Entezam, A.; Usdin, K. X inactivation plays a major role in the gender bias in somatic expansion in a mouse model of the fragile X-related disorders: implications for the mechanism of repeat expansion. *Hum. Mol. Genet.* **2014**, *23*, 4985–4994. [[CrossRef](#)] [[PubMed](#)]
17. Zhao, X.N.; Usdin, K. Timing of expansion of fragile X premutation alleles during intergenerational transmission in a mouse model of the fragile X-related disorders. *Front. Genet.* **2018**, *9*, 314. [[CrossRef](#)] [[PubMed](#)]
18. Wang, N.; Satirapod, C.; Ohguchi, Y.; Park, E.S.; Woods, D.C.; Tilly, J.L. Genetic studies in mice directly link oocytes produced during adulthood to ovarian function and natural fertility. *Sci. Rep.* **2017**, *7*, 10011. [[CrossRef](#)] [[PubMed](#)]
19. Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium. Identification of genetic factors that modify clinical onset of Huntington's disease. *Cell* **2015**, *162*, 516–526. [[CrossRef](#)]
20. Bettencourt, C.; Hensman-Moss, D.; Flower, M.; Wiethoff, S.; Brice, A.; Goizet, C.; Stevanin, G.; Koutsis, G.; Karadima, G.; Panas, M.; et al. DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. *Ann. Neurol.* **2016**, *79*, 983–990. [[CrossRef](#)] [[PubMed](#)]
21. Hensman Moss, D.J.; Pardinias, A.F.; Langbehn, D.; Lo, K.; Leavitt, B.R.; Roos, R.; Durr, A.; Mead, S.; TRACK-HD Investigators; REGISTRY Investigators; et al. Identification of genetic variants associated with Huntington's disease progression: A genome-wide association study. *Lancet Neurol.* **2017**, *16*, 701–711. [[CrossRef](#)]
22. Zhao, X.N.; Usdin, K. The Repeat Expansion Diseases: The dark side of DNA repair. *DNA Repair (Amst.)* **2015**, *32*, 96–105. [[CrossRef](#)] [[PubMed](#)]
23. Zhao, X.N.; Usdin, K. Ups and downs: mechanisms of repeat instability in the fragile X-related disorders. *Genes* **2016**, *7*, 70. [[CrossRef](#)] [[PubMed](#)]
24. Zhao, X.N.; Kumari, D.; Gupta, S.; Wu, D.; Evanitsky, M.; Yang, W.; Usdin, K. Mutsbeta generates both expansions and contractions in a mouse model of the Fragile X-associated disorders. *Hum. Mol. Genet.* **2015**, *24*, 7087–7096. [[CrossRef](#)] [[PubMed](#)]
25. Zhao, X.N.; Lokanga, R.; Allette, K.; Gazy, I.; Wu, D.; Usdin, K. A MutSbeta-dependent contribution of MutSalpha to repeat expansions in fragile X premutation mice? *PLoS Genet.* **2016**, *12*, e1006190. [[CrossRef](#)] [[PubMed](#)]
26. Lokanga, R.A.; Entezam, A.; Kumari, D.; Yudkin, D.; Qin, M.; Smith, C.B.; Usdin, K. Somatic expansion in mouse and human carriers of fragile X premutation alleles. *Hum. Mutat.* **2013**, *34*, 157–166. [[CrossRef](#)] [[PubMed](#)]
27. Gazy, I.; Hayward, B.; Potapova, S.; Zhao, X.; Usdin, K. Double-strand break repair plays a role in repeat instability in a fragile X mouse model. *DNA Repair (Amst.)* **2019**, *74*, 63–69. [[CrossRef](#)] [[PubMed](#)]
28. Lokanga, R.A.; Senejani, A.G.; Sweasy, J.B.; Usdin, K. Heterozygosity for a hypomorphic Polbeta mutation reduces the expansion frequency in a mouse model of the Fragile X-related disorders. *PLoS Genet.* **2015**, *11*, e1005181. [[CrossRef](#)] [[PubMed](#)]
29. Mollersen, L.; Rowe, A.D.; Larsen, E.; Rognes, T.; Klungland, A. Continuous and periodic expansion of CAG repeats in Huntington's disease R6/1 mice. *PLoS Genet.* **2010**, *6*, e1001242. [[CrossRef](#)] [[PubMed](#)]
30. Zhao, X.; Zhang, Y.; Wilkins, K.; Edelman, W.; Usdin, K. MutLgamma promotes repeat expansion in a Fragile X mouse model while EXO1 is protective. *PLoS Genet.* **2018**, *14*, e1007719. [[CrossRef](#)] [[PubMed](#)]
31. Hayward, B.E.; Zhou, Y.; Kumari, D.; Usdin, K. A set of assays for the comprehensive analysis of FMR1 alleles in the fragile X-related disorders. *J. Mol. Diagn.* **2016**, *18*, 762–774. [[CrossRef](#)] [[PubMed](#)]
32. Dobkin, C.S.; Nolin, S.L.; Cohen, I.; Sudhalter, V.; Bialer, M.G.; Ding, X.H.; Jenkins, E.C.; Zhong, N.; Brown, W.T. Tissue differences in fragile X mosaics: Mosaicism in blood cells may differ greatly from skin. *Am. J. Med. Genet.* **1996**, *64*, 296–301. [[CrossRef](#)]
33. Maddalena, A.; Yadavish, K.N.; Spence, W.C.; Howard-Peebles, P.N. A fragile X mosaic male with a cryptic full mutation detected in epithelium but not in blood. *Am. J. Med. Genet.* **1996**, *64*, 309–312. [[CrossRef](#)]

34. Taylor, A.K.; Tassone, F.; Dyer, P.N.; Hersch, S.M.; Harris, J.B.; Greenough, W.T.; Hagerman, R.J. Tissue heterogeneity of the *FMRI* mutation in a high-functioning male with fragile X syndrome. *Am. J. Med. Genet.* **1999**, *84*, 233–239. [[CrossRef](#)]
35. MacKenzie, J.J.; Sumargo, I.; Taylor, S.A. A cryptic full mutation in a male with a classical fragile X phenotype. *Clin. Genet.* **2006**, *70*, 39–42. [[CrossRef](#)] [[PubMed](#)]
36. Pretto, D.I.; Mendoza-Morales, G.; Lo, J.; Cao, R.; Hadd, A.; Latham, G.J.; Durbin-Johnson, B.; Hagerman, R.; Tassone, F. CGG allele size somatic mosaicism and methylation in *FMRI* premutation alleles. *J. Med. Genet.* **2014**, *51*, 309–318. [[CrossRef](#)] [[PubMed](#)]
37. Jiraanont, P.; Kumar, M.; Tang, H.T.; Espinal, G.; Hagerman, P.J.; Hagerman, R.J.; Chutabhakdikul, N.; Tassone, F. Size and methylation mosaicism in males with fragile X syndrome. *Expert Rev. Mol. Diagn.* **2017**, *17*, 1023–1032. [[CrossRef](#)] [[PubMed](#)]
38. Fernandez, E.; Gennaro, E.; Pirozzi, F.; Baldo, C.; Forzano, F.; Turolla, L.; Faravelli, F.; Gastaldo, D.; Coviello, D.; Grasso, M.; et al. FXS-like phenotype in two unrelated patients carrying a methylated premutation of the *FMRI* gene. *Front. Genet.* **2018**, *9*, 442. [[CrossRef](#)] [[PubMed](#)]
39. Mailick, M.R.; Movaghar, A.; Hong, J.; Greenberg, J.S.; DaWalt, L.S.; Zhou, L.; Jackson, J.; Rathouz, P.J.; Baker, M.W.; Brilliant, M.; et al. Health profiles of mosaic versus non-mosaic *FMRI* premutation carrier mothers of children with fragile X syndrome. *Front. Genet.* **2018**, *9*, 173. [[CrossRef](#)] [[PubMed](#)]
40. Nolin, S.L.; Houck, G.E., Jr.; Gargano, A.D.; Blumstein, H.; Dobkin, C.S.; Brown, W.T. *FMRI* CGG-repeat instability in single sperm and lymphocytes of fragile-X premutation males. *Am. J. Hum. Genet.* **1999**, *65*, 680–688. [[CrossRef](#)] [[PubMed](#)]
41. Alvarez-Mora, M.I.; Guitart, M.; Rodriguez-Revenga, L.; Madrigal, I.; Gabau, E.; Mila, M. Paternal transmission of a *FMRI* full mutation allele. *Am. J. Med. Genet. A* **2017**, *173*, 2795–2797. [[CrossRef](#)] [[PubMed](#)]
42. Lokanga, R.A.; Zhao, X.N.; Usdin, K. The mismatch repair protein MSH2 is rate limiting for repeat expansion in a fragile X premutation mouse model. *Hum. Mutat.* **2014**, *35*, 129–136. [[CrossRef](#)] [[PubMed](#)]
43. Reyes-Quizoz, M.E.; Jesus, S.; Ramos, I.; Garcia, A.E.; Martinez, R.; Mir, P.; Pintado, E. Tissue-specific size and methylation analysis in two fragile X families: Contribution to the clinical phenotype. *J. Mol. Genet. Med.* **2016**, *10*. [[CrossRef](#)]
44. Zhao, X.N.; Usdin, K. Gender and cell-type-specific effects of the transcription-coupled repair protein, ERCC6/CSB, on repeat expansion in a mouse model of the fragile X-related disorders. *Hum. Mutat.* **2014**, *35*, 341–349. [[CrossRef](#)] [[PubMed](#)]
45. Sullivan, A.K.; Crawford, D.C.; Scott, E.H.; Leslie, M.L.; Sherman, S.L. Paternally transmitted *FMRI* alleles are less stable than maternally transmitted alleles in the common and intermediate size range. *Am. J. Hum. Genet.* **2002**, *70*, 1532–1544. [[CrossRef](#)] [[PubMed](#)]
46. Fry, M.; Loeb, L.A. The fragile X syndrome d(CGG)n nucleotide repeats form a stable tetrahelical structure. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 4950–4954. [[CrossRef](#)] [[PubMed](#)]
47. Mitas, M.; Yu, A.; Dill, J.; Haworth, I.S. The trinucleotide repeat sequence d(CGG)15 forms a heat-stable hairpin containing Gsyn. Ganti base pairs. *Biochemistry* **1995**, *34*, 12803–12811. [[CrossRef](#)] [[PubMed](#)]
48. Nadel, Y.; Weisman-Shomer, P.; Fry, M. The fragile X syndrome single strand d(CGG)n nucleotide repeats readily fold back to form unimolecular hairpin structures. *J. Biol. Chem.* **1995**, *270*, 28970–28977. [[CrossRef](#)] [[PubMed](#)]
49. Usdin, K.; Woodford, K.J. CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis in vitro. *Nucleic Acids Res.* **1995**, *23*, 4202–4209. [[CrossRef](#)] [[PubMed](#)]
50. Yu, A.; Barron, M.D.; Romero, R.M.; Christy, M.; Gold, B.; Dai, J.; Gray, D.M.; Haworth, I.S.; Mitas, M. At physiological pH, d(CCG)15 forms a hairpin containing protonated cytosines and a distorted helix. *Biochemistry* **1997**, *36*, 3687–3699. [[CrossRef](#)] [[PubMed](#)]
51. Voineagu, I.; Surka, C.F.; Shishkin, A.A.; Krasilnikova, M.M.; Mirkin, S.M. Replisome stalling and stabilization at CGG repeats, which are responsible for chromosomal fragility. *Nat. Struct. Mol. Biol.* **2009**, *16*, 226–228. [[CrossRef](#)] [[PubMed](#)]
52. Yudkin, D.; Hayward, B.E.; Aladjem, M.I.; Kumari, D.; Usdin, K. Chromosome fragility and the abnormal replication of the *FMRI* locus in fragile X syndrome. *Hum. Mol. Genet.* **2014**, *23*, 2940–2952. [[CrossRef](#)] [[PubMed](#)]
53. Zhou, Y.; Kumari, D.; Sciascia, N.; Usdin, K. CGG-repeat dynamics and *FMRI* gene silencing in fragile X syndrome stem cells and stem cell-derived neurons. *Mol. Autism* **2016**, *7*, 42. [[CrossRef](#)] [[PubMed](#)]

54. Drost, J.B.; Lee, W.R. Biological basis of germline mutation: comparisons of spontaneous germline mutation rates among drosophila, mouse, and human. *Environ. Mol. Mutagen.* **1995**, *25* (Suppl. 26), 48–64. [[CrossRef](#)]
55. Crow, J.F. The origins, patterns and implications of human spontaneous mutation. *Nat Rev Genet* **2000**, *1*, 40–47. [[CrossRef](#)] [[PubMed](#)]
56. Malter, H.E.; Iber, J.C.; Willemsen, R.; de Graaff, E.; Tarleton, J.C.; Leisti, J.; Warren, S.T.; Oostra, B.A. Characterization of the full fragile X syndrome mutation in fetal gametes. *Nat. Genet.* **1997**, *15*, 165–169. [[CrossRef](#)] [[PubMed](#)]
57. Reyniers, E.; Vits, L.; De Boulle, K.; Van Roy, B.; Van Velzen, D.; de Graaff, E.; Verkerk, A.J.; Jorens, H.Z.; Darby, J.K.; Oostra, B.; et al. The full mutation in the *FMR-1* gene of male fragile X patients is absent in their sperm. *Nat. Genet.* **1993**, *4*, 143–146. [[CrossRef](#)] [[PubMed](#)]
58. Rousseau, F.; Robb, L.J.; Rouillard, P.; Der Kaloustian, V.M. No mental retardation in a man with 40% abnormal methylation at the *FMR-1* locus and transmission of sperm cell mutations as premutations. *Hum. Mol. Genet.* **1994**, *3*, 927–930. [[CrossRef](#)] [[PubMed](#)]
59. Entezam, A.; Lokanga, A.R.; Le, W.; Hoffman, G.; Usdin, K. Potassium bromate, a potent DNA oxidizing agent, exacerbates germline repeat expansion in a fragile X premutation mouse model. *Hum. Mutat.* **2010**, *31*, 611–616. [[CrossRef](#)] [[PubMed](#)]
60. Zhao, X.N.; Usdin, K. FAN1 protects against repeat expansions in a Fragile X mouse model. *DNA Repair. (Amst.)* **2018**, *69*, 1–5. [[CrossRef](#)] [[PubMed](#)]
61. Chen, X.; Mariappan, S.V.; Catasti, P.; Ratliff, R.; Moyzis, R.K.; Laayoun, A.; Smith, S.S.; Bradbury, E.M.; Gupta, G. Hairpins are formed by the single DNA strands of the fragile X triplet repeats: structure and biological implications. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 5199–5203. [[CrossRef](#)] [[PubMed](#)]
62. Mariappan, S.V.; Catasti, P.; Chen, X.; Ratliff, R.; Moyzis, R.K.; Bradbury, E.M.; Gupta, G. Solution structures of the individual single strands of the fragile X DNA triplets (GCC)<sub>n</sub>(GGC)<sub>n</sub>. *Nucleic Acids Res.* **1996**, *24*, 784–792. [[CrossRef](#)] [[PubMed](#)]
63. Zhao, X.N.; Usdin, K. The transcription-coupled repair protein ERCC6/CSB also protects against repeat expansion in a mouse model of the fragile X premutation. *Hum. Mutat.* **2015**, *36*, 482–487. [[CrossRef](#)] [[PubMed](#)]
64. Tran, H.T.; Degtyareva, N.P.; Koloteva, N.N.; Sugino, A.; Masumoto, H.; Gordenin, D.A.; Resnick, M.A. Replication slippage between distant short repeats in *Saccharomyces cerevisiae* depends on the direction of replication and the *RAD50* and *RAD52* genes. *Mol. Cell. Biol.* **1995**, *15*, 5607–5617. [[CrossRef](#)] [[PubMed](#)]
65. Hirst, M.C.; White, P.J. Cloned human *FMR1* trinucleotide repeats exhibit a length- and orientation-dependent instability suggestive of in vivo lagging strand secondary structure. *Nucleic. Acids Res.* **1998**, *26*, 2353–2358. [[CrossRef](#)] [[PubMed](#)]
66. Bichara, M.; Wagner, J.; Lambert, I.B. Mechanisms of tandem repeat instability in bacteria. *Mutat. Res.* **2006**, *598*, 144–163. [[CrossRef](#)] [[PubMed](#)]
67. Bissler, J.J. DNA inverted repeats and human disease. *Front. Biosci.* **1998**, *3*, 408–418. [[CrossRef](#)]
68. Lovett, S.T. Encoded errors: Mutations and rearrangements mediated by misalignment at repetitive DNA sequences. *Mol. Microbiol.* **2004**, *52*, 1243–1253. [[CrossRef](#)] [[PubMed](#)]
69. Kim, N.; Cho, J.E.; Li, Y.C.; Jinks-Robertson, S. RNA/DNA hybrids initiate quasi-palindrome-associated mutations in highly transcribed yeast DNA. *PLoS Genet* **2013**, *9*, e1003924. [[CrossRef](#)] [[PubMed](#)]
70. Nolin, S.L.; Sah, S.; Glicksman, A.; Sherman, S.L.; Allen, E.; Berry-Kravis, E.; Tassone, F.; Yrigollen, C.; Cronister, A.; Jodah, M.; et al. Fragile X AGG analysis provides new risk predictions for 45–69 repeat alleles. *Am. J. Med. Genet. A* **2013**, *161A*, 771–778. [[CrossRef](#)] [[PubMed](#)]
71. Martins, S.; Pearson, C.E.; Coutinho, P.; Provost, S.; Amorim, A.; Dube, M.P.; Sequeiros, J.; Rouleau, G.A. Modifiers of (CAG)<sub>n</sub> instability in Machado-Joseph disease (MJD/SCA3) transmissions: an association study with DNA replication, repair and recombination genes. *Hum. Genet.* **2014**, *133*, 1311–1318. [[CrossRef](#)] [[PubMed](#)]
72. Morales, F.; Vasquez, M.; Santamaria, C.; Cuenca, P.; Corrales, E.; Monckton, D.G. A polymorphism in the *MSH3* mismatch repair gene is associated with the levels of somatic instability of the expanded CTG repeat in the blood DNA of myotonic dystrophy type 1 patients. *DNA Repair (Amst.)* **2016**, *40*, 57–66. [[CrossRef](#)] [[PubMed](#)]
73. Entezam, A.; Usdin, K. ATR protects the genome against CGG.CCG-repeat expansion in fragile X premutation mice. *Nucleic Acids Res.* **2008**, *36*, 1050–1056. [[CrossRef](#)] [[PubMed](#)]

74. Entezam, A.; Usdin, K. ATM and ATR protect the genome against two different types of tandem repeat instability in fragile X premutation mice. *Nucleic Acids Res.* **2009**, *37*, 6371–6377. [[CrossRef](#)] [[PubMed](#)]
75. Wheeler, V.C.; Lebel, L.A.; Vrbancac, V.; Teed, A.; te Riele, H.; MacDonald, M.E. Mismatch repair gene *Msh2* modifies the timing of early disease in Hdh(Q111) striatum. *Hum. Mol. Genet.* **2003**, *12*, 273–281. [[CrossRef](#)] [[PubMed](#)]
76. Morales, F.; Couto, J.M.; Higham, C.F.; Hogg, G.; Cuenca, P.; Braidia, C.; Wilson, R.H.; Adam, B.; del Valle, G.; Brian, R.; et al. Somatic instability of the expanded CTG triplet repeat in myotonic dystrophy type 1 is a heritable quantitative trait and modifier of disease severity. *Hum. Mol. Genet.* **2012**, *21*, 3558–3567. [[CrossRef](#)] [[PubMed](#)]
77. Budworth, H.; Harris, F.R.; Williams, P.; Lee, D.Y.; Holt, A.; Pahnke, J.; Szczesny, B.; Acevedo-Torres, K.; Ayala-Pena, S.; McMurray, C.T. Suppression of somatic expansion delays the onset of pathophysiology in a mouse model of Huntington's disease. *PLoS Genet.* **2015**, *11*, e1005267. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



## OPEN Both cis and trans-acting genetic factors drive somatic instability in female carriers of the *FMR1* premutation

Ye Hyun Hwang<sup>1,6</sup>, Bruce Eliot Hayward<sup>2,6</sup>, Marwa Zafarullah<sup>1</sup>, Jay Kumar<sup>1</sup>, Blythe Durbin Johnson<sup>3</sup>, Peter Holmans<sup>4</sup>, Karen Usdin<sup>2,5</sup> & Flora Tassone<sup>1,5</sup>✉

The fragile X mental retardation (*FMR1*) gene contains an expansion-prone CGG repeat within its 5' UTR. Alleles with 55–200 repeats are known as premutation (PM) alleles and confer risk for one or more of the *FMR1* premutation (PM) disorders that include Fragile X-associated Tremor/Ataxia Syndrome (FXTAS), Fragile X-associated Primary Ovarian Insufficiency (FXPOI), and Fragile X-Associated Neuropsychiatric Disorders (FXAND). PM alleles expand on intergenerational transmission, with the children of PM mothers being at risk of inheriting alleles with > 200 CGG repeats (full mutation FM) alleles and thus developing Fragile X Syndrome (FXS). PM alleles can be somatically unstable. This can lead to individuals being mosaic for multiple size alleles. Here, we describe a detailed evaluation of somatic mosaicism in a large cohort of female PM carriers and show that 94% display some evidence of somatic instability with the presence of a series of expanded alleles that differ from the next allele by a single repeat unit. Using two different metrics for instability that we have developed, we show that, as with intergenerational instability, there is a direct relationship between the extent of somatic expansion and the number of CGG repeats in the originally inherited allele and an inverse relationship with the number of AGG interruptions. Expansions are progressive as evidenced by a positive correlation with age and by examination of blood samples from the same individual taken at different time points. Our data also suggests the existence of other genetic or environmental factors that affect the extent of somatic expansion. Importantly, the analysis of candidate single nucleotide polymorphisms (SNPs) suggests that two DNA repair factors, *FAN1* and *MSH3*, may be modifiers of somatic expansion risk in the PM population as observed in other repeat expansion disorders.

Over 35 different inherited genetic disorders are caused by the expansion of a specific short tandem repeat tract<sup>1</sup>. In these repeat expansion disorders, the repeat is unstable showing a strong expansion bias. The *FMR1* disorders or Fragile X-related disorders (FXDs), are members of this group that result from the presence of an evolutionarily conserved, but expansion-prone, CGG repeat tract at the 5' end of the transcriptional unit of the X-linked *FMR1* gene. The repeats are situated upstream of the open reading frame for FMRP, an RNA binding protein important for the regulation of translation in post-synaptic neurons in response to synaptic activation. The repeats are thought to modulate mGluR-dependent enhancement of FMRP synthesis via non-AUG initiated (RAN) translation through the repeat tract<sup>2</sup>. Premutation alleles (PM) have 55–200 repeats and are associated with a risk of developing one or more of the PM associated disorders, Fragile X-associated tremor/ataxia syndrome (FXTAS), Fragile X-associated primary ovarian insufficiency (FXPOI), and Fragile X-associated neuropsychiatric disorders (FXAND)<sup>3–6</sup>. Pathology is thought to arise from some deleterious effect of the excess

<sup>1</sup>Department of Biochemistry and Molecular Medicine, University of California Davis, School of Medicine, Sacramento, CA 95817, USA. <sup>2</sup>Laboratory of Molecular and Cellular Biology, National Institute of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA. <sup>3</sup>Department of Public Health Sciences, University of California, Davis, School of Medicine, Sacramento, CA 95817, USA. <sup>4</sup>Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurology, School of Medicine, Cardiff University, Cardiff, UK. <sup>5</sup>MIND Institute, University of California Davis Medical Center, Sacramento, CA 95817, USA. <sup>6</sup>These authors contributed equally: Ye Hyun Hwang and Bruce Eliot Hayward. ✉email: karenu@nih.gov; ftassone@ucdavis.edu

number of repeats in the *FMR1* transcript<sup>7</sup>. Carriers of PM alleles are also at risk of transmitting larger alleles to their children, with increasing CGG repeat number being associated with increased risk<sup>8</sup>. In particular, female PM carriers with ~90 CGG repeats, have a >90% probability of transmitting alleles with >200 CGG repeats to their children. Such alleles are known as full mutation (FM) alleles and result in Fragile X syndrome (FXS), a neurodevelopmental disorder that is the most common inherited form of intellectual disability and the most common monogenic cause of autism spectrum disorder. Pathology in this instance is thought to be related to the repeat-mediated silencing of the *FMR1* promoter<sup>9</sup>. The prevalence of the PM allele among the general population is 1:110–200 females and 1:430 males. However, the PM disorders have a variable penetrance with 40–75% of males and 8–16% of females developing FXTAS<sup>10,11</sup> and ~20% of females developing FXPOI<sup>12,13</sup>.

The increased use of higher resolution techniques for the analysis of PM alleles has demonstrated that some carriers of PM alleles show somatic repeat size mosaicism, i.e., the presence of two or more alleles of different sizes in a particular tissue. Previous studies of mosaicism have focused on individuals containing a combination of multiple discrete alleles often in both the PM and FM range<sup>14–23</sup>. The origin of the smaller alleles is uncertain, but likely reflects contractions of larger alleles. The second type of mosaicism is also present in PM carriers, in which multiple alleles differing by a single repeat are seen in some individuals<sup>24</sup>. This form of mosaicism is reminiscent of the products of somatic expansion seen in an FXD mouse model and in humans with other repeat expansion diseases<sup>25</sup>. Molecular modeling of these products suggests that they arise via small but frequent events that accumulate over the lifetime of the individual<sup>26</sup>. In an FXD mouse model, the frequency with which these events occur differs between tissues and cell types. While this phenomenon has not been extensively examined in the *FMR1* disorders, it has been reported to occur in humans for other repeat expansion diseases such as Huntington's Disease (HD) and Myotonic Dystrophy type 1 (DM1)<sup>27–29</sup>. The extent of this somatic expansion has been shown to be affected by repeat length and purity as well as a variety of genetic factors with the extent of expansion affecting the age of onset and severity of many of these diseases<sup>29–39</sup>. This study represents the first study of the somatic instability of the *FMR1* repeat in a large cohort of female PM carriers.

## Materials and methods

**Study.** Peripheral blood was collected from a total of 426 PM female participants after signing an informed consent form and using a protocol approved by the UC Davis Institutional Review Board.

For the analysis of the correlation of a subset of molecular measures, data from the entire cohort of 426 females were used. For the analysis of the correlation between instability and molecular measures, data from a subset consisting of 384 participants was used. Some individuals were excluded from this subset because the quality of the capillary electrophoresis trace was too poor to allow calculation of instability ( $n = 19$ ), no AR value was available ( $n = 1$ ), or the allele corresponding to that on the inactive X could not be identified ( $n = 8$ ). Individuals with an activation ratio (AR, defined as the percentage of cells carrying the normal allele on the active X chromosome) of >0.8 ( $n = 14$ ) who showed no evidence of expansion were also excluded since in these individuals the proportion of alleles able to expand would be relatively small and thus any expansion, should it occur, would be difficult to detect.

For the study of changes in pre-mutation allele stability over time, a subset of 24 female PM participants was selected, based on the availability of at least two blood draws taken a minimum of 2 years apart (mean 6.7; SD 2.9). The age mean was 46.7 (SD 19.5); the mean of the CGG repeats (based on the draw at the first visit) was 100.1 (SD 27.2) (Supplementary Table 1).

**CGG sizing, methylation status, AGG interruptions, and SNP selection.** Genomic DNA (gDNA) was isolated from 3 ml of peripheral blood by using the Gentra Puregene Blood Kit (Qiagen, Valencia, CA, United States). CGG repeat allele size and methylation status were assessed using a combination of PCR and Southern Blot analysis. A PCR that specifically targeted *FMR1* amplification (AmplideX PCR/CE, Asuragen, Inc.) was used to determine CGG repeat length and PCR products were visualized by CE and analyzed as previously reported<sup>40</sup>. Southern blotting was performed using the Stb12.3 *FMR1* specific chemiluminescent intronic probe, as detailed in Ref.<sup>41</sup>. Briefly, 10 µg of isolated gDNA was digested with *EcoRI* and *NruI*, run on an agarose gel, transferred to a nylon membrane, and hybridized with the *FMR1*-specific dig-labeled Stb12.3. Southern Blot analysis was also used to determine the methylation status of the *FMR1* alleles (Activation ratio, AR, and the percent of methylation) as previously described<sup>42</sup>.

To visualize the methylation status of alleles by capillary electrophoresis a modified version of the assay described in Ref.<sup>43</sup> was employed. Briefly, 600 ng of genomic DNA was placed in a 40 µl volume of 50 mM Tris-HCl pH 9.0, 1.75 mM MgCl<sub>2</sub>, 22 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, and 1 µl of *HindIII* restriction enzyme were added. This was divided into two equal aliquots and 0.5 µl of *HpaII* restriction enzyme was added to one. Digestion was allowed to proceed overnight at 37 °C. 5 µl of each digest was then made to 20 µl containing 50 mM Tris-HCl pH 9.0, 1.75 mM MgCl<sub>2</sub>, 22 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 2.5 M betaine, 2% DMSO, 0.5 µM each primer, 0.2 mM dATP and dTTP, 0.475 mM dCTP and dGTP, and 0.75U of KAPA2G Robust HotStart polymerase. The PCR conditions were 98 °C for 3 min, 32 cycles of 98 °C for 30 s, 65 °C for 30 s, and 72 °C for 210 s, followed by 72 °C for 10 min. The primers used are:

Not FraxC: AGTTCAGCGGCCGCGCTCAGCTCCGTTTCGGTTTCACTTCGGGT.

Not FraxR4: FAM-CAAGTCGCGGCCGCTTGTAGAAAGCGCCATTGGAGCCCCGCA.

The number of AGG interruptions was determined by using a triplet primed PCR protocol as described in Ref.<sup>8</sup>, visualized by CE, and analyzed with Gene Mapper software. The number of AGG interruptions in a sample was determined based on the number of sharp depressions visualized by capillary electrophoresis (CE) images<sup>8</sup>.

A total of ten single nucleotide polymorphisms (SNPs) were investigated in a subset of 384 PM female participants for whom the extent of somatic instability could be reliably determined. The choice of SNPs was based



on their significant association with instability in other trinucleotide repeat expansion disorders<sup>38</sup>. SNP analysis was performed using the Taqman Single Nucleotide Polymorphism Allele Discrimination Assay for sample genotyping (Applied Biosystems, Inc., Foster City, CA). Predesigned TaqMan assays were used for genotyping. Briefly, probes were mixed with TaqMan Master Mix in a ratio of 2.5 TaqMan Master Mix to 0.125  $\mu$ l of SNP probe per well, and aliquoted into plates containing 50–100 ng of genomic DNA. A visualization of the cluster plots was performed for each plate to ensure the absence of poor clustering of the SNP. Internal positive and negative controls with all the known genotypes for each SNP were included in each plate. Genotypes were determined using Applied Biosystems automated Taqman genotyping software, SDS v2.1. Genotype data were blind for statistical analysis.

***FMR1* mRNA expression levels.** Total RNA was isolated from 2.5 ml of peripheral blood collected in PAXgene Blood RNA tubes using the PAXgene Blood RNA Kit (Qiagen, Valencia, CA, United States) and quantified using the Agilent 2100 Bioanalyzer system. RNA isolation was performed in a clean and RNA-designated area. cDNA was synthesized as previously described<sup>44</sup>. *FMR1* transcript levels were measured by performing reverse transcription followed by real-time PCRs (qRT-PCR). qRT-PCR was performed using both Assays-On-Demand from Applied Biosystems (Applied Biosystems, Foster City, CA, United States) and custom-designed TaqMan primers and probe assays<sup>44</sup>.

**Measurement of instability.** Two different metrics for the extent degree of expansion were used. Since the expansion is limited to the active X chromosome, the smaller alleles represented by Peak 1 represent the originally inherited allele. Our primary measure of expansion,  $\Delta R_{pts}$ , is the difference in the number of repeats in a repeat profile between the modal expanded allele (Peak 2) and modal stable allele (Peak 1). Since in males X inactivation does not occur, we adapted a second metric from Ref.<sup>26</sup> which is based on the increase in the dispersion of the allele populations in the PCR profile. This was calculated by first identifying the modal peaks of the stable (Peak 1) and unstable (Peak 2) allele populations. The RFU values of the peaks exceeding a threshold value ( $\geq 0.2 \times$  RFU of modal peak) in each population were then converted into a histogram which was treated as being derived from a normal distribution and the standard deviation of that distribution became the dispersion (D) value. To minimize the contribution of alleles in Peak 1 to the dispersion of Peak 2 (D2) and vice versa, we determined the dispersion metric of Peak 2 (D2) by using only Peak 2 and peaks lying to the right of it. Similarly, the dispersion of Peak 1 (D1) was calculated by using only Peak 1 and peaks lying to the left of it.

To determine the proportion of alleles that expand, both the area under the stable peaks in a PCR profile (StableArea) and the area under the curve of the unstable peaks (UnstableArea) were calculated. The proportion of alleles that expand (AUC2) is given by  $UnstableArea / (UnstableArea + StableArea)$  and the proportion of alleles that are stable (AUC1) is then  $1 - AUC2$ .

**Statistical analysis.** Statistical analysis was used to determine the correlation between the *FMR1* molecular measures, instability, age, CGG repeat size, AGG interruption, *FMR1* mRNA, and AR. *FMR1* mRNA expression was analyzed by CGG repeat number using linear regression, adjusting for activation ratio (AR) by including this as a covariate. The largest CGG repeat number was used for subjects with different numbers of CGG repeats reported. The above analyses were conducted in R version 4.0.5 (2021-03-31). The overall correlation of factors with instability (as measured by Peak2 – Peak1) was determined using the CORR Procedure, along with the generation of Pearson correlation coefficients. Relationships of individual factors with instability were determined through GLM Procedure. Association of repeat expansion with genetic and other risk factors was tested by negative binomial regression, using the `glm.nb()` function in R. We estimated the variance inflation factors for each variable in R using the `VIF()` function in the 'regclass' package. The VIFs ranged from 1.13 (AGG) to 2.97 (Peak1), which are comfortably below the cutoff of 5 commonly used to indicate problematic collinearity<sup>45</sup>.

## Results

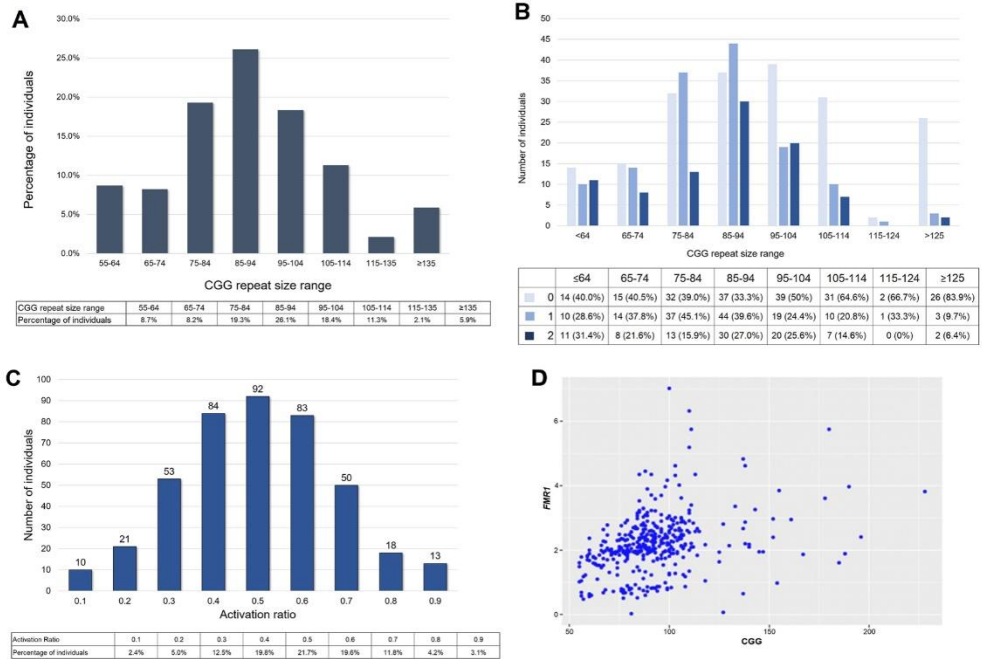
**Study participants.** Blood samples were collected from a total of 426 female PM carriers. The studies and all protocols were carried out in accordance with the Institutional Review Board at the University of California, Davis. All participants gave written informed consent before participating in the study in line with the Declaration of Helsinki. Capillary electrophoresis PCR profiles were determined for the PM alleles in everyone as previously described<sup>40</sup>. Standard practice is to report the number of repeats present in the most common allele as the individual's repeat number. The number of AGG interruptions was determined by triplet-primed PCR as previously described<sup>8</sup>. The activation ratio (AR), the fraction of normal alleles that are located on the active X chromosome was determined by Southern blot analysis<sup>42</sup>. The *FMR1* mRNA levels were determined by real-time PCR as described previously<sup>44</sup>. The ages of the participants in this study at the time their blood was drawn, their CGG repeat number, number of AGG interruptions, AR, and *FMR1* mRNA levels are shown in Table 1.

**Characterization of somatic expansion.** The CGG repeat number showed a normal distribution in our study population (Fig. 1A). The proportion of alleles with no interruptions increased from 40% for alleles with  $\leq 64$  repeats to >80% for alleles with  $\geq 125$  repeats (Fig. 1B). The AR for the study participants was also normally distributed with a mean of  $\sim 0.5$  (Fig. 1C), as previously reported<sup>46</sup>. There was no significant association of repeat size with AR. Consistent with previous reports, higher levels of *FMR1* mRNA were associated with larger repeat lengths (Fig. 1D) even after correction for AR  $p < 0.0001$ .

A variety of different repeat PCR profiles were seen. Some females showed a single sharp and asymmetric PCR profile with a small number of PCR products smaller than the modal allele (Fig. 2A). This is like the PCR profile seen in the blood of very young female PM mice or in the tissue of mice with mutations that block

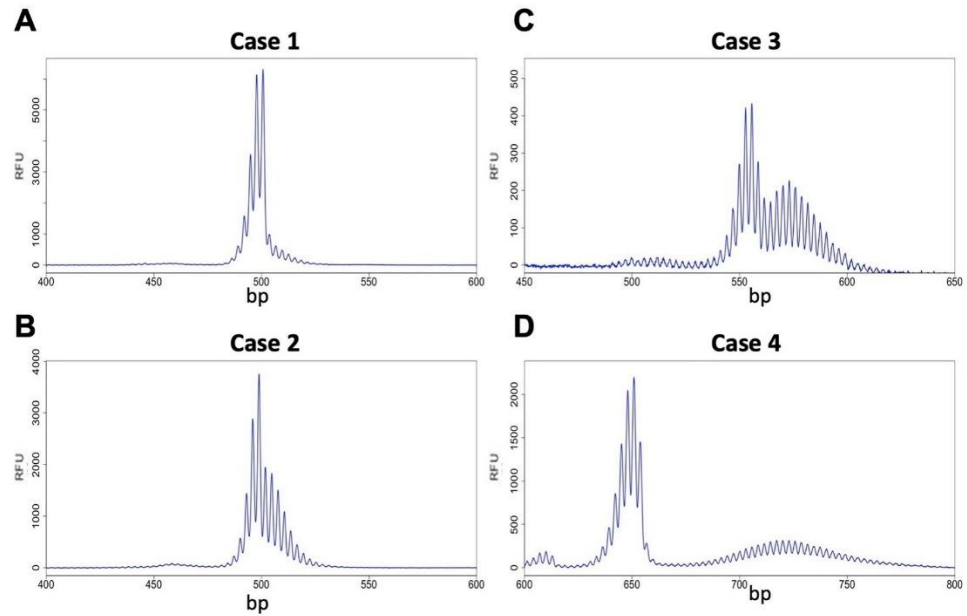
Molecular measures	Total Group, n = 426			Subset, n = 384			Unstable, n = 361	Stable, n = 23	Welch Two Sample t-test	Actual value
	n	Mean	Std. Dev	n	Mean	Std. Dev				
CGG repeat	425**	92.08	22.94	383**	90.40	19.00	91.69 ± 0.98	70.13 ± 2.65	< 0.0001	~ 3e-8
AGG	426	0.75	0.78	384	0.78	0.79	0.72 ± 0.04	1.65 ± 0.12	< 0.0001	~ 6e-8
AR	424	0.54	0.17	382	0.53	0.16	0.53 ± 0.008	0.58 ± 0.04	0.135153	
<i>FMRI</i> mRNA	401	2.18	0.91	361	2.21	0.89	2.24 ± 0.05	1.60 ± 0.18	0.001696	
Age	423	42.49	17.18	381	42.01	16.84	42.80 ± 0.87	29.76 ± 3.99	0.003882	
AUC1	413	0.71	0.23	384	0.71	0.21	0.69 ± 0.01	0.99 ± 0.007	< 0.0001	~ 8e-60
AUC2	413	0.29	0.23	384	0.29	0.21	0.31 ± 0.01	0.01 ± 0.007	< 0.0001	~ 9e-60
D1	412	1.41	0.61	384	1.36	0.32	1.40 ± 0.02	0.77 ± 0.007	< 0.0001	~ 4e-120
D2	411	2.44	2.09	383	2.50	2.01	2.66 ± 0.10	0 ± 0	< 0.0001	~ 2e-83
<b>AGG</b>										
0		197 (46.2%)*			171 (44.5%)*		170 (47.1%)*	1 (4.3%)*		
1		138 (32.4%)*			127 (33.1%)*		121 (33.5%)*	6 (26.1%)*		
2		91 (21.4%)*			86 (22.4%)*		70 (19.4%)*	16 (69.6%)*		

**Table 1.** Molecular measures of the 426 and of the subset of 384 female PM carrier groups. \*Percentage of females relative to the total number, presenting with 0, 1 or 2 AGG interruptions. \*\*Number of females for whom the CGG repeat allele size was included (one participant was removed as she was a double heterozygous- two premutation alleles).



**Figure 1.** Molecular measures of the study population (n = 426). (A) Distribution of CGG repeat numbers in the premutation allele. (B) Distribution of AGG interruptions by allele size. A number of individuals and the percentage of individuals are indicated per each category (0, 1, or 2 AGG interruptions within the premutation allele). (C) Activation Ratio (AR) distribution in the female participants (n = 424). (D) Correlation of CGG repeat number and *FMRI* mRNA level after correction for AR (n = 401, p < 0.0001).

somatic expansion<sup>47,48</sup>. As such, this PCR profile likely reflects a stable allele population with little, or no, somatic expansion, and with some, if not all, of the peaks smaller than the modal allele representing PCR “stutter”. Other individuals showed PCR profiles in which a “shoulder” was seen corresponding to alleles larger than the modal allele (Fig. 2B). The third group of women had a clear bimodal distribution of allele populations with the smaller allele population showing a narrow distribution of allele sizes and the larger allele population showing a broader distribution (Fig. 2C,D). These profiles are like those seen in older female PM mice with a genetic background



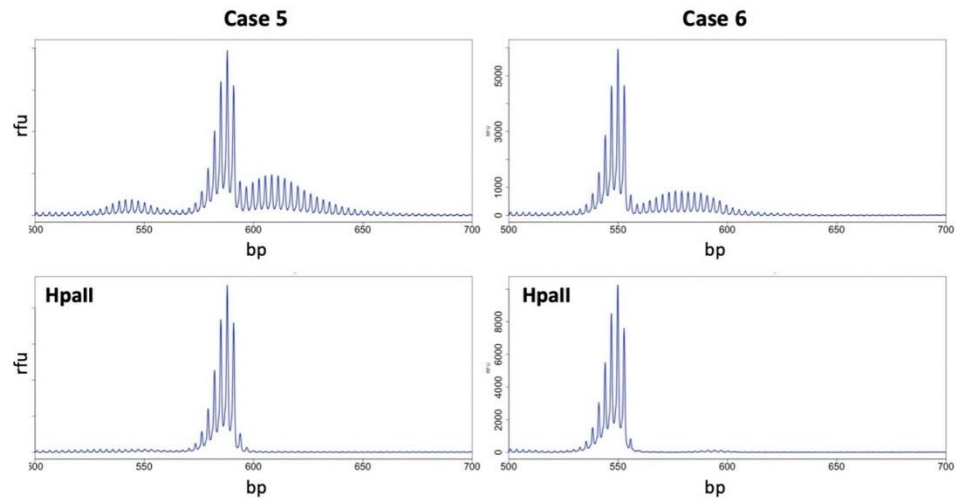
**Figure 2.** Examples of different types of PCR profiles observed in female PM carriers. Capillary electrophoretograms for 4 different females focusing on the PM allele size range showing the profiles for a stable allele (case 1) and three alleles with increasing levels of somatic expansion.

permissive to somatic expansion. In mice, the smaller of the two allele populations in older animals is similar in size to the alleles present in the tail at 3 weeks of age, an approximate measure of the number of repeats in the originally inherited allele, and the size of this population does not change over time. In contrast, the larger of the two allele populations tend to have a modal repeat number that increases with the age of the animal and thus reflects alleles that have expanded or gained repeats during the animal's lifetime<sup>49</sup>.

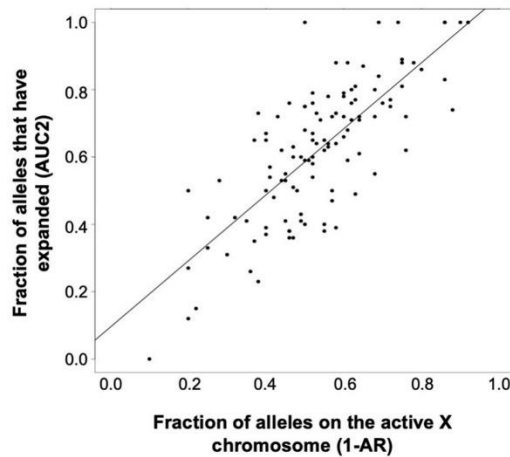
Interestingly, as in mice, *HpaII* pre-digestion of the PCR template from women with evidence of alleles larger than the modal allele eliminates such alleles from the PCR profile resulting in the production of a unimodal PCR profile characteristic of stable alleles (Fig. 3). Since *HpaII* is a methylation-sensitive enzyme with recognition sites within the amplicon used for PCR analysis of the repeat, pre-digestion eliminates any PCR template derived from an active X chromosome. Thus, the disappearance of these products after *HpaII* digestion suggests that they are derived from the active X chromosome. We interpret this to mean that these products represent expanded alleles with expansions being limited to the active X as in mice.

The association between expansion and the presence of the PM allele on the active X is supported by the fact that there is a direct relationship between the fraction of alleles that expand, as assessed by an estimation of the area under the curve of the expanded allele (AUC2) and the fraction of alleles where the PM is on the active X ( $1 - AR$ ) (Fig. 4). Thus, the allele population with the smaller repeat number corresponds to unexpanded alleles on the inactive X, with the modal repeat number likely reflecting the repeat number present on the originally inherited allele. This is consistent with our previous more limited analysis<sup>49</sup> and suggests that expansions are limited to the active X chromosome, as they are in mice<sup>47</sup>. This indicates that transcription or a euchromatin configuration is required for these expansions.

To investigate the PM allele stability over time, a subset of 24 female participants with specimens available from multiple blood draws, was selected. In 20 of the cases examined, the time between draws was < 10 years. Eight participants showed changes in CGG repeat number (1–12 CGGs; Supplementary Table 1 and Fig. 5). The remaining sixteen individuals (66.7%) showed no evidence of change in their repeat PCR profile between draws, regardless of the age at first sampling and the time between draws. Of these, 11 had < 96 CGG repeats and five had alleles > 96 CGG, with three of the alleles > 96 repeats having AGG interruptions. The other eight individuals showed evidence of a change in the PCR profile with an increase in the modal number of CGG repeats seen in the larger of the two allele populations. Seven of these individuals had inherited alleles with > 96 CGG repeats and no AGG interruptions. A female with ~144 CGG repeats in her expanded allele at the first blood draw at two years of age, showed an allele representing a gain of ~8 repeats relative to her originally inherited allele (Fig. 5A). She had alleles with a mean repeat number of ~147 CGG repeats at the second draw two years later i.e., the gain of three repeats in 2 years (Fig. 5B) shows the PCR profile of a female with ~160 CGG repeats on her expanded allele at the first blood draw at eight years of age, 19 repeats more than the original allele. At the second blood draw six years later, the expanded alleles had gained an average of an additional 11 CGG repeats. In addition, as we previously described in an FXD mouse model<sup>24</sup>, the size distribution of expanded alleles broadens with



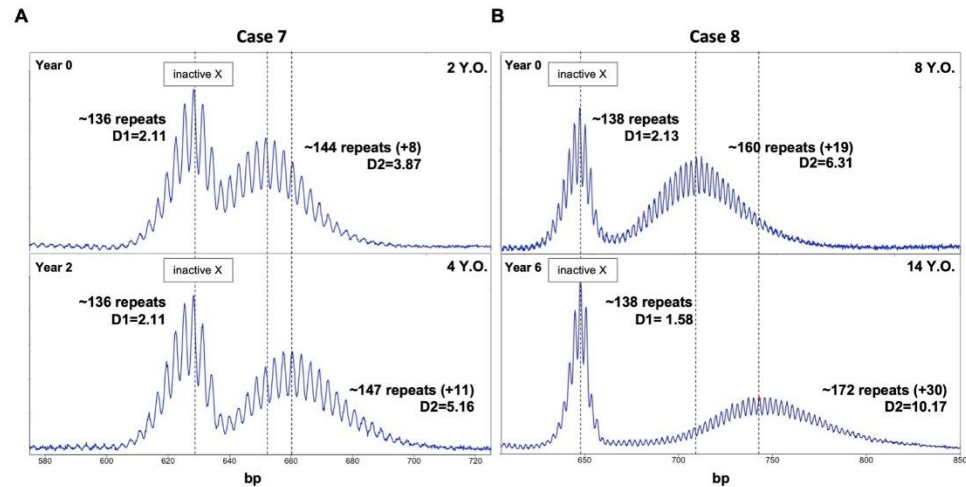
**Figure 3.** Expansion is limited to PM alleles on the active X chromosome. PCR profiles of two PM female carriers without (top panels) and with (bottom panels) *HpaII* digestion show the loss of the larger of the two allele populations on predigestion of the template with *HpaII* which preferentially eliminates alleles on the active X chromosome.



**Figure 4.** Correlation between the percentage of PM alleles on the active X chromosome and the fraction of alleles that have expanded.

age. This is consistent with mathematical modeling which suggests that each expansion event adds one-to-two repeats<sup>26</sup>. As a result, over time the dispersion of the population of expanding alleles,  $D_2$ , increases.

**Relationship between the extent of expansion, AR, AGG, age, and the dispersion of the expanded alleles..** The fact that the smaller of the two alleles corresponds to the originally inherited allele and the larger corresponds to those that have expanded would suggest that the difference in the modal number of repeats of the expanded and stable peaks, a metric we call  $\Delta R_{pts}$ , reflects the extent of somatic expansion. We used this metric to examine the relationship between the extent of expansion and AGG number, AR, and age. For this purpose, we excluded alleles with  $AR > 0.8$  that showed no evidence of expansion on the grounds that the absence of a detectable second peak might reflect expansions present at levels below the limit of detection by capillary electrophoresis, as could occur if extensive expansion had happened.



**Figure 5.** CE analysis shows the changes in the extent of somatic expansion over time in peripheral blood cells from two female PM carriers. The X-axis indicates the number of base pairs, and the Y-axis indicates relative fluorescence intensity.

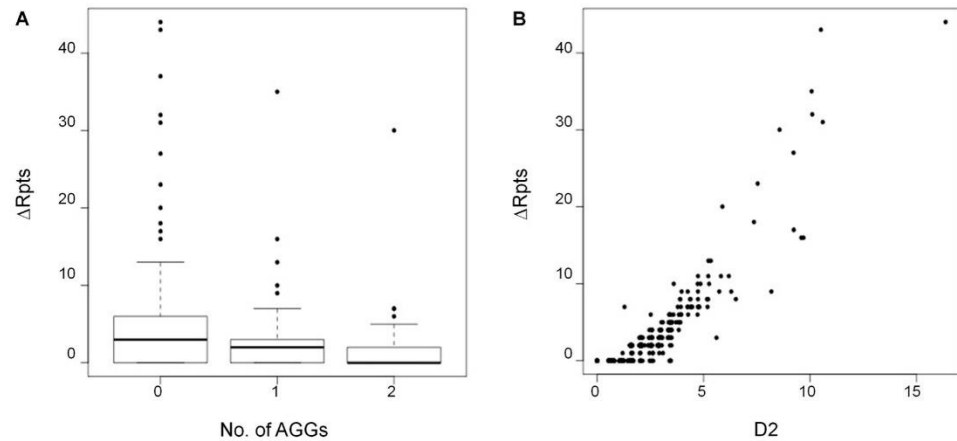
	Estimate	Std. Error	p-value
D1	-0.06	0.17	0.74
D2	0.34	0.02	<2e-16
Peak 1	0.02	0.004	1.40E-07
AGG	-0.23	0.06	0.0001
Age	0.01	0.003	1.19E-05
AR	0.49	0.28	0.08
Amount of FMRI	0.000412	0.05	0.99

**Table 2.** Relationship of  $\Delta Rpts$  to D1, D2, and other molecular measures. The values in this table refer to a multivariable negative binomial regression of  $\Delta Rpts$  on all of the molecular measures simultaneously.

In addition, we excluded poor quality capillary electrophoresis traces and individuals where the stable peak could not be identified leaving us with 384 individuals. We then used  $\Delta Rpts$  as a measure of expansion and performed negative binomial regression of this on the initial repeat number, AGG, AR, age, and the fraction of stable vs unstable alleles (represented by the area under the curve (AUC) of peak 1 and peak 2). We found a significant association between  $\Delta Rpts$  and the size of the original allele along with a significant direct relationship with age (Table 2).

There is also an inverse relationship between  $\Delta Rpts$  and the number of AGG interruptions (Fig. 6A) which is consistent with the stabilizing effect of AGGs observed on intergenerational transmission<sup>50,51</sup>. Since the dispersion about the mean of the expanding alleles increases with increasing expansion, we also tested the association of the  $\Delta Rpts$  metric with a measure of the dispersion of the stable (D1) and unstable alleles (D2). There was a significant association between the  $\Delta Rpts$  metric and D2 (Table 2). This is consistent with the data shown in (Fig. 6B) in which the heterogeneity of the expanding allele population increases with time. There was no association with D1 consistent with the fact that the size distribution of the stable allele population shows no increase over time. There was also no relationship between instability and the amount of *FMRI* transcript after correction for the initial repeat number, AGG, AR, and age.

**Genetic factors affecting the expansion.** Genome-Wide Association Studies (GWAS) have identified a number of single nucleotide polymorphisms (SNPs) that are significantly associated with the risk of somatic expansion or age of disease onset in various other Repeat Expansion Diseases<sup>29-38</sup>. To assess whether some of the same SNPs were associated with somatic expansion risk in our PM population, we examined the association of the  $\Delta Rpts$  metric with 10 single nucleotide polymorphisms (SNPs) previously found to be associated with a variation in the age of onset, disease severity or extent of somatic expansion in studies of other Repeat Expansion Diseases. Of the selected ten SNPs chosen and reported in Table 3, two, rs701383 and rs150393409, showed a significant association with the extent of instability, although neither of them would survive correction for multiple testing.



**Figure 6.** Relationship between instability ( $\Delta Rpts$ ) and the number of AGG interruptions (A) or the D2 metric (B) showing the inverse relationship between  $\Delta Rpts$  and the number of AGG interruptions and the direct relationship with D2.

SNP	Candidate modifier gene(s) (and distance in kb)	Test allele	Effect (Std. Error)	p-value	HW p-val
rs1650742	<i>MSH3</i> (0), <i>DHFR</i> (40.1)	T	0.12 (0.08)	0.12	0.37
rs1799977	<i>MLH1</i> (0)	G	-0.03 (0.08)	0.73	0.93
rs274883	<i>LIG1</i> (0)	G	-0.001 (0.09)	0.99	0.31
rs34017474	<i>FAN1</i> (0), <i>MTMR10</i> (0)	T	-0.02 (0.07)	0.78	0.8
rs35811129	<i>FAN1</i> (6.04) & <i>MTMR10</i> (0)	G	0.02 (0.08)	0.78	0.38
rs3791767	<i>PMS1</i> (8.8)	C	-0.07 (0.09)	0.4	0.06
rs701383	<i>DHFR</i> (8.77), <i>MSH3</i> (37.2)	G	-0.22 (0.08)	0.007	0.86
rs74302792	<i>PMS2</i> (31.3)	T	-0.05 (0.11)	0.61	0.81
rs145821638	<i>LIG1</i> (0)	C	-0.11 (0.53)	0.84	0.94
rs150393409	<i>FAN1</i> (0)	G	-0.70 (0.31)	0.02	0.74

**Table 3.** Correlation between SNPs associated with different repair genes and allele instability.

## Discussion

In this study, we describe the first large-scale characterization of somatic expansion in female premutation allele carriers. We show that most PM carriers show some degree of somatic expansion in blood as evidenced by their PCR profile and by the serial sampling of a subset of individuals. The extent of this expansion is related to the CGG-repeat number and inversely related to the number of AGG interruptions as with intergenerational expansions<sup>14,46,52</sup>. There was also a relationship between the extent of expansion and age, consistent with the observation of a maternal age effect on the risk of a female PM carrier having a child with an FM allele<sup>49-51</sup>. We also showed that the extent of expansion correlates with the proportion of the PM allele that is on the active X chromosome (Fig. 3). This is consistent with the fact that expansion in humans requires transcription or open chromatin as it does in mice<sup>47</sup>. While expansions were not seen on the inactive X chromosome, we observed a relationship between AR and the extent of expansion of the allele on the active X. No evidence of CGG repeat allele contractions was seen in this data set, although the occurrence of low-frequency contraction events or contraction events that generate heterogeneous deletion products cannot be definitively excluded.

The measurement of somatic expansion in females is facilitated by the fact that expansion is limited to alleles on the active X chromosome and thus that the size of the inherited allele can be inferred from the size of the allele on the inactive X. However, this is not possible in males. Our demonstration that the extent of expansion as measured by  $\Delta Rpts$  shows a direct relationship with DM2, the dispersion of the expanded allele about the mean, suggests that the DM metric could be useful for examining somatic expansion in male PM carriers.

The demonstration of the association of the rs701383 SNP with the extent of somatic expansion is of interest since this SNP has located 8.77 kb from the dihydrofolate reductase (*DHFR*) gene and 37.2 kb from *MSH3*, whose gene product is important for mismatch repair and is required for both somatic and germline expansion in the mouse model of FXDs<sup>1</sup>. rs701383 is an eQTL for *MSH3* in GTEx, that is significant in several tissues (minimum  $p = 1.5 \times 10^{-71}$  in cultured fibroblasts) with the minor allele (A) at rs701383 being associated with higher expression of *MSH3*<sup>32</sup>. rs701383 is an eQTL for *DHFR* in artery ( $p = 6.7 \times 10^{-22}$ ) and nerve ( $p = 5.9 \times 10^{-19}$ ) but the association is only weak in whole blood ( $p = 1.3 \times 10^{-8}$  compared to  $2.8 \times 10^{-63}$  for *MSH3*). The minor allele

at this SNP is associated with an earlier age at onset of HD ( $p = 5.46 \times 10^{-10}$ )<sup>38</sup> and increasing somatic instability in HD and DM1<sup>32</sup>.

The rs150393409 SNP is located within *FAN1*, a DNA repair gene that encodes a nuclease FAN1 that protects against expansion in the FXD mouse<sup>53,54</sup>. This SNP results in the substitution of Arg for His at amino acid 507 in FAN1, a change predicted to be deleterious or damaging in SIFT and PolyPhen, respectively. The directionality of the observed effect of the rs150393409 SNP would be consistent with FAN1 normally protecting against repeat expansion in women with the PM as well. Thus, although studies of larger cohorts are needed, our data suggest that genetic factors that affect somatic expansion in women with the PM are consistent with data from a mouse model of the FXDs and with other Repeat Expansion Diseases. This similarity between humans and mice with respect to the genetic factors involved in somatic expansion supports the idea that the FXD mouse model can provide useful insights into the expansion process in human PM carriers. The fact that the same SNPs are associated with disease risk in other Repeat Expansion Diseases lends weight to the idea that these diseases share a common underlying mutational mechanism.

It is notable that expansion can be readily detected in the blood of many PM human carriers. In an FXD mouse model, blood shows much less expansion than the brain<sup>48</sup>. A similar difference between the extent of expansion in blood and brain has been reported in other Repeat Expansion Diseases<sup>55–58</sup>. Thus, in PM carriers where expansion can be detected in blood, the extent of expansion in the brain maybe even larger. Since there is a direct relationship between repeat number and FXTAS age of onset<sup>39</sup>, this raises the possibility that the propensity to undergo somatic expansion could contribute to the variable penetrance of FXTAS pathology seen in PM carriers. Furthermore, since in the FXD mouse model the same genetic factors that affect expansion risk in somatic cells affect expansion in the germline, the genetic factors identified in this study as potential modifiers of somatic expansion risk, may also be modifiers of intergenerational expansion risk. These factors may account for some of the variances in expansion risk that are not explained by repeat number or the number of AGG interruptions<sup>14</sup>. Thus, a better understanding of the full range of genetic factors affecting expansion risk may contribute to better assessments of disease risk in PM carriers as well as the risk of transmission of FXS.

### Data availability

Data and results generated from this project will be fully available from corresponding author upon request. Biological samples from subjects included in this study will be available under MTA agreement accordingly to the University of California, Davis policy.

Received: 23 January 2022; Accepted: 18 May 2022

Published online: 21 June 2022

### References

- Zhao, X.-N. & Usdin, K. The repeat expansion diseases: The dark side of DNA repair. *DNA Repair* **32**, 96–105 (2015).
- Rodriguez, C. M. *et al.* A native function for RAN translation and CGG repeats in regulating fragile X protein synthesis. *Nat. Neurosci.* **23**, 386–397 (2020).
- Rajaratnam, A. *et al.* Fragile X syndrome and fragile X-associated disorders. *F1000Research* **6**, 2112 (2017).
- Hagerman, R. J. *et al.* Fragile X-associated neuropsychiatric disorders (FXAND). *Front. Psychiatry* **9**, 564 (2018).
- Leehey, M. A. *et al.* FMR1 CGG repeat length predicts motor dysfunction in premutation carriers. *Neurology* **70**, 1397–1402 (2008).
- Loesch, D. Z. *et al.* Psychological status in female carriers of premutation FMR1 allele showing a complex relationship with the size of CGG expansion. *Clin. Genet.* **87**, 173–178 (2015).
- Hagerman, R. J. & Hagerman, P. Fragile X-associated tremor/ataxia syndrome—features, mechanisms and management. *Nat. Rev. Neurol.* **12**, 403–412 (2016).
- Yrigollen, C. M., Tassone, F., Durbin-Johnson, B. & Tassone, F. The role of AGG interruptions in the transcription of FMR1 premutation alleles. *PLoS One* **6**, e21728 (2011).
- Hagerman, R. J. *et al.* Fragile X syndrome. *Nat. Rev. Dis. Primers.* **3**, 1–19 (2017).
- Tassone, F. *et al.* FMR1 CGG allele size and prevalence ascertained through newborn screening in the United States. *Genome Med.* **4**, 100 (2012).
- Jacquemont, S. Penetrance of the fragile X-Associated tremor/ataxia syndrome in a premutation carrier population. *JAMA* **291**, 460 (2004).
- Allen, E. G. *et al.* Refining the risk for fragile X-associated primary ovarian insufficiency (FXPOI) by FMR1 CGG repeat size. *Genet. Med.* **23**, 1648–1655 (2021).
- Allen, E. G. *et al.* Examination of reproductive aging milestones among women who carry the FMR1 premutation. *Hum. Reprod.* **22**, 2142–2152 (2007).
- Nolin, S. L. *et al.* Fragile X AGG analysis provides new risk predictions for 45–69 repeat alleles. *Am. J. Med. Genet. A* **161A**, 771–778 (2013).
- Pretto, D. *et al.* Clinical and molecular implications of mosaicism in FMR1 full mutations. *Front. Genet.* **5**, 318 (2014).
- Cohen, I. L. *et al.* Mosaicism for the FMR1 gene influences adaptive skills development in fragile X-affected males. *Am. J. Med. Genet.* **64**, 365–369 (1996).
- Dobkin, C. S. *et al.* Tissue differences in fragile X mosaics: Mosaicism in blood cells may differ greatly from skin. *Am. J. Med. Genet.* **64**, 296–301 (1996).
- Han, X.-D., Powell, B. R., Phalin, J. L. & Chehab, F. F. Mosaicism for a full mutation, premutation, and deletion of the CGG repeats results in 22% FMRP and elevated FMR1 mRNA levels in a high-functioning fragile X male. *Am. J. Med. Genet. A* **140A**, 1463–1471 (2006).
- Jiraanont, P. *et al.* Size and methylation mosaicism in males with Fragile X syndrome. *Expert Rev. Mol. Diagn.* **17**, 1023–1032 (2017).
- Field, M. *et al.* Significantly elevated FMR1 mRNA and mosaicism for methylated premutation and full mutation alleles in two brothers with autism features referred for fragile X testing. *Int. J. Mol. Sci.* **20**, 3907 (2019).
- Saldarriaga, W., González-Teshima, L. Y., Forero-Forero, J. V., Tang, H.-T. & Tassone, F. Mosaicism in fragile X syndrome: A family case series. *J. Intellect. Disabil.* <https://doi.org/10.1177/1744629521995346> (2021).
- Schmucker, B. & Seidel, J. Mosaicism for a full mutation and a normal size allele in two fragile X males. *Am. J. Med. Genet.* **84**, 221–225 (1999).

23. Mailick, M. R. *et al.* Health profiles of mosaic versus non-mosaic FMR1 premutation carrier mothers of children with fragile X syndrome. *Front. Genet.* **9**, 173 (2018).
24. Zhao, X.-N. *et al.* Mutsβ generates both expansions and contractions in a mouse model of the Fragile X-associated disorders. *Hum. Mol. Genet.* <https://doi.org/10.1093/hmg/ddv408> (2015).
25. Lokanga, R. A. *et al.* Somatic expansion in mouse and human carriers of fragile X premutation alleles. *Hum. Mutat.* **34**, 157–166 (2013).
26. Møllersen, L., Rowe, A. D., Larsen, E., Rognes, T. & Klungland, A. Continuous and periodic expansion of CAG repeats in Huntington's disease R6/1 mice. *PLoS Genet.* **6**, e1001242 (2010).
27. Pinto, R. M. *et al.* Patterns of CAG repeat instability in the central nervous system and periphery in Huntington's disease and in spinocerebellar ataxia type 1. *Hum. Mol. Genet.* **29**, 2551–2567 (2020).
28. Monckton, D. G., Wong, L. J., Ashizawa, T. & Caskey, C. T. Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: Small pool PCR analyses. *Hum. Mol. Genet.* **4**, 1–8 (1995).
29. Wong, L. J., Ashizawa, T., Monckton, D. G., Caskey, C. T. & Richards, C. S. Somatic heterogeneity of the CTG repeat in myotonic dystrophy is age and size dependent. *Am. J. Hum. Genet.* **56**, 114–122 (1995).
30. Ciosi, M. *et al.* A genetic association study of glutamine-encoding DNA sequence structures, somatic CAG expansion, and DNA repair gene variants, with Huntington disease clinical outcomes. *EBioMedicine* **48**, 568–580 (2019).
31. Kim, K.-H. *et al.* Genetic and functional analyses point to FAN1 as the source of multiple Huntington disease modifier effects. *Am. J. Hum. Genet.* **107**, 96–110 (2020).
32. Flower, M. *Genetic Variation in DNA Repair Proteins Modifies the Course of Huntington's Disease.* (2019).
33. Morales, F. *et al.* A polymorphism in the MSH3 mismatch repair gene is associated with the levels of somatic instability of the expanded CTG repeat in the blood DNA of myotonic dystrophy type 1 patients. *DNA Repair* **40**, 57–66 (2016).
34. Cumming, S. A. *et al.* De novo repeat interruptions are associated with reduced somatic instability and mild or absent clinical features in myotonic dystrophy type 1. *Eur. J. Hum. Genet.* **26**, 1635–1647 (2018).
35. Morales, F. *et al.* Longitudinal increases in somatic mosaicism of the expanded CTG repeat in myotonic dystrophy type 1 are associated with variation in age-at-onset. *Hum. Mol. Genet.* **29**, 2496–2507 (2020).
36. Veitch, N. J. *et al.* Inherited CAG/CTG allele length is a major modifier of somatic mutation length variability in Huntington disease. *DNA Repair* **6**, 789–796 (2007).
37. Bettencourt, C. *et al.* DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. *Ann. Neurol.* **79**, 983–990 (2016).
38. CAG repeat not polyglutamine length determines timing of Huntington's disease Onset. *Cell* **178**, 887–900.e14 (2019).
39. Tassone, F. *et al.* CGG repeat length correlates with age of onset of motor signs of the fragile X-associated tremor/ataxia syndrome (FXTAS). *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **144B**, 566–569 (2007).
40. Filipovic-Sadic, S. *et al.* A novel FMR1 PCR method for the routine detection of low abundance expanded alleles and full mutations in fragile X syndrome. *Clin. Chem.* **56**, 399–408 (2010).
41. Tassone, F., Pan, R., Amiri, K., Taylor, A. K. & Hagerman, P. J. A rapid polymerase chain reaction-based screening method for identification of all expanded alleles of the fragile X (FMR1) gene in newborn and high-risk populations. *J. Mol. Diagn.* **10**, 43–49 (2008).
42. Tassone, F. *et al.* FMRP expression as a potential prognostic indicator in fragile X syndrome. *Am J Med Genet.* **84**, 250–261 (1999).
43. Hayward, B. E. & Usdin, K. Assays for determining repeat number, methylation status, and AGG interruptions in the fragile X-related disorders. *Methods Mol. Biol.* **1942**, 49–59 (2019).
44. Tassone, F. *et al.* Elevated levels of FMR1 mRNA in carrier males: A new mechanism of involvement in the fragile-X syndrome. *Am. J. Hum. Genet.* **66**, 6–15 (2000).
45. James, G., Witten, D., Hastie, T. & Tibshirani, R. *An Introduction to Statistical Learning: with Applications in R.* (Springer Science & Business Media, 2013).
46. Yrigollen, C. M. *et al.* AGG interruptions within the maternal FMR1 gene reduce the risk of offspring with fragile X syndrome. *Genet. Med.* **14**, 729–736 (2012).
47. Lokanga, R. A., Zhao, X.-N. & Usdin, K. The mismatch repair protein MSH2 is rate limiting for repeat expansion in a fragile X premutation mouse model. *Hum. Mutat.* **35**, 129–136 (2014).
48. Zhao, X., Zhang, Y., Wilkins, K., Edelmann, W. & Usdin, K. MutLy promotes repeat expansion in a Fragile X mouse model while EXO1 is protective. *PLoS Genet.* **14**, e1007719 (2018).
49. Zhao, X. *et al.* Repeat instability in the Fragile X-related disorders: Lessons from a mouse model. *Brain Sci.* **9**, 52 (2019).
50. Nolin, S. L. *et al.* Expansions and contractions of the FMR1 CGG repeat in 5,508 transmissions of normal, intermediate, and premutation alleles. *Am. J. Med. Genet. A* **179**, 1148–1156 (2019).
51. Yrigollen, C. M. *et al.* AGG interruptions and maternal age affect FMR1 CGG repeat allele stability during transmission. *J. Neurodev. Disord.* **6**, 24 (2014).
52. Latham, G. J., Coppinger, J., Hadd, A. G. & Nolin, S. L. The role of AGG interruptions in fragile X repeat expansions: A twenty-year perspective. *Front. Genet.* **5**, 244 (2014).
53. Xiao-Nan Zhao, K. U. FAN1 protects against repeat expansions in a Fragile X mouse model. *DNA Repair* **69**, 1 (2018).
54. Zhao, X. *et al.* Modifiers of somatic repeat instability in mouse models of Friedreich Ataxia and the Fragile X-related disorders: Implications for the Mechanism of somatic expansion in Huntington's disease. *J. Huntington's Dis.* **10**, 149–163 (2021).
55. Ballester-Lopez, A. *et al.* Preliminary findings on CTG expansion determination in different tissues from patients with myotonic dystrophy type 1. *Genes* **11**, 1321 (2020).
56. Kacher, R. *et al.* Propensity for somatic expansion increases over the course of life in Huntington disease. *Elife* **10**, e64674 (2021).
57. Corrales, E. *et al.* Analysis of mutational dynamics at the DMPK (CTG)<sub>n</sub> locus identifies saliva as a suitable DNA sample source for genetic analysis in myotonic dystrophy type 1. *PLoS One* **14**, e0216407 (2019).
58. Kennedy, L. *et al.* Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. *Hum. Mol. Genet.* **12**, 3359–3367 (2003).

### Author contributions

Conceptualization: E.T., K.U. Writing, original draft preparation: Y.H., and B.H. Writing, review, and editing: Y.H., B.H., M.Z., J.K., B.D.J., P.H., K.U., and F.T. Methodology and Analysis: Y.H., B.H., K.U., B.D.J., P.H., and F.T. Manuscript writing and editing: all authors have read and agreed to the published version of the manuscript.

### Competing interests

Dr. Flora Tassone received funds from Azrieli Foundation and Zynerva for studies in Fragile X syndrome. The other authors declare no conflict of interest.



### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-14183-0>.

**Correspondence** and requests for materials should be addressed to K.U. or F.T.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

Case #	CGG size of premutation alleles	AGG	Age at IV	Age at FV	Change in CGG repeats
CASE 1	95	0	53	60	1 repeat in 7 years
CASE 2	93	1	49	56	0 repeats in 7 yrs
CASE 3	83, 101	0	33	39	2 repeats in 6 years
CASE 4	96	0	47	50	0 repeats in 3 yrs
CASE 5	73	0	58	70	0 repeats in 12 yrs
CASE 6	198	0	6	12	1 repeat in 6 years
CASE 7	136	0	2	4	3 repeats in 2 yrs
CASE 8	138	0	8	14	12 repeats in 6 years
CASE 9	104	0	40	46	0 repeats in 6 yrs
CASE 10	100	0	35	42	1 repeat in 7 yrs
CASE 11	89	1	57	65	0 repeats in 8 yrs
CASE 12	74	1	61	68	0 repeats in 7 yrs
CASE 13	100	1	29	33	0 repeats in 4 years
CASE 14	98	0	57	63	0 repeats in 6 yrs
CASE 15	103	2	56	62	0 repeats in 6 yrs
CASE 16	97	2	51	53	0 repeats in 2 yrs
CASE 17	82	1	47	51	0 repeats in 4 yrs
CASE 18	94	2	67	76	0 yrs in 9 yrs
CASE 19	96	1	65	75	0 repeats in 10 yrs
CASE 20	107	0	42	48	2 repeats in 6 yrs
CASE 21	110	2	63	71	3 repeats in 7 years
CASE 22	69	1	47	60	0 repeats in 13 yrs
CASE 23	62	0	11	22	0 repeats in 11 yrs
CASE 24	89	2	31	35	0 repeats in 4 yrs

**Supplementary Table 1: PM Females drawn over time (n=24)**

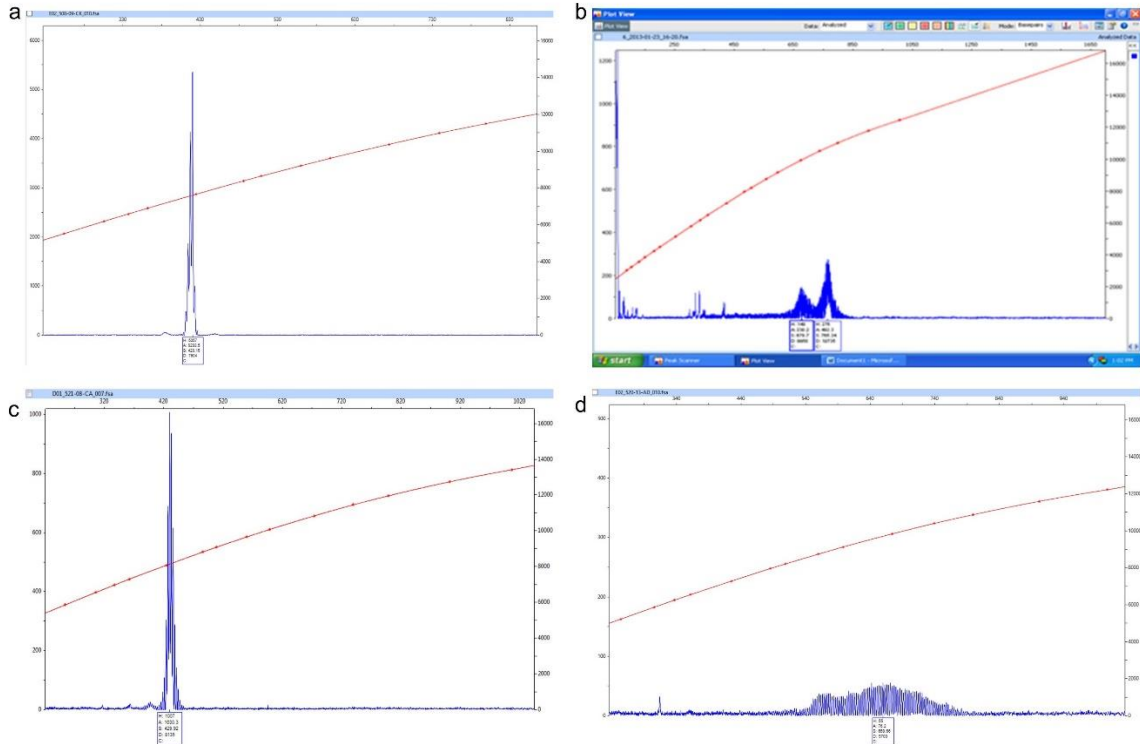
A total of 24 PM females were drawn for the study in the change in CGG repeats over time within individuals. Among these individuals, 8 individuals experienced an increase in repeats over time, and 16 did not, with an overall ratio of 1:2 individuals who experience change in repeat size over time. Most of the changes in repeat size occurred in the unmethylated regions of the PM allele.

## VI. Factors Associated with CGG Repeat Instability and Mosaicism in *FMR1* Premutation Males

Males with Fragile X syndrome are known to display various degrees of CGG instability leading to mosaicism, which can translate into clinical differences between individuals who display different levels of mosaicism [Helderman-van den Enden et al, 1999; Saldarriaga et al, 2021]. However, it is not clear if mosaicism also affects individuals with a premutation although a few cases have been reported in males [Saldarriaga et al, 2021]. Importantly, it is currently unknown which molecular factors could inform us if PM maybe at higher risk of allele instability and thus, expansion and mosaicism, which could ultimately affect the phenotypes observed in premutation carriers.

As described in [Hwang and Hayward et al, 2022], peripheral blood leukocytes were collected from 454 male PM individuals, using similar criteria as in the female study, to determine which molecular factors are associated with the risk of somatic expansion. Molecular factors including CGG repeats, AGG interruptions, *FMR1* mRNA, and age were included in this investigation. Furthermore, among the 454 males, a subset of 50 individuals were chosen based on the availability of multiple biological samples to look for any changes in CGG repeat size over time.

Preliminary PCR data indicate that most PM males display a broader size range of CGG repeats as depicted in Figure 6. Interestingly, PM males do not display the shoulder pattern observed in PM females [Hwang and Hayward et al, 2022]. Those who do not display a mosaic pattern have a clear dominant peak (Figure 7a), but often little expansion (Figure 7b). Many often present with a broader range of CGG alleles, which indicate the presence of multiple alleles (Figure 7c) and in several cases, an extreme instability in the CGG allele pattern can be observed (Figure 7d). For example, an individual could express a very large range of CGG size repeats from 30 to 200 CGG repeats with no existing dominant peak size, as shown in Figure 6d. This situation makes difficult to define the level of the degree of instability such that the male categorization of mosaicism is more qualitative than quantitative.

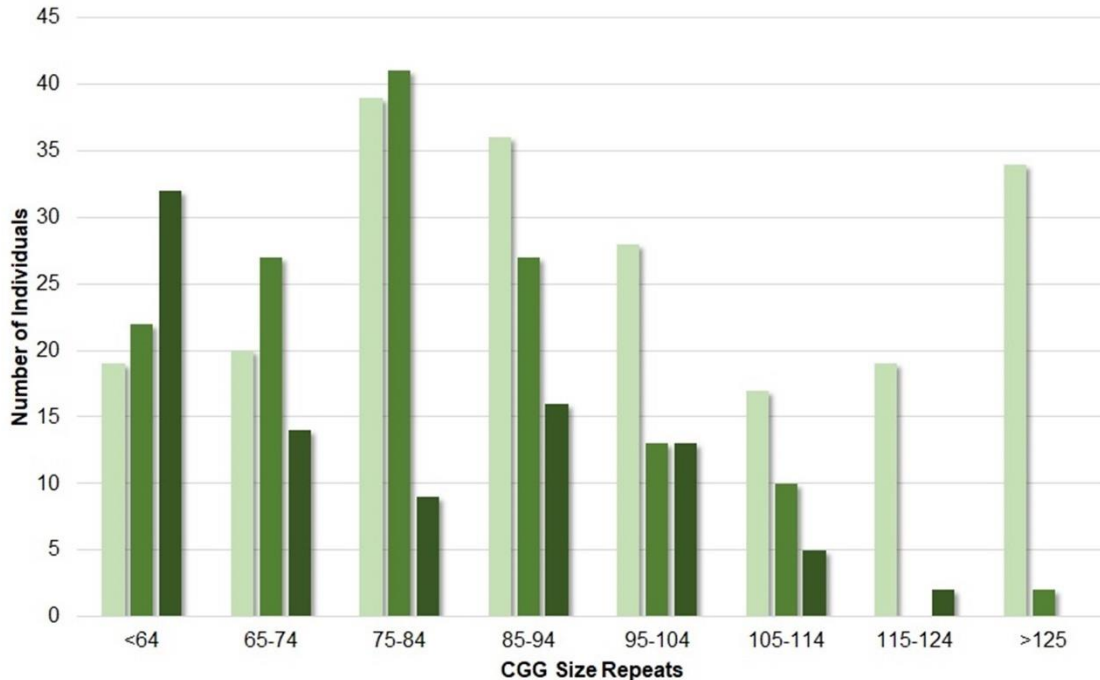


**Figure 7: Observed patterns of PCR profiles in PM males**

Examples of different types of PCR profiles observed in PM males. Figure 7a displays an individual with a stable PM allele, and Figures 7b, 7c, and 7d show individuals with various degrees of increased levels of somatic expansion. Panel 7d depicts an individual with broad CGG expansion, in which no dominant peak is observable and thus the degree of instability for this individual is difficult to quantify based on CGG repeat size.

### Correlation between CGG repeats and AGG interruptions in PM Males

CGG repeats and AGG interruptions were determined for 445 PM males as described in [Hwang and Hayward et al, 2022]. We observed a negative relationship between CGG repeat size and AGG interruptions (Figure 8). In our cohort, the overall CGG repeat size distribution resembles a binomial distribution with most individuals falling between 75 to 84 CGG repeats. As expected, increased CGG repeats was associated with fewer AGG interruptions.



0 AGG	19 (26.03%)	20 (32.79%)	39 (43.82%)	36 (45.57%)	28 (51.85%)	17 (53.13%)	19 (90.48%)	34 (94.44%)
1 AGG	22 (30.14%)	27 (44.26%)	41 (46.07%)	27 (34.18%)	13 (24.07%)	10 (31.25%)	0 (0.00%)	2 (5.56%)
2 AGG	32 (43.84%)	14 (22.95%)	9 (10.11%)	16 (20.25%)	13 (24.07%)	5 (15.63%)	2 (9.52%)	0 (0.00%)

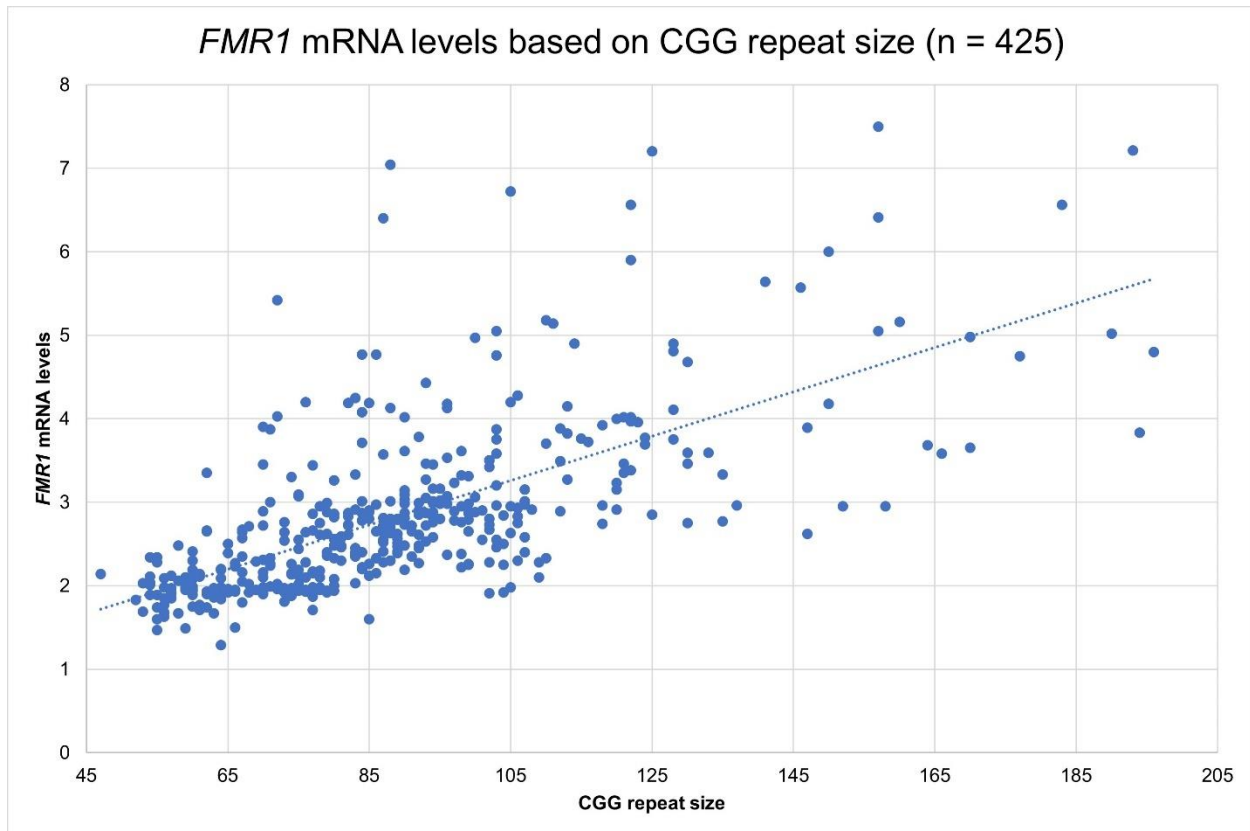
**Figure 8: AGG distribution as function of CGG repeat number in PM males (n=445)**

This chart depicts the distribution and percentage of CGG allele size in PM males sorted by CGG repeat range and number of AGG interruptions (n=445), with 0 AGG, 1 AGG, and 2 AGG separated into subgroups per each CGG size category. The percentages in the chart indicate the proportion of individuals possessing 0, 1, or 2 AGG interruptions in each size category.

About 47% of the participants (n=212, total = 445) had no AGG interruptions, with the actual distribution of individuals without AGG interruptions skewed toward the larger CGG repeats. Unlike PM females, no PM male individuals in our cohort with >125 CGG repeats carried more than 1 AGG interruption. It is worth noting that most individuals with ≤64 CGG repeats had 2 AGG interruptions, and most individuals in this group did not display allele instability, and thus somatic mosaicism, which is consistent with the notion that AGG interruptions prevent CGG expansion intergenerationally [Yrigollen et al, 2014].

#### Correlation between CGG repeats and *FMR1* mRNA in PM males

An initial subgroup of 428 PM males were selected based on existing CGG and *FMR1* mRNA data. Three individuals were then removed due to broad expansion preventing categorization of CGG repeat size. Among the remaining 425 PM males, we observe increased levels of *FMR1* mRNA levels with increased CGG repeat size (Figure 9). These results support the observation of prior studies on the association between *FMR1* mRNA levels and CGG repeats in PM carriers first reported in [Tassone et al, 2000]. This correlation is also seen in our PM female cohort [Hwang and Hayward et al, 2022].



**Figure 9. *FMR1* mRNA expression levels as function of CGG repeat size (n = 425)**

The scatter plot depicts the positive correlation between CGG repeat allele size and *FMR1* mRNA expression levels. Greater the number of CGG repeats, higher is the expression level of *FMR1* mRNA (line:  $y = 0.0265x + 0.4747$ ).

### Male PM Instability Correlation with Molecular Factors

Measures of *FMR1* locus molecular factors were compared to allele instability to see if any were significantly association. Although not all PM male samples have been fully analyzed and sorted into mosaicism categories yet, we were able to garner some information involving correlation of mosaicism with molecular factors from the current available data.

Individuals chosen for this study were sorted based on presence or absence of mosaicism in their PM allele, thus, based on whether the individual displayed the presence of allele expansion on the CE plot. Out of 454 PM male individuals, a total of 381 individuals were selected based on data about mosaicism, CGG repeat, AGG interruptions, *FMR1* mRNA levels, and age, was available at the time of analysis. Thus, the final number of individuals in each molecular measure subgroup is different from the original number of individuals depending on the data available.

<b>Molecular Measures</b>	<b>Total</b>	<b>n (M)</b>	<b>Mean (M)</b>	<b>Std. Dev. (M)</b>	<b>n (NM)</b>	<b>Mean (NM)</b>	<b>Std. Dev. (NM)</b>	<b>p-value</b>
<b>CGG</b>	381	257	98.22*	26.03*	124	70.69	14.80	2.4324E-32
<b>AGG</b>	380	257	0.55	0.71	123	1.02	0.82	1.8738E-07
<b>FMR1 mRNA</b>	355	242	3.13	1.14	113	2.34	0.65	3.8277E-15
<b>Age</b>	320	224	52.69	21.99	96	42.03	27.41	0.00094792

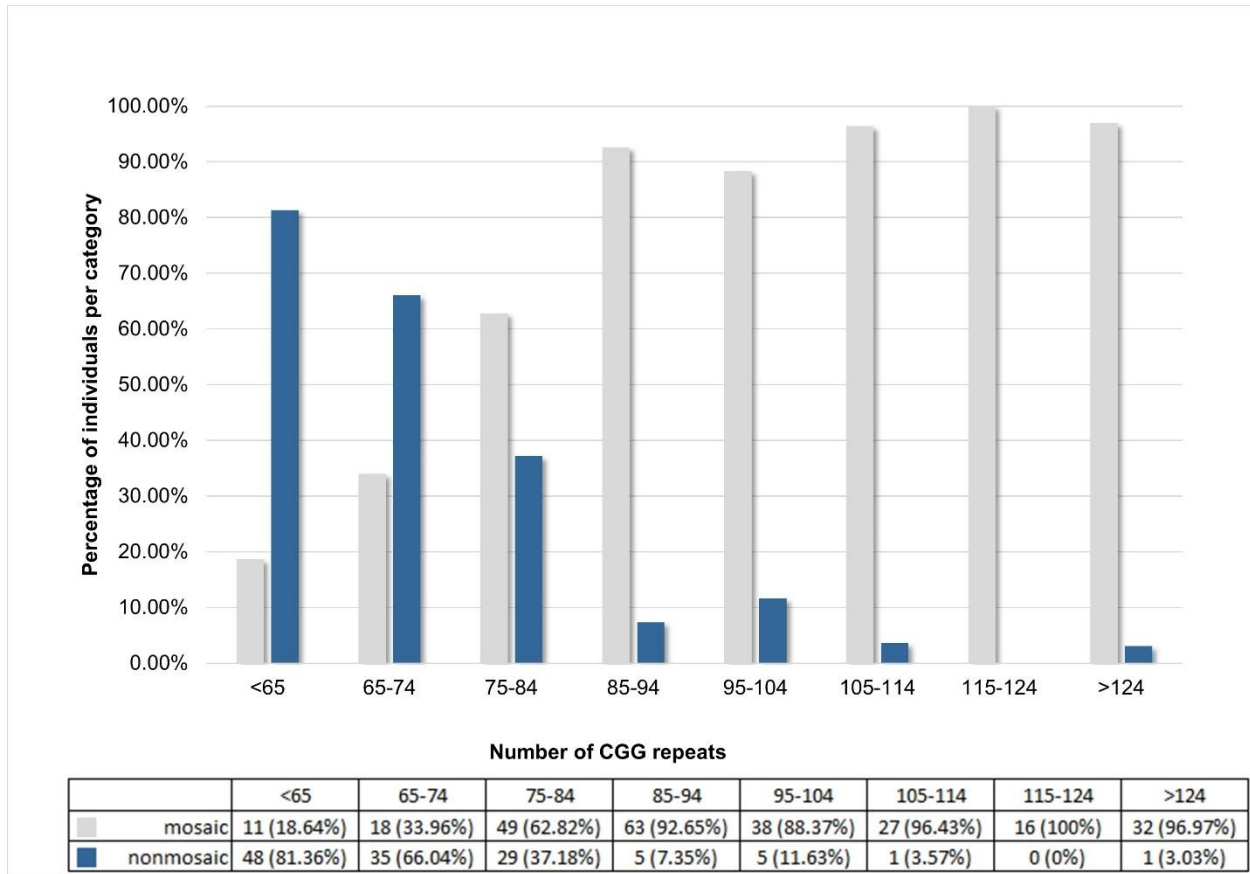
**Table 1. Molecular measures and age in the PM male cohort**

The total individuals in each molecular measurement group were divided into mosaic (*M*) and non-mosaic (*NM*) subgroups. CGG mean and CGG standard deviation (Std. Dev.) of the mosaic group were calculated using 378 individuals (\*), as three individuals had broad CGG expansion size and had to be excluded from analysis.

Each molecular measurement groups were divided into mosaic and non-mosaic groups for this study (Table 1). On average, participants in the mosaic group were older, had a higher CGG repeat number, higher *FMR1* mRNA levels, and had lower number of AGG interruptions than the non-mosaic group. The p-values were calculated using a two-tailed Unequal variance test which shows that the mosaic and non-mosaic groups significantly differ from each other and thus mosaic and non-mosaic individuals could be distinguished from each other based on these molecular measures.

When investigating the potential correlation between CGG repeats and mosaicism, three individuals were removed from analysis due to broad CGG repeat size expansion preventing identification of a distinct PM allele peak size. For the 378 individuals included in this analysis, we find a positive correlation between CGG repeat category and presence of mosaicism (Figure 10). The proportion of individuals with mosaicism dramatically increases once the *FMR1* allele is greater than >75 CGG repeats. Almost all individuals with an allele  $\geq 115$  CGG repeats present with allele instability and mosaicism (97.96%), whereas most individuals with an allele <75 CGG repeats had no mosaicism (74.1071%), indicating that CGG repeat correlates with mosaicism status mostly because the presence of 1 or 2 AGG interruptions.

A much larger percentage of individuals with  $\leq 64$  CGG repeats do not display mosaicism; in contrast, almost all individuals with  $\geq 115$  repeats displayed mosaicism, with only one exception. This agrees with the positive association of CGG repeats with instability also found in the PM female group [Hwang and Hayward et al., 2022].

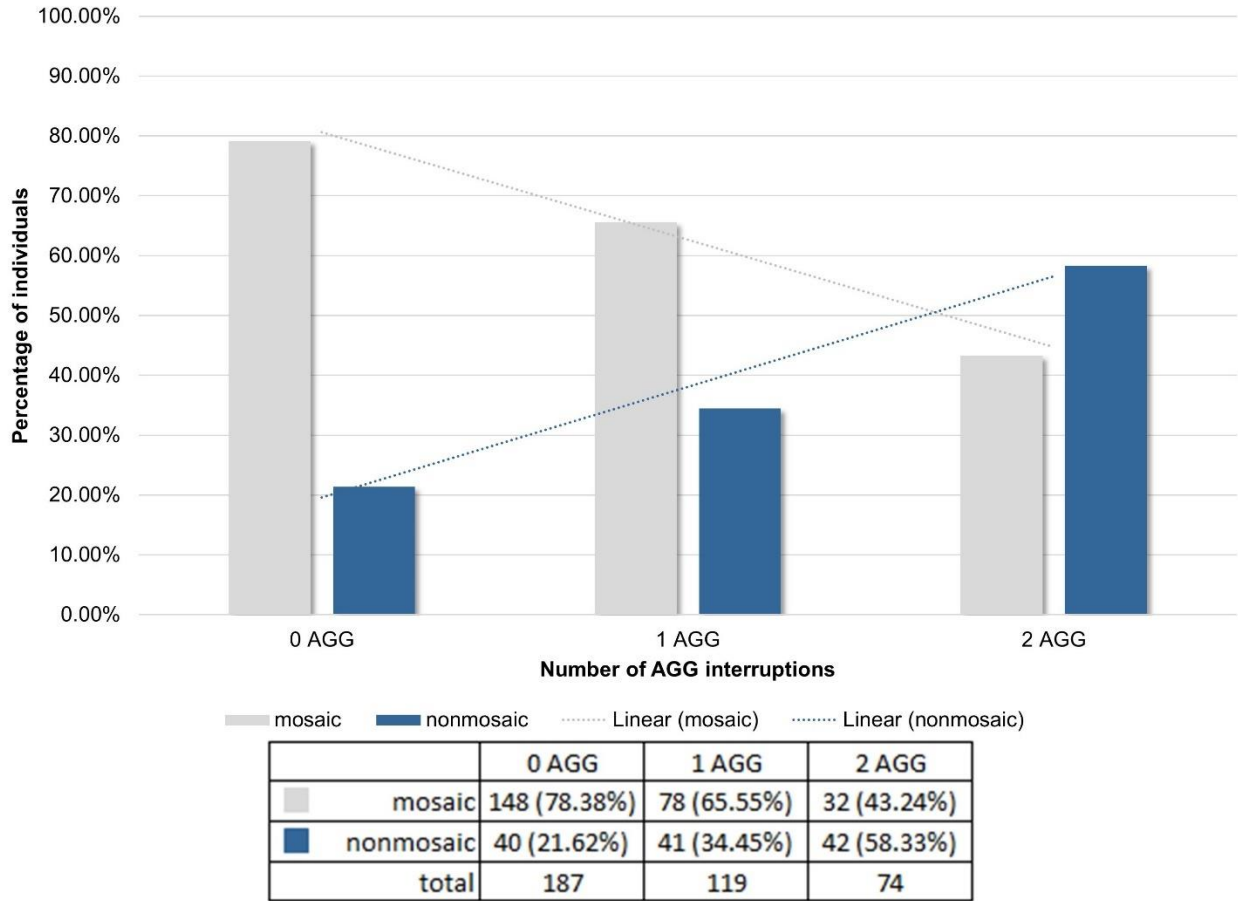


**Figure 10: Percentage of mosaicism as function of the CGG repeat number (n = 378)**

The chart depicts presence of mosaicism based on CGG repeat size. The percentage of mosaic and non-mosaic individuals are based on CGG repeat size range. The percentages in the parentheses indicate the proportion of mosaic (grey) or non-mosaic (blue) individuals calculated in each CGG size range category.

The negative association between presence of mosaicism and AGG is very clear when visualized as shown in (Figure 11). As the number of AGG interruptions increase, the overall percentage of individuals with mosaicism decrease. Conversely, the lack of AGG interruptions correlate with a higher percentage of individuals presenting with mosaicism. This negative correlation between AGG and instability was also observed in PM females and was also previously reported [Hwang and Hayward et al, 2022, Yrigollen et al, 2012].



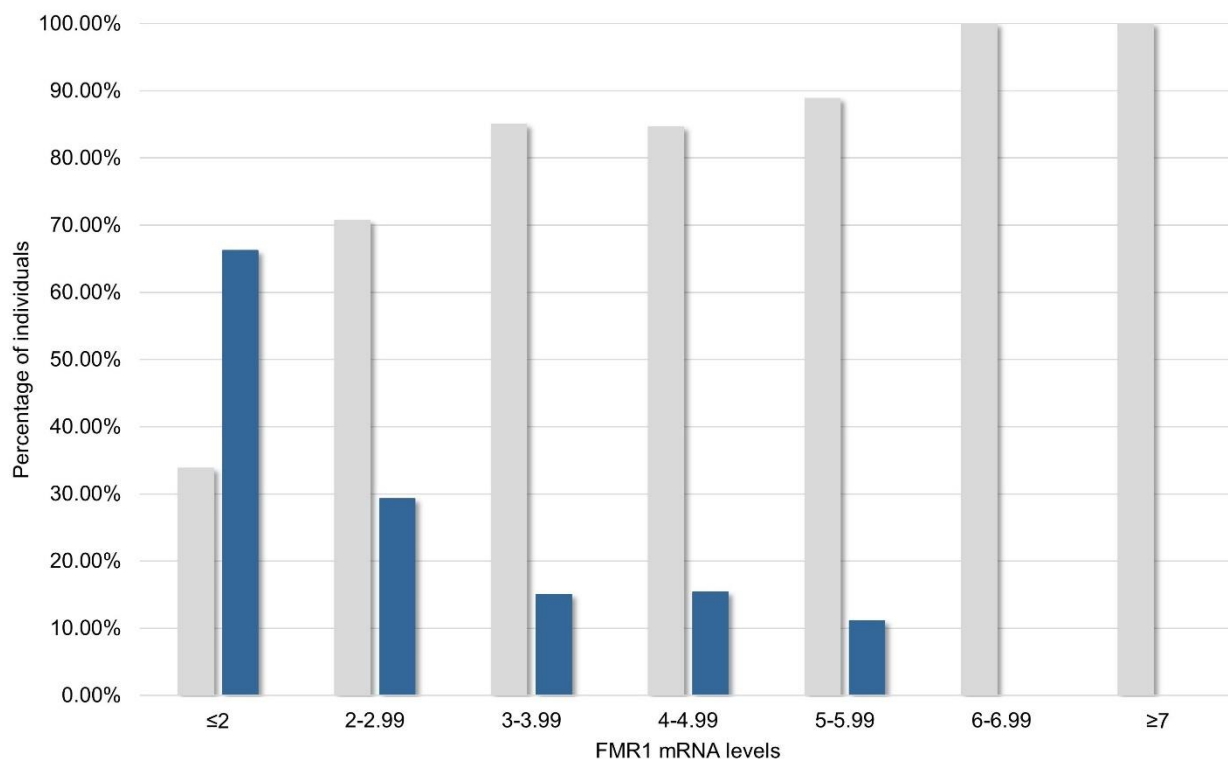


**Figure 11: Percentage of mosaicism as function of the AGG interruptions (n=380)**

The chart depicts the proportion of mosaic individuals based on the number of AGG interruptions. Numbers in parentheses indicate percentage of individuals with (grey) or without (blue) mosaicism for 0, 1, or 2 AGG interruptions. Likelihood of mosaicism decreases the more AGG interruptions an individual carries ( $y = -0.1795x + 0.9854$ ) along with the increase of allele stability ( $y = 0.1847x + 0.0112$ ).

Furthermore, through the two-tail unequal variance test, we observed a significant difference between the CGG size in the mosaic vs non-mosaic group ( $p = 2.432e-32$ ), which is in line with the skewed percentage distribution of mosaicism based on CGG repeat size. This is also the case when testing for significant difference in number of AGG interruptions between the two groups ( $p=1.874e-7$ ). Thus, individuals with mosaicism can be relatively distinguished from those without.

For *FMR1* mRNA correlation with allele instability in PM males, we find that increased *FMR1* mRNA expression levels strongly follows the presence of mosaicism (Figure 12). Among all samples, most individuals overall had between 2 – 3 folds higher *FMR1* mRNA expression levels than controls. Most individuals with  $\leq 2$  fold mRNA levels were non-mosaic, despite the lower number of non-mosaic individuals overall. Conversely, the percentage of individuals without mosaicism drops to 0% when the *FMR1* mRNA levels are approximately  $\geq 6$  fold. The two-tail test also shows that the difference between the mosaic and non-mosaic groups is significant as well ( $p= 3.83e-15$ ).



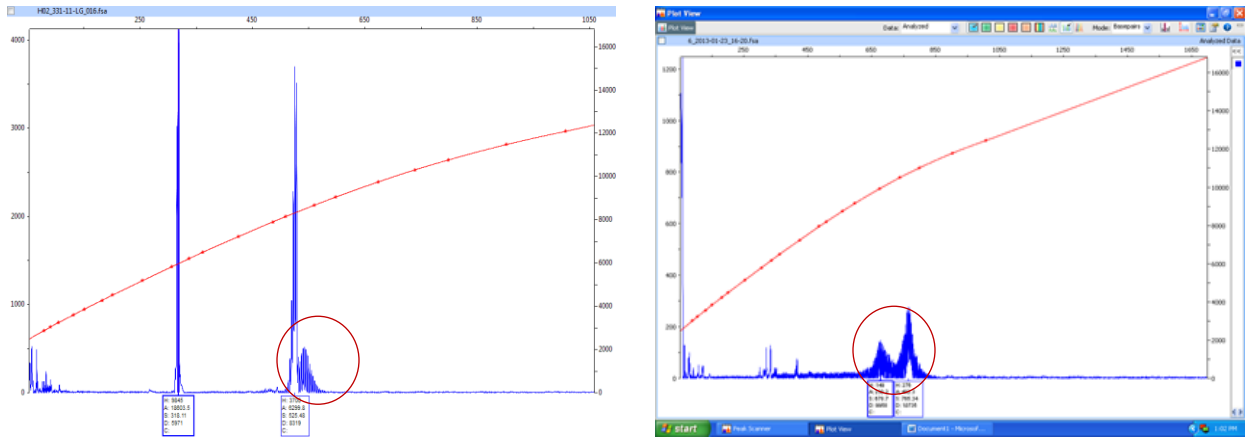
	<2	2-2.99	3-3.99	4-4.99	5-5.99	6-6.99	≥7
mosaic	24 (33.80%)	128 (70.72%)	51 (85.00%)	22 (84.62%)	8 (88.89%)	6 (100.00%)	4 (100.00%)
nonmosaic	47 (66.20%)	53 (29.28%)	9 (15.00%)	4 (15.38%)	1 (11.11%)	0 (0.00%)	0 (0.00%)
total	71	181	60	26	9	6	4

**Figure 12. Percentage of mosaicism as function of *FMR1* mRNA (n = 357)**

This chart shows the percentage of mosaic individuals based on the levels of *FMR1* mRNA. Numbers in parentheses show percent of individuals with (in grey) or without mosaicism (in blue) for each mRNA level category.

### Expansion and Methylation Patterns in PM Males

As males possess only one X-chromosome, the single *FMR1* PM allele they possess tends to display a wider range of expansion. Unlike PM females, many PM males do not have a clear dominant peak (Figure 13), so for this male study, the presence of mosaicism was determined qualitatively (presence or absence of multiple expanded peaks) as opposed to quantitatively (comparing measurement of stable peak area to unstable peak area).



**Figure 13: PCR profiles comparison between PM Female and PM Males**

PCR profile of a PM female (left) and of PM male (right) displaying bot mosaicism. On an average the PM female displays two peaks representing each *FMR1* allele on the X-chromosome, and the instability is clearly rendered as a shoulder smear pattern on the right side (circled) of the PM allele. In contrast, the PM male case displayed on the right panel shows two major broad peaks (circled) connected and do not display the shoulder smear pattern.

#### Allele instability over time in PM males

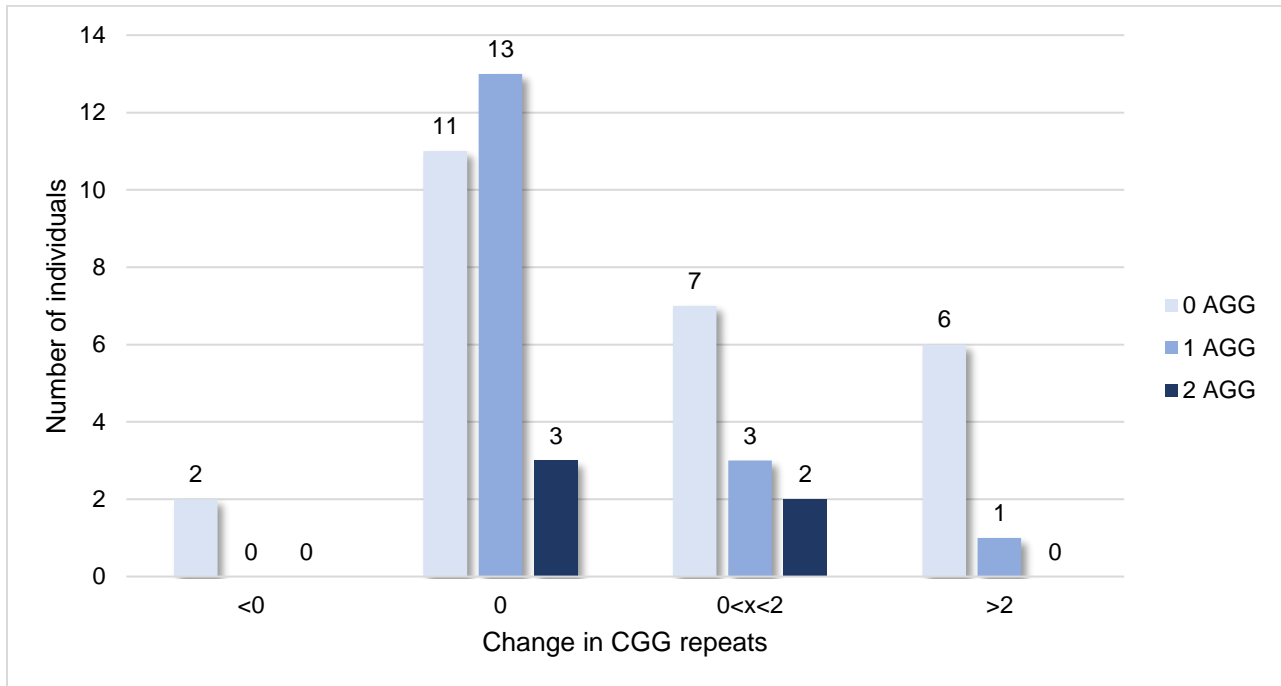
A total of 50 PM males were selected based on availability of multiple blood draw samples over the course of time with a minimum year gap, between first and last draw, of 4 years (Table 2). Molecular measures such as AGG interruptions, and age at time of first and last blood draw were also gathered for this analysis.

Case #	CGG size of premutation alleles	AGG	Age at IV	Age at FV	Change in CGG repeats
CASE 1	83	1	57	68	0 repeats in 11 yrs
CASE 2	69	0	71	81	0 repeats in 11 yrs
CASE 3	65	0	35	42	0 repeats in 7 yrs
CASE 4	82	0	53	63	0 repeats in 10 yrs
CASE 5	79	1	49	57	0 repeats in 8 yrs
CASE 6	66	0	74	84	0 repeats in 10 yrs
CASE 7	57	2	5	14	0 repeats in 10 yrs
CASE 8	broad smear	1	61	65	broad smear
CASE 9	76	1	18	23	0 repeats in 5 yrs
CASE 10	90	1	9	16	0 repeats in 7 yrs
CASE 11	56	1	41	49	0 repeats in 9 yrs
CASE 12	88	1	70	75	1 repeat in 5 years
CASE 13	98	2	60	71	1 repeat in 11 years
CASE 14	91	1	53	59	0 repeats in 6 yrs
CASE 15	83	2	64	70	0 repeats in 6 yrs
CASE 16	141	0	8	14	6 repeats in 6 years
CASE 17	82	1	71	77	0 repeats in 6 yrs
CASE 18	104	1	60	69	0 repeats in 9 yrs
CASE 19	122	0	40	51	2 repeats in 11 years
CASE 20	69	0	4	10	0 repeats in 6 yrs
CASE 21	84	0	68	83	0 repeats in 15 yrs
CASE 22	55	2	5	12	0 repeats in 6 yrs
CASE 23	69	0	67	80	1 repeat in 12 years
CASE 24	110	0	72	78	1 repeat in 5 years
CASE 25	65	1	6	11	0 repeats in 6 yrs
CASE 26	85	1	65	75	0 repeats in 11 yrs
CASE 27	85	1	55	69	1 repeat in 14 years
CASE 28	75	1	59	71	0 repeats in 12 yrs
CASE 29	149	0	11	18	6 repeats in 7 years
CASE 30	81	1	42	54	0 repeats in 6 yrs
CASE 31	78	1	32	41	0 repeats in 12 yrs
CASE 32	101	1	72	78	2 repeats in 8 years
CASE 33	70	0	11	17	0 repeats in 6 yrs
CASE 34	63	0	7	13	0 repeats in 6 yrs
CASE 35	109	1	50	65	5 repeats in 14 years
CASE 36	149,176	0	7	12	4 & 8 repeats in 4 years
CASE 37	77	0	2	12	0 repeats in 9 yrs
CASE 38	154, 175	0	7	12	2 & 7 repeats in 4 years
CASE 39	52,68	0	59	70	1 repeat in 11 years
CASE 40	84	0	52	60	-4 repeats in 8 years
CASE 41	159	0	1	9	9 repeats in 8 years
CASE 42	98	0	61	74	1 repeat in 13 years
CASE 43	111	0	66	73	1 repeat in 7 years
CASE 44	85	0	55	64	-5 repeats in 8 years
CASE 45	75	0	64	73	1 repeat in 8 years
CASE 46	80	0	70	76	0 repeats in 6 yrs
CASE 47	64	0	62	68	0 repeats in 5 yrs
CASE 48	broad smear	0	54	58	broad smear
CASE 49	101	2	66	75	1 repeat in 9 years
CASE 50	122	0	41	49	3 repeats in 9 years

**Table 2: Change in allele size overtime in PM males (n=50)**

50 individuals were chosen based on the availability of multiple draws over time. Most of the changes in repeat size resulted in broadened expansion of the PM allele, and individuals with 2 distinct allele peaks tended to experience greater expansion in the larger allele peak. 2 participants showed a decrease in allele size of 4 and 5 CGG repeats respectively. Two participants showed a broad CGG repeat allele size range that prevented a conclusive analysis.

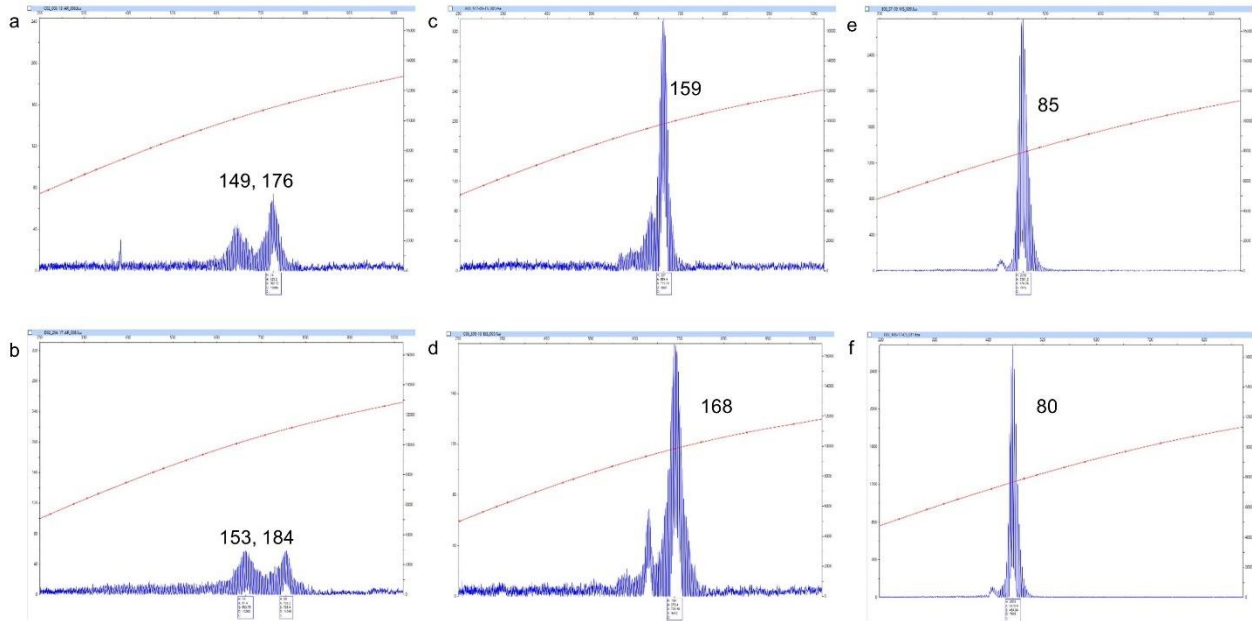
gDNA isolated from multiple blood draws were analyzed with *FMR1* specific primers- their PCR profiles were compared to determine the presence of potential expansions or contractions among the different *FMR1* alleles. Two individuals had unclear results in CGG size measurement as they showed a broad range of repeats (one such individual had a range of 30-200 CGG repeat size) and were thus excluded from the final analysis (Figure 14). Over half of the individuals examined had no change in CGG repeats over time (56.25%, n = 27). Out of the individuals who experienced a decrease in repeat size (n = 2) or an increase of greater than 2 CGG repeats (n = 5), none had 2 AGG interruptions.



**Figure 14: Change in CGG repeat number based on the number of AGG interruptions**

This bar chart represents the distribution of individuals (n=48) based on the degree of change in CGG repeat size over time. Individuals were sorted based on the change in CGG repeats between first and last sample draw and were further categorized into subgroups based on number of AGG interruptions.

Fewer PM males observed in this study had a clearly distinct PM allele, and thus change in CGG size was measured using the tallest peak in the CE plot. If an individual had two distinct peaks, both peaks were considered and measured for change in CGG repeat size. Over half (56.25%, n = 27) of the 50 individuals experienced no change in CGG size, consistent with what we observed in the PM females [Hwang and Hayward et al, 2022]. Within individuals with distinct double peaks, the larger CGG repeat size change occurred on the largest premutation allele (Figure 15), consistent with the notion that they tend to be more unstable.



**Figure 15: Change in CGG repeats over time in three PM males**

These six images depict the three different types of CGG repeat change over time observed in PM males. Figures 15a and 15b show the PCR profiles of an individual who displays 2 main alleles which increased in size over time (4 and 8 CGG repeats over 4 years). Panel 15c and 15d show an increase in size of 9 CGG repeats over the course of 8 years. Figures 15e and 15f show the PCR profiles from an individual who experienced a contraction of five CGG repeats after 8 years.

#### SNP correlation with allele instability in PM males

In my recent study [Hwang and Hayward et al, 2022], we observed that trans molecular factors potentially contribute to allele instability in *FMR1* PM female carriers. Specifically, genes involved in DNA repair contribute to instability as it has been observed in other trinucleotide-repeat disorders including Huntington's and Myotonic Dystrophy 1 [Ciosi et al, 2019; Kim et al, 2020; Cumming et al, 2018; Flower et al, 2016; Bettencourt et al, 2016; Morales et al, 2020, Genetic Modifiers of Huntington's Disease Consortium, 2019; Morales et al, 2016; Tomé et al, 2013].

In my recent study on PM females [Hwang and Hayward et al, 2022], we observed that out of 10 investigated genes involved in DNA repair and known to contribute to other trinucleotide repeat disorders, *MSH3* and *FAN1* were significantly correlated with PM allele instability. *MSH3* is a DNA mismatch repair related gene which is associated with somatic instability levels in CTG repeat expansion in Myotonic Dystrophy 1, albeit with a sex-dependent effect [Morales et al, 2016], and *FAN1* is a repair nuclease known to protect against *FMR1* somatic expansions as seen in a Fragile X mouse model [Zhao et al, 2018].

The SNPs investigated which significantly correlated with allele instability were rs150393409 (*FAN1*) and rs701383 (*MSH3*). rs150393409 codes for a missense mutation in the endonuclease involved with *FAN1* [Kim et al, 2020], and rs701383 is a candidate SNP potentially involved as a CAG somatic expansion modifier in Huntington's Disease (HD) [Genetic Modifiers of

Huntington's Disease Consortium, 2019]. Their potential correlation with the instability observed in PM males is currently under investigation.

Out of 454 PM males, we currently have SNP data on a total of 446 individuals, as 8 individuals did not have available DNA samples for analysis. Among these individuals, 380 individuals showed allelic instability and thus, mosaicism.

So far, I found that within rs701383, 19 individuals have the AA genotype, 167 have the AG genotype, and 258 have the GG genotype, with a total of 444 individuals analyzed. For rs150393409, we find that 6 have the AG genotype and 439 have the GG genotype. Among those with known mosaicism status, we find that for rs701383, 11 out of 16 AA individuals are mosaic, 99 out of 143 AG individuals are mosaic, and 146 out of 220 GG individuals are mosaic. Whereas with rs150393409, 4 out of 5 AG individuals are mosaic and 252 out of 374 GG individuals are mosaic. However, these statistics and results will be subject to change as I continue to collect and categorize more individuals by mosaicism status as this study continues. Analysis of the role of these genes in *FMR1* allele instability is also in progress.

## VII. Discussion

In this study, a total of 426 and 454 specimens derived respectively from male and female premutation carriers, were investigated to determine which molecular factors were significantly associated with *FMR1* allelic instability, a new phenomenon that has not been previously reported – this study is the first to demonstrate the presence of somatic instability in female [Hwang and Hayward et al, 2022] and in male PM carriers (manuscript in preparation).

Like previous reports which focused on the effects of CGG repeats and AGG interruptions on expansion during intergenerational transmission of the *FMR1* allele e [Yrigollen et al 2014, Yrigollen et al 2012, Yrigollen et al 2011], we found that CGG repeat allele size and AGG interruptions are also significantly correlated with *FMR1* somatic instability. This implies that molecular factors at the *FMR1* locus can be used to predict whether an individual is more likely to undergo CGG allele expansion. CGG size repeats may change and increase in both human blood samples and in multiple premutation mouse tissues, and our preliminary data suggest that age should also be considered as potential risk factor for *FMR1* PM allele expansion regardless of gender.

The allele instability seen as a “shoulder effect” has been observed only in female PM carriers and not in PM males, maybe influenced by AR. This bimodal PM allele profile has been observed in both mice and humans [Zhao et al, 2019]. Upon being treated with HpaII, the shoulder effect pattern disappears, implying that these unstable alleles are unmethylated. It is yet unknown what type of clinical effect this phenomenon has on individuals who possess it, though it can be speculated that this pattern could contribute to more severe phenotypes in mosaic individuals who display this pattern than those who do not, due to increased toxicity of the PM allele as they are unmethylated and therefore, active.

Although not all tested SNPs in this study correlated with PM allele instability, SNP variants within two DNA repair-related genes, MSH3 and FAN1, significant correlated with *FMR1* allele instability. SNP variants associated with MSH3 and FAN1 genes were inversely correlated with allele instability in female premutation carriers and could be considered protective against allele expansion within *FMR1* gene. This falls in line with what has been known previously about these two genes – FAN1 is a DNA repair gene that is known to act as a modifier for CAG repeats in HD [Kim et al, 2020; Goold et al, 2019], and MSH3 is another DNA repair gene associated with CTG expansion in myotonic dystrophy and HD [Morales et al, 2016, Tomé et al, 2013]. Thus, inverse correlation between MSH3 and FAN1 with *FMR1* PM expansion implies that external genetic factors can influence allele instability, in a few trinucleotide repeats.

It is interesting to note that a significantly larger number of PM carriers, both males and females are mosaics, which implies that mosaicism tends to be the norm rather than the exception among PM individuals. Quite remarkably, a greater proportion of individuals with lower CGG repeat sizes tend to be non-mosaics despite the lower number of non-mosaic individuals overall and thus more likely to possess more stable alleles, most likely due to presence of 1 or 2 AGG interruptions which stabilize expansion. CGG repeat size in PM individuals prevent allele expansion in these individuals.



## VIII. Conclusion

In summary, the purpose of this study was to look at stability of the *FMR1* CGG repeat in male and female premutation carriers, also overtime and, to identify molecular factors that potentially contribute to *FMR1* CGG repeat instability and somatic mosaicism. Through this investigation, we demonstrated that both cis and trans molecular factors play a role in *FMR1* allele instability.

Through analysis of 426 females and 454 males, observed associations between different molecular factors at the *FMR1* locus, including CGG repeat number and AGG interruptions with allele instability and to investigate which factors are best correlated with individuals at a higher risk of expansion, which can potentially lead to worse phenotypic outcomes. We found that CGG repeats and AGG interruptions are associated with repeat size and allele instability in both males and females. CGG and AGG interruptions, as well as CGG and *FMR1* mRNA, were also correlated with one another regardless of gender status.

Although the risk factors that involve expansion is alike in males and females, the frequency of the expansion and the pattern profiles observed in each gender are different. Females with mosaicism tend to display, mainly, 1 PM allele being clearly distinguished as the original CGG repeat size plus a serial of additional peaks, indicating the presence of unstable premutation allele of different sizes (shoulder effect); interestingly these unstable alleles are always unmethylated. On the contrary, males with mosaicism more often display a broader expansion profile characterized by several peaks with the original CGG size alleles being, in many cases, difficult to discern, compared to females and unmethylated and therefore active. Thus, a distinction of PM allele instability profile based on gender exists, though how this may translate to the observed phenotypes in this category of individuals is unknown and although was not included in this study, it is currently investigated.

The investigation of change in CGG size over time shows that about half of individuals, regardless of gender, show changes in CGG repeat size. Indeed, ~40% of females and ~44% of males displayed change in repeat size over time. This implies that there may other factors involved in the instability process, including aging that may increase the risk of somatic expansion in an individual.

Allelic variants in genes involved in DNA repair, such as MSH3, and FAN1 correlate with allele instability in the *FMR1* gene. Variants in the MSH3 have been previously found to have a higher affinity with instability in other trinucleotide expansion disorders than the ones used in this study. Nonetheless, the two of the ten investigated DNA repair gene variants were significantly correlated with somatic instability alludes to possibly investigating other DNA repair gene related SNPs which may contribute to allele instability.

Altogether, these findings could be useful to identify which individuals are at greater risk of somatic expansion and potential worsening the FX-related phenotypes. As some individuals also experienced a greater degree of expansion while aging whereas others did not, future studies can focus on differences in molecular and other environmental factors between individuals that displayed size repeat changes, compared to those who had not, to determine

which factors can predict higher risk of expansion through aging and thus, a more severe phenotype.

## VIII. Bibliography

1. Allen, E. G., Charen, K., Hipp, H. S., Shubeck, L., Amin, A., He, W., Hunter, J. E., & Sherman, S. L. (2020). **Clustering of comorbid conditions among women who carry an *FMR1* premutation.** *Genetics in medicine: official journal of the American College of Medical Genetics*, 22(4), 758–766. <https://doi.org/10.1038/s41436-019-0733-5>
2. Bagni, C., Tassone, F., Neri, G., & Hagerman, R. (2012). **Fragile X syndrome: causes, diagnosis, mechanisms, and therapeutics.** *The Journal of clinical investigation*, 122(12), 4314–4322. <https://doi.org/10.1172/JCI63141>
3. Bailey, D. B., Jr, Hatton, D. D., Skinner, M., & Mesibov, G. (2001). **Autistic behavior, *FMR1* protein, and developmental trajectories in young males with fragile X syndrome.** *Journal of autism and developmental disorders*, 31(2), 165–174. <https://doi.org/10.1023/a:1010747131386>
4. Bettencourt, C., Hensman-Moss, D., Flower, M., Wiethoff, S., Brice, A., Goizet, C., Stevanin, G., Koutsis, G., Karadima, G., Panas, M., Yescas-Gómez, P., García-Velázquez, L. E., Alonso-Vilatela, M. E., Lima, M., Raposo, M., Traynor, B., Sweeney, M., Wood, N., Giunti, P., SPATAX Network, ... Jones, L. (2016). **DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases.** *Annals of neurology*, 79(6), 983–990. <https://doi.org/10.1002/ana.24656>
5. Ciosi, M., Maxwell, A., Cumming, S. A., Hensman Moss, D. J., Alshammari, A. M., Flower, M. D., Durr, A., Leavitt, B. R., Roos, R., TRACK-HD team, Enroll-HD team, Holmans, P., Jones, L., Langbehn, D. R., Kwak, S., Tabrizi, S. J., & Monckton, D. G. (2019). **A genetic association study of glutamine-encoding DNA sequence structures, somatic CAG expansion, and DNA repair gene variants, with Huntington disease clinical outcomes.** *EBioMedicine*, 48, 568–580. <https://doi.org/10.1016/j.ebiom.2019.09.020>
6. Cumming, S. A., Hamilton, M. J., Robb, Y., Gregory, H., McWilliam, C., Cooper, A., Adam, B., McGhie, J., Hamilton, G., Herzyk, P., Tschannen, M. R., Worthey, E., Petty, R., Ballantyne, B., Scottish Myotonic Dystrophy Consortium, Warner, J., Farrugia, M. E., Longman, C., & Monckton, D. G. (2018). **De novo repeat interruptions are associated with reduced somatic instability and mild or absent clinical features in myotonic dystrophy type 1.** *European journal of human genetics: EJHG*, 26(11), 1635–1647. <https://doi.org/10.1038/s41431-018-0156-9>
7. Eichler, E. E., Hammond, H. A., Macpherson, J. N., Ward, P. A., & Nelson, D. L. (1995). **Population survey of the human *FMR1* CGG repeat substructure suggests biased**

- polarity for the loss of AGG interruptions.** *Human molecular genetics*, 4(12), 2199–2208. <https://doi.org/10.1093/hmg/4.12.2199>
8. Eichler, E. E., Richards, S., Gibbs, R. A., & Nelson, D. L. (1993). **Fine structure of the human *FMR1* gene.** *Human molecular genetics*, 2(8), 1147–1153. <https://doi.org/10.1093/hmg/2.8.1147>
  9. Flower, M. *Genetic Variation in DNA Repair Proteins Modifies the Course of Huntington's Disease.* (2019).
  10. Freed, D., Stevens, E. L., & Pevsner, J. (2014). **Somatic mosaicism in the human genome.** *Genes*, 5(4), 1064–1094. <https://doi.org/10.3390/genes5041064>
  11. Gazy, I., Hayward, B., Potapova, S., Zhao, X., & Usdin, K. (2019). **Double-strand break repair plays a role in repeat instability in a fragile X mouse model.** *DNA repair*, 74, 63–69. <https://doi.org/10.1016/j.dnarep.2018.12.004>
  12. Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium. Electronic address: [gusella@helix.mgh.harvard.edu](mailto:gusella@helix.mgh.harvard.edu), & Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium (2019). **CAG Repeat Not Polyglutamine Length Determines Timing of Huntington's Disease Onset.** *Cell*, 178(4), 887–900.e14. <https://doi.org/10.1016/j.cell.2019.06.036>
  13. Goold, R., Flower, M., Moss, D. H., Medway, C., Wood-Kaczmar, A., Andre, R., Farshim, P., Bates, G. P., Holmans, P., Jones, L., & Tabrizi, S. J. (2019). **FAN1 modifies Huntington's disease progression by stabilizing the expanded HTT CAG repeat.** *Human molecular genetics*, 28(4), 650–661.
  14. Hagerman, R. J., Protic, D., Rajaratnam, A., Salcedo-Arellano, M. J., Aydin, E. Y., & Schneider, A. (2018). **Fragile X-Associated Neuropsychiatric Disorders (FXAND).** *Frontiers in psychiatry*, 9, 564. <https://doi.org/10.3389/fpsy.2018.00564>
  15. Hagerman, P. J., & Hagerman, R. J. (2015). **Fragile X-associated tremor/ataxia syndrome.** *Annals of the New York Academy of Sciences*, 1338(1), 58–70. <https://doi.org/10.1111/nyas.12693>
  16. Hansen, R. S., Gartler, S. M., Scott, C. R., Chen, S. H., & Laird, C. D. (1992). **Methylation analysis of CGG sites in the CpG island of the human *FMR1* gene.** *Human molecular genetics*, 1(8), 571–578. <https://doi.org/10.1093/hmg/1.8.571>
  17. Helderma-van den Enden, A. T., Maaswinkel-Mooij, P. D., Hoogendoorn, E., Willemsen, R., Maat-Kievit, J. A., Losekoot, M., & Oostra, B. A. (1999). **Monozygotic twin brothers with the fragile X syndrome: different CGG repeats and different mental capacities.** *Journal of medical genetics*, 36(3), 253–257.

18. Hinton, V. J., Brown, W. T., Wisniewski, K., & Rudelli, R. D. (1991). **Analysis of neocortex in three males with the fragile X syndrome.** *American journal of medical genetics*, 41(3), 289–294. <https://doi.org/10.1002/ajmg.1320410306>
19. Hunter, J., Rivero-Arias, O., Angelov, A., Kim, E., Fotheringham, I., & Leal, J. (2014). **Epidemiology of fragile X syndrome: a systematic review and meta-analysis.** *American journal of medical genetics. Part A*, 164A (7), 1648–1658. <https://doi.org/10.1002/ajmg.a.36511>
20. Hwang, Y. H., Hayward, B. E., Zafarullah, M., Kumar, J., Durbin Johnson, B., Holmans, P., Usdin, K., & Tassone, F. (2022). **Both cis and trans-acting genetic factors drive somatic instability in female carriers of the *FMR1* premutation.** *Scientific reports*, 12(1), 10419. <https://doi.org/10.1038/s41598-022-14183-0>
21. Jacquemont, S., Hagerman, R. J., Leehey, M. A., Hall, D. A., Levine, R. A., Brunberg, J. A., Zhang, L., Jardini, T., Gane, L. W., Harris, S. W., Herman, K., Grigsby, J., Greco, C. M., Berry-Kravis, E., Tassone, F., & Hagerman, P. J. (2004). **Penetrance of the fragile X-associated tremor/ataxia syndrome in a premutation carrier population.** *JAMA*, 291(4), 460–469. <https://doi.org/10.1001/jama.291.4.460>
22. Kaytor, M. D., & Orr, H. T. (2001). **RNA targets of the fragile X protein.** *Cell*, 107(5), 555–557. [https://doi.org/10.1016/s0092-8674\(01\)00590-6](https://doi.org/10.1016/s0092-8674(01)00590-6)
23. Kim, K. H., Hong, E. P., Shin, J. W., Chao, M. J., Loupe, J., Gillis, T., Mysore, J. S., Holmans, P., Jones, L., Orth, M., Monckton, D. G., Long, J. D., Kwak, S., Lee, R., Gusella, J. F., MacDonald, M. E., & Lee, J. M. (2020). **Genetic and Functional Analyses Point to *FAN1* as the Source of Multiple Huntington Disease Modifier Effects.** *American journal of human genetics*, 107(1), 96–110. <https://doi.org/10.1016/j.ajhg.2020.05.012>
24. Lokanga, R.A., Zhao, X. N., Entezam, A., & Usdin, K. (2014). **X inactivation plays a major role in the gender bias in somatic expansion in a mouse model of the fragile X-related disorders: implications for the mechanism of repeat expansion.** *Human molecular genetics*, 23(18), 4985–4994. <https://doi.org/10.1093/hmg/ddu213>
25. Lokanga, R. A., Entezam, A., Kumari, D., Yudkin, D., Qin, M., Smith, C. B., & Usdin, K. (2013). **Somatic expansion in mouse and human carriers of fragile X premutation alleles.** *Human mutation*, 34(1), 157–166. <https://doi.org/10.1002/humu.22177>
26. Lu, R., Wang, H., Liang, Z., Ku, L., O'donnell, W. T., Li, W., Warren, S. T., & Feng, Y. (2004). **The fragile X protein controls microtubule-associated protein 1B translation and microtubule stability in brain neuron development.** *Proceedings of*

- the National Academy of Sciences of the United States of America*, 101(42), 15201–15206. <https://doi.org/10.1073/pnas.0404995101>
27. Maddalena, A., Richards, C. S., McGinniss, M. J., Brothman, A., Desnick, R. J., Grier, R. E., Hirsch, B., Jacky, P., McDowell, G. A., Popovich, B., Watson, M., & Wolff, D. J. (2001). **Technical standards and guidelines for fragile X: the first of a series of disease-specific supplements to the Standards and Guidelines for Clinical Genetics Laboratories of the American College of Medical Genetics. Quality Assurance Subcommittee of the Laboratory Practice Committee.** *Genetics in medicine : official journal of the American College of Medical Genetics*, 3(3), 200–205. <https://doi.org/10.1097/00125817-200105000-00010>
28. Mailick, M. R., Movaghar, A., Hong, J., Greenberg, J. S., DaWalt, L. S., Zhou, L., Jackson, J., Rathouz, P. J., Baker, M. W., Brilliant, M., Page, D., & Berry-Kravis, E. (2018). **Health Profiles of Mosaic Versus Non-mosaic *FMR1* Premutation Carrier Mothers of Children With Fragile X Syndrome.** *Frontiers in genetics*, 9, 173. <https://doi.org/10.3389/fgene.2018.00173>
29. Meng, L., Kaufmann, W. E., Frye, R. E., Ong, K., Kaminski, J. W., Velinov, M., & Berry-Kravis, E. (2022). **The association between mosaicism type and cognitive and behavioral functioning among males with fragile X syndrome.** *American journal of medical genetics. Part A*, 188(3), 858–866. <https://doi.org/10.1002/ajmg.a.62594>
30. Morales, F., Vásquez, M., Corrales, E., Vindas-Smith, R., Santamaría-Ulloa, C., Zhang, B., Siritto, M., Estecio, M. R., Krahe, R., & Monckton, D. G. (2020). **Longitudinal increases in somatic mosaicism of the expanded CTG repeat in myotonic dystrophy type 1 are associated with variation in age-at-onset.** *Human molecular genetics*, 29(15), 2496–2507. <https://doi.org/10.1093/hmg/ddaa123>
31. Morales, F., Vásquez, M., Santamaría, C., Cuenca, P., Corrales, E., & Monckton, D. G. (2016). **A polymorphism in the MSH3 mismatch repair gene is associated with the levels of somatic instability of the expanded CTG repeat in the blood DNA of myotonic dystrophy type 1 patients.** *DNA repair*, 40, 57–66. <https://doi.org/10.1016/j.dnarep.2016.01.001>
32. Moss, D., Pardiñas, A. F., Langbehn, D., Lo, K., Leavitt, B. R., Roos, R., Durr, A., Mead, S., TRACK-HD investigators, REGISTRY investigators, Holmans, P., Jones, L., & Tabrizi, S. J. (2017). **Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study.** *The Lancet. Neurology*, 16(9), 701–711. [https://doi.org/10.1016/S1474-4422\(17\)30161-8](https://doi.org/10.1016/S1474-4422(17)30161-8)

33. Naumann, A., Hochstein, N., Weber, S., Fanning, E., & Doerfler, W. (2009). **A distinct DNA-methylation boundary in the 5'- upstream sequence of the *FMR1* promoter binds nuclear proteins and is lost in fragile X syndrome.** *American journal of human genetics*, *85*(5), 606–616. <https://doi.org/10.1016/j.ajhg.2009.09.018>
34. Nimchinsky, E. A., Oberlander, A. M., & Svoboda, K. (2001). **Abnormal development of dendritic spines in *FMR1* knock-out mice.** *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *21*(14), 5139–5146. <https://doi.org/10.1523/JNEUROSCI.21-14-05139.2001>
35. Nolin, S. L., Sah, S., Glicksman, A., Sherman, S. L., Allen, E., Berry-Kravis, E., Tassone, F., Yrigollen, C., Cronister, A., Jodah, M., Ersalesi, N., Dobkin, C., Brown, W. T., Shroff, R., Latham, G. J., & Hadd, A. G. (2013). **Fragile X AGG analysis provides new risk predictions for 45-69 repeat alleles.** *American journal of medical genetics. Part A*, *161A*(4), 771–778. <https://doi.org/10.1002/ajmg.a.35833>
36. Pretto, D. I., Eid, J. S., Yrigollen, C. M., Tang, H. T., Loomis, E. W., Raske, C., Durbin-Johnson, B., Hagerman, P. J., & Tassone, F. (2015). **Differential increases of specific *FMR1* mRNA isoforms in premutation carriers.** *Journal of medical genetics*, *52*(1), 42–52. <https://doi.org/10.1136/jmedgenet-2014-102593>
37. Pretto, D. I., Mendoza-Morales, G., Lo, J., Cao, R., Hadd, A., Latham, G. J., Durbin-Johnson, B., Hagerman, R., & Tassone, F. (2014). **CGG allele size somatic mosaicism and methylation in *FMR1* premutation alleles.** *Journal of medical genetics*, *51*(5), 309–318. <https://doi.org/10.1136/jmedgenet-2013-102021>
38. Pretto, D., Yrigollen, C. M., Tang, H. T., Williamson, J., Espinal, G., Iwahashi, C. K., Durbin-Johnson, B., Hagerman, R. J., Hagerman, P. J., & Tassone, F. (2014 (2)). **Clinical and molecular implications of mosaicism in *FMR1* full mutations.** *Frontiers in genetics*, *5*, 318.
39. Primerano, B., Tassone, F., Hagerman, R. J., Hagerman, P., Amaldi, F., & Bagni, C. (2002). **Reduced *FMR1* mRNA translation efficiency in fragile X patients with premutations.** *RNA (New York, N.Y.)*, *8*(12), 1482–1488.
40. Rajaratnam, A., Shergill, J., Salcedo-Arellano, M., Saldarriaga, W., Duan, X., & Hagerman, R. (2017). **Fragile X syndrome and fragile X-associated disorders.** *F1000Research*, *6*, 2112. <https://doi.org/10.12688/f1000research.11885.1>
41. Roth, M., Ronco, L., Cadavid, D., Durbin-Johnson, B., Hagerman, R. J., & Tassone, F. (2021). **FMRP Levels in Human Peripheral Blood Leukocytes Correlates with**

- Intellectual Disability.** *Diagnostics (Basel, Switzerland)*, 11(10), 1780.  
<https://doi.org/10.3390/diagnostics11101780>
42. Saldarriaga, W., González-Teshima, L. Y., Forero-Forero, J. V., Tang, H. T., & Tassone, F. (2021). **Mosaicism in Fragile X syndrome: A family case series.** *Journal of intellectual disabilities : JOID*, 1744629521995346. Advance online publication.  
<https://doi.org/10.1177/1744629521995346>
43. Sutcliffe, J. S., Nelson, D. L., Zhang, F., Pieretti, M., Caskey, C. T., Saxe, D., & Warren, S. T. (1992). **DNA methylation represses *FMR-1* transcription in fragile X syndrome.** *Human molecular genetics*, 1(6), 397–400.  
<https://doi.org/10.1093/hmg/1.6.397>
44. Tassone, F., Hagerman, R. J., Taylor, A. K., Gane, L. W., Godfrey, T. E., & Hagerman, P. J. (2000). **Elevated levels of *FMR1* mRNA in carrier males: a new mechanism of involvement in the fragile-X syndrome.** *American journal of human genetics*, 66(1), 6–15. <https://doi.org/10.1086/302720>
45. Tassone, F., Hagerman, R. J., Iklé, D. N., Dyer, P. N., Lampe, M., Willemsen, R., Oostra, B. A., & Taylor, A. K. (1999). **FMRP expression as a potential prognostic indicator in fragile X syndrome.** *American journal of medical genetics*, 84(3), 250–261.
46. Taylor, A. K., Tassone, F., Dyer, P. N., Hersch, S. M., Harris, J. B., Greenough, W. T., & Hagerman, R. J. (1999). Tissue heterogeneity of the *FMR1* mutation in a high-functioning male with fragile X syndrome. *American journal of medical genetics*, 84(3), 233–239.
47. Tomé, S., Simard, J. P., Slean, M. M., Holt, I., Morris, G. E., Wojciechowicz, K., te Riele, H., & Pearson, C. E. (2013). **Tissue-specific mismatch repair protein expression: MSH3 is higher than MSH6 in multiple mouse tissues.** *DNA repair*, 12(1), 46–52.  
<https://doi.org/10.1016/j.dnarep.2012.10.006>
48. Tseng, E., Tang, H. T., AlOlaby, R. R., Hickey, L., & Tassone, F. (2017). **Altered expression of the *FMR1* splicing variants landscape in premutation carriers.** *Biochimica et biophysica acta. Gene regulatory mechanisms*, 1860(11), 1117–1126. <https://doi.org/10.1016/j.bbagr.2017.08.007>
49. Usdin, K., House, N. C., & Freudenreich, C. H. (2015). **Repeat instability during DNA repair: Insights from model systems.** *Critical reviews in biochemistry and molecular biology*, 50(2), 142–167. <https://doi.org/10.3109/10409238.2014.999192>
50. Wöhrle, D., Kennerknecht, I., Wolf, M., Enders, H., Schwemmler, 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(7), 1147–1153.

<https://doi.org/10.1093/hmg/4.7.1147>

51. Wöhrle, D., Hennig, I., Vogel, W., & Steinbach, P. (1993). **Mitotic stability of fragile X mutations in differentiated cells indicates early post-conceptual trinucleotide repeat expansion.** *Nature genetics*, 4(2), 140–142. <https://doi.org/10.1038/ng0693-140>
52. Yrigollen, C. M., Martorell, L., Durbin-Johnson, B., Naudo, M., Genoves, J., Murgia, A., Polli, R., Zhou, L., Barbouth, D., Rupchock, A., Finucane, B., Latham, G. J., Hadd, A., Berry-Kravis, E., & Tassone, F. (2014). **AGG interruptions and maternal age affect FMR1 CGG repeat allele stability during transmission.** *Journal of neurodevelopmental disorders*, 6(1), 24. <https://doi.org/10.1186/1866-1955-6-24>
53. Yrigollen, C. M., Mendoza-Morales, G., Hagerman, R., & Tassone, F. (2013). **Transmission of an FMR1 premutation allele in a large family identified through newborn screening: the role of AGG interruptions.** *Journal of human genetics*, 58(8), 553–559. <https://doi.org/10.1038/jhg.2013.50>
54. Yrigollen, C. M., Durbin-Johnson, B., Gane, L., Nelson, D. L., Hagerman, R., Hagerman, P. J., & Tassone, F. (2012). **AGG interruptions within the maternal FMR1 gene reduce the risk of offspring with fragile X syndrome.** *Genetics in medicine: official journal of the American College of Medical Genetics*, 14(8), 729–736. <https://doi.org/10.1038/gim.2012.34>
55. Yrigollen, C. M., Tassone, F., Durbin-Johnson, B., & Tassone, F. (2011). **The role of AGG interruptions in the transcription of FMR1 premutation alleles.** *PloS one*, 6(7), e21728. <https://doi.org/10.1371/journal.pone.0021728>
56. Zhao, X., Gazy, I., Hayward, B., Pintado, E., Hwang, Y. H., Tassone, F., & Usdin, K. (2019). **Repeat Instability in the Fragile X-Related Disorders: Lessons from a Mouse Model.** *Brain sciences*, 9(3), 52. <https://doi.org/10.3390/brainsci9030052>
57. Zhao, X. N., & Usdin, K. (2018). **FAN1 protects against repeat expansions in a Fragile X mouse model.** *DNA repair*, 69, 1–5. <https://doi.org/10.1016/j.dnarep.2018.07.001>
58. Zhao, X. N., & Usdin, K. (2016). **Ups and Downs: Mechanisms of Repeat Instability in the Fragile X-Related Disorders.** *Genes*, 7(9), 70. <https://doi.org/10.3390/genes7090070>

59. Zhao, X. N., Kumari, D., Gupta, S., Wu, D., Evanitsky, M., Yang, W., & Usdin, K. (2015). **Mut $\beta$  generates both expansions and contractions in a mouse model of the Fragile X-associated disorders.** *Human molecular genetics*, 24(24), 7087–7096. <https://doi.org/10.1093/hmg/ddv408>
60. Zhou, Y., Kumari, D., Sciascia, N., & Usdin, K. (2016). **CGG-repeat dynamics and *FMR1* gene silencing in fragile X syndrome stem cells and stem cell-derived neurons.** *Molecular autism*, 7, 42. <https://doi.org/10.1186/s13229-016-0105-9>