Expression analysis of the Fragile X messenger ribonucleoprotein 1 (*FMR1*) gene in *FMR1*-associated disorders

By

JAMIE RANDOL

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Integrative Genetics and Genomics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Paul J. Hagerman, Chair

Frédéric L. Chédin

Jan A. Nolta

Committee in Charge

2023

Acknowledgements

Choose the lab, not the school. So advised James Murray, my major professor during the pursuit of my master's degree, when he learned that I was applying to doctoral programs. He counseled me to focus on finding labs conducting research that interested me, rather than fixating on the universities themselves. Accordingly, I chose to pursue my doctoral degree at UC Davis due to the molecular work of three professors: Paul Hagerman, a renowned expert in *FMR1*-associated disorders; Frédéric Chédin, one of the world's leading authorities on R-loop biology; and David Segal, a titan in genome engineering for neurologic disorders. I feel blessed to have worked with both Paul and Fred during my predoctoral career. Paul served as my primary mentor, while I was co-mentored by Fred during the R-loop project. Dave, if you are reading this, please keep me in mind! I would be thrilled to collaborate on a project one day and realize my dream of working with all three researchers who played a pivotal role in my decision to stay at UC Davis following my master's degree.

After years of graduate school, "choose the lab" has taken on a deeper meaning to me than merely choosing stimulating research. It also encapsulates the idea of choosing the people with whom I work every day. Although I did not fully realize it at the time I joined Paul's lab, I chose very well. I have been happy and fulfilled as a graduate student because of the genuinely kind and supportive people surrounding me. I thank my immeasurably patient major professor, **Paul Hagerman**. He is knowledgeable, speedy with feedback, and authentically cares about his students.

I also thank **Glenda Espinal**, lab manager extraordinaire, and **Lisa Makhoul**, the seemingly omniscient research coordinator and executive assistant to Paul. Every graduate student deserves their own versions of Glenda and Lisa. As lab manager, Glenda facilitated experiments, provided helpful tips based on a wealth of experience and sharp insights, and

fiercely protected the lab members under her care. Lisa has been my go-to person when I have questions outside of the wet lab. If she does not know the answer, she knows how to get it.

Nichole Holm, **Lisa Ma**, and **Ellen Osborn** completed their advanced degrees alongside me in Paul's lab. They often provided emotional and mental health support and a safe space to discuss scientific ideas. They helped make this challenging time a happy one. Thank you, my friends. I also thank **Malgorzata Bzymek** for her scientific advice and conversations during her time in Hagerman Lab.

I received support and assistance outside of Hagerman Lab as well. I thank **Kyoungmi Kim**, professor of biostatistics, and **Matthew Ponzini**, senior statistician, in the UC Davis School of Medicine. They devised several of the statistical approaches in this dissertation and spent many hours explaining them to me. I am much more confident in statistics due to their efforts.

I sincerely thank Professor **Frédéric Chédin** for co-mentoring me and introducing me to **Lionel Sanz**, a Ph.D. scientist in the Chédin Lab during the R-loop project. Fred taught me that biological science with human donor cells is messy and that I should not expect perfection. The goal is to answer the question being investigated. Lio trained me in measuring R-loops and gave extensively of his time, advice, and labor, especially when he completed measurements on additional human samples. Thank you, Lio.

I thank **Flora Tassone**, a UC Davis MIND Institute professor who studies *FMR1*-related disorders. Flora generously shared her data, methodologies, and advice. I value her input and collaboration. I also thank **Randi Hagerman**, a medical doctor and the director of the MIND Institute. Her clinical work brought in the patients whose samples I studied for this dissertation. Importantly, I acknowledge those individuals who donated their time and tissues. Our research would not be possible without your contributions.

I am grateful for the support of my graduate program, particularly from **Sean Burgess**, Integrative Genetics and Genomics (IGG) co-chair, and **Najwa Marrush**, IGG graduate coordinator. They have helped me navigate multiple unique and time-sensitive situations.

Thank you to **Denneal Jamison-McClung**, the director of the Designated Emphasis in Biotechnology (DEB) program, for her attentive support and flexibility. She advised and provided feedback on my biotech internship application materials.

Lastly, I thank the collaborative and responsive staff and researchers at UC Davis. Every time I have reached out for assistance, I received it, whether it involved borrowing reagents and materials, sharing protocols, or seeking answers to administrative questions. Your support smoothed my academic journey.

Abstract

The human fragile X messenger ribonucleoprotein 1 (*FMR1*) gene contains an expansion-prone trinucleotide CGG repeat that alters gene expression, giving rise to various central nervous system (CNS)-centric disorders. Alleles with repeats between ~55-200 CGGs, termed premutation alleles, predispose an individual to the adult-onset neurodegenerative disorder, Fragile X-associated tremor/ataxia syndrome (FXTAS). Typically, premutation alleles result in increased transcription, a slight decrease in protein product (FMRP), and cellular toxicity. Unfortunately, the exact molecular mechanism underlying FXTAS pathogenesis is unknown and no targeted treatments exist. *FMR1* alleles harboring expansions of more than 200 CGG repeats, termed full mutation alleles, trigger gene silencing. The resultant loss of FMRP expression causes Fragile X syndrome (FXS), the most common single gene form of autism and heritable intellectual disability. This dissertation describes investigations of aspects of *FMR1* gene expression in FXTAS and FXS to assess models of pathogenesis and to further evaluate the interplay of various relevant molecular features, respectively.

Chapter 1 of this dissertation addresses one aspect of the DNA damage repair (DDR) model of FXTAS pathogenesis, which proposes that unresolved DDR events triggered by R-loop formation result in cellular toxicity and death. R-loops are generated during transcription when nascent mRNA hybridizes with the template DNA strand, exposing the single stranded DNA to damage. DNA-RNA immunoprecipitation (DRIP)-qPCR was used to quantify *FMR1* R-loop abundance in primary fibroblasts from control and premutation carriers with and without FXTAS. Importantly, this study is the first known effort to characterize R-loops in FXTAS patients. *FMR1* was determined to be a low to moderate R-loop producing locus with an average steady-state frequency of ~2.5%. Furthermore, no associations between steady-state R-loop frequency and CGG repeat size, *FMR1* mRNA level, or FXTAS were found. Therefore,

R-loop frequency may not substantially trigger DDR in FXTAS although the DDR model more broadly should be further investigated.

Chapter 2 of this dissertation studies the effect of CGG repeat size and methylation status on *FMR1* protein (FMRP) production in FXS. FMRP expression in control and full mutation primary peripheral blood mononuclear cells (PBMCs) was quantified by time-resolved fluorescence resonance energy transfer (TR-FRET). *FMR1* mRNA and methylation status were quantified by qRT-PCR and southern blot, respectively. Most full mutation PBMCs displayed a high degree of size and methylation mosaicism, which was necessary for significant FMRP production. Interestingly, mRNA and repeat size of unmethylated alleles successfully modeled non-significant levels of FMRP, suggesting that FMRP was present, but below the significance threshold for many samples. Thus, a more sensitive assay may provide deeper insights into the interaction between these molecular characteristics. The results presented here highlight the continued need to investigate the molecular mechanisms by which altered expression leads to *FMR1*-associated disorders.

Table of Contents

List of Tables and Figures

Chapter 1

R-loop frequency does not associate with FXTAS in patient fibroblasts

Jamie L. Randol, Lionel A. Sanz, Paul J. Hagerman, Frédéric L. Chédin

Contributions

Lionel Sanz provided training in the DRIP-qPCR protocol and advice when challenges arose. He also completed DRIP on three primary fibroblasts lines for this study. Paul Hagerman provided the primary fibroblast cells, supported experiments to modify the DRIP protocol, and helped inform the discussion. Frédéric Chédin welcomed me to train on and conduct DRIP in his lab and consulted on data analyses and figures. All other aspects of this project, including study design, DRIP for six primary fibroblasts, figure generation, statistical analyses, and discussion were completed by Jamie Randol. I thank Glenda Espinal and Cris Iwahashi for establishing the primary fibroblasts in Hagerman Lab that were eventually used in this study. I also thank Flora Tassone who supplied the qRT-PCR protocol along with the necessary primer and probe sets. She also compared fibroblast genotype data in this study to PBMC genotype data generated by her lab from the same individuals to confirm genotype determinations. Finally, thank you to Robert Stolz who compared the R-loop energetics of control and premutation *FMR1* alleles.

This research was sponsored by NIH R01 grants GM113929 and HD036071, and Randi and Paul Hagerman's grant from the Azrieli Foundation.

Introduction

FMR1-associated disorders

The human *Fragile X messenger ribonucleoprotein 1 gene* (*FMR1*) contains a CGG trinucleotide repeat in its 5' untranslated region (UTR). Therefore, the repeat is transcribed into

mRNA, but is not translated into *FMR1* protein (FMRP). Intriguingly, expansion of the CGG repeat to different degrees results in different heritable disorders based on gene expression. Non-pathogenic control alleles comprise ~5-45 CGGs. Full mutation alleles contain greater than 200 CGGs, which generally results in gene silencing and the neurodevelopmental disorder, Fragile X Syndrome (FXS), which is the leading genetic cause of intellectual disability (R. J. Hagerman et al., 1994; Pieretti et al., 1991; Richter & Zhao, 2021; Schmitt et al., 2019). As an aside, Chapter 2 of this dissertation discusses protein production in FXS in more detail.

Premutation alleles contain ~55-200 CGG repeats and generally result in a moderate to high increase (2 to 8-fold) in transcription with no change or a seemingly paradoxical modest decrease in FMRP (R. J. Hagerman & Hagerman, 2016; Tassone, Hagerman, Taylor, Gane, et al., 2000; Tassone, Hagerman, Taylor, Mills, et al., 2000). Premutation alleles are associated with multiple disorders: Fragile X-Associated Primary Ovarian Insufficiency (FXPOI), defined as menopause before the age of 40 years (Fink et al., 2018); Fragile X-Associated Neuropsychiatric Disorders (FXAND), an umbrella term for many neuropsychiatric disorders, including depression and anxiety (R. J. Hagerman et al., 2018); and Fragile X-Associated Tremor/Ataxia Syndrome (FXTAS), an adult-onset neurodegenerative disorder, generally with onset after 50-55 years of age (Cabal-Herrera et al., 2020; R. J. Hagerman & Hagerman, 2021). Chapter 1 of this dissertation specifically addresses FXTAS.

FXTAS prevalence and symptoms

Globally, as many as 1 in 150 women and 1 in 400 men carry the premutation allele that predisposes individuals to FXTAS. Of these, up to 20% of women and 75% of men will develop FXTAS in their lifetime (R. J. Hagerman & Hagerman, 2016, 2021; Jacquemont et al., 2004; Rodriguez-Revenga et al., 2009). Of note, FXTAS tends to be less common and less severe in women due to the presence of a second X-chromosome with a control allele that is active in a subpopulation of neurons and can, therefore, provide a protective effect.

Common symptoms of FXTAS, which are generally associated with brain atrophy and white matter disease, include intention tremor, parkinsonian features, cerebellar ataxia, deficit in memory and executive function, depression, and anxiety (R. J. Hagerman et al., 2001; R. J. Hagerman & Hagerman, 2016; Jacquemont et al., 2003). Additionally, many patients experience serious co-morbidities, such as congestive heart failure, hypertension, migraine, neuropathy, thyroid disease, fibromyalgia, sleep apnea, central pain sensitivity syndrome, and autonomic dysfunctions, including incontinence and impotence (Cabal-Herrera et al., 2020; R. J. Hagerman & Hagerman, 2016; Jacquemont et al., 2003).

Its prevalence and debilitating symptoms make FXTAS a global health concern. Unfortunately, despite Hagerman and colleagues first reporting FXTAS in male premutation carriers in 2001 and approximately two decades of research uncovering small parts of the molecular puzzle, the exact molecular mechanism of FXTAS pathogenesis is unknown and no specific targeted treatment for FXTAS exists (R. J. Hagerman et al., 2001; R. J. Hagerman & Hagerman, 2021). Instead, management includes supporting a healthy lifestyle and treating symptoms.

Models of FXTAS pathogenesis

A model of FXTAS pathogenesis that has gained attention in recent years is the repeat-associated non-ATG (RAN) translation model (Todd et al., 2013). Expanded CGG repeat mRNA is known to form secondary structures, such as hairpins and G-quadruplexes, that impede ribosomal scanning during translation (Feng et al., 1995; Ludwig et al., 2011; Yousuf et al., 2022). Therefore, this model proposes that translation of expanded *FMR1* mRNA is partially shifted to out-of-frame, non-canonical start sites upstream of the expanded CGG repeat, resulting in several peptides of repetitive amino acids. One such peptide, termed FMRpolyG, is composed of repeating glycine amino acids and is speculated to be toxic (Glineburg et al., 2018; Sellier et al., 2017; Todd et al., 2013).

Buijsen and colleagues identified the presence of FMRpolyG in patient brain inclusions, a hallmark feature of FXTAS, via immunostaining (Buijsen et al., 2014). However, recent evidence suggests that the FMRpolyG peptide, although detected by immunofluorescence, exists in only trace amounts (~10-40 ppm) (Ma et al., 2019). To date, endogenous FMRpolyG in FXTAS brains has only been detected via direct protein sequencing in minute quantities (0 - 0.04% molar abundance of total protein) following significant enrichment (Ma et al., 2019). Moreover, no FMRpolyG was detected by protein sequencing in whole brain lysates or nuclear preparations, in multiple studies (Holm et al., 2021; Ma et al., 2019). Thus, it is unclear how such small quantities of FMRpolyG could initiate FXTAS neuropathology. The RAN model has been discussed in detail elsewhere (P. J. Hagerman, 2013; Holm et al., 2021; Ma et al., 2019) and will not be considered further in this study.

Transcription-based models of FXTAS pathogenesis

Evidence to support a transcription-based pathogenic trigger is strong. Significantly, the CGG repeats that give rise to FXTAS are located in the transcribed, but untranslated 5' region of the gene. Additionally, premutation carriers have a two to 10-fold elevation in *FMR1* mRNA with only modest reduction in *FMR1* protein (FMRP) in peripheral blood mononuclear cells (PBMCs), an accessible cell type (Tassone, Hagerman, Taylor, Gane, et al., 2000). Furthermore, cellular toxicity only results from the increased transcription of premutation repeats; neither 100-fold overexpression of a control repeat nor the presence of non-transcribing premutation repeat DNA cause cellular toxicity (Hoem et al., 2011). Nevertheless, how mRNA triggers toxicity is largely unknown.

One model of *FMR1* mRNA toxicity is the mRNA sequestration model, in which expanded *FMR1* mRNA binds excessively to various RNA binding proteins leading to a functional deficit of those proteins (R. J. Hagerman & Hagerman, 2016). This paradigm is true of the heritable disorder, myotonic dystrophy type 1, which primarily results in progressive muscle

atrophy and weakness. The *myotonic dystrophy protein kinase* gene (*DMPK*) contains a CTG trinucleotide repeat in its 3'-UTR. mRNA from expanded alleles of *DMPK* sequester RNA processing factors, like muscleblind-like 1 protein, that block normal cellular function and result in splicing defects (Lanni & Pearson, 2019). In FXTAS, evidence suggests that DGCR8 protein, which forms a complex with Drosha to mediate microRNA (miRNA) production, is sequestered by expanded CGG-repeat *FMR1* mRNA, resulting in a global depression in the miRNA landscape (Sellier et al., 2013a). Other candidates for sequestration include Sam68, Purα, and hnRNPA2/B1 (Iwahashi et al., 2006; Jin et al., 2007; Sellier et al., 2010; Sofola et al., 2007). However, only DGCR8 is known to display a CGG-repeat-dependence of the binding affinity of the protein to the CGG-repeat-containing RNA (Sellier et al., 2013b).

The DNA damage repair (DDR) model of FXTAS

A second, non-exclusive model of *FMR1* mRNA toxicity in FXTAS is the DNA damage repair (DDR) model, which proposes that co-transcriptional R-loop formation results in a chronic DNA damage response (DDR) at the *FMR1* CGG repeat, which leads to cellular toxicity and ultimately cell death (R. J. Hagerman & Hagerman, 2016; Loomis et al., 2014). R-loops are three-stranded nucleic acid structures resulting from a G-rich RNA transcript hybridizing with its C-rich DNA template, creating a displaced single-stranded DNA (Crossley et al., 2019).

R-loops form ubiquitously throughout the genome, playing roles in gene expression, DNA replication and repair, and immunoglobulin class switching, among other roles (Sollier & Cimprich, 2015). However, aberrant R-loop regulation can lead to DNA damage and the DDR response, genomic instability, and disease by exposing the displaced DNA strange to damage, among other mechanisms (Beletskii & Bhagwat, 1996; Costantino & Koshland, 2015, 2018; Crossley et al., 2019; Groh & Gromak, 2014; Huertas & Aguilera, 2003; Skourti-Stathaki & Proudfoot, 2014; Sollier & Cimprich, 2015). For example, abnormal elevation of R-loops in vertebrate lymphoblasts results in double strand breaks (DSBs) and gross chromosomal

rearrangements (Li & Manley, 2005). The remainder of this chapter will address the R-loop instigated DDR model of FXTAS pathogenesis.

DNA damage and cellular stress in FXTAS

DNA damage has been linked to neurodegeneration, a key feature of FXTAS brains (Coppedè & Migliore, 2015). Therefore, it is not surprising that in various premutation and FXTAS model systems cells express markers of cellular toxicity and DNA damage, including altered Lamin A/C architecture; decreased cell viability; and increased ɣH2AX, oxidative stress, mitochondrial dysfunction, and calcium dysregulation (R. J. Hagerman & Hagerman, 2016; Hoem et al., 2011; Robin et al., 2017; Ross-Inta et al., 2010). For instance, in an inducible episomal model, transcribed *FMR1* premutation repeats, not control repeats, triggered the DSB repair pathway (Hoem et al., 2011).

Furthermore, protein analyses of the hallmark intranuclear inclusions in FXTAS brains reveal the presence of proteins in the autophagy pathway that point to cellular distress, such as αβ-crystallin, heat shock protein (hsp) 70, hsp27, and p62 (Iwahashi et al., 2006; Kurosawa et al., 2016). Finally, immunostaining in the Hagerman Lab has identified the presence of key DDR proteins (pATM, Ku86, ɣH2AX, 53BP1 and Mre11) in inclusions from FXTAS cortical neurons (Hoem et al., 2011; unpublished data).

R-loops at FMR1

Co-transcriptional R-loop initiation is favored at G clusters that carry three or four guanines sequentially (GGGG) (Roy & Lieber, 2009). Furthermore, R-loop elongation is favored in DNA that has a high GC skew, in which guanine nucleotides appear more frequently on the non-template strand, especially downstream of transcriptional start sites (Ginno et al., 2012). For example, DNA with G clusters formed greater than 20-fold more R-loops compared with dispersed guanines at the same G density (Roy & Lieber, 2009). Accordingly, CpG island (CGI)

promoters, which are GC-rich and GC-skewed and serve as promoters for ~60% of human genes, are common sites for R-loop formation. Indeed, R-loops at CGIs are thought to protect these genes from DNA methylation and epigenetic silencing (Ginno et al., 2012).

Given that *FMR1* possesses a CGI promoter, has multiple G clusters following its transcriptional start sites, and has a high GC content with a high GC skew across its 5' end, it is predicted to form R-loops (Loomis et al., 2014). Therefore, it is not surprising that R-loops have been detected at the endogenous *FMR1* locus. *FMR1* R-loops across the CGG repeat have been identified in individuals with expanded repeats in fibroblasts, patient-derived lymphoblastoid lines, human embryonic stem cells, and in an inducible episomal premutation model (Colak et al., 2014; Groh et al., 2014; Loomis et al., 2014).

R-loops in neurodegenerative and FMR1-associated disorders

R-loops have been implicated in multiple neurodegenerative disorders. For example, different mutations in the *SETX* gene, which purportedly resolves R-loops, are associated with two neurodegenerative disorders that cause neuronal and muscle cell death: ataxia oculomotor apraxia type 2 (AOA2) and amyotrophic lateral sclerosis type 4 (ALS) (Costantino & Koshland, 2015; Groh & Gromak, 2014; Moreira et al., 2004). Additionally, some evidence suggests that ataxia telangiectasia syndrome (AT), another neurodegenerative disorder, may be caused by aberrant R-loops that generate DSBs. Importantly, this has been shown in nonproliferating, post-mitotic neurons (Sordet et al., 2009, 2010).

R-loops have also been proposed to contribute to *FMR1*-related disorders. Specifically, R-loops in full mutation alleles may trigger the gene silencing that causes FXS (Colak et al., 2014; Groh et al., 2014). This is particularly interesting given that R-loops in the control and premutation range are suspected to protect the gene from methylation-coupled silencing (Ginno et al., 2012; Loomis et al., 2014). Furthermore, *FMR1* repeat instability that predisposes premutation and full mutation allele carriers to disease may result, in part, from R-loops.

Secondary structures, such as G-quadruplexes and hairpins, in the displaced ssDNA can stabilize the R-loop and cause DSBs, nicking, and other problems that induce aberrant recombination or repair, leading to variability in repeat size (Freudenreich, 2018; Groh & Gromak, 2014; McIvor et al., 2010). Similarly, Lin and colleagues showed that R-loops over CTG repeats engender repeat instability (Lin et al., 2010).

Study rationale and hypothesis

Considering the strong evidence that FXTAS is caused by a toxic expanded *FMR1* mRNA mechanism, the purported role of R-loops in both neurodegenerative and repeat expansion disorders, and the known presence of R-loops at *FMR1*, we sought to quantify *FMR1* R-loop frequency in premutation carriers with and without FXTAS compared to controls in an accessible cell type: primary epidermal fibroblasts. We hypothesized that R-loop frequency would be higher in premutation carriers compared to controls, with the highest levels in patients with FXTAS. Contrary to expectations, R-loop frequencies in FXTAS patients and controls were indistinguishable, while premutation carriers without FXTAS may present with elevated levels. Given this, the R-loop instigated DDR model of FXTAS pathogenesis may need to be reassessed.

Materials and Methods

Establishment of primary fibroblasts

Three mm, full thickness skin biopsies from the left posterior shoulder were collected from nine individuals recruited to the University of California Davis Health MIND Institute's Fragile X Clinical and Research Center between 2007 and 2012 as participants in studies of fragile X-associated disorders. Protocols were approved by the Institutional Review Board of the University of California, Davis. Participants submitted written consent for their biopsies to be

used in further research beyond the study for which they were originally collected. The biopsies were stored in RPMI 1640 (Gibco, Thermo Fisher Scientific, Waltham, MA, USA) culture medium at 4°C until processing within 24 hours. In aseptic conditions, biopsies were sliced into approximately 15 pieces and distributed to three culture dishes containing 60:40 RPMI 1640:Amnio MAX^{TM}-C100 Basal Medium (Gibco) with 15% Amnio MAX^{TM}-C100 Supplement (Gibco). Over the course of 2-3 weeks, fibroblasts were permitted to grow out from the cut biopsy pieces. These fibroblasts continued to be expanded up to passage seven. Cells were then cryopreserved using 10% dimethyl sulfoxide (DMSO) in 60:40 medium for future use.

Primary fibroblasts

As *FMR1* is an X-linked gene, primary fibroblasts in this study were taken exclusively from males to eliminate the confounding effects of the control *FMR1* allele in females. Fibroblasts were separated into three allele categories: control (n = 3), premutation without FXTAS (premutation; n = 3), and premutation with stage 4 FXTAS (FXTAS; n = 3). Control fibroblasts had a mean age of 62.3 ± 1.2 years at the time of biopsy. Premutation fibroblasts had a mean age of 64 ± 1.7 years. Finally, FXTAS fibroblasts had a mean age of 65 ± 3.5 years. No significant difference in ages among allele categories was identified by one-way ANOVA ($p = 0.422$). Table 1.1 identifies primary fibroblast characteristics.

***FMR1* CGG repeat size genotyping**

Primary fibroblasts were grown in culture flasks to 80% confluency then lysed in SDS and Proteinase K at 37°C overnight followed by phenol-chloroform extraction using MaXtrack phase lock tubes (Qiagen, Valencia, CA, USA) and ethanol precipitation. Rehydrated DNA was then fragmented by HindIII and EcoRI followed by PCR across the *FMR1* CGG repeat according to (Hayward et al., 2016). PCR products were then run on an agarose gel and sized

using Fluorchem 8000 software (Fig. 1.1). CGG repeat size was calculated as (fragment size – 269)/3. See Table 1.1 for repeat sizes for all primary fibroblasts.

FMR1 mRNA expression levels via qRT-PCR

Total RNA was isolated from primary fibroblasts grown in culture to 80% confluency and quantified via NanoDrop. cDNA synthesis and determination of *FMR1* mRNA expression levels were performed using real-time PCR (qRT-PCR) as described in (Tassone, Hagerman, Taylor, Gane, et al., 2000). The β-glucuronidase (*GUSB*) reference gene was used to correct *FMR1* mRNA levels. See Table 1.1 for *FMR1/GUSB* mRNA levels for all primary fibroblasts.

R-loop frequency via DNA-RNA Immunoprecipitation (DRIP)

Fibroblast cells were lysed by SDS and Proteinase K treatment at 37°C overnight followed by phenol-chloroform extraction using MaXtrack phase lock tubes (Qiagen, Valencia, CA, USA) and ethanol precipitation. Rehydrated DNA was treated overnight at 37°C using HindIII, EcoRI, BsrGI, and XbaI (NEB, Ipswich, MA, USA) to sufficiently fragment genomic DNA and avoid cutting the *FMR1* locus. Another round of phenol-chloroform extraction and ethanol precipitation in the presence of glycogen was performed to purify fragmented DNA from the digest reaction. A subset of the digested gDNA was stored as input for later qPCR analysis. Subsequently, R-loops were immunoprecipitated (IP) by adding 4.4 ug of digested DNA to 10 ug of RNA:DNA hybrid-recognizing S9.6 antibody (a gift from the Chédin Lab at University of California Davis, USA). The antibody-R-loop complexes were incubated with rotation for 16 hours at 4°C then bound by Protein G Magnetic beads (NEB, Ipswich, MA, USA) and thoroughly washed. The bound fraction was eluted by Proteinase K digestion and purified by a final round of phenol-chloroform extraction and ethanol precipitation.

qPCR was performed using Sso Universal SYBR green supermix (Bio-Rad, Hercules, CA, USA) for 40 cycles at 60 °C targeting up to four loci with 1-2 qPCR replicates per IP. *FMR1*

was the locus of interest. *RPL13A* served as a high-frequency positive control locus and generally shows R-loop frequencies corresponding to ~10-12% of input (Sanz et al., 2016). The *EGR1* locus serves as one of two negative controls. A segment of the *SNRPN* gene serves as the second negative control. Negative control loci typically have R-loops in the range of 0.01-0.1% input (Sanz et al., 2016). The *EGR1* locus (also referred to as 83-84 for the primers used to amplify it) showed larger background R-loop frequency in this study because one restriction enzyme was eliminated from the standard enzyme cocktail to prevent cutting *FMR1*.

The relative enrichment of R-loops (Output) at each locus was determined as the percent of fragmented gDNA before immunoprecipitation (Input) using the following formula:

$$\%\ Input = \frac{\frac{1}{9} \times 2^{(-Cq\ Output)}}{2^{(-Cq\ Input)}} \times\ 100\%$$

Output was corrected for relative loading, as 9-fold more DNA was immunoprecipitated compared to the Input.

Data processing

Primary fibroblasts were grown for DRIP on 1-2 occasions. Therefore, standard deviation (SD) and standard error of the mean (SEM) of R-loop frequency, which require a minimum of three replicates, were not calculable for separate growth events. Each growth resulted in 1-4 IPs per fibroblast line. And each IP was assessed with 1-2 qPCR replicates per target locus. As a result, to maximize the number of fibroblast lines with calculable SDs and SEMs, statistical analyses were run at the IP level, despite some growth events having variable number of IPs for the same line. Of note, primary fibroblast 1034-09 was assessed with a single IP from a single growth event. Also of note, Dr. Lionel Sanz, previously from the Chédin Lab, performed a subset of the DRIP runs to complete these data. All IPs, whether processed by Jamie Randol or Lionel Sanz, were included together.

RStudio software (version 2023.03.0+386) was used to quantify R-loop frequency and generate figures. R-loop frequency was represented four ways: 1) raw % input (see calculation above), 2) relative to the positive control locus *RPL13A* to account for differences in extraction efficiency, 3) raw % input corrected to *FMR1* mRNA level, and 4) relative R-loops corrected to *FMR1* mRNA level. SD for mRNA-corrected R-loops was calculated by error propagation. See Table 1.2 for all quantifications of R-loop frequency.

Results

FMR1 R-loop frequency is independent of CGG repeat size

R-loop frequency at *FMR1*, *RPL13A* (positive control), *EGR1* (83-84; negative control 1), and *SNRPN* (negative control 2) were plotted against *FMR1* CGG repeat size. Linear regression analyses showed that both raw (Fig. 1.2) and relative (Fig. 1.3) *FMR1* R-loops were independent of *FMR1* CGG repeat size between 20 and 119 repeats in human epidermal fibroblasts (Raw: n = 9, p-value = 0.49; Relative: n = 9, p = 0.55). That is, R-loops at *FMR1* did not increase in frequency as allele size increased. *FMR1* R-loops among all allele categories ranged from 1.91 - 3.26 % input, with a mean of 2.53 ± 0.15 (SEM). Relative *FMR1* R-loops ranged from 0.19 - 0.46 the level of those at the *RPL13A* positive control locus, with a mean of 0.28 ± 0.03 (SEM). Essentially, on average, *FMR1* R-loops were 28% the level of *RPL13A* R-loops. Importantly, the frequency of *RPL13A* R-loops is also independent from CGG repeat size, bolstering their use as a normalizer for *FMR1* R-loops (Fig. 1.2; n = 9, p-value = 0.67). Expectedly, the *SNRPN* negative control locus was also independent of CGG repeat size for both raw and relative R-loops (Fig. 1.2; n = 5, Raw p-value = 0.90, Relative p-value = 0.99).

However, the *EGR1* (83-84) negative control locus did slightly, albeit significantly, increase as CGG repeat size increased, but only for raw R-loop frequency (n = 6, p-value = 0.04). Relative R-loops at *EGR1* did not significantly increase with CGG repeat size (n = 6, p-

value = 0.22). Given this unexpected behavior, the *SNRPN* negative control locus was used partway through the experiment as a true negative control. Its R-loops were 0.04 ± 0.00 (mean ± SEM) and did not vary with CGG repeat size.

FMR1 R-loop frequency is independent of *FMR1* mRNA level

As R-loops are a product of transcription, *FMR1* mRNA levels were measured for all nine primary fibroblasts in this study. A one-way ANOVA by allele category showed no statistical differences in *FMR1* mRNA among the three allele types (p-value = 0.168) (Fig. 1.4). However, FXTAS samples tended to have higher mRNA, by visual inspection. *FMR1* R-loops were plotted against *FMR1* mRNA levels, and a linear regression was performed. Interestingly, neither raw (Fig. 1.5) nor relative (Fig. 1.6) *FMR1* R-loop frequency associated with *FMR1* mRNA level, though both approached significance (Raw: n = 9, p-value = 0.063; Relative: n = 9, p-value = 0.12).

Next, *FMR1 R-loops* were corrected to their respective *FMR1* mRNA levels (R-loop/mRNA) and plotted by their CGG repeat size. A linear regression analysis was performed. Again, no significant association between corrected raw (Fig. 1.7) nor corrected relative (Fig. 1.8) *FMR1* R-loops and CGG repeat size was observed (Corrected raw: n = 9, p-value = 0.81; Corrected relative: n = 9, p-value = 0.97).

Relative *FMR1* R-loop frequency is significantly higher in premutation fibroblasts

While linear regression analyses of all nine primary fibroblasts together revealed no significant associations between *FMR1* R-loops and CGG repeat size nor *FMR1* mRNA level, a one-way ANOVA of relative *FMR1* R-loops by allele category found a significant difference among allele types (n = 9, p-value = 0.036) (Fig. 1.9). Pairwise differences were investigated with Tukey's Honest Significant Difference. R-loop frequency in premutation samples (n = 3) as a group was found to be almost significantly higher than those in both control (n = 3, p-value =

0.054) and FXTAS (n = 3, p-value = 0.051) samples. No difference was found between FXTAS and control samples (p-value = 1.00). The same analysis in *FMR1* mRNA-corrected relative R-loop frequency approached significance, again, with premutation fibroblasts having the largest R-loop frequency (p-value = 0.070) (Fig. 1.10). Indeed, premutation samples (in red) almost appear as a separate group above the best fit lines in figures 1.6 and 1.8. Of note, the frequency of raw *FMR1* R-loops, whether or not corrected to *FMR1* mRNA, showed no significant differences among allele categories by one-way ANOVA (Raw: p-value = 0.251; Corrected raw: p-value = 0.226).

Discussion

The exact molecular mechanism of pathogenesis underlying FXTAS is still unknown (R. J. Hagerman & Hagerman, 2021). One model of FXTAS pathogenesis posits that R-loop formation in premutation repeats results in an unresolved DDR at the *FMR1* locus, which ultimately leads to cellular toxicity and death (R. J. Hagerman & Hagerman, 2016; Loomis et al., 2014). In the current study, we sought to investigate a facet of the R-loop instigated DDR model by quantifying the frequency of R-loops at *FMR1* in primary human fibroblasts from individuals with control alleles and premutation alleles, with and without FXTAS.

The current study identified that *FMR1* is a medium to low R-loop producing locus. R-loop-positive loci typically have R-loops in the range of ~1-12% input when quantified via DRIP-qPCR (Sanz et al., 2016; personal comments of Prof. Frédéric Chédin). The average raw R-loop frequency across all nine primary fibroblasts in this study was 2.53 ± 0.44% input (mean ± SD), slightly higher than the lower limit. In contrast, the highly positive control locus, *RPL13A*, had an average R-loop frequency of 9.53 ± 1.30% input (mean ± SD) for all nine primary fibroblasts. In addition to raw % input, the current study also reported R-loop frequency relative

to the highly positive *RPL13A* locus. The mean relative *FMR1* R-loop frequency was 0.28 ± 0.08 (mean ± SD), signifying that, on average, *FMR1* R-loops are 28% the level of *RPL13A* R-loops.

Three previous studies measuring R-loop frequency at *FMR1* employed different cell types, allele sizes, and/or assay and reporting methods, making quantitative comparisons between these studies impossible. Colak et al. (2014) measured R-loop frequency in human embryonic stem cells (hESCs) that carried full mutation alleles with greater than 400 CGGs both before and after transcriptional silencing of *FMR1*. Moreover, they used the chromatin isolation by RNA purification (ChIRP) technique, in which crosslinked mRNA is pulled down via *FMR1*-specific biotinylated probes followed by qPCR to quantify the degree of bound *FMR1* promoter DNA relative to input (no pulldown). RNase H, which digests RNA-DNA hybrids, was used to verify that the ChIRP signal was indeed mostly R-loop. Two active full mutation hESCs showed extremely low R-loop frequencies: ~0.07 and ~0.095% input. Interestingly, two premutation hESCs with 70 and 73 CGGs and a control hESC were assayed as well and no evidence of R-loops at the *FMR1* promoter was found. Yet, Loomis and colleagues detected R-loops at *FMR1* in control fibroblasts using a different assay (Loomis et al., 2014). Therefore, the lack of R-loops in (Colak et al., 2014) may be due to the stage of development, cell type, or assay employed.

Similarly, (Groh et al., 2014) quantified R-loop frequency for a full mutation *FMR1* allele. However, they used transformed human lymphoblastoid cells and reactivated the silenced full allele with 5'aza-2'-deoxycytidine to 25% the level of control cells. Additionally, Groh and colleagues quantified R-loops using DNA immuno-precipitation (DIP) rather than DRIP. DIP is similar to DRIP in that both use the S9.6 antibody to pull down RNA-DNA hybrids. However, apart from some buffer and incubation changes, the main discrepancy is that DIP uses sonication rather than restriction enzyme digestion to fragment genomic DNA prior to immunoprecipitation. All DIP signals were calculated relative to that of untreated control cells at exon 1. R-loop frequency was similar in both control and untreated full mutation cells, yet ~3.5-fold higher in reactivated full mutation lymphoblastoid cells. As all data is presented relative to

the control locus, it is unclear what the raw R-loop frequency was. However, untreated full mutation cells are silenced and therefore would not produce co-transcriptional R-loops. Therefore, presumably, R-loop frequency of control lymphoblastoid cells is also quite low or non-existent using DIP with their primers.

Finally, (Loomis et al., 2014) measured R-loop frequency in primary fibroblasts from individuals with control and premutation alleles, similar to the current study. However, no distinction between premutation samples with and without FXTAS was made. While, Loomis and colleagues used DRIP-qPCR, as in the current study, they reported R-loop frequencies relative to a known negative locus that does not produce R-loops. They did not report raw % input. Despite high variability, premutation fibroblasts as a group (n = 4) had significantly more R-loops than control fibroblasts (n = 3). Premutation samples had a mean R-loop frequency of 9 ± 3.9 fold enrichment (mean ± SD), with a range of 2.9 – 13.4; while control samples had a frequency of 4.2 ± 2.4 fold enrichment, with a range of 1.6 – 8.6.

In summary, *FMR1* R-loop frequency measurements across multiple studies to date are challenging to compare due to discrepancies in cell type, allele class, and methodology, and due to large variation between samples. However, all show that R-loops can form at *FMR1*. Moving forward, reporting raw % inputs, whether for DRIP, DIP, or ChIRP, would help correlate R-loop data from different studies with the benefit of providing the absolute signal strength and thus an indication of the reaction efficiency (Chédin et al., 2021).

The main finding from this study is that there is no evidence that FXTAS leads to increased *FMR1* R-loop frequency at steady-state. Indeed, *FMR1* R-loop frequency was independent of CGG repeat size and *FMR1* mRNA level in linear regression analyses (Figs. 1.2, 1.3, 1.5, and 1.6). Consistently, *FMR1* R-loop frequency corrected to *FMR1* mRNA level was also independent of CGG repeat size (Figs. 1.7 and 1.8). This result was unexpected for four reasons. First, previous studies found that active expanded repeat samples have higher R-loop frequencies than control samples, as discussed above. Briefly, neither (Colak et al., 2014) nor

(Groh et al., 2014) identified significant R-loops in control samples, while active full mutation alleles did produce detectable R-loop signals. Furthermore, (Loomis et al., 2014) found significantly higher R-loop frequencies in premutation fibroblasts compared to controls. Of note, all studies had small sample sizes. For example, Groh and colleagues used a single CGG expansion sample (Groh et al., 2014). And Loomis and colleagues showed a high degree of variation in R-loop frequency among all seven fibroblasts. One control fibroblast, for instance, had R-loop frequency approximately equal to that of three of the four premutation samples (Loomis et al., 2014). Therefore, previous studies may not best represent the spectrum of R-loop frequencies at *FMR1*.

Second, transcription of premutation alleles is partially shifted upstream compared to control alleles (Beilina et al., 2004; Tassone et al., 2011). Upstream *FMR1* sequence contains more G-clusters, which favor R-loop nucleation (Roy & Lieber, 2009). Third, premutation alleles (~55-200 CGGs), by definition, contain more guanine-rich sequence than their control counterparts (~5-45 CGGs). Moreover, that G-richness is skewed to the non-template strand. Both are favorable predictors of R-loop elongation (Ginno et al., 2012; Roy & Lieber, 2009). Additionally, G-rich RNA hybridized to C-rich DNA, as is the case with *FMR1* R-loops, is more stable than double-stranded DNA (Ratmeyer et al., 1994; Roberts & Crothers, 1992). It follows that longer CGG repeats could result in more stable R-loops. Thus, R-loop frequency could, but not necessarily, mirror R-loop stability.

Counterintuitively, some recent evidence indicates that longer CGG repeat tracts may be less favored to form R-loops. Stolz and colleagues developed a statistical mechanical equilibrium model of how DNA topology and base sequence stabilize R-loops based on the energy of various base pairings, superhelicity, and torsional winding of the displaced DNA strand (Stolz et al., 2019). Dr. Robert Stolz was kind enough to run control and premutation *FMR1* sequences through his model and reported that longer CGGs are less favored to form R-loops due to the relative unfavorable energetics associated with the CG dinucleotide

(unpublished personal comment). This may explain, in part, why R-loop frequency was not significantly higher in premutation samples in the current study. However, predicting energies for non-Watson-Crick base pairs is challenging and results from models that rely on those energies should be assessed carefully (Zuber & Mathews, 2019).

Fourth and finally, co-transcriptional R-loops have been shown to increase as transcription increases. In an inducible episomal model, higher levels of transcription resulted in higher R-loop formation, proportional to the concentration of inducer for both a control and a premutation repeat (Loomis et al., 2014). Given that expanded repeats produce ~2 to 8-fold more *FMR1* mRNA than control alleles in peripheral blood mononuclear cells (PBMCs), we predicted that samples with premutation alleles would produce more R-loops as well (Tassone, Hagerman, Loesch, Lachiewicz, et al., 2000; Tassone, Hagerman, Taylor, Gane, et al., 2000).

However, unexpectedly, no significant difference in *FMR1* mRNA was found in this study by one-way ANOVA among fibroblasts from controls (n = 3) and premutation carriers with (n = 3) and without (n = 3) FXTAS (Fig. 1.4). This could be due to natural variation among the population, a small sample size, and the choice of cell type. Previously, (Garcia-Arocena et al., 2010) reported a modest increase in transcription with high variability for premutation fibroblasts: 2.25 ± 1.05-fold increase relative to controls (mean ± SD) for asymptomatic premutation carriers and 2.95 ± 1.19 for carriers with FXTAS. This parallels the 2.5-fold overexpression of *FMR1* mRNA in the frontal cortex of FXTAS patients (Garcia-Arocena et al., 2010). Therefore, while differences in *FMR1* mRNA in fibroblasts are less extreme than in PBMCs, they more closely reflect changes observed in patient brains and may be more relevant for the study of FXTAS R-loops. Regardless, data from the current study show that for the same level of transcription, there is no significant differences in R-loop frequency for FXTAS and control fibroblasts.

Another finding of the current study is that premutation fibroblasts without FXTAS had significantly higher relative R-loop frequency compared to both controls and premutation

samples with FXTAS by one-way ANOVA (Fig. 1.9). Notably, the level of significance was small. The p-value of 0.036 was just under the significance level of 0.05 when comparing all three groups simultaneously. Additionally, no significant differences between the three groups were found when pairwise differences were investigated with Tukey's Honest Significant Difference (HSD). Although, more frequent R-loops in asymptomatic premutation samples approached significance compared to both controls and samples with FXTAS (Fig. 1.9). Moreover, no other quantification of R-loops (raw, raw corrected to mRNA, or relative corrected to mRNA) was significant by one-way ANOVA. Taken together, this indicates that the finding of higher premutation R-loops is not robust.

Regardless, the finding should not be ignored. Overall, relative R-loops (raw R-loops normalized to the highly positive *RPL13A* locus) improved the mean variation in R-loop frequency, presumably by accounting for variable extraction and IP efficiencies among samples. This indicates that relative R-loops may be more appropriate than raw R-loops to use for statistical analyses, such as one-way ANOVA. And while no clear mechanism by which asymptomatic premutation carriers produce more abundant R-loops compared to those with FXTAS is forthcoming, we can broadly speculate on potential processes. For example, premutation carriers with and without FXTAS may possess a different milieu of R-loop binding partners that stabilize the *FMR1* R-loop in asymptomatic carriers or help resolve the R-loop in patients with FXTAS. It is also conceivable that R-loops may help protect against FXTAS. It has already been posited that *FMR1* R-loops protect the premutation allele from epigenetic silencing that would lead to FXS (Loomis et al., 2014). Likewise, they may not be the villains we suspect when it comes to FXTAS. (Castillo-Guzman & Chédin, 2021) defined two classes of R-loops: promoter-paused (Class I) and elongation-associated (Class II). They propose that Class I R-loops are more likely to cause genomic instability and cellular toxicity. Interestingly, *FMR1* R-loops more closely resemble Class II elongation-associated R-loops that tend to be longer

(200bp – 2kb), have lower frequencies (1-10%), and be less pathological than class I R-loops (<60bp and >10% frequency).

Study strengths

There are a couple notable strengths to the current study. To the best of our knowledge, this is the first report of R-loop frequencies in patient samples from premutation carriers with and without FXTAS. Previously *FMR1* R-loop studies focused on full mutation alleles or premutation samples without distinguishing clinical outcome.

Another strength is the use of a modified DRIP protocol to improve ease and reproducibility for new DRIP users. The current study substituted protein A/G agarose beads used in the original DRIP protocol from the Chédin Lab with protein G magnetic beads to capture the S9.6 during immunoprecipitation. Magnetic beads are less susceptible to error, especially with inexperienced DRIP users. S9.6-bound R-loops are captured by a magnet on the side of the tube via the magnetic beads, which allows for complete removal of supernatant and subsequent washes from the tube bottom. Alternately, agarose beads require spinning at low to medium speed for short lengths of time to avoid being compressed. Since the beads are spun to the tube bottom, it is impossible to completely remove the supernatant. Moreover, less experienced users can easily aspirate some agarose beads, and thus R-loops, while removing supernatant. When tested in the Hagerman Lab, protein G magnetic beads performed equal to or better than protein A/G agarose beads (unpublished data). Interestingly, when tested in the Chédin Lab by DRIP expert Dr. Lionel Sanz, protein A/G agarose beads performed better (unpublished data). The difference likely lies in user experience, supporting the use of magnetic beads for those new to DRIP.

Study limitations

There are several limitations to this study. First is the small sample size of three primary fibroblasts per allele category. DRIP (and the highly similar DIP) is a highly published method of R-loop quantification (Castillo-Guzman & Chédin, 2021; Chédin et al., 2021; Groh et al., 2014; Loomis et al., 2014; Sanz et al., 2016). However, it is a multiday, labor-intensive method with a high learning curve. Moreover, when followed by qPCR, DRIP is low throughput. A year of effort to reduce the time and labor of the assay yielded poorer R-loop frequencies when compared to the original DRIP method. Therefore, efforts to modify the protocol were abandoned and the study proceeded with the original DRIP assay, except for the substitution of a magnetic bead step. As such, few samples could be processed at one time. Unfortunately, funding for the study ended after only two samples per allele class were quantified for 1-2 growth events. Dr. Lionel Sanz, a DRIP expert previously from the Chédin Lab, volunteered to complete DRIP on additional samples to achieve three primary fibroblasts per allele category, enabling statistical analyses.

Another consequence of few samples per allele category is that sample repeat sizes did not cover the full premutation range (~55-200 CGGs). The current study examined premutation and FXTAS samples with between 65 and 119 CGG repeats. Therefore, we cannot comment on R-loop frequencies between 119 and 200 CGGs.

A second limitation of the study is the low number of replicates used to determine R-loop frequency for each sample. One to two (1-2) qPCR replicates were used for each of one to four (1-4) IPs per each of one to two (1-2) fibroblast growth events. As a result, R-loop frequency for one primary fibroblast (1034-09) was determined from only two qPCR replicates from a single IP from a single growth event. More accurate determinations of R-loop frequency require more replicates at all levels (growth, IP, and qPCR). Unsurprisingly, some fibroblasts had large variation in R-loop frequency, which is not uncommon (Chédin et al., 2021). Unfortunately, this large variation made identifying statistically significant differences between groups challenging.

Regardless, large variation is not uncommon when processing biological samples. Similar variation was observed in fibroblast R-loop frequencies in (Loomis et al., 2014).

A third limitation of this study is the lack of RNase H and spike-in controls to validate R-loop recovery and both extraction and IP efficiency, respectively. RNase H is an endonuclease that degrades the RNA moiety of R-loops and other RNA:DNA hybrids. Therefore, if RNase H is applied to genomic DNA prior to immunoprecipitation by the S9.6 antibody that recognizes RNA:DNA hybrids, little to no signal should be detected by DRIP-qPCR, validating that the signal quantified in the absence of RNase H is a true R-loop signal. RNase H controls are considered mandatory by many in the field (Chédin et al., 2021; Crossley et al., 2020). Unfortunately, this study ended before RNase H controls could be tested. However, previous research on primary fibroblasts in the Hagerman Lab at UC Davis showed that R-loops at *FMR1* were significantly reduced by the addition of RNase H (Loomis et al., 2014). Therefore, R-loop frequencies quantified in this study are assumed to represent true R-loop signal.

A spike-in RNA:DNA control would serve as a true measure of extraction and immunoprecipitation efficiency. A known volume or amount of an *in vitro* transcribed plasmid with a propensity to form R-loops (such as pFC53 from the Chédin Lab) could be added to all samples during sample preparation and carried through the experimental workflow, including immunoprecipitation. Measuring spiked-in R-loops via unique qPCR primers before and after the workflow would provide a ratio by which to correct for efficiency of each IP. More details on a variation of this method can be found in (Crossley et al., 2020).

Rather than correct to a spike-in control, the current study corrected to the raw percent input of the highly positive *RPL13A* locus, which generally has R-loop frequencies of ~8-12% (personal comments from Dr. Lionel Sanz) (Sanz et al., 2016). Table 1.2 shows that all nine primary fibroblasts had similar raw *FMR1* R-loop frequencies close to 2.5 % input regardless of allele category. Indeed, a one-way ANOVA found no significant difference in raw *FMR1* % input by allele category. However, Table 1.2 also shows that premutation fibroblasts had lower raw

RPL13A % input, which points to a poorer extraction or immunoprecipitation compared to FXTAS and control samples. Once R-loops for all samples were corrected to their corresponding raw *RPL13A* % input, premutation samples had significantly higher relative *FMR1* R-loop frequencies by one-way ANOVA. Moreover, relative R-loops corrected to the *RPL13A* locus had reduced coefficients of variance (CVs), on average, compared to raw R-loop frequencies. For example, the mean CV of raw *FMR1* R-loops was 25.4%, while that of relative *FMR1* R-loops was 21.6%.

Finally, it is unclear how well R-loop frequencies in fibroblasts reflect those in the brain. As FXTAS is a neurodegenerative disorder, it would be ideal to measure R-loops features in the human brain. However, understandably, accessing human brain samples from living patients is not feasible. Post-mortem brain is also challenging to acquire and when collected can suffer from degradation due to variable storage conditions and collection periods after death. It is unclear how stable R-loops are in post-mortem brain in the presence of endogenous RNAses and other enzymes that can unravel them. For example, in cell culture, the half-life of R-loops at *RPL13A* is ~10 minutes after transcription is halted by the drug 5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) (Castillo-Guzman & Chédin, 2021; Sanz et al., 2016). Consequently, post-mortem brain may not be a useful tissue for R-loop analyses.

Indeed, DRIP-qPCR was attempted twice on eight post-mortem brain samples: four controls and four FXTAS cortices. All replicates from all samples at *RPL13A* (positive locus), *FMR1*, *83-84* (negative locus 1), and *SNRPN* (negative locus 2) had R-loop frequencies of ~1 % input (unpublished data). Potentially, R-loops had deteriorated in the post-mortem brain tissue and non-specific binding in brain produced a background level of 1 %. However, without further optimization and analysis, we cannot say. Of note, a fibroblast line run alongside these brain samples produced expected R-loop frequencies.

The challenges of working with brain necessitate the continued use of more accessible tissues and cell types to study *FMR1* disorders, like FXTAS. Tissue specific differences in allele

size and methylation status of *FMR1* alleles have been found in premutation carriers (Tassone et al., 1999). Moreover, hallmark inclusions have yet to be observed in primary fibroblasts from patients with FXTAS (Garcia-Arocena et al., 2010). Nevertheless, fibroblasts have been shown to recapitulate some molecular features of FXTAS (Garcia-Arocena et al., 2010) and may help identify population-level trends despite not having predictive power for an individual (Kim et al., 2019).

Implications and future directions

The current study sought to evaluate the R-loop-instigated DDR model of FXTAS pathogenesis by quantifying R-loop frequency in primary fibroblasts from controls and premutation carriers with and without FXTAS. We hypothesized that premutation carriers would form more frequent R-loops compared to controls and that carriers with FXTAS would have the highest levels. This hypothesis was disproved. The main finding is that there is little to no evidence that FXTAS pathogenesis is associated with R-loop frequency.

These data call for a re-evaluation of the role of R-loop formation in the DDR model. The R-loop instigated DDR model poses that R-loops in premutation carriers cause increase DNA damage that, if unresolved, leads to cellular toxicity and death. Aside from control and FXTAS fibroblasts forming statistically equal levels of R-loops, inclusion localization also calls the DDR model into question. In this model, the hallmark intranuclear inclusions of FXTAS are composed of aggregated DDR proteins bound to the *FMR1* R-loop or the locus directly surrounding the R-loop, which explains the unique features of FXTAS that inclusions are always intranuclear and almost exclusively singular. However, recent evidence shows that FXTAS inclusions are uncoupled from the *FMR1* locus in primary fibroblasts, primary lymphocytes, and brain smears from human frontal cortices (Ma et al., 2019). Taken together, these data suggest that the DDR model should be reworked. DDR and cellular toxicity typical of FXTAS, for example, may result from calcium dysregulation instead of R-loop formation (Robin et al., 2017).

Nevertheless, there are still many avenues by which to explore the role of R-loops in FXTAS. While frequency does not play a clear role in FXTAS, other R-loop features may. R-loop length, structure, stability, and binding partners should be evaluated. The displaced non-template single-stranded DNA has been shown to influence R-loop architecture and thus pathogenicity in various contexts (Carrasco-Salas et al., 2019). The *FMR1* CGG repeats alone and in R-loops is known to generate higher-order structures, such as stem-loops/hairpins and G-quadruplexes (Chen et al., 1995; Gacy et al., 1995; Loomis et al., 2014). Predictably, expanded premutation R-loop structures tend to be more complex than controls (Loomis et al., 2014) Additionally, any R-loop across the CGG repeat will necessarily be longer in premutation cells harboring an expanded allele, as was shown by Loomis and colleagues (Loomis et al., 2014). Finally, some indirect evidence suggests that R-loops in premutation alleles may be more stable and have a negative effect on *FMR1* transcription (Derbis et al., 2021). Of interest, the R-loop frequency in the current study was unaffected by CGG repeat size and mRNA level, which indirectly suggests that premutation and FXTAS samples do not form more stable R-loops. However, further direct measures of R-loop stability are needed to form a conclusion.

More stable, complex, or longer R-loops could increase the probability of being hit by DNA damaging agents, nucleases, or a replication fork (Sollier & Cimprich, 2015). Moreover, G-quadruplexes have been implicated in various regulatory roles as well as in neurodegenerative diseases and can be recognized by DNA damage pathway proteins (Mishra et al., 2016; Simone et al., 2015; Tanaka et al., 2013). Taken together, we cannot fully dismiss the role of R-loops in FXTAS.

Finally, R-loops may be regulated differently in proliferating cells compared to post-mitotic neurons, which may be particularly sensitive to low levels of R-loops. While studying R-loops in patient brain samples is fraught (see Study Limitations above), neurons derived from induced pluripotent stem cells (iPSCs) could be key. The Hagerman Lab has generated several

monoallelic iPSCs in the control and premutation range that could be induced to neurons for this purpose (Liu et al., 2012).

In conclusion, the current study is the only known study of R-loop frequency in premutation carriers that accounts for FXTAS clinical outcome. Contrary to our predictions, there was no evidence that FXTAS samples generated more frequent R-loops at *FMR1*. However, more rigorous studies with larger sample sizes and relevant cell types are needed to conclusively elucidate the role R-loops may play in premutation carriers with and without FXTAS.

Tables

Allele category	Primary fibroblast	FXTAS stage	Age (yr)	*FMR1* (CGG)n	*FMR1* mRNA ± SEM
Control	1049-08	-	63	22	0.177 ± 0.009
	1005-08	-	61	30	0.27 ± 0.013
	1051-11	-	63	32	0.223 ± 0.019
Premutation	1007-12	-	62	65	0.219 ± 0.009
	1011-09	-	65	89	0.205 ± 0.005
	1014-07	-	65	107	0.233 ± 0.002
FXTAS	1034-09	4	61	85	0.274 ± 0.008
	1028-09	4	67	107	0.244 ± 0.024
	1051-08	4	67	119	0.289 ± 0.039

Table 1.1. Primary fibroblast characteristics. Primary fibroblasts were separated into three categories: Control, Premutation no FXTAS (Premutation), and Premutation with FXTAS (FXTAS). *FMR1* mRNA was quantified by qPCR as the ratio of *FMR1* to *GUSB* signal. SEM = standard error of the mean. A one-way ANOVA among allele categories showed no significant difference in age (p-value = 0.422).

Allele category	Primary fibroblast	Target	IPs	R-loop frequency (mean ± SEM)			
				Not corrected to mRNA		*FMR1* mRNA-corrected	
				Raw	Relative	Raw	Relative
Control	1049-08	*RPL13A*	4	10.126 ± 0.732	1 ± 0	-	-
		FMR1	4	2.531 ± 0.461	0.244 ± 0.029	14.296 ± 5.695	1.379 ± 0.397
		83-84	4	0.354 ± 0.070	0.034 ± 0.005	-	-
	1005-08	*RPL13A*	2	9.904 ± 0.528	1 ± 0	-	-
		FMR1	2	1.911 ± 0.031	0.193 ± 0.007	7.08 ± 0.862	0.716 ± 0.094
		SNRPN	2	0.049 ± 0.000	0.005 ± 0	-	-
	1051-11	*RPL13A*	4	10.776 ± 1.286	1 ± 0	-	-
		FMR1	4	2.626 ± 0.404	0.259 ± 0.059	11.795 ± 4.171	1.164 ± 0.565
		83-84	4	0.577 ± 0.040	0.057 ± 0.009	-	-
Premutation	1007-12	*RPL13A*	6	7.266 ± 0.324	1 ± 0	-	-
		FMR1	6	2.458 ± 0.239	0.337 ± 0.025	11.228 ± 2.812	1.539 ± 0.309
		83-84	4	0.807 ± 0.108	0.106 ± 0.010	-	-
		SNRPN	2	0.052 ± 0.032	0.008 ± 0.005	-	-
	1011-09	*RPL13A*	5	7.623 ± 0.894	1 ± 0	-	-
		FMR1	5	3.265 ± 0.340	0.464 ± 0.084	15.962 ± 3.801	2.268 ± 0.929
		83-84	4	0.834 ± 0.243	0.113 ± 0.026	-	-
		SNRPN	1	0.068	0.007	-	-
	1014-07	*RPL13A*	2	9.401 ± 2.815	1 ± 0	-	-
		FMR1	2	2.939 ± 1.113	0.305 ± 0.027	12.636 ± 6.774	1.309 ± 0.169
		SNRPN	2	0.048 ± 0.003	0.006 ± 0.001	-	-
FXTAS	1034-09	*RPL13A*	1	10.463	1	-	-
		FMR1	1	2.096	0.200	7.644	0.731
		SNRPN	1	0.039	0.004	-	-
	1028-09	*RPL13A*	4	10.897 ± 0.627	1 ± 0	-	-
		FMR1	4	2.834 ± 0.022	0.262 ± 0.013	11.598 ± 2.242	1.073 ± 0.232
		83-84	4	0.914 ± 0.074	0.084 ± 0.003	-	-
	1051-08	*RPL13A*	6	9.344 ± 1.175	1 ± 0	-	-
		FMR1	6	2.112 ± 0.275	0.227 ± 0.016	7.32 ± 3.071	0.788 ± 0.253
		83-84	6	0.756 ± 0.164	0.075 ± 0.010	-	-

Table 1.2. Quantification of R-loop frequency in primary fibroblast. Primary fibroblasts were separated into three categories: Control, Premutation without FXTAS (Premutation), and Premutation with FXTAS (FXTAS). R-loop frequency at up to four target loci (*RPL13A*: positive control; *FMR1*: gene of interest; 83-84: nominal negative control; *SNRPN*: true negative control) was quantified by DRIP (See Methods). The resulting values were represented as either “Raw” (as a percent of the input DNA) or “Relative” (raw values normalized to the that of the positive control locus). Additionally, both raw and relative *FMR1* R-loops were subsequently corrected to *FMR1* mRNA levels. *FMR1* mRNA was quantified by qPCR as the ratio of *FMR1* to *GUSB* signal. SEM = standard error of the mean. SEM was not calculated for samples with only 1 immunoprecipitation event (IP).

Figures

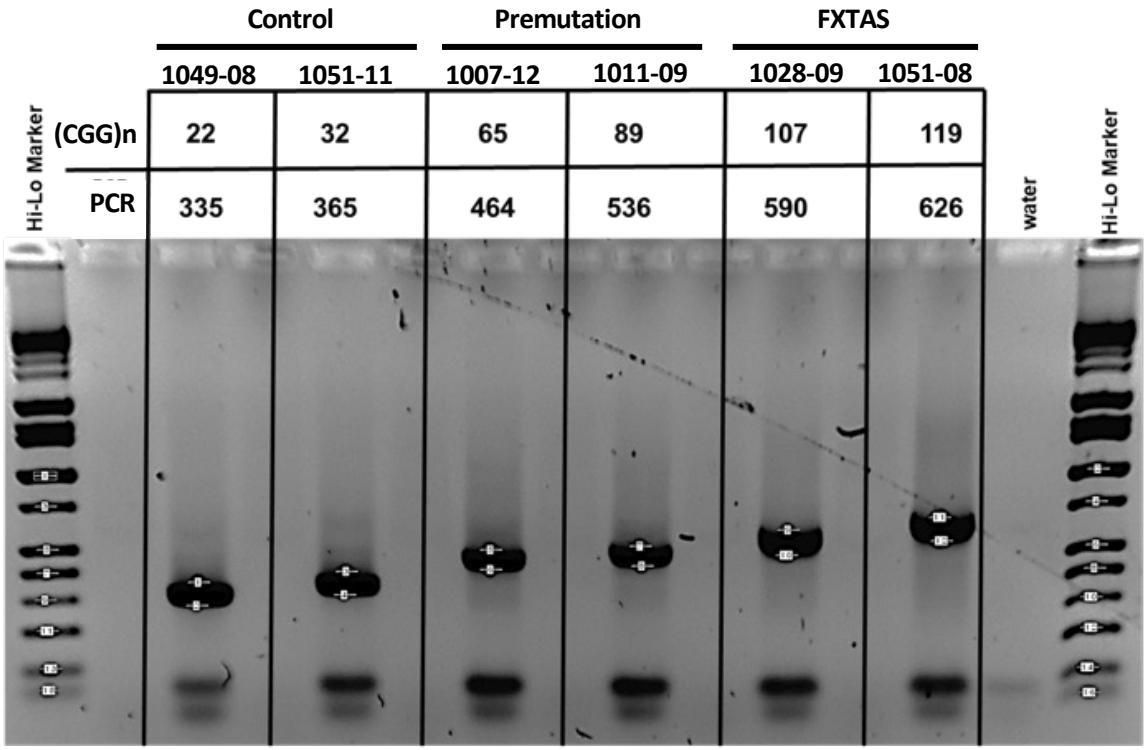


Fig. 1.1. *FMR1* CGG repeat length for primary fibroblasts. PCR through the *FMR1* CGG repeat was performed, and the products run on an agarose gel. Repeat size was calculated as (fragment size – 269)/3. A subset of fibroblasts are shown here.

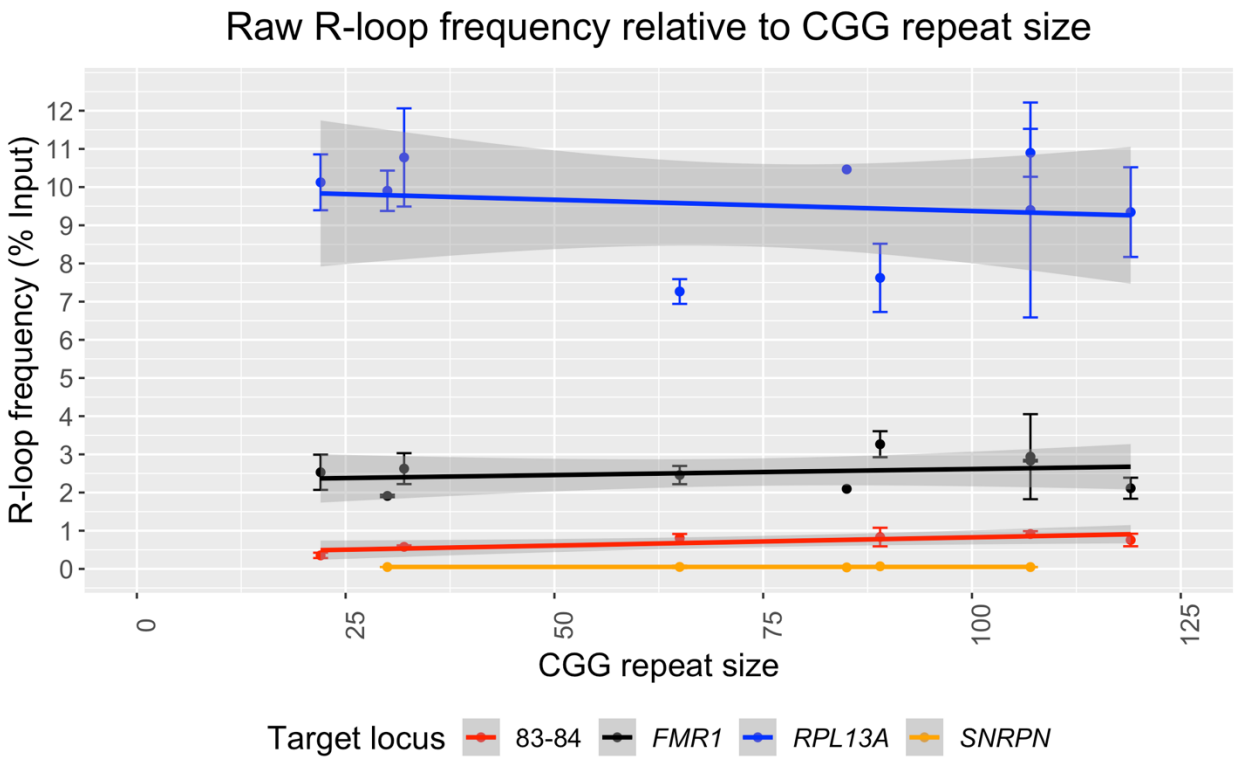


Fig. 1.2. Raw R-loop frequency relative to CGG repeat size in human fibroblasts. R-loop frequency (mean *R*- SEM), measured as raw percent input (see Methods), was plotted against *FMR1* CGG repeat size and a linear regression was fitted for each target locus. No significant association between R-loop frequency and repeat size was detected for our target of interest, *FMR1* (n = 9, Estimate/slope = 0.0032 % Input/CGG repeat, p-value = 0.49). Nor was any significant association detected for the positive control locus *RPL13A* (n = 9; Estimate/slope = -0.0059 % Input/CGG repeat, p-value = 0.67) and the negative control locus *SNRPN* (n = 5, Estimate/slope = 0.000028 % Input/CGG repeat, p-value = 0.90). However, a significant association between R-loop frequency and CGG repeat size was detected for the negative control locus 83-84 (n = 6, Estimate/slope = 0.0043 % Input/CGG repeat, p-value = 0.042).

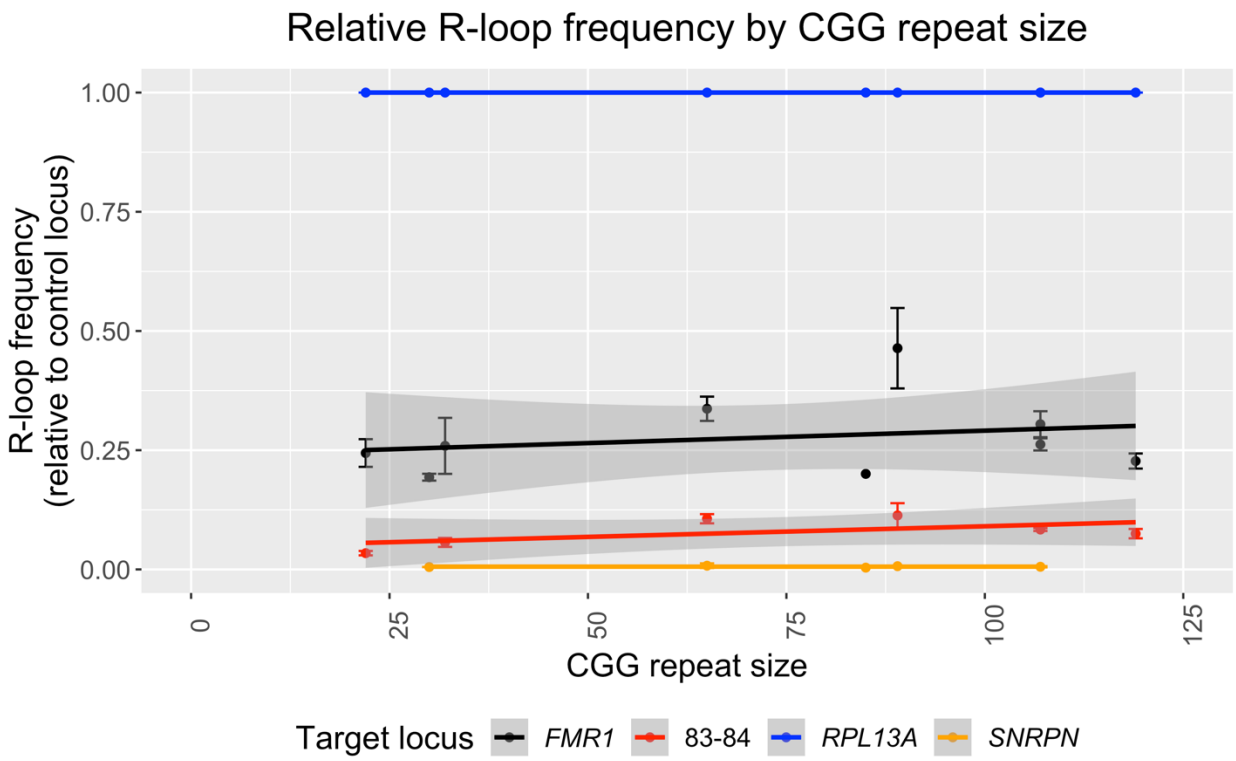


Fig. 1.3. Relative R-loop frequency does not associate with CGG repeat size in fibroblasts. R-loop frequency (mean ± SEM), measured as percent input relative to that of the positive control locus *RPL13A* (see Methods), was plotted against *FMR1* CGG repeat size and a linear regression was fitted for each target locus. No significant association between relative R-loop frequency and repeat size was detected for any locus assayed (*FMR1:* n = 9, Estimate/slope = 0.00052 relative R-loop units/CGG repeat, p-value = 0.55; *RPL13A:* n = 9; perfect fit; 83-84: n = 6, Estimate/slope = 0.00045 relative R-loop units/CGG repeat, p-value = 0.22; *SNRPN*: n = 5, Estimate/slope = 0.00000043 relative R-loop units/CGG repeat, p-value = 0.99)

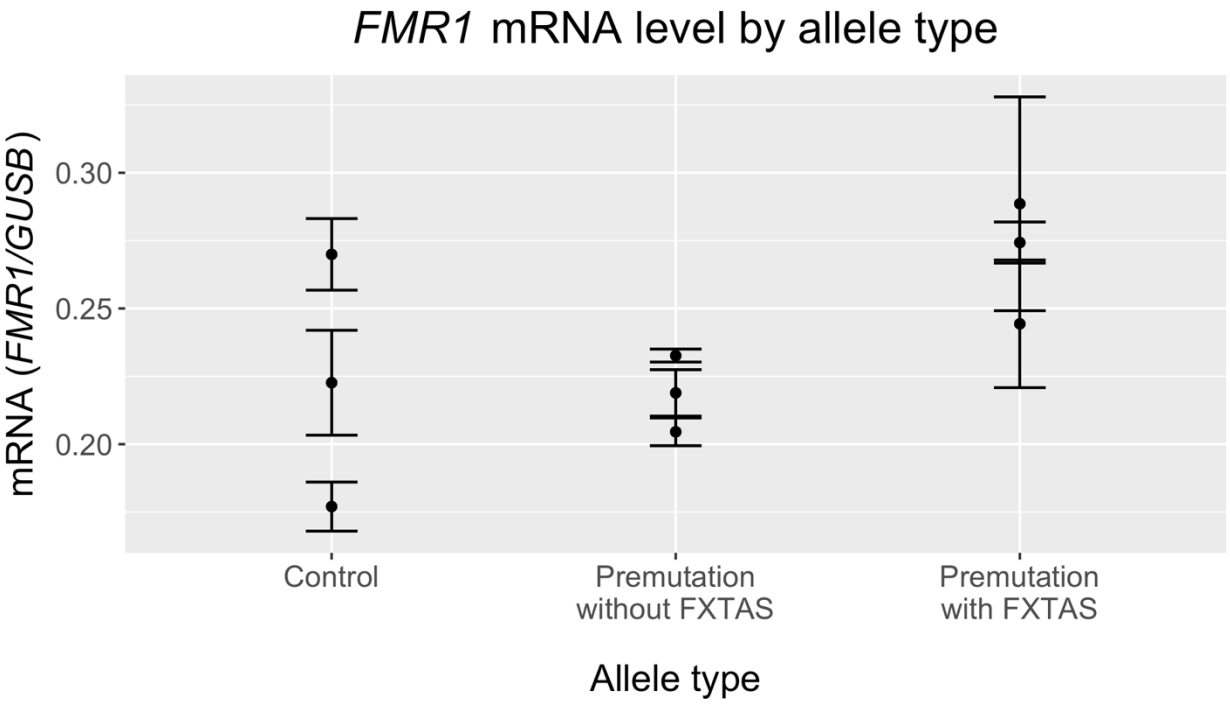


Fig. 1.4. mRNA levels are not statistically different across samples with control and premutation alleles. A one-way ANOVA was performed on *FMR1* mRNA levels (mean ± SEM), measured relative to the GUSB housekeeping gene (see Methods), and no statistical difference was found among fibroblasts with different allele types ($p = 0.168$).

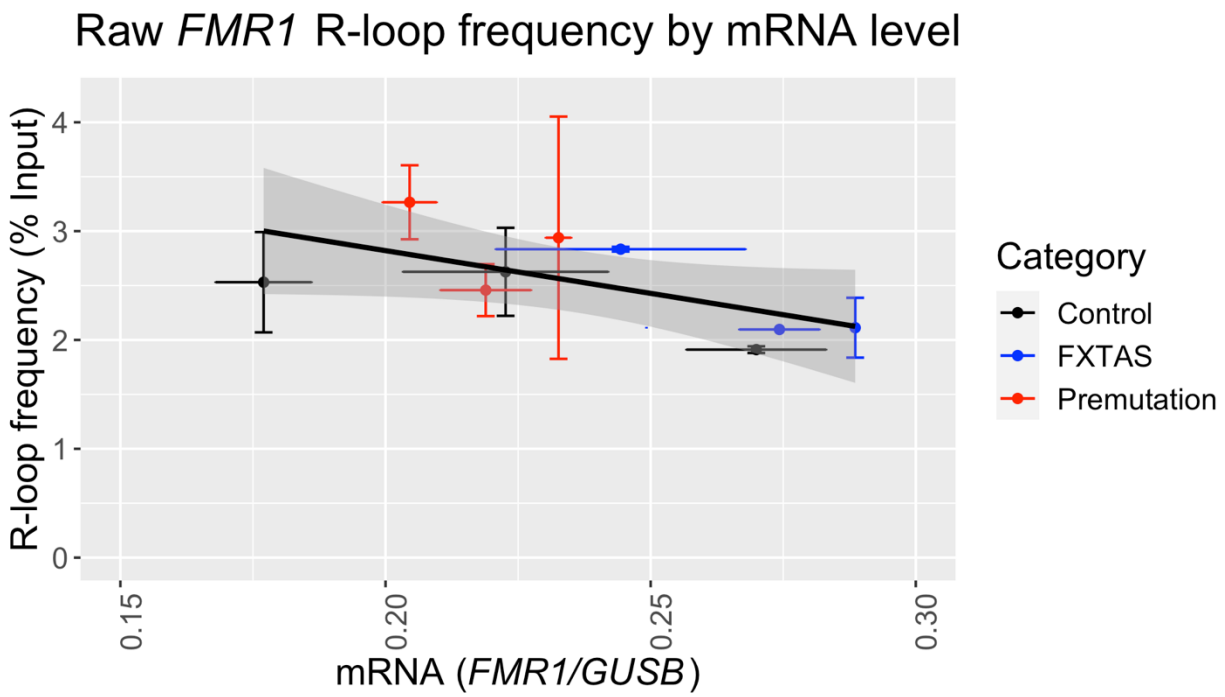


Fig. 1.5. Raw *FMR1* R-loop frequency does not associate with *FMR1* mRNA levels in fibroblasts. Raw *FMR1* R-loop frequency (mean ± SEM), measured as raw percent input (see Methods), was plotted against *FMR1* mRNA level (mean ± SEM), measured relative to the GUSB housekeeping gene. A linear regression was fitted. The association of raw R-loop frequency and mRNA level approaches significance (n = 9, Estimate/slope = -7.85 unit of % Input/unit of relative mRNA, p-value = 0.063). Notably, linear regression analysis does not account for confounding factors, such as *FMR1* CGG repeat size. Samples were color coded by allele category: black = control, blue = premutation with FXTAS, red = premutation without FXTAS.

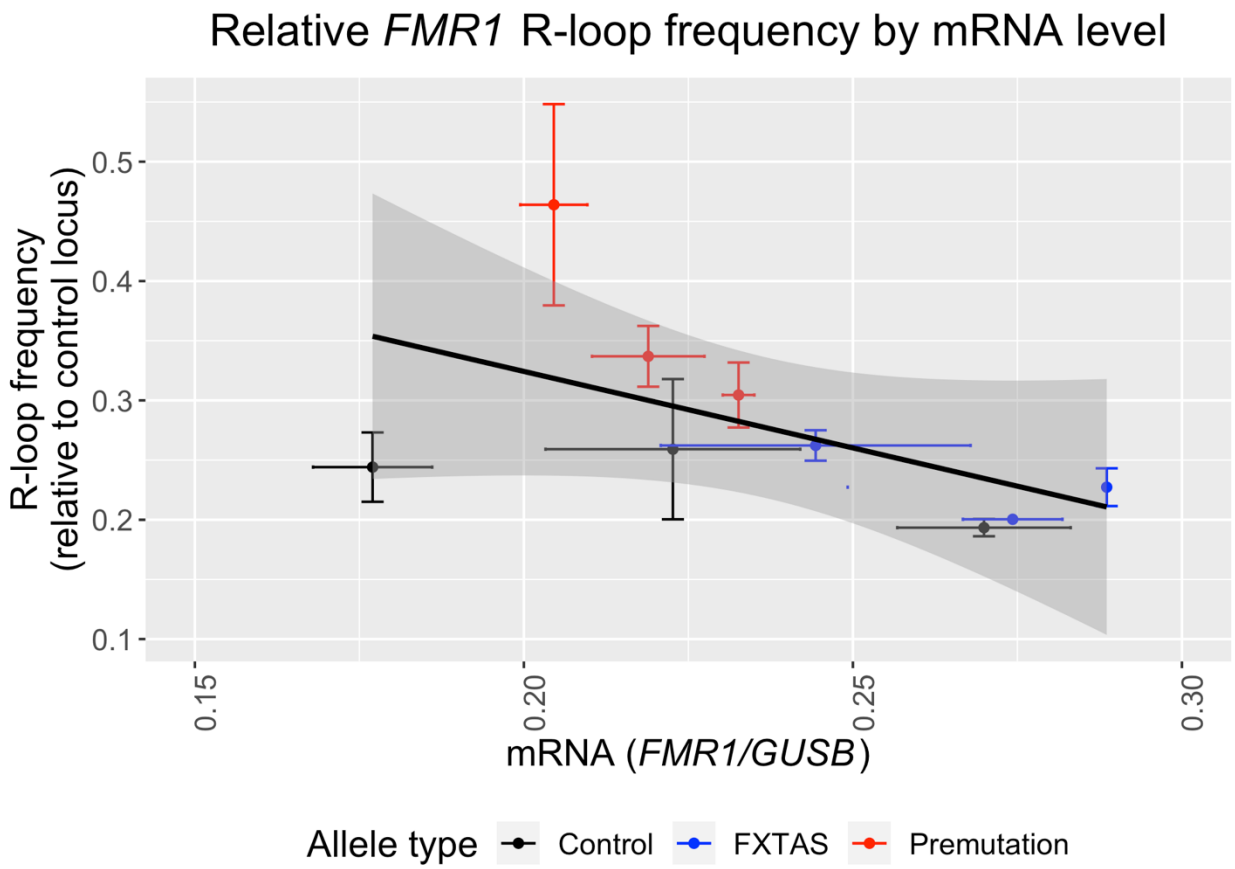


Fig. 1.6. Relative *FMR1* R-loop frequency does not associate with *FMR1* mRNA levels in fibroblasts. Relative *FMR1* R-loop frequency (mean ± SEM), measured as percent input relative to that of the positive control locus *RPL13A* (see Methods), was plotted against *FMR1* mRNA level (mean ± SEM), measured relative to the GUSB housekeeping gene. A linear regression was fitted. The association of relative R-loop frequency and mRNA level is not significant (n = 9, Estimate/slope = -1.28 unit of % Input/unit of relative mRNA, p-value = 0.12). Notably, linear regression analysis does not account for confounding factors, such as *FMR1* CGG repeat size. Samples were color coded by allele category: black = control, blue = premutation with FXTAS, red = premutation without FXTAS.

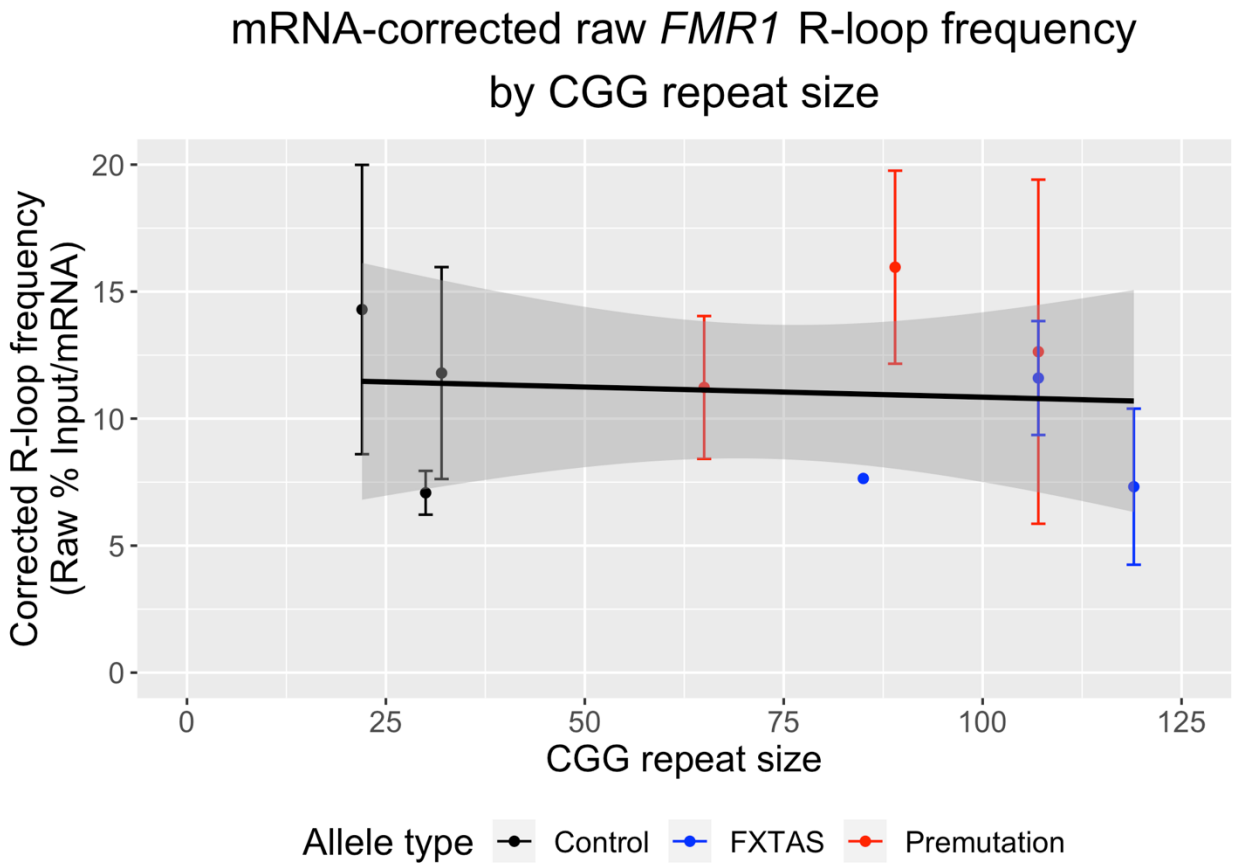


Fig 1.7. Raw *FMR1* R-loop frequency does not associate with CGG repeat size, even when corrected for RNA level. Raw *FMR1* R-loop frequency was measured as percent input (See Methods), then corrected to *FMR1* mRNA level (mean ± SD), and plotted against *FMR1* CGG repeat size. A linear regression was fitted. The association between mRNA-corrected raw *FMR1* R-loops and CGG repeat size is not significant (n=9, Estimate/slope = -0.0080 corrected R-loop units/CGG repeat, p-value = 0.81). Samples were color coded by allele category: black = control, blue = premutation with FXTAS, red = premutation without FXTAS.

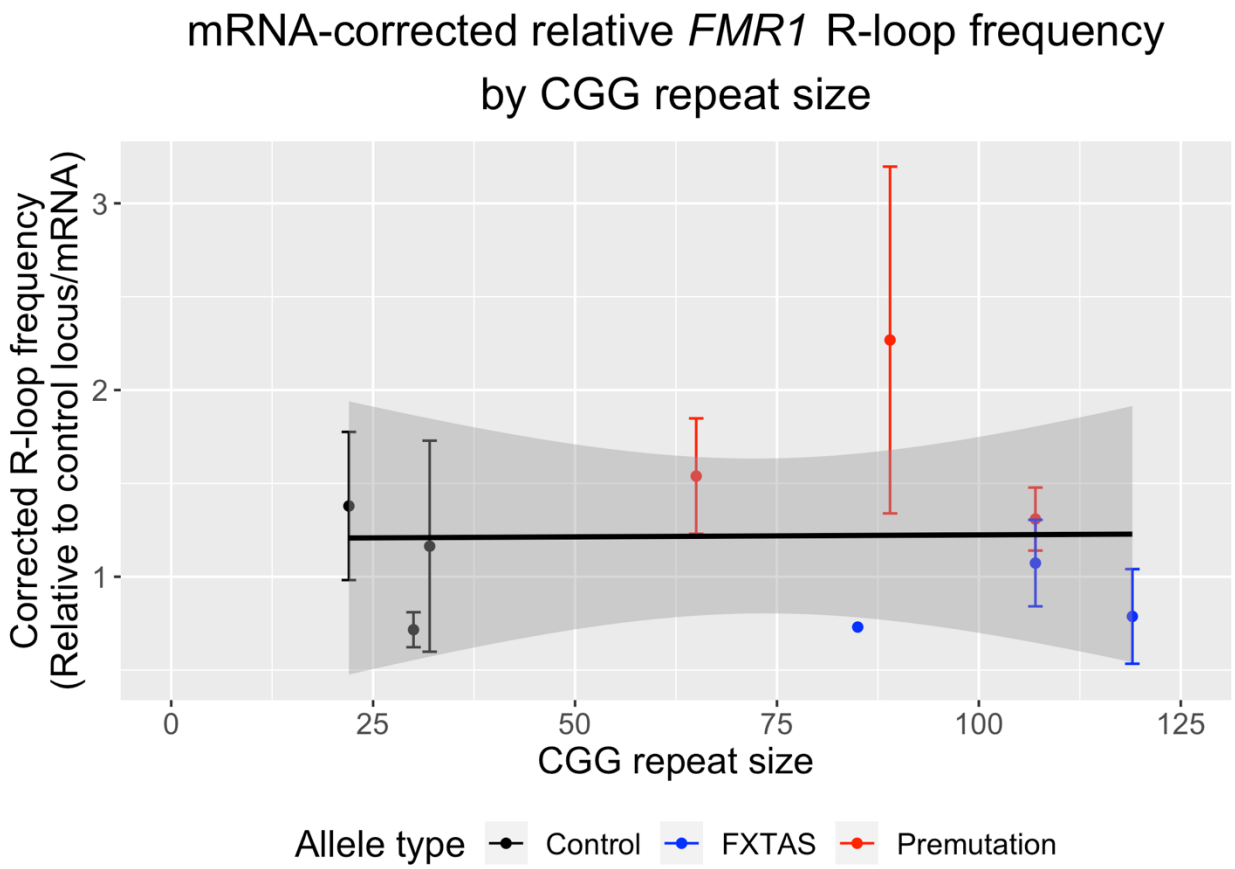


Fig 1.8. Relative *FMR1* R-loop frequency does not associate with CGG repeat size, even when corrected for RNA level. Relative *FMR1* R-loop frequency was measured as percent input relative to that of the positive control locus *RPL13A* (See Methods), then corrected to *FMR1* mRNA level (mean ± SD), and plotted against *FMR1* CGG repeat size. A linear regression was fitted. The association between mRNA-corrected relative *FMR1* R-loops and CGG repeat size is not significant (n=9; Estimate/slope = 0.00021 corrected R-loop units/CGG repeat, p-value = 0.97). Samples were color coded by allele category: black = control, blue = premutation with FXTAS, red = premutation without FXTAS.

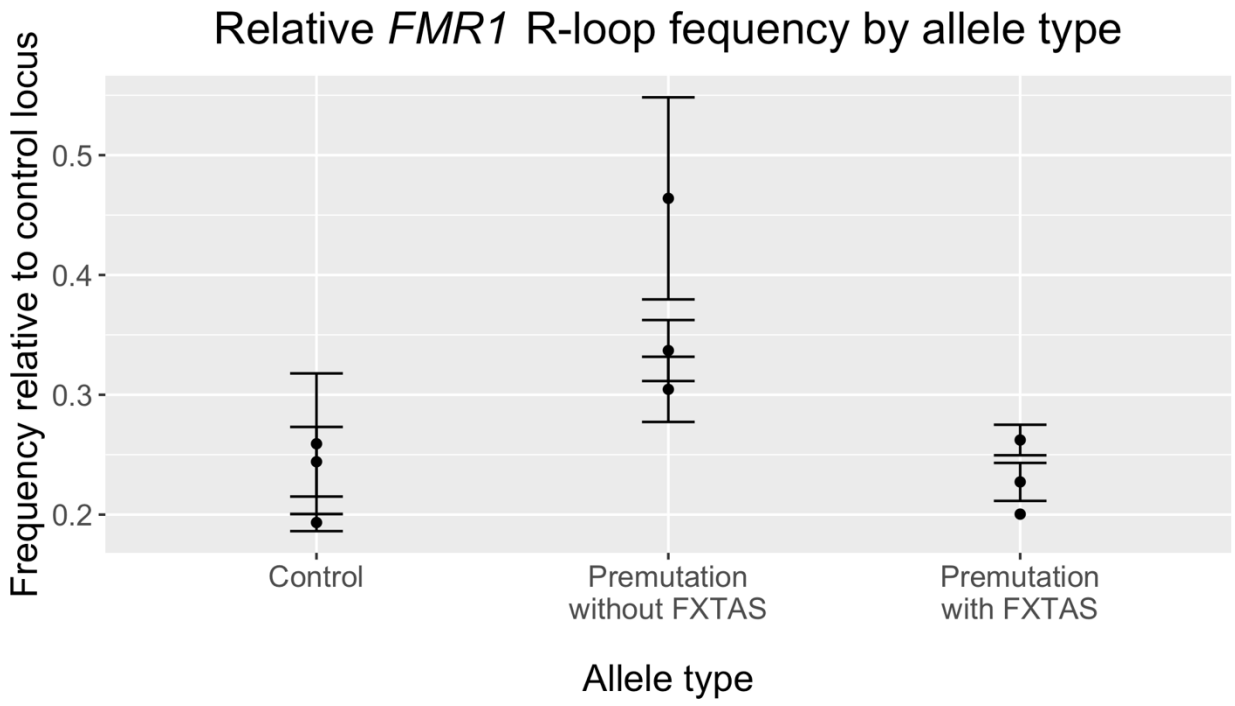


Fig. 1.9. Relative R-loops at *FMR1* are significantly different in premutation fibroblasts. A one-way ANOVA was performed on R-loops at *FMR1*, measured as percent input relative to that of the positive control locus *RPL13A* (see Methods) and a statistical difference was found between allele types in fibroblasts (p = 0.036). Pairwise differences in ANOVA were investigated with Tukey's Honest Significant Difference. Premutation samples (n = 3) as a group were found to be almost significant from both control (n = 3, p-value = 0.054) and FXTAS (n = 3, p-value = 0.051) samples. No difference was found between FXTAS and control samples (p-value = 1.00).

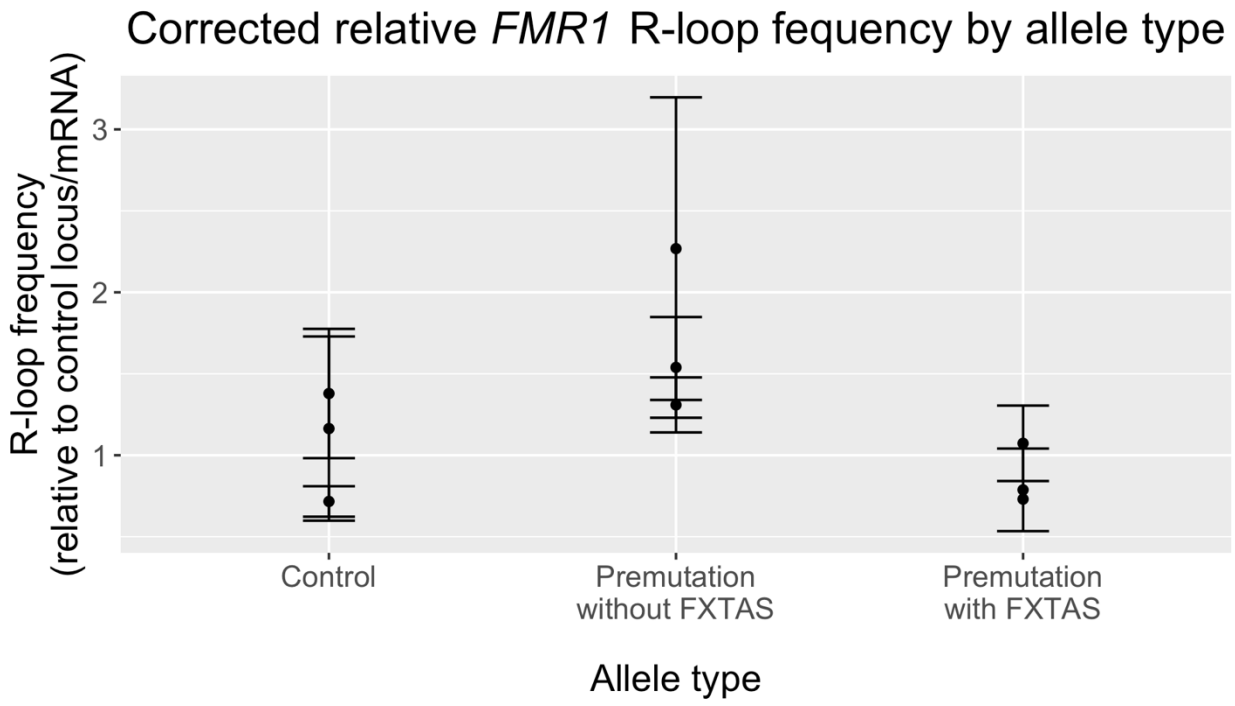


Fig. 1.10. Relative *FMR1* R-loops corrected for mRNA level are not statistically different across samples with control and premutation alleles. A one-way ANOVA was performed on R-loops at *FMR1*, measured as percent input relative to that of the positive control locus *RPL13A* then corrected to *FMR1* mRNA level (see Methods) and no statistical difference was found between allele types (n = 3 for all groups) in fibroblasts, though differences approached significance (p-value = 0.070). Pairwise differences in ANOVA were investigated with Tukey's Honest Significant Difference. Premutation samples were found to be almost significant from FXTAS samples (p-value = 0.067). However, no significant differences were detected between premutation and control samples (p-value = 0.17), nor control and FXTAS samples (p-value = 0.75).

References

Beilina, A., Tassone, F., Schwartz, P. H., Sahota, P., & Hagerman, P. J. (2004). Redistribution of transcription start sites within the FMR1 promoter region with expansion of the downstream CGG-repeat element. *Human Molecular Genetics*, *13*(5), 543–549. https://doi.org/10.1093/hmg/ddh053

Beletskii, A., & Bhagwat, A. S. (1996). Transcription-induced mutations: increase in C to T mutations in the nontranscribed strand during transcription in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(24), 13919–13924. https://doi.org/10.1073/PNAS.93.24.13919

Buijsen, R. A., Sellier, C., Severijnen, L.-A. W., Oulad-Abdelghani, M., Verhagen, R. F., Berman, R. F., Charlet-Berguerand, N., Willemsen, R., & Hukema, R. K. (2014). FMRpolyG-positive inclusions in CNS and non-CNS organs of a fragile X premutation carrier with fragile X-associated tremor/ataxia syndrome. In *Acta Neuropathologica Communications* (Vol. 2, Issue 1). https://doi.org/10.1186/s40478-014-0162-2

Cabal-Herrera, A. M., Tassanakijpanich, N., Salcedo-Arellano, M. J., & Hagerman, R. J. (2020). Fragile X-associated tremor/ataxia syndrome (FXTAS): Pathophysiology and clinical implications. *International Journal of Molecular Sciences*, *21*(12), 1–23. https://doi.org/10.3390/ijms21124391

Carrasco-Salas, Y., Malapert, A., Sulthana, S., Molcrette, B., Ea Chazot-Franguiadakis, L. ´, Bernard, P., Fréd´, F., Chédin, F., Chédin, C., Faivre-Moskalenko, C., & Vanoosthuyse, V. (2019). The extruded non-template strand determines the architecture of R-loops. *Nucleic Acids Research*, *1*. https://doi.org/10.1093/nar/gkz341

Castillo-Guzman, D., & Chédin, F. (2021). Defining R-loop classes and their contributions to genome instability. *DNA Repair*, *106*, 103182. https://doi.org/10.1016/J.DNAREP.2021.103182

Chédin, F., Hartono, S. R., Sanz, L. A., & Vanoosthuyse, V. (2021). Best practices for the visualization, mapping, and manipulation of R-loops. *The EMBO Journal*, *40*(4), e106394. https://doi.org/10.15252/embj.2020106394

Chen, X., Mariappan, S. V., Catasti, P., Ratliff, R., Moyzis, R. K., Laayoun, A., Smith, S. S., Bradbury, E. M., & Gupta, G. (1995). Hairpins are formed by the single DNA strands of the fragile X triplet repeats: structure and biological implications. *Proceedings of the National Academy of Sciences*, *92*(11), 5199–5203.

Colak, D., Zaninovic, N., Cohen, M. S., Rosenwaks, Z., Yang, W.-Y., Gerhardt, J., Disney, M. D., & Jaffrey, S. R. (2014). Promoter-Bound Trinucleotide Repeat mRNA Drives Epigenetic Silencing in Fragile X Syndrome. *Science*, *343*(6174), 1002–1005. https://doi.org/10.1126/science.1245831

Coppedè, F., & Migliore, L. (2015). DNA damage in neurodegenerative diseases. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, *776*, 84–97. https://doi.org/10.1016/j.mrfmmm.2014.11.010

Costantino, L., & Koshland, D. (2015). The Yin and Yang of R-loop biology. *Current Opinion in Cell Biology*, *34*, 39–45. https://doi.org/10.1016/j.ceb.2015.04.008

Costantino, L., & Koshland, D. (2018). Genome-wide Map of R-Loop-Induced Damage Reveals How a Subset of R-Loops Contributes to Genomic Instability. *Molecular Cell*, *71*(4), 487-497.e3. https://doi.org/10.1016/j.molcel.2018.06.037

Crossley, M. P., Bocek, M., & Cimprich, K. A. (2019). R-Loops as Cellular Regulators and Genomic Threats. *Molecular Cell*, *73*(3), 398–411. https://doi.org/10.1016/J.MOLCEL.2019.01.024

Crossley, M. P., Bocek, M. J., Hamperl, S., Swigut, T., & Cimprich, K. A. (2020). qDRIP: a method to quantitatively assess RNA-DNA hybrid formation genome-wide. *Nucleic Acids Research*, *48*(14). https://doi.org/10.1093/nar/gkaa500

Derbis, M., Kul, E., Niewiadomska, D., Sekrecki, M., Piasecka, A., Taylor, K., Hukema, R. K., Stork, O., & Sobczak, K. (2021). Short antisense oligonucleotides alleviate the pleiotropic toxicity of RNA harboring expanded CGG repeats. *Nature Communications*, *12*(1), 1265. https://doi.org/10.1038/s41467-021-21021-w

Feng, Y., Zhang, F., Lokey, L. K., Chastain, J. L., Lakkis, L., Eberhart, D., & Warren, S. T. (1995). Translational suppression by trinucleotide repeat expansion at FMR1. *Science*, *268*(5211), 731–734. https://doi.org/10.1126/science.7732383

Fink, D. A., Nelson, L. M., Pyeritz, R., Johnson, J., Sherman, S. L., Cohen, Y., & Elizur, S. E. (2018). Fragile X Associated Primary Ovarian Insufficiency (FXPOI): Case Report and Literature Review. *Frontiers in Genetics*, *9*. https://doi.org/10.3389/FGENE.2018.00529

Freudenreich, C. H. (2018). R-loops: targets for nuclease cleavage and repeat instability. *Current Genetics*, *64*(4), 789–794. https://doi.org/10.1007/s00294-018-0806-z

Gacy, M. A., Goellner, G., Juranić, N., Macura, S., & McMurray, C. T. (1995). Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell*, *81*(4), 533–540.

Garcia-Arocena, D., Yang, J. E., Brouwer, J. R., Tassone, F., Iwahashi, C., Berry-Kravis, E. M., Goetz, C. G., Sumis, A. M., Zhou, L., Nguyen, D. V, Campos, L., Howell, E., Ludwig, A., Greco, C., Willemsen, R., Hagerman, R. J., & Hagerman, P. J. (2010). Fibroblast phenotype in male carriers of FMR1 premutation alleles. *Human Molecular Genetics*, *19*(2), 299–312. https://doi.org/10.1093/hmg/ddp497

Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I., & Chédin, F. (2012). R-Loop Formation Is a Distinctive Characteristic of Unmethylated Human CpG Island Promoters. *Molecular Cell*, *45*(6), 814–825. https://doi.org/10.1016/j.molcel.2012.01.017

Glineburg, M. R., Todd, P. K., Charlet-Berguerand, N., & Sellier, C. (2018). Repeat-associated non-AUG (RAN) translation and other molecular mechanisms in Fragile X Tremor Ataxia Syndrome. *Brain Research*, *1693*, 43–54. https://doi.org/10.1016/J.BRAINRES.2018.02.006

Groh, M., & Gromak, N. (2014). Out of Balance: R-loops in Human Disease. *PLoS Genetics*, *10*(9). https://doi.org/10.1371/journal.pgen.1004630

Groh, M., Lufino, M. M. P., Wade-Martins, R., & Gromak, N. (2014). R-loops Associated with Triplet Repeat Expansions Promote Gene Silencing in Friedreich Ataxia and Fragile X Syndrome. *PLoS Genetics*, *10*(5). https://doi.org/10.1371/journal.pgen.1004318

Hagerman, P. J. (2013). Fragile X-associated tremor/ataxia syndrome (FXTAS): Pathology and mechanisms. In *Acta Neuropathologica* (Vol. 126, Issue 1, pp. 1–19). https://doi.org/10.1007/s00401-013-1138-1

Hagerman, R. J., & Hagerman, P. J. (2021). FXTAS: Pathophysiology and Management. *Current Opinion in Neurology*, *34*(4), 541. https://doi.org/10.1097/WCO.0000000000000954

Hagerman, R. J., & Hagerman, P. J. (2016). Fragile X-associated tremor/ataxia syndrome — features, mechanisms and management. *Nature Reviews Neurology*, *12*(7), 403–412. https://doi.org/10.1038/nrneurol.2016.82

Hagerman, R. J., Hull, C. E., Safanda, J. F., Carpenter, I., Staley, L. W., O'Connor, R. A., Seydel, C., Mazzocco, M. M. M., Snow, K., Thibodeau, S. N., Kuhl, D., Nelson, D. L., Caskey, C. T., & Taylor, A. K. (1994). High functioning fragile X males: Demonstration of an unmethylated fully expanded FMR-1 mutation associated with protein expression. *American Journal of Medical Genetics*, *51*(4), 298–308. https://doi.org/10.1002/ajmg.1320510404

Hagerman, R. J., Leehey, M., Heinrichs, W., Tassone, F., Wilson, R., Hills, J., Grigsby, J., Gage, B., & Hagerman, P. J. (2001). Intention tremor, parkinsonism, and generalized brain

atrophy in male carriers of fragile X. *Neurology*, *57*(1), 127–130. https://doi.org/10.1212/WNL.57.1.127

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*. https://doi.org/10.3389/FPSYT.2018.00564

Hayward, B. E., Zhou, Y., Kumari, D., & Usdin, K. (2016). A Set of Assays for the Comprehensive Analysis of FMR1 Alleles in the Fragile X–Related Disorders. *Journal of Molecular Diagnostics*, *18*(5), 762–774. https://doi.org/10.1016/j.jmoldx.2016.06.001

Hoem, G., Raske, C. R., Garcia-Arocena, D., Tassone, F., Sanchez, E., Ludwig, A. L., Iwahashi, C. K., Kumar, M., Yang, J. E., & Hagerman, P. J. (2011). *CGG-repeat length threshold for FMR1 RNA pathogenesis in a cellular model for FXTAS*. https://doi.org/10.1093/hmg/ddr101

Holm, K. N., Herren, A. W., Taylor, S. L., Randol, J. L., Kim, K., Espinal, G., Martiínez-Cerdeño, V., Pessah, I. N., Hagerman, R. J., & Hagerman, P. J. (2021). Human Cerebral Cortex Proteome of Fragile X-Associated Tremor/Ataxia Syndrome. *Frontiers in Molecular Biosciences*, *7*, 600840. https://doi.org/10.3389/FMOLB.2020.600840/FULL

Huertas, P., & Aguilera, A. (2003). Cotranscriptionally Formed DNA:RNA Hybrids Mediate Transcription Elongation Impairment and Transcription-Associated Recombination. *Molecular Cell*, *12*, 711–721. https://ac.els-cdn.com/S1097276503003198/1-s2.0-S1097276503003198-main.pdf?_tid=f892d4d1-37bb-4591-b6bb-7e3307606205&acdnat=1538096444_db2715e28d4d2ba8dd998837f606fb92

Iwahashi, C. K., Yasui, D. H., An, H. J., Greco, C. M., Tassone, F., Nannen, K., Babineau, B., Lebrilla, C. B., Hagerman, R. J., & Hagerman, P. J. (2006). Protein composition of the intranuclear inclusions of FXTAS. *Brain*, *129*(1), 256–271. https://doi.org/10.1093/brain/awh650

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

Jacquemont, S., Hagerman, R. J., Leehey, M., Grigsby, J., Zhang, L., Brunberg, J. A., Greco, C., Des Portes, V., Jardini, T., Levine, R., Berry-Kravis, E., Brown, W. T., Schaeffer, S., Kissel, J., Tassone, F., & Hagerman, P. J. (2003). Fragile X Premutation Tremor/Ataxia Syndrome: Molecular, Clinical, and Neuroimaging Correlates. *The American Journal of Human Genetics*, *72*(4), 869–878. https://doi.org/10.1086/374321

Jin, P., Duan, R., Qurashi, A., Qin, Y., Tian, D., Rosser, T. C., Liu, H., Feng, Y., & Warren, S. T. (2007). Pur a Binds to rCGG Repeats and Modulates Repeat-Mediated Neurodegeneration in a Drosophila Model of Fragile X Tremor/Ataxia Syndrome. *Neuron*, *55*(4), 556–564. https://doi.org/10.1016/j.neuron.2007.07.020

Kim, K., Hessl, D., Randol, J. L., Espinal, G. M., Schneider, A., Protic, D., Aydin, E. Y., Hagerman, R. J., & Hagerman, P. J. (2019). Association between IQ and FMR1 protein (FMRP) across the spectrum of CGG repeat expansions. *PLoS ONE*, *14*(12). https://doi.org/10.1371/journal.pone.0226811

Kurosawa, M., Matsumoto, G., Sumikura, H., Hatsuta, H., Murayama, S., Sakurai, T., Shimogori, T., Hattori, N., & Nukina, N. (2016). Serine 403-phosphorylated p62/SQSTM1 immunoreactivity in inclusions of neurodegenerative diseases. *Neuroscience Research*, *103*, 64–70. https://doi.org/10.1016/J.NEURES.2015.08.002

Lanni, S., & Pearson, C. E. (2019). Molecular genetics of congenital myotonic dystrophy. *Neurobiology of Disease*, *132*, 104533. https://doi.org/10.1016/J.NBD.2019.104533

Li, X., & Manley, J. L. (2005). Inactivation of the SR Protein Splicing Factor ASF/SF2 Results in Genomic Instability. *Cell*, *122*, 365–378. https://doi.org/10.1016/j.cell.2005.06.008

Lin, Y., Dent, S. Y. R., Wilson, J. H., Wells, R. D., & Napierala, M. (2010). R loops stimulate genetic instability of CTG.CAG repeats. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(2), 692–697. https://doi.org/10.1073/PNAS.0909740107

Liu, J., Koscielska, K. A., Cao, Z., Hulsizer, S., Grace, N., Mitchell, G., Nacey, C., Githinji, J., McGee, J., Garcia-Arocena, D., Hagerman, R. J., Nolta, J., Pessah, I. N., & Hagerman, P. J. (2012). Signaling defects in iPSC-derived fragile X premutation neurons. *Human Molecular Genetics*, *21*(17), 3795–3805. https://doi.org/10.1093/hmg/dds207

Loomis, E. W., Sanz, L. A., Chédin, F., & Hagerman, P. J. (2014). Transcription-Associated R-Loop Formation across the Human FMR1 CGG-Repeat Region. *PLoS Genetics*, *10*(4). https://doi.org/10.1371/journal.pgen.1004294

Ludwig, A. L., Hershey, J. W., & Hagerman, P. J. (2011). Initiation of translation of the FMR1 mRNA occurs predominantly through 5′end-dependent ribosomal scanning. *J. Mol. Biol*, *407*(1), 21–34. https://doi.org/10.1016/j.jmb.2011.01.006

Ma, L., Herren, A. W., Espinal, G., Randol, J., Mclaughlin, B., Martinez-Cerdeño, V., Pessah, I. N., Hagerman, R. J., & Hagerman, P. J. (2019). Composition of the Intranuclear Inclusions of Fragile X-associated Tremor/Ataxia Syndrome. *Acta Neuropathologica Communications*, *7*(143). https://doi.org/10.1186/s40478-019-0796-1

McIvor, E. I., Polak, U., & Napierala, M. (2010). New insights into repeat instability: Role of RNA•DNA hybrids. *RNA Biology*, *7*(5), 551–558. https://doi.org/10.4161/rna.7.5.12745

Mishra, S. K., Tawani, A., Mishra, A., & Kumar, A. (2016). G4IPDB: A database for G-quadruplex structure forming nucleic acid interacting proteins. *Scientific Reports*, *6*(1), 38144. https://doi.org/10.1038/srep38144

Moreira, M.-C., Klur, S., Watanabe, M., Németh, A. H., Le Ber, I., Moniz, J.-C., Tranchant, C., Aubourg, P., Tazir, M., Schöls, L., Pandolfo, M., Schulz, J. B., Pouget, J., Calvas, P., Shizuka-Ikeda, M., Shoji, M., Tanaka, M., Izatt, L., Shaw, C. E., … Koenig, M. (2004). Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. *Nature Genetics*, *36*(3), 225–227. https://doi.org/10.1038/ng1303

Pieretti, M., Zhang, F., Fu, Y. H., Warren, S. T., Oostra, B. A., Caskey, C. T., & Nelson, D. L. (1991). Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell*, *66*(4), 817–822. https://doi.org/10.1016/0092-8674(91)90125-I

Ratmeyer, L., Zhong, Y. Y., Wilson, W. D., Vinayak, R., & Zon, G. (1994). Sequence Specific Thermodynamic and Structural Properties for DNA·RNA Duplexes. *Biochemistry*, *33*(17), 5298–5304. https://doi.org/10.1021/BI00183A037/ASSET/BI00183A037.FP.PNG_V03

Richter, J. D., & Zhao, X. (2021). The molecular biology of FMRP: new insights into fragile X syndrome. In *Nature Reviews Neuroscience* (Vol. 22, Issue 4, pp. 209–222). Nature Publishing Group. https://doi.org/10.1038/s41583-021-00432-0

Roberts, R. W., & Crothers, D. M. (1992). Stability and Properties of Double and Triple Helices: Dramatic Effects of RNA or DNA Backbone Composition. In *New Series* (Vol. 258, Issue 5087). https://www.jstor.org/stable/pdf/2880278.pdf?refreqid=excelsior%3Aa9cdc66eaa7d04f41b16966f535792c8

Robin, G., López, J. R., Espinal, G. M., Hulsizer, S., Hagerman, P. J., & Pessah, I. N. (2017). Calcium dysregulation and Cdk5-ATM pathway involved in a mouse model of fragile X-associated tremor/ataxia syndrome. *Human Molecular Genetics*, *26*(14), 2649–2666. https://doi.org/10.1093/hmg/ddx148

Rodriguez-Revenga, L., Madrigal, I., Pagonabarraga, J., Xunclà, M., Badenas, C., Kulisevsky, J., Gomez, B., & Milà, M. (2009). Penetrance of FMR1 premutation associated pathologies in fragile X syndrome families. *European Journal of Human Genetics 2009 17:10*, *17*(10), 1359–1362. https://doi.org/10.1038/ejhg.2009.51

Ross-Inta, C., Omanska-Klusek, A., Wong, S., Barrow, C., Garcia-Arocena, D., Iwahashi, C., Berry-Kravis, E., Hagerman, R. J., Hagerman, P. J., & Giulivi, C. (2010). Evidence of mitochondrial dysfunction in fragile X-associated tremor/ataxia syndrome. *The Biochemical Journal*, *429*(3), 545–552. https://doi.org/10.1042/BJ20091960

Roy, D., & Lieber, M. R. (2009). G clustering is important for the initiation of transcription-induced R-loops in vitro, whereas high G density without clustering is sufficient thereafter. *Molecular and Cellular Biology*, *29*(11), 3124–3133. https://doi.org/10.1128/MCB.00139-09

Sanz, L. A., Hartono, S. R., Lim, Y. W., Steyaert, S., Rajpurkar, A., Ginno, P. A., Xu, X., & Chédin, F. (2016). Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. *Molecular Cell*, *63*(1), 167–178. https://doi.org/10.1016/j.molcel.2016.05.032

Schmitt, L. M., Shaffer, R. C., Hessl, D., & Erickson, C. (2019). Executive Function in Fragile X Syndrome: A Systematic Review. *Brain Sciences 2019, Vol. 9, Page 15*, *9*(1), 15. https://doi.org/10.3390/BRAINSCI9010015

Sellier, C., Buijsen, R. A. M., He, F., Martinat, C., Todd, P. K., Charlet-berguerand, N., Sellier, C., Buijsen, R. A. M., He, F., Natla, S., Jung, L., Tropel, P., Gaucherot, A., Martinat, C., & Todd, P. K. (2017). Translation of Expanded CGG Repeats into FMRpolyG Is Pathogenic and May Contribute to Fragile X Tremor Ataxia Syndrome. *Neuron*, *93*(2), 331–347. https://doi.org/10.1016/j.neuron.2016.12.016

Sellier, C., Freyermuth, F., Tabet, R., Tran, T., He, F., Ruffenach, F., Alunni, V., Moine, H., Thibault, C., Page, A., Tassone, F., Willemsen, R., Disney, M. D., Hagerman, P. J., Todd, P. K., & Charlet-Berguerand, N. (2013a). Sequestration of DROSHA and DGCR8 by Expanded CGG RNA Repeats Alters MicroRNA Processing in Fragile X-Associated Tremor/Ataxia Syndrome. *CellReports*, *3*, 869–880. https://doi.org/10.1016/j.celrep.2013.02.004

Sellier, C., Freyermuth, F., Tabet, R., Tran, T., He, F., Ruffenach, F., Alunni, V., Moine, H., Thibault, C., Page, A., Tassone, F., Willemsen, R., Disney, M. D., Hagerman, P. J., Todd, P. K., & Charlet-Berguerand, N. (2013b). Sequestration of DROSHA and DGCR8 by expanded CGG RNA repeats alters microRNA processing in fragile X-associated tremor/ataxia syndrome. *Cell Reports*, *3*(3), 869–880. https://doi.org/10.1016/j.celrep.2013.02.004

Sellier, C., Rau, F., Liu, Y., Tassone, F., Hukema, R. K., Gattoni, R., Schneider, A., Richard, S., Willemsen, R., Elliott, D. J., Hagerman, P. J., & Charlet-Berguerand, N. (2010). Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. *The EMBO Journal*, *29*(7), 1248–1261. https://doi.org/10.1038/emboj.2010.21

Simone, R., Fratta, P., Neidle, S., Parkinson, G. N., & Isaacs, A. M. (2015). G-quadruplexes: Emerging roles in neurodegenerative diseases and the non-coding transcriptome. *FEBS Letters*, *589*, 1653–1668. https://doi.org/10.1016/j.febslet.2015.05.003

Skourti-Stathaki, K., & Proudfoot, N. J. (2014). A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes & Development*, *28*(13), 1384–1396. https://doi.org/10.1101/gad.242990.114

Sofola, O. A., Jin, P., Qin, Y., Duan, R., Liu, H., de Haro, M., Nelson, D. L., & Botas, J. (2007). RNA-Binding Proteins hnRNP A2/B1 and CUGBP1 Suppress Fragile X CGG Premutation Repeat-Induced Neurodegeneration in a Drosophila Model of FXTAS. *Neuron*, *55*(4), 565–571. https://doi.org/10.1016/j.neuron.2007.07.021

Sollier, J., & Cimprich, K. A. (2015). Breaking bad: R-loops and genome integrity. *Trends in Cell Biology*, *25*(9), 514–522. https://doi.org/10.1016/j.tcb.2015.05.003

Sordet, O., Nakamura, A. J., Redon, C. E., & Pommier, Y. (2010). DNA double-strand breaks and ATM activation by transcription-blocking DNA lesions. *Cell Cycle*, *9*(2), 274–278. https://doi.org/10.4161/cc.9.2.10506

Sordet, O., Redon, C. E., Guirouilh-Barbat, J., Smith, S., Solier, S., Douarre, C., Conti, C., Nakamura, A. J., Das, B. B., Nicolas, E., Kohn, K. W., Bonner, W. M., & Pommier, Y. (2009). Ataxia telangiectasia mutated activation by transcription- and topoisomerase I-induced DNA double-strand breaks. *EMBO Reports*, *10*(8), 887–893. https://doi.org/10.1038/embor.2009.97

Stolz, R., Sulthana, S., Hartono, S. R., Malig, M., Benham, C. J., & Chedin, F. (2019). Interplay between DNA sequence and negative superhelicity drives R-loop structures. *Proceedings of the National Academy of Sciences*, *116*(13), 6260–6269. https://doi.org/10.1073/pnas.1819476116

Tanaka, A., Choi, J., Kim, S. K., & Majima, T. (2013). Interaction of G-Quadruplex with RecA Protein Studied in Bulk Phase and at the Single-Molecule Level. *The Journal of Physical Chemistry B*, *117*, 6711–6717. https://doi.org/10.1021/jp4036277

Tassone, F., De Rubeis, S., Carosi, C., La Fata, G., Serpa, G., Raske, C., Willemsen, R., Hagerman, P. J., & Bagni, C. (2011). Differential usage of transcriptional start sites and polyadenylation sites in FMR1 premutation alleles. *Nucleic Acids Research*, *39*(14), 6172–6185. https://doi.org/10.1093/nar/gkr100

Tassone, F., Hagerman, R. J., Loesch, D. Z., Lachiewicz, A., Taylor, A. K., & Hagerman, P. J. (2000). Fragile X males with unmethylated, full mutation trinucleotide repeat expansions have elevated levels of FMR1 messenger RNA. *American Journal of Medical Genetics*, *94*(3), 232–236. https://doi.org/10.1002/1096-8628(20000918)94:3<232::AID-AJMG9>3.0.CO;2-H

Tassone, F., Hagerman, R. J., Taylor, A. K., Gane, L. W., Godfrey, T. E., & Hagerman, P. J. (2000). Elevated levels of FMR1 mRNA 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

Tassone, F., Hagerman, R. J., Taylor, A. K., Mills, J. B., Harris, S. W., Gane, L. W., & Hagerman, P. (2000). Clinical involvement and protein expression in individuals with the FMR1 premutation. *American Journal of Medical Genetics*, *91*(2), 144–152. https://doi.org/10.1002/(SICI)1096-8628(20000313)91:2<144::AID-AJMG14>3.0.CO;2-V

Tassone, F., Longshore, J., Zunich, J., Steinbach, P., Salat, U., & Taylor, A. K. (1999). Tissue-specific methylation differences in a fragile X premutation carrier. *Clinical Genetics*, *55*(5), 346–352. https://doi.org/10.1034/j.1399-0004.1999.550508.x

Todd, P. K., Oh, S. Y., Krans, A., He, F., Sellier, C., Frazer, M., Renoux, A. J., Chen, K. chun, Scaglione, K. M., Basrur, V., Elenitoba-Johnson, K., Vonsattel, J. P., Louis, E. D., Sutton, M. A., Taylor, J. P., Mills, R. E., Charlet-Berguerand, N., & Paulson, H. L. (2013). CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron*. https://doi.org/10.1016/j.neuron.2013.03.026

Yousuf, A., Ahmed, N., & Qurashi, A. (2022). Non-canonical DNA/RNA structures associated with the pathogenesis of Fragile X-associated tremor/ataxia syndrome and Fragile X syndrome. In *Frontiers in Genetics* (Vol. 13, p. 866021). Front Genet. https://doi.org/10.3389/fgene.2022.866021

Zuber, J., & Mathews, D. H. (2019). Estimating uncertainty in predicted folding free energy changes of RNA secondary structures. *RNA*, *25*(6), 747–754. https://doi.org/10.1261/RNA.069203.118

Chapter 2:

FMRP in Fragile X Syndrome (FXS) patient peripheral blood mononuclear cells (PBMCs)

Jamie L. Randol, Kyoungmi Kim, Matt D. Ponzini, Flora Tassone, Alexandria K. Falcon, Paul J. Hagerman

Contributions

Kyoungmi Kim and Matt Ponzini performed nested linear mixed-effects modeling and suggested the use of one-sample t-tests. Flora Tassone provided data for *FMR1* CGG repeat size, methylation status, and mRNA levels generated by her lab. Alexandria Falcon assisted in cell lysis and BCA quantification of total protein. I thank Glenda Espinal for sharing her experiences with the Cisbio TR-FRET protocol and giving her opinions when I encountered challenges. I thank Randi Hagerman and her team for drawing blood samples. All other protocols, processing, and designs were completed by Jamie Randol with support from Paul Hagerman.

This research was sponsored by Randi Hagerman's NIH grant R01 HD036071; Kyoungmi Kim's NIH grant U54 HD079125, supporting the MIND Institute Intellectual and Developmental Disabilities Research Center; and Randi and Paul Hagerman's grant from the Azreli Foundation.

Introduction

Fragile X syndrome (FXS) is an X-linked neurodevelopmental disorder that is caused in the vast majority of cases by reduced or absent expression of the protein product (FMRP) of the *Fragile X messenger ribonucleoprotein 1* (*FMR1*) gene (Hagerman et al., 1994; Pieretti et al., 1991; Richter & Zhao, 2021; Schmitt et al., 2019). Expansion of a non-coding trinucleotide

(CGG) microsatellite located in the 5' untranslated region (UTR) of *FMR1* above 200 repeats (full mutation) triggers hypermethylation and transcriptional silencing, with consequent reduced or absent FMRP translation. FMRP is an RNA-binding protein with myriad binding partners and functions in the brain. It regulates synaptic plasticity, neural development, and cognitive function, primarily by regulating the translation, transport, and stability of many mRNAs (Richter & Zhao, 2021). Unsurprisingly, the loss of FMRP results in intellectual disability and, often, attention-deficit/hyperactivity disorder (ADHD), anxiety, autism, and other neurological and behavioral symptoms. In fact, full mutation alleles are the most common cause of inherited intellectual impairment. Predictably, low FMRP levels are associated with the severity of FXS (Budimirovic et al., 2020; Graef et al., 2020; Kim et al., 2019; Loesch et al., 2004; Pretto, Yrigollen, et al., 2014) and other neurological disorders, such as autism, bipolar disorder, depression, and schizophrenia (Fatemi & Folsom, 2011).

Interestingly, changes in *FMR1* CGG repeat length are known to cause various other disorders as well. Alleles with between 6 and 44 CGGs are generally regarded as having a normal *FMR1* phenotype, with rare exceptions for various point or insertion/deletion mutations within the coding region of the gene (Carroll et al., 2020; Sitzmann et al., 2018). Those with repeats between 55-200 CGGs are termed "premutation," given their propensity to expand to full mutation alleles within one generation (Nolin et al., 2019; Yrigollen et al., 2012). Unlike full mutations, premutations are generally unmethylated and frequently produce excess mRNA and relatively normal to slightly reduced levels of FMRP (Devys et al., 1993; Pieretti et al., 1991; Tassone, Hagerman, Taylor, Mills, et al., 2000). Therefore, most individuals with premutation alleles have neurodevelopmental trajectories associated with normal cognitive function. However, carriers of premutation alleles are predisposed in adulthood to develop other physical and psychiatric disorders, including fragile X primary ovarian insufficiency (FXPOI) (Fink et al., 2018); fragile X adult onset neuropsychiatric disorder (FXAND) (Hagerman et al., 2018); and fragile X-associated tremor/ataxia syndrome (FXTAS) (Cabal-Herrera et al., 2020; Hagerman &

Hagerman, 2021; Robertson et al., 2016), a late adult-onset, neurodegenerative disorder with core features of intention tremor, cerebellar ataxia, parkinsonism, and cognitive decline.

Individuals with FXS are typically mosaic, possessing multiple alleles that can cross allele size boundaries; it is not uncommon for an individual with FXS to have unmethylated normal or premutation alleles in addition to multiple full mutation alleles. Moreover, methylation mosaicism is also a frequent occurrence. Some full mutation alleles can be fully or partially unmethylated (Jiraanont et al., 2017). The large degree of mosaicism contributes to varying levels of FMRP production and is a confounding factor in determining the exact relationship between CGG repeat size, methylation, transcription, and translation at the *FMR1* locus, particularly in the full mutation range.

Given the allelic complexity associated with FXS, accurate measurement of FMRP levels in patients with FXS is essential for better understanding the relationship between FMRP levels and the FXS phenotype, and for assessment of potential treatments for the disorder. The earliest methods for quantifying FMRP levels, including western blotting, immunocytochemistry, or immunohistochemistry, were low-throughput, only semiquantitative, and relatively labor-intensive (Lafauci et al., 2016). The introduction of assays using two unique antibodies to simultaneously bind FMRP improved both sensitivity and specificity (Iwahashi et al., 2009; Kim et al., 2019; Lafauci et al., 2013, 2016; Roth et al., 2021; Schutzius et al., 2013).

One such assay, time-resolved fluorescence resonance energy transfer (TR-FRET), utilizes a donor fluorophore conjugated to one unique anti-FMRP antibody and a second (acceptor) fluorophore conjugated to a second unique anti-FMRP antibody. The two fluorophores are brought into close proximity only when both antibodies are bound to the same FMRP molecule. Upon excitation, the donor fluorophore transfers energy to the acceptor fluorophore. TR-FRET measures this long-lived energy transfer after short-lived background fluorescence decays, thereby improving sensitivity and specificity over traditional FRET. Another advantage of TR-FRET is that it can occur in homogenous cell lysate, eliminating the

need for multiple wash or separation steps. Therefore, TR-FRET for FMRP can easily be scaled up for measuring large sample sizes or screening compounds relevant for the treatment of FXS and/or other *FMR1* disorders (Bidinosti et al., 2012; Lafauci et al., 2016; Schutzius et al., 2013).

The current study exploits TR-FRET to explore FMRP production from FXS individuals using a large sample size in an accessible and diagnostically important cell type: peripheral blood mononuclear cells (PBMCs). Here we report FMRP levels determined by homogenous TR-FRET from 390 observations of 293 individuals with and without FXS and continue to explore its relationship to *FMR1* repeat size, methylation status, and mRNA production. We found a high degree of size and methylation mosaicism and that significant levels of FMRP were only produced from such mosaics. Additionally, a reduction in translation efficiency was observed as repeat size of the smallest allele increased in the full mutation range. Despite evidence of excess mRNA production in the full mutation range, no PBMCs with their smallest allele above 273 CGGs produced significant FMRP.

Materials and Methods

Participants and sample storage

Blood samples used in this study were drawn exclusively from males; this restriction is necessary to eliminate the confounding effects of normal FMRP expression from the fraction of active, normal-repeat *FMR1* alleles (activation ratio) in females. Peripheral blood mononuclear cells (PBMCs) were collected by venipuncture at the UC Davis MIND Institute, under protocols approved by the Institutional Review Board and ethics committee at the University of California, Davis. Between 2006 and 2017, bloods were drawn from a total of 339 individuals: 172 having been diagnosed previously with FXS, 155 typically developing controls, and 12 with unknown *FMR1* allele status.

Originally, 461 frozen PBMC tubes from 339 individuals were processed. Frozen PBMC tubes included technical replicates (multiple aliquots from the same blood draw) and biological replicates (same individual, different date of blood draw). Seventy-one tubes were eliminated, resulting in 390 observations from 293 individuals: 154 diagnosed with FXS and 139 typically developing controls. The final 390 tubes were collected between 2010 and 2017. Tubes were eliminated due to protein degradation from either cell lysis during sample collection or from cryopreservation before the routine use of CPT tubes (n = 49), insufficient protein for analysis (n = 2), loading errors (n = 16), and extreme standard deviation (n = 4). See Table 2.1 for descriptive statistics of all samples (each blood draw from all participants was represented only once) by allele class, CGG repeat size, and age.

PBMC collection and storage

Peripheral blood mononuclear cells (PBMCs) were collected by venipuncture in BD Vacutainer CPT tubes (BD, Franklin Lakes, NJ; catalog number 362761) and processed following the manufacturer's protocol to obtain mostly (~70-90%) lymphocytes. Cells were resuspended in RPMI 1640 (Gibco) with 10% dimethyl sulfoxide (DMSO) and partitioned into one to three aliquots for cryopreservation in liquid nitrogen until needed.

CGG genotyping and methylation status

Genomic DNA was isolated from PBMCs using standard procedures (Qiagen, Valencia, CA). CGG repeat sizing was carried out by a combination of both PCR and Southern Blot analysis as previously reported (Filipovic-Sadic et al., 2010; Tassone et al., 2008). For Southern Blot analysis, 10 µg of DNA was digested with EcoRI and NruI, fixed on a nylon membrane and hybridized with the *FMR1* genomic probe StB12.3, labeled with Dig-11-dUTP by PCR (PCR Dig Synthesis Kit; Roche Diagnostics) following the protocol as previously described (Tassone et al., 2008). PCR analyses were performed using *FMR1* specific primers (AmplideX PCR/CE,

Asuragen, Inc.) and amplicons were visualized by capillary electrophoresis, and analyzed using Gene Mapper software (Filipovic-Sadic et al., 2010). Methylation status, including the percentage of methylation (% of methylated alleles), was determined by densitometric analysis of Southern blot images as described in (Tassone, Longshore, et al., 1999).

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 2,100 Bioanalyzer system. cDNA synthesis and determination of *FMR1* mRNA expression levels were performed using real-time PCR (qRT-PCR) as described in (Tassone, Hagerman, Taylor, Gane, et al., 2000). Three references gene were used, including β-glucuronidase (*GUS*), hydroxymethylbilane synthase (*HMBS*), and hypoxanthine-guanine phosphoribosyltransferase (*HGPRT*). Details are as described in (Tassone, Hagerman, Taylor, Gane, et al., 2000). Relative RNA was calculated by normalizing to the mean *FMR1* mRNA value of control samples in this study.

Homogenous TR-FRET assay for FMRP measurement

Previously, we found that FMRP was somewhat unstable in formerly frozen PBMC isolates (data not shown), most likely due to protease activity from lysed residual granulocytes or other cell types during cell collection and storage. Therefore, to avoid protein degradation, 200 µl of Dulbecco's phosphate-buffered saline (DPBS; Gibco, catalog number 14190144) supplemented with Roche cOmpleteTM Ultra Protease Inhibitor Tablets (MilliporeSigma, Burlington, MA; catalog number 5892791001) was added to approximately 1 ml of frozen PBMCs and to a fibroblast fiducial (1062-09) that were partially thawed by hand-warming to prevent DPBS from freezing on contact. PBMCs were then fully thawed in a 37 °C dry bath with intermittent gentle vortexing to mix protease inhibitor with thawing cells.

Cells were pelleted by centrifugation at 1500×g for five minutes at 4°C and washed with 100-200 µl DPBS with protease inhibitor. Cells were briefly spun to re-pellet and the supernatant was carefully aspirated. Cells were lysed in 85 µl of 1X Cisbio Human FMRP lysis buffer supplemented with cOmplete™ protease inhibitor, 0.25 U/µl Benzonase (Millipore Sigma, Burlington, MA; catalog number 70664-3), and 2 mM $MgCl_2$. Pellets were disrupted by pipetting followed by rotation at room temperature for 2-3 hours. Lysates were then spun at 16,000×g for 3-6 minutes to pellet any debris or unlysed cells. The resulting supernatants were used to perform total protein concentration analysis in triplicate using Pierce™ BCA Protein Assay (Thermo Fisher Scientific, Rockford, IL; catalog number 23225) with an 8-fold dilution to preserve sample.

The fluorescence resonance energy transfer (FRET) method was used to quantify FMRP using the Cisbio Human FMRP assay (CisbioUS, Bedford, MA; catalog number 63ADK038PEC0) following the manufacturer's protocol. Individual lysates were diluted to two protein concentrations differing by a factor of two, each within the range of 0.75 to 6.3 µg total protein in supplemented lysis buffer. Ten microliters of each total protein concentration were loaded in quadruplicate in a 384-well Opti-Plate (Perkin Elmer, Boston, MA; catalog number 6007290). Ten microliters of homogenous time-resolved fluorescence (HTRF) technology pre-mixed antibodies was added to each well. The FRET plate was rocked overnight for 18 hours at room temperature, and then read on the PerkinElmer VictorX5.

The CisBio FRET assay takes advantage of HTRF technology in which, upon excitation at ~620 nm, a Eu^{2+}-Cryptate-conjugated donor transfers energy to a d2-conjugated acceptor when both monoclonal antibodies are brought into proximity of one another by binding to separate locations on FMRP. FRET measurements occurred over a 400 µs window after a 50 µs delay to allow decay of short-lived (ns) background fluorescence, such as from direct excitation of the acceptor. Readings at 615 nm (donor) and 665 nm (acceptor) were taken, and

ratios calculated as [$\text{Ratio} = \frac{\text{Signal 665 nm}}{\text{Signal 615 nm}} \text{ x } 10^4$]. The fractional change in this ratio (ΔF%) [$\Delta F\% = \frac{\text{Ratio Sample} - \text{Ratio lysis buffer}}{\text{Ratio lysis buffer}} \text{ x } 100$] was computed and used to determined relative FMRP concentrations (below).

FMRP quantification

Calculations

FMRP levels were quantified by interpolating ΔF% on a standard curve using a fibroblast fiducial line (1062-09) run alongside PBMC samples from study participants. The same fibroblast fiducial was used in the FRET analysis of (Kim et al., 2019). The fiducial was grown in ten T-175 flasks to 80-90% confluency then pooled and cryopreserved at 5×10^5 cells/ml. Samples with ΔF% > 65 were interpolated using a four-factor fit generated from 0 to 3.5 µg total protein of the fiducial to account for non-linearity of the model. However, for samples with ΔF% ≤ 65, interpolations used a linear fit generated for 0 to 0.4 µg of the fiducial. This allowed for negative ΔF% replicates to be interpolated and more accurate FMRP determination for samples with FMRP at or near zero. A ΔF% of 65 was chosen as the value to unite the two models since the interpolated FMRP values for both models were approximately equal for this ΔF% value. Next, interpolated FMRP values were corrected for PBMC total protein loaded (FMRP/ µg). Finally, all corrected FMRP values were normalized to the mean corrected FMRP for individuals with control alleles and known mRNA levels ($FMRP_{rel}$).

Outlier removal

During FMRP quantification, we noted some wells with unexpected wavelength readings or large variation in predicted protein. As a result, we tested the electronic pipette that was used for repeated dispensing and noticed sporadic anomalous ejections, despite a majority of correct ejections. Therefore, outliers were assessed at three stages during FMRP quantification. At all

stages, values were considered extreme outliers and removed from analysis if they were less than or greater than 3-fold the interquartile range (IQR) from the first (Q1) or third (Q3) quartile, respectively: outlier $< Q1 - 3 * IQR$ or outlier $> Q3 + 3 * IQR$. First, wells from the 384 well plate were removed based on their 615 nm wavelength readings. These represented technical pipetting issues in which no conjugate or double conjugate was added to a FRET well. Next, a well was removed from analysis if its corresponding $FMRP_{rel}$ was an extreme outlier among the 8 measurements for its sample. Finally, a sample itself (all 8 measurements) was removed from analysis if its $FMRP_{rel}$ standard deviation was an extreme outlier among the standard deviations for all 390 observations in the current study. This generally resulted from pipetting issues in which more or less than 10 µl total protein was added to a FRET well and for which some replicates showed FMRP within the full mutation range while other replicates showed FMRP within the normal range.

Protein significance

Because the number of replicates within each sample is small (n= 3-8), with some having non-normal distributions by the Shapiro-Wilk normality test, FMRP significance was determined via one-sided, one-sample t-tests on FRET ratios after correcting to two types of negative controls. FRET ratios, rather than ΔF%, were used to determine FMRP significance as these are the raw readings from the Victor X5 before further uncertainty is introduced by interpolation at the low end of the standard curve.

First, wells containing only lysis buffer in the absence of total protein were used to determine background fluorescence on a plate-to-plate basis. Second, an *FMR1* deletion PBMC sample (195-13) was run to identify background fluorescence in the presence of total protein, but absence of FMRP. Sample FRET ratios were corrected first to the median FRET ratio for its plate's lysis buffer control and then to the median lysis buffer-corrected FRET ratio of the deletion control. This was done separately for the two protein concentrations for each sample

that varied from 0.75 to 6.3 µg per well. A sample was determined to have significant FMRP (FMRP(+)) if the higher concentration had a p-value < 0.05. That is, samples whose lower concentration was significant while the higher concentration was non-significant were considered non-significant overall (FMRP(-)) as the assay is more variable for lower concentrations of total protein.

Statistical analyses

To assess the effects of mRNA, unmethylated CGG repeats, methylated CGG repeats, fraction methylated, and/or age on FMRP, regression analyses were performed using nested linear mixed-effects models to incorporate nested data structures for FMRP (technical replicates nested within biological replicates). The models included mRNA, age, unmethylated CGG repeats, methylated CGG repeats, and fraction methylated as fixed effects, a random intercept for biological and technical replicates, and a random slope for age. The median CGG repeat size - calculated using all CGG repeats, unmethylated and methylated CGG repeats separately, and the lower bound of smears plus one-quarter the range of the smears - was used in the regression analyses. The first quartile of the smear range was used to more heavily weight smaller alleles more likely to contribute to FMRP. We fitted the following sequential models: 1) the model that fitted the effects of mRNA and unmethylated CGG repeats on FMRP, controlling for age (aka Model 1), 2) the model that fitted the effects of mRNA and methylated CGG repeats on FMRP, controlling for age and fraction of methylated CGG repeats (aka Model 2), and 3) the model that fitted the effects of mRNA, unmethylated CGG repeats, and methylated CGG repeats on FMRP, controlling for age and fraction of methylated CGG repeats (aka Model 3). Finally, likelihood ratio tests (LRT) were performed to compare the sequential models. All the analyses were conducted using the open-source R software (version 4.2.1).

Results

PBMCs contain lower levels of FMRP relative to fibroblasts

A control dermal fibroblast line was used as a fiducial to generate a standard curve for relative FMRP levels, both for PBMC samples in this study and for dermal fibroblasts in (Kim et al., 2019). Interpolation on this standard curve followed by correction for total protein loaded produced a mean of 0.26 FMRP/µg for PBMCs with control alleles. That is, control PBMCs produce approximately 4-fold less FMRP for the same total protein than the control fibroblast fiducial line. Subsequent analyses were performed on FMRP values normalized to the mean of control PBMCs (0.26 FMRP/ µg was normalized to 1.0 as the control PBMC mean).

Assessing the accuracy of the FRET FMRP assay

The accuracy of FMRP determinations were assessed by coefficient of variation (CV) [CV (%) = 100 × σ/$FMRP_{rel;}$ σ = standard deviation] (Fig. 2.1). $FMRP_{rel}$ levels greater than or equal to 0.5 generally had CV values less than 25% and corresponded to control samples only. Samples were then separated by FMRP significance based on one-sample t-tests of corrected FRET ratios (See Methods). For samples with significant protein and $FMRP_{rel}$ below 0.25, CV values ranged from ~38-110% and corresponded to non-control samples only. Samples with non-significant protein had much larger CVs, reflecting small and/or negative interpolated values of FMRP. Given that FMRP is approximately 4-fold lower in PBMCs compared to fibroblasts, the current results were not unexpected. Kim and colleagues (Kim et al., 2019) found that dermal fibroblasts with $FMRP_{rel}$ of 0.2 generally had CVs less than 20% and fibroblasts with $FMRP_{rel}$ below 0.1 had CVs up to 100%. That is, $FMRP_{rel}$ of 0.05 in fibroblasts corresponds to ~0.2 in PBMCs and both FMRP levels can approach CVs of ~100% in their respective studies.

Biological and technical replicates

The FRET assay reproducibility for PBMCs was assessed by comparing FMRP significance (See Methods) and standard error of the mean (SEM) among biological replicates from the same individual and technical replicates from the same blood draw. Of 27 individuals with multiple blood draws, only 2 (7.4%) had biological replicates that differed in significance of FMRP. That is, at least one biological replicate was FMRP(+) and at least one was FMRP(-) (Fig. 2.2). Both individuals had non-control alleles whose corresponding relative FMRP values occurred at the transition in ability to significantly detect FMRP, indicating that the samples likely have protein, but at levels difficult to detect by significance testing. Similarly, of 68 individuals with multiple measurements of the same blood draw, only 4 (5.9%) had technical replicates that differed in FMRP significance (Fig. 2.3). Again, all 4 occurred at the transition between non-significant and significant FMRP.

SEMs were calculated and presented as a percent of the replicate group mean (percent variability) for both biological and technical replicates. SEM was used instead of standard deviation to account for different sample sizes (range of 2- 4) among replicate groups. The percent variability followed trends similar to those of CVs for all samples (Fig. 2.1, Fig. 2.4). For both biological and technical replicates, $FMRP_{rel}$ levels greater than or equal to 0.5 generally had variability values less than 25% and corresponded to control samples only. For samples with $FMRP_{rel}$ below 0.25, variability values ranged from -194% to 109% and corresponded to non-control samples only. High variability is consistent with small values of FMRP. A relative FMRP of 0.25 is an FMRP level that is 25% that of control PBMCs. However, given that PBMCs contain ~4-fold less FMRP than fibroblasts, 0.25 would correspond to only ~6% the level of FMRP in control fibroblasts, indicating that the assay is detecting very low levels of FMRP. Finally, no plate-effect was detected when performing t-tests between technical replicates run on different plates (data not shown).

FMRP levels are independent of year of blood draw and age

Biological (same individual, different blood draw) and technical replicates (same individual, same blood draw run on a different day) were randomly eliminated to generate a sample set with one observation per individual. Relative FMRP was plotted against the date of blood draw for patients with known blood draw dates (Fig. 2.5). Linear regression analysis showed that FMRP levels are independent of the date of blood draw in individuals with control (n = 135 unique individuals with known blood draw dates, p = 0.07213) and non-control (n = 143 unique individuals with known blood draw dates, p = 0.143) alleles, suggesting that FMRP is stable in cryopreserved whole PBMCs stored in liquid nitrogen for nine years, from 2010 to 2019 when PBMCs were assayed. Therefore, year of draw was not considered as a source of variation to bias results in subsequent statistical analyses. Relative FMRP was also plotted by age at time of draw (Fig. 2.6). Linear regression analysis showed no association between FMRP levels and age in individuals with control (n = 138 unique individuals, p = 0.4733) and non-control (n = 155 unique individuals, p = 0.1277) alleles.

Patients with full mutation alleles generally have complex genotypes

Individuals with full mutation alleles generally are mosaic for number of alleles, size, and/or methylation. That is, individuals with nominally full mutation alleles tend to have multiple alleles whose CGG repeats can span allele classes and be distinctly methylated. (Table 2.1, Fig. 2.7). Of 153 individuals with expanded CGG-repeat alleles covering the full mutation range, only 26 (~17%) possessed a single detectable allele. Of the 127 remaining individuals, there were 19 with methylation mosaicism, one who was mosaic for allele class, 41 who were mosaic for both methylation and allele class, and 66 with up to eight discrete full mutation methylated alleles. Discrete allele sizes ranged from 13 to 1400 CGGs (Table 2.1).

Additional complexity was created by the presence of smears and degree of methylation. A smear is a quasi-continuous series of PCR products that differ in length by only a few CGG

repeats such that they appear as a smear rather than a discrete band when run on a gel or an electrophoretogram. Of the 153 individuals with nominally full mutation alleles, 20 had unmethylated smears and 4 had methylated smears. In 61 samples with unmethylated alleles, the fraction of unmethylated alleles ranged from 5% to 95% of all alleles present. Smear allele sizes ranged between 30 and 1540 CGGs (Table 2.1). Unsurprisingly, expanded-repeat samples with the largest significant FMRP levels were those with size and/or methylation mosaicism. Interestingly, not all samples with detectable RNA produced detectable FMRP (Fig. 2.7).

Both methylated and large full mutation alleles produce mRNA

Methylation mosaics with only full mutation alleles (methylation mosaic) can produce mRNA: 0 to 2.28-fold relative to the mean of control samples (Fig. 2.8). Sample P03-10, with an unmethylated allele at 250 CGGs, produced the highest relative *FMR1* mRNA levels at 2.28 ± 0.085 (mean ± SEM), suggesting that smaller full mutation alleles can produce large quantities of mRNA, even above the range for samples with control alleles (0.47 – 1.43) and similar to the excess mRNA production in the premutation range (Tassone, Hagerman, Chamberlain, & Hagerman, 2000; Tassone, Hagerman, Taylor, Gane, et al., 2000; Tassone, Hagerman, Loesch, Lachiewicz, et al., 2000). Larger full mutation alleles can produce mRNA as well, though to a lesser degree. For example, sample P06-32 had an unmethylated 500 CGG repeat plus a methylated 860 repeat and still produced some mRNA (0.37 ± 0.046 relative to control mean).

Additionally, in the absence of evident unmethylated alleles, some full mutation alleles are still capable of producing detectable mRNA: 0 to 0.30-fold relative to the control mean (Fig. 2.8). Sample P08-19 had large, methylated alleles of sizes 760 and 870 CGG repeats, yet produced 0.08 ± 0.0046 relative mRNA.

Large full mutation alleles produce little to no FMRP, rarely approaching control levels despite excess mRNA

Despite the ability of full mutation alleles to produce *FMR1* mRNA, they make little to no detectable levels of FMRP. Samples with any form of full mutation allele have low levels of FMRP, thus forming a distinct group from samples with control alleles (Fig. 2.9). However, rarely, size and methylation mosaic samples approached the lower bound of FMRP levels produced from control alleles. This resulted in a small range of overlap between the highest FMRP levels for size and methylation mosaics (max = 0.49) and the lower bound of control samples (min = 0.39). Notably, excess mRNA did not guarantee higher levels of FMRP. Several samples have near-zero FMRP levels despite a three-fold excess of *FMR1* mRNA relative to control means (Fig. 2.9).

One-sided, one-sample t-tests on corrected FRET ratios (See Methods) resulted in only 27 of 199 (~14%) non-control samples with significant FMRP (Fig. 2.7). Non-control FMRP-positive samples were exclusively size and/or methylation mosaics. Twenty were both size and methylation mosaics, many of which contained unmethylated premutation alleles. Four had methylated premutation alleles. Only three samples with purely full mutation alleles had evidence of FMRP: P03-10, P10-04, and P12-30 (Fig. 2.7). Notably, all three contain unmethylated full mutation alleles. No sample containing only methylated full mutation alleles produced significant FMRP. P03-10 had methylated alleles between 470 and 800 CGGs and one unmethylated allele at approximately 250 CGGs. It produced *FMR1* mRNA and FMRP at levels of 2.28 ± 0.085 and 0.15 ± 0.032, respectively, relative to control means. P10-04 had a methylated 220 CGG repeat and an unmethylated 200 CGG repeat. It produced 1.60 ± 0.031 *FMR1* mRNA, but only 0.09 ± 0.036 FMRP, compared to the control means. P12-30 had methylated alleles between 273 and 810 CGGs and one unmethylated allele at 340 CGGs. Its relative *FMR1* mRNA was 0.25 ± 0.015 and relative FMRP levels were 0.25 ± 0.037.

There is a significant association between *FMR1* mRNA and FMRP in non-control PBMCs

To further characterize the effect of *FMR1* mRNA on FMRP production, linear regression analysis was performed. Biological and technical replicates were randomly eliminated to generate one observation per individual. A linear regression was then performed separately on control and non-control samples, with relative *FMR1* mRNA against relative FMRP (Fig. 2.10). No significant association between *FMR1* mRNA and FMRP in samples with control alleles was found (n = 134 unique individuals with both mRNA and FMRP values, slope = -0.041, p = 0.73). However, there was a significant association between relative *FMR1* mRNA and FMRP in non-control samples (n = 148 unique individuals with both mRNA and FMRP values, slope = 0.059 relative FMRP units per relative mRNA unit, $p = 1.21 \times 10^{-9}$).

This association in non-control samples was driven by those with the lowest FMRP values. Visual inspection of a Loess regression on mRNA vs. FMRP in all non-control samples with available mRNA data (n = 191) shows that between 0 and ~0.5 relative mRNA, increasing mRNA leads to increasing protein. However, above ~0.5 relative mRNA, relative FMRP levels plateau at ~0.25, at which point further increases of *FMR1* mRNA no longer influence FMRP levels (Fig. 2.11). Linear regression on FMRP-positive and FMRP-negative non-control samples supported the visual inspection (Fig. 2.12). Non-control samples were separated by FMRP significance (See Methods), and linear regression analysis was repeated to understand which samples are driving the association between mRNA and FMRP. FMRP(+) samples no longer had a significant association between *FMR1* mRNA and FMRP (n = 18, slope = -0.0024, p = 0.93). However, levels of FMRP(-) samples were positively associated with *FMR1* mRNA levels (n = 131, slope = 0.028, p = 0.011), suggesting that samples with non-significant FMRP (by TR-FRET analysis) do produce some FMRP, albeit at levels too low for significance testing on an individual sample basis.

Translation efficiency negatively associates with the smallest CGG repeat size in samples with full mutation alleles

Considering the ability of full mutation alleles to produce mRNA, but little protein, we next examined the ratio of relative FMRP to relative *FMR1* mRNA to approximate the efficiency of protein production in FMRP(+) individuals. Biological and technical replicates were randomly eliminated to generate one observation per individual. The ratio of relative FMRP to mRNA was plotted against the smallest CGG repeat size regardless of methylation status (Fig. 2.13). The smallest CGG repeat size was chosen to represent an individual's allele most likely to contribute to protein production. A linear regression was fitted for individuals with either control or non-control (expanded CGG-repeat) alleles, respectively. No association was detected for samples with control alleles (n = 134 unique individuals with both FMRP and mRNA data, slope= 0.0065 relative efficiency units per CGG, p-value = 0.31). However, a significant negative association was observed for non-control samples (n = 18 unique FMRP(+) individuals, slope= -0.006 relative efficiency units per CGG, p-value = 0.045). That is, as the minimum CGG repeat size increased, less FMRP was detected for the same quantity of mRNA, suggesting a decrease in translation efficiency for alleles with larger repeats, in agreement with earlier studies (Kenneson et al., 2001; Primerano et al., 2002; Schneider et al., 2020; Tassone, Hagerman, et al., 1999; Tassone, Hagerman, Taylor, Gane, et al., 2000).

No significant FMRP production was detected for alleles greater than 273 CGG repeats

To estimate the largest allele capable of producing FMRP *in vivo* in the current study, each sample was represented by its smallest (lower-bound) CGG repeat size, regardless of methylation status, and plotted against relative FMRP (Fig. 2.14). The lower-bound CGG repeat size was chosen to represent an individual's allele most likely to contribute to protein production. Sample P12-30 had the largest lower-bound CGG repeat, at 273 CGGs, while still producing significant FMRP. That is, no sample with its lower-bound allele larger than 273 CGG repeats

(from sample P12-30) produced significant levels of protein. Notably, P12-30's 273 CGG allele was methylated. P12-30 also possessed an unmethylated 340 CGG repeat. It is unclear from this data which allele or combination of alleles produced the *FMR1* mRNA and subsequent FMRP.

FMR1 mRNA significantly affects FMRP, and unmethylated CGG repeat size trends toward significance in nested mixed-effects models

Three nested linear mixed-effects models were fitted to examine the effects of mRNA level, the median unmethylated CGG repeat size, the median methylated CGG repeat size, and fraction of methylated alleles in a hierarchical manner to account for individuals with either biological or technical replicates while controlling for patient age (see Methods).

Model 1: $FMRP = Age + mRNA + CGG_{unmethylated}$

Model 2: $FMRP = Age + mRNA + CGG_{methylated} + Fraction\ Methylated$

Model 3: $FMRP = Age + mRNA + CGG_{unmethylated} + CGG_{methylated} + Fraction\ Methylated$

All three models showed a significant dependence of FMRP on *FMR1* mRNA level in a linear fashion for non-control samples with median allele sizes in the full mutation range ($p < 0.02$ for all; Table 2.2, Fig. 2.15). A likelihood-ratio test (LRT) comparing the models with and without the unmethylated CGG repeat size (Model 2 vs. Model 3) suggested that the size of the median unmethylated CGG repeat significantly affects FMRP levels, when accounting for the median methylated CGG repeat size, fraction of methylated alleles, patient age, and mRNA level ($p = 0.0368$). Models 1 and 3 differ by inclusion of both the methylated CGG repeat and the fraction of methylated alleles. An LRT comparing Models 1 and 3 suggested that the size of the methylated CGG repeat and the fraction of methylated alleles had no significant impact on the relation of mRNA level with FMRP levels, while accounting for the unmethylated CGG repeat, and patient age ($p = 0.6079$). This is likely because total mRNA level encompasses the effect of the fraction of methylated (and thus unmethylated) alleles. Moreover, methylated CGGs

likely contribute little to protein, as evidenced by the fact that no sample containing only methylated full mutation alleles had significant levels of FMRP (Fig. 2.7). Therefore, LRTs comparing all three models suggested that the variation in FMRP production is most explained by mRNA level and the unmethylated CGG repeat size, while controlling for age. That is, FMRP is significantly affected by mRNA level and the median repeat size of an individual's unmethylated alleles.

When fixing an individual's age and median size of the unmethylated CGG repeat size with the population average, doubling the level of relative *FMR1* mRNA resulted in an 11.3% (2*0.0565*100% = 11.3%) increase in relative FMRP for individuals with alleles in the full mutation range, on average (See mRNA slope of Model 1 in Table 2.2).

The length of the unmethylated CGG repeat was identified as a factor affecting FMRP production via LTR testing between Models 2 and 3. However, when analyzing Models 1 and 3 independently, neither showed significance of the median unmethylated CGG repeat size, though it trended toward significance in both (Model 1: $p = 0.0596$; Model 3: $p = 0.0513$). While fixing an individual's age and their relative *FMR1* mRNA level at their average values in Model 1, doubling the length of the median unmethylated CGG repeat resulted in a 0.04% (2*(-0.0002)*100% = -0.04%) reduction in relative FMRP in individuals with alleles in the full mutation range, on average (see slope of $CGG_{unmethylated}$ in Model 1 of Table 2.2).

Discussion

The *Fragile X messenger ribonucleoprotein 1* (*FMR1*) gene contains a trinucleotide CGG repeat of variable size in its 5'-UTR. Repeat expansions of greater than 200 CGGs are associated with Fragile X syndrome (FXS), a neurodevelopmental disorder resulting from little to no expression of *FMR1* protein (FMRP). Reduced expression can result from either transcriptional silencing of *FMR1* or decreased translation efficiency of *FMR1* mRNA in the

presence of expanded repeats. Here, we assess FMRP levels by homogeneous time-resolved fluorescence resonance energy transfer (TR-FRET) in previously frozen peripheral blood mononuclear cells (PBMCs) from control individuals (CGG_n = 6-44) and individuals with FXS that possess at least one full mutation allele (CGG_n > 200). The term “non-control” is used to refer to alleles from FXS patients to encompass those with size mosaicism.

Neither the years PBMCs were stored in liquid nitrogen (controls: p = 0.072; non-controls: p = 0.14) nor an individual’s age at the time of blood-draw (controls: p = 0.47; non-controls: p = 0.13) affected FMRP levels. That is, linear regression analyses showed no significant differences from a slope of 0 for either condition (Fig 2.5, 2.6). Moreover, the slope of FMRP relative to year of draw (Fig 2.5) is slightly negative in samples with control alleles, with samples more recently collected and stored showing slightly less FMRP, dissuading the interpretation that storage of PBMCs in liquid nitrogen degrades FMRP over time.

An age-effect was not observed for individuals with either control (1-80 years) or non-control (0-57 years) alleles in the current study. However, using a Luminex assay on dried blood spots, LaFauci and colleagues showed a decline in FMRP in individuals with control alleles between infancy and preteen ages that leveled off in adulthood, with seven-fold higher expression in neonates compared to adults (Adayev et al., 2014; Lafauci et al., 2013). A similar trend has been observed in mice. Adayev et al (2021) found highest levels of FMRP in blood from 3-week-old mice and observed a 2.5-fold decrease in expression by seven-weeks of age (Adayev et al., 2021). That research group also found an age-dependent decrease in mouse FMRP over weeks three to 14 in three of four brain regions assayed (Adayev et al., 2021). Likewise, other groups have observed an age-dependent decrease in FMRP in the brain of both male (Singh et al., 2007) and female (Singh & Prasad, 2008) mice. Therefore, statistical modeling from the current study accounted for an individual’s age at blood draw, despite not detecting an association in the current study.

Another finding from this study is the large degree of mosaicism observed in patients with FXS. Mosaicism was observed in three categories: 1) the number of alleles within a mutation class, 2) the number of alleles across mutation classes (size mosaicism), and 3) methylation status of each allele (methylation mosaicism). Size and methylation mosaicism for samples with full mutation alleles, especially in individuals variably affected by FXS, have been observed in many studies (Boggs et al., 2022; Budimirovic et al., 2020; Genç et al., 2000; Loesch et al., 2012; McConkie-Rosell et al., 1993; Nolin et al., 1994; Pretto, Yrigollen, et al., 2014; Roth et al., 2021; Tassone, Hagerman, Loesch, Lachiewicz, et al., 2000). In 2022, Meng and colleagues also observed mosaicism in blood samples, but to a lesser degree (Meng et al., 2022).

The current study of FXS patients found that a single methylated, full mutation allele was detected in only 17% of individuals. All other patients with full mutation alleles had a combination of methylation or size mosaicism, including multiple full mutation alleles. However, size mosaicism is generally defined as possessing alleles that cross class boundaries, rather than multiple alleles within the same class. Considering this definition, 60% of patients in the current study showed no size or methylation mosaicism. Similarly, Meng et al found that 69% of 487 male FXS patients lacked mosaicism in blood (Meng et al., 2022). Moreover, both Meng et al (2022) and this study found similar degrees of methylation mosaicism at approximately 12% (Meng et al., 2022). In contrast, Budimirovic et al found methylation mosaicism in more than half of their blood and buccal samples (Budimirovic et al., 2020). Individuals with size mosaicism alone represented a smaller proportion in this study (<1%) compared to Meng et al 2022 (11%). Accordingly, size and methylation mosaics were a larger percentage of the population in this study (27%) than in the Meng study (8%). Regardless of the exact percentages, this study supports previous findings of the high degree of methylation and size mosaicism in FXS patients.

Size mosaicism is likely due to the instability of the repeat track outside of the control allele range in both meiotically and mitotically dividing cells (Usdin et al., 2014; Zhao & Usdin,

2016). The mechanisms for repeat instability at *FMR1* are not well understood, but likely result from secondary structures of expanded repeats that disrupt DNA replication, repair, and/or recombination (Tabolacci et al., 2022). Repeat number, parental sex, and the presence of AGG interruptions within the CGG repeat have all been shown to influence *FMR1* repeat instability (Nolin et al., 2019; Tabolacci et al., 2022). While expansions and contractions, to a lesser degree, occur in any allele class, premutation alleles are particularly unstable, often expanding into large full mutation alleles in a single generation (Zhao and Usdin 2016, Nolin et al 2019).

How mosaicism, both size and methylation, affect *FMR1* expression has been surveyed. Generally, mosaicism in individuals with full mutation alleles is associated with increased expression and improved cognitive function, given that alleles in the control or premutation range and unmethylated full mutation alleles may contribute to FMRP production (Budimirovic et al., 2020; De Vries et al., 1996; Jacquemont et al., 2018; Kim et al., 2019; Meng et al., 2022; Pretto, Yrigollen, et al., 2014; Tassone, Hagerman, et al., 1999; Taylor et al., 1999). Indeed, there is increasing evidence, contrary to previous dogma, that unmethylated full mutation alleles are actively transcribed (Pretto, Yrigollen, et al., 2014; Tassone, Hagerman, Chamberlain, & Hagerman, 2000). However, more details are needed about which specific alleles can express *FMR1* mRNA and FMRP and to what degree.

In the current study, we found that only samples with size and/or methylation mosaicism produced significant levels of FMRP (Fig. 2.7). Identifying significant levels of FMRP in methylation mosaics confirmed that full mutation alleles express in their native cellular environment. For example, sample P12-30 contained an unmethylated 340 CGG repeat along with methylated alleles of sizes 273, 408, and 810 repeats. This sample produced 0.25 ± 0.015 relative mRNA and 0.25 ± 0.037 relative FMRP. That is, P12-30 achieved one-fourth the level of FMRP with one-fourth the level of *FMR1* mRNA compared to the control population.

Furthermore, FMRP was detectable in many full mutation-only samples but did not achieve statistical significance. Despite this, we found that in the absence of evident

unmethylated alleles, some methylated full mutation alleles are still capable of producing detectable *FMR1* mRNA: 0 to 0.30-fold relative to the control mean (Fig. 2.8). For example, sample P08-19 had large, methylated alleles of sizes 760 and 870 CGG repeats, yet produced 0.06 ± 0.0046 relative mRNA. Currently, the fraction of unmethylated alleles is determined by Southern blot and is based on the methylation status of a single locus (See Methods); however, the degree of methylation along the length of each allele is unknown. Therefore, the degree of methylation may be important to identify which alleles can be transcribed and, potentially, which isoforms thereof.

No association between mRNA and FMRP in control samples was observed (Fig. 2.10). This could be due to the broad variation in FMRP from the control population, which is consistent with previous slot-blot (Kenneson et al., 2001) and Luminex immunoassay (Lafauci et al., 2013) results. However, we did find an association in samples with non-control alleles as assessed via linear regression and via nested linear mixed-effects models in the current study (Figs. 2.10, 2.15). Interestingly, a LOESS fit, a locally weighted polynomial regression, showed that FMRP increased with increasing mRNA up to ~0.5-fold relative mRNA, after which no association between mRNA and FMRP was observed (Fig. 2.11). Therefore, the dependence of FMRP on mRNA level in non-control samples was driven by those with the lowest FMRP levels, which generally failed significance testing (Fig. 2.12). Likely, these samples produced small amounts of FMRP below the level of significance for an individual sample, which nevertheless, uncover an association between mRNA and FMRP when taken as whole. A more sensitive assay will be needed to tease apart the relationship between CGG repeat size, methylation status, mRNA, and FMRP levels moving forward.

Some methylation mosaic FXS samples produce excess *FMR1* mRNA. Control levels of relative mRNA ranged from 0.46 to 1.43 (n = 187 samples), while those for methylation mosaics with FXS ranged from 0 to 2.28 (n = 23 samples) (Fig. 2.8). Accordingly, the highest mRNA level for methylation mosaic FXS patients was ~1.6-fold larger than that for the control

population. This finding is consistent with observations of increased *FMR1* mRNA levels, sometimes five- to ten-fold, in carriers of premutation alleles (Kenneson et al., 2001; Ludwig et al., 2014; Tassone, Hagerman, Taylor, Gane, et al., 2000). However, here we observe that even alleles in the full mutation range can overexpress *FMR1* mRNA. For example, sample P03-10 had an unmethylated 250 CGG repeat along with methylated repeats at 470, 620, and 800. Yet, it produced 2.28 relative mRNA. This is in keeping with a case study of a FXS patient with fully unmethylated full mutation alleles who produced 7-fold excess *FMR1* mRNA (Santa María et al., 2014) and with a mouse study that found a continuing increase in *FMR1* mRNA in mice with full mutation alleles up to the 250 CGG maximum assayed (Ludwig et al., 2014).

However, despite the capacity to produce excess *FMR1* mRNA, FMRP levels remain low in FXS patients (Santa María et al., 2014) (Fig. 2.9). For example, sample P03-10 discussed above had excess mRNA yet only produced ~15% the FMRP level as controls. This is likely due to the difficulty of translation machinery in traversing secondary structure in the *FMR1* mRNA during ribosomal scanning as CGG repeats expand (Feng et al., 1995; Kenneson et al., 2001; Ludwig et al., 2011; Tassone et al., 2001). *In vitro* transcribed CGG repeats have been shown to produce hairpin-like structures with both C-G and G-G base pair bonding and even tetraplex structures resulting from guanine quartets between two such hairpins (Tabolacci et al., 2022; Weisman-Shomer et al., 2000; Zumwalt et al., 2007). Indeed, the ratio of relative FMRP to relative mRNA, a measure of translation efficiency, decreased as the size of the smallest CGG repeat allele in a sample increased in the current study (Fig. 2.13). Others have also found decreased translation efficiency in expanded repeats starting in the intermediate and premutation range by measuring levels of FMRP and *FMR1* mRNA or the percent of *FMR1* mRNA associated with polysomes (Kenneson et al., 2001; Peprah et al., 2009; Primerano et al., 2002). Therefore, it is unsurprising that translation efficiency would continue to decrease in the full mutation range.

Still, the question of what the largest CGG repeat capable of producing FMRP is remains unanswered. In the current study, no sample with a minimum allele size above 273 CGG repeats produced significant levels of FMRP in its native cellular environment (Fig. 2.14). Similarly, Feng and colleagues identified low, but detectable levels of FMRP for a fibroblast clone with 285 CGG repeats via western blot (Feng et al., 1995). However, examining large repeat sizes in the absence of methylation and size mosaicism will more directly answer this question.

A corollary to this question is how much FMRP can be produced by large repeat alleles that express mRNA. As seen in this study, even non-control samples can produce some FMRP, though levels do not approach those of the normal range, unless the sample was mosaic and also possessed repeats in the normal or premutation range (Fig. 2.9). A correlation between repeat size, especially in the premutation range and 1) percent of methylation, 2) *FMR1* mRNA levels, and 3) more clinical involvement has been found (Budimirovic et al., 2020; Pretto, Mendoza-Morales, et al., 2014). And the level of FMRP is directly related to the degree of cognitive and neurodevelopmental impairment (Loesch et al., 2003; Menon et al., 2004). In an unmethylated context, could full mutation alleles produce sufficient FMRP to improve cognitive function?

Our determinations of FMRP are consistent with previous recent studies measuring FMRP in PBMCs. Roth et al 2021 determined PBMC FMRP concentration from a recombinant FMRP (rFMRP) standard curve using a chemiluminescent assay (Roth et al., 2021). Boggs et al 2022 measured FMRP concentration in dried blood spots via an optimized Luminex assay (Boggs et al., 2022). The current study used a TR-FRET assay. To compare results among all three research studies, we normalized values to the mean of reported control samples (Roth: 50.8 fmol/µg, n = 4; Boggs: 28.9 pM, n = 24). Relative FMRP in blood was similar for all three studies. For full mutation only samples, Roth and colleagues measured 6.2% (3.14 fmol/µg, n = 1), Boggs and colleagues measured 2.1% (0.6 pM, n = 36), and the current study measured

5.2% (n = 117). For mosaic samples, the Roth group obtained 26.3% (13.38 fmol/µg, n = 1), the Boggs group obtained 13.1% (3.8pM, n = 18), and the current study obtained 14.2% for all mosaics (n = 80) and 24.0% for mosaics with significant FMRP (n = 27; See Methods).

Study strengths

There are some notable strengths to this study. First, a large sample size of 390 PBMCs from 293 individuals was used. For comparison, other recent studies using sensitive detection methods for FMRP measured between 18 and 187 total participants (Boggs et al., 2022; Budimirovic et al., 2020; Roth et al., 2021). In particular, we had a large FXS sample size (n= 154), which allowed us to examine FMRP in the full mutation range via linear regression and mixed effects modeling independent of patterns in the control or premutation range. Of note, Adayev and colleagues measured FMRP in dried blood spots (DBS) using a quantitative Luminex assay in 2,000 infants. However, importantly, no FXS individuals were assayed and storage time of the DBS significantly reduced FMRP detection (Adayev et al., 2014).

Another strength is the improved recovery of FMRP from PBMCs due to three factors. First, retention of FMRP at the time of blood collection was aided by the introduction of CPT tubes with a polyester gel barrier in ~ 2007. The gel barrier reduces lysis and degradation of protein during sample preparation and subsequent cryopreservation by sequestering denser blood components with lytic agents beneath the barrier and preserving plasma and monocytes above the barrier. Second, the introduction of protease inhibitors during cell thawing helps preserve FMRP by inactivating proteases that were freed from ruptured cells during cryo-recovery. Third, FMRP was assayed in fresh lysate. Recombinant FMRP has been reported to form aggregates and to precipitate from solution (Iwahashi et al., 2009). However, via western blot analysis, we have found native FMRP degradation and changes in concentration in lysate, presumably from aggregation or precipitation, after freeze-thawing as well (data not shown), contrary to other reports that found no degradation in SDS buffer after 6 months at -80°C (Lessard et al., 2012).

To improve the accurate detection of FMRP, fluorescent signals were interpolated on a standard curve generated by serial dilution of a fibroblast line with a control allele. The cellular matrix can interfere with fluorescent antibody assays, altering the background signal and antibody specificity (Zachary et al., 2009). Therefore, FMRP was quantified within a cellular matrix rather than using recombinant FMRP in a buffer (Boggs et al., 2022; Roth et al., 2021). Unless the matrix effect is considered (Boggs et al., 2022), previous studies may have overestimated FMRP expression, especially in blood.

Finally, TR-FRET, in our hands, is a sensitive, medium- to high-throughput assay (depending on application) compared to past techniques used to assess FMRP in cells. Early methods to identify FMRP after identification of the *FMR1* gene in 1991 include immunohistochemistry (De Vries et al., 2003; Devys et al., 1993) and immunoblotting techniques (Kenneson et al., 2001; Ludwig et al., 2014; Pretto, Mendoza-Morales, et al., 2014; Verheij et al., 1993). Both assay types are low throughput, relatively labor-intensive, and often semiquantitative. Western blotting, for example, is limited by transfer efficiency and exposure methods and is particularly challenging for detecting low levels of FMRP present in full mutation samples.

In 1995, Willemsen and colleagues introduced a rapid and inexpensive immunocytochemical (IC) approach to screen for FXS in one to two drops of blood by scoring PBMCs as either positive or negative for FMRP (Willemsen et al., 1995). Interestingly, not all control PBMCs stained positively for FMRP. In fact, individuals with control alleles had as low as 40% FMRP-positive cells. Their approach is therefore less useful in detecting females with full mutation or premutation alleles, as they can have up to 80% FMRP-positive PBMCs (De Vries et al., 1998). Moreover, the IC approach cannot distinguish cells that express FMRP to different degrees and how that impacts overall FMRP levels (Tassone et al., 2001; Willemsen et al., 1997). Finally, that method involves operator subjectivity in assigning which cells are positive and which are negative for FMRP.

Capture sandwich immunoassays improved the sensitivity and specificity of detecting FMRP by requiring the binding of two specific anti-FMRP antibodies. TR-FRET uses fluorophore-labeled antibodies that recognize close epitopes on FMRP. When the donor flour is brought into close proximity via binding to the acceptor fluor, a detectable transfer of resonance energy occurs. An advantage of FRET assays is the one-step nature in which both antibodies are incubated simultaneously in cell lysate without separation or wash steps, making it amenable to high-throughput screening (Bidinosti et al., 2012; Lafauci et al., 2016).

However, other capture sandwich immunoassays exist. In 2009, a sandwich Enzyme-Linked ImmunoSorbent Assay (ELISA) was developed (Iwahashi et al., 2009). In 2013, a Luminex-based capture immunoassay was created by a separate group using independently generated antibodies (Lafauci et al., 2013). And in 2021, Roth and colleagues introduced an electro-chemiluminescent ELISA immunoassay (Roth et al., 2021).

Interestingly, despite relative FMRP values in their chemiluminescent assay being similar to those in the current study using Cisbio TR-FRET, Roth and colleagues found that in their hands the Cisbio TR-FRET assay had a lower limit of detection (LLOD) 100-fold less sensitive. That is, they found a LLOD of 85 fmol/µg in cell lysate and 7fmol for rFMRP. The reported LLOD for their electrochemiluminescent assay was 0.07 fmol per reaction (Roth et al., 2021).

Study limitations

There are several limitations of this study. First is the underdetermination of complex molecular features, a circumstance that plagues *FMR1*-related research. For example, methylation status of various alleles is determined by assessing a single locus within *FMR1.* It is not clear how methylation of that locus affects expression, as even methylated alleles sometimes produce mRNA. Additionally, identifying all alleles in a sample is challenging given the presence of smears or cryptic alleles that represent a minor but contributing fraction. Due to this complexity, smears are often ignored, and alleles are simplified to the smallest allele or the

most prevalent allele. For modeling, this study calculated the lower and upper limit of smears and represented them as a unique allele by using the first quartile of the smear range. While it is important to include smear in analyses, how to best represent them is still in unclear. Furthermore, total *FMR1* mRNA levels were assessed for this study. The contribution and impact of different *FMR1* mRNA isoforms was not addressed.

Second, as FXS is an X-linked disorder, only males were used in this study. Females were not analyzed due to the confounding effect of X-inactivation. Third, FMRP was calculated relative to the mean of control samples, limiting our understanding of absolute FMRP levels. However, no standard method for measuring FMRP exists, even among projects that report absolute FMRP values. That is, there are varied detection assays and reporting strategies: fmol/µg (Roth et al., 2021), pM per reaction (Boggs et al., 2022; Roth et al., 2021), pg/10^6 platelets (Lessard et al., 2012), or pg/ng gDNA (Budimirovic et al., 2020). Consequently, comparing FMRP levels, sensitivity, and reproducibility across these studies is challenging and may require conversion to relative FMRP, as demonstrated in this discussion.

Fourth, it is unclear how well FMRP levels in PBMCs reflect those in the brain and therefore how useful PBMC data is to increase the understanding of neurodevelopmental disorders like FXS. Few studies have examined the correlation between blood FMRP and postmortem brain tissue FMRP in the same subject. Pretto and colleagues showed differences in methylation status in blood versus brain in the premutation range (Pretto, Mendoza-Morales, et al., 2014). Tassone and colleagues showed tissue-specific methylation differences in premutation carriers, even when the size of alleles remained the same (Tassone, Longshore, et al., 1999). The obvious reason for this dearth of human brain FMRP research is the difficulty of accessing and working with human brain, especially in live patients. Measuring FMRP in peripheral tissue in an individual human may not have predictive power or provide a clear understanding of brain function, but it does provide overarching population-level trends that make these efforts worthwhile (Kim et al., 2019).

A final limit we would like to address is the larger than expected differences in FMRP levels between biological and technical replicates, even among samples with control alleles, despite previous reports that FMRP levels in PBMCs fluctuated little over an 11-month period (Iwahashi et al., 2009). An optimized TR-FRET assay for PBMCs should employ a high FMRP-producing fiducial obtained from a large blood draw from a control individual to produce many aliquots. Additionally, a true negative control within a cellular matrix should be established. For example, DM is a lymphoblastoid cell line with an *FMR1* deletion. Therefore, it can serve as a true representation of all background signal in the presence of the cellular matrix and in the absence of FMRP. It can also be used to serially dilute the fiducial control sample to obtain varying concentrations of FMRP within a cellular matrix. These added controls will allow for better assessment of inter-plate variation.

Implications and future directions

Here we have performed an analysis of FMRP levels in PBMCs of individuals with FXS compared to controls using TR-FRET while accounting for allele length and methylation status. The current study supports many previous findings. Continued research into allele size, methylation, mRNA synthesis, and protein production would benefit from simplifying their complex relationship. Unlike in humans, the *Fmr1* gene in mice does not undergo methylation and transcriptional silencing in the presence of full mutation expansions above 200 CGG repeats, at least up to 250 (Brouwer et al., 2007). Therefore, the relationship between expanded CGG repeat length and FMRP levels in mice could be addressed without the confounding effects of variable degrees of silencing methylation seen in human samples. Furthermore, allelic mosaicism could be simplified by selecting single-cell clones to generate a population of cells with a single allele size.

Additionally, improved determination of molecular features would provide more detail that could help clarify their relationships. For example, methylation of *FMR1* can be assessed

along the length of the molecule rather than at one locus. Currently, members of our lab are assessing methylation on a molecule-by-molecule basis using the characteristic increase in nucleotide incorporation time (IPD: inter pulse duration) during polymerization following detection of a methylated cytosine in the template strand using PacBio sequencing. This may help elucidate which alleles are capable of transcriptional activity.

The field would also benefit from a better understanding of which mRNA and FMRP isoforms are generated by various full mutation alleles and how they impact cognition. Alternative splicing (AS) is a mechanism used by over 95% of human genes to generate different protein variations from a single strand of genetic code by reassembling mRNA (Pan et al., 2008). Past research has shown that AS affects the expression and function of FMRP (Banerjee et al., 2010; Brackett et al., 2013; Sittler et al., 1996; Verkerk et al., 1993). It is also known that there is a repeat length association with expression and distribution of FMRP isoforms in premutation carriers (Pretto et al., 2015; Tseng et al., 2017). For example, *FMR1* mRNA isoforms *Iso4/4b* and *Iso10/10b* are elevated in individuals with expanded premutation repeats (55-200 CGGs) (Pretto et al., 2015; Zafarullah et al., 2020). Therefore, full mutation alleles (> 200 CGGs) that express mRNA and protein may also demonstrate a preference for different isoforms as well.

Potentially, different *FMR1* isoforms may use different translation methods. Translation of the full-length mRNA transcript primarily occurs via 5'end-dependent ribosomal scanning (Ludwig et al., 2011). However, other AS transcripts could potentially favor the use of internal ribosomal entry sties (IRESs) that bypass the higher-order RNA structures generated by expanded CGG repeats. Isoforms *Iso4/4b* and *Iso10/10b* generate a truncated FMRP and are associated with progression of the neurodegenerative disorder Fragile X associated tremor/ataxia syndrome (FXTAS) in individuals with premutation alleles (Zafarullah et al., 2020). Accordingly, the relationship between full mutation alleles, FMRP isoforms, and neurological changes should be investigated in the full mutation range as well.

Their relationship may impact how FXS is treated. A common proposed method to treat FXS is to epigenetically reactive the *FMR1* gene by removing silencing methylation (Kumari et al., 2019). The current work underscores several concerns with this approach. Aside from the unknowns surrounding full mutation *FMR1* isoforms and their neurological consequences, translation efficiency is reduced as CGG repeat size increases. Therefore, it is unclear whether activated full mutation alleles are cable of producing sufficient FMRP for normal cognitive function. Though some evidence exists to show that they might (Hagerman et al., 1994). Additionally, activated full mutation alleles can produce excess mRNA, which may lead to RNA toxicity and the neurological issues seen in individuals with premutation alleles who develop FXTAS and other *FMR1* premutation disorders. In fact, reports of unmethylated full mutation carriers displaying FXTAS features and symptoms already exist (Loesch et al., 2012; Pretto et al., 2013; Santa María et al., 2014).

In conclusion, the current study used the largest sample size of PBMCs, to the best of our knowledge, to assess FMRP levels in FXS individuals using a detection method in the more sensitive sandwich immunoassay class. The pattern of relative FMRP production by allele type in the current study closely mirrors results from studies using various other assays. However, in our hands, many FXS samples with the lowest levels of FMRP did not achieve statistical significance using the TR-FRET assay, despite evidence from linear regression and statistical modeling that they likely contain small amounts of FMRP. Therefore, the TR-FRET assay should be further optimized, or another assay chosen when measuring samples predicted to contain low levels of FMRP.

Tables

Allele class	N	Age (yr)					CGG repeat size					FMRPrel				
		Min	Max	Mean	Median	SD	Min	Max	Mean	Median	SD	Min	Max	Mean	Median	SD
Control	152	1	80	33.9	30.5	21.4	19	46	29.7	30	4.8	0.39	1.87	0.99	0.98	0.26
Non-control	170	0	57	11.1	8.5	10.2	13	1400	542.5	505	282.4	-0.05	0.49	0.09	0.07	0.09
Full mutation	97	1	44	10.7	9	9.2	200	1400	610	590	245.3	-0.05	0.42	0.05	0.04	0.06
Methylation mosaic	22	1	32	10.6	5.5	10.2	200	1270	482.9	380	253.4	-0.02	0.25	0.09	0.08	0.06
Size mosaic	4	9	53	29	27	18.4	170	1240	532.8	440	364.9	0.02	0.15	0.11	0.12	0.06
Size and methylation mosaic	45	0	57	10.5	8	10.6	13	1540	477	440	314.7	0.01	0.49	0.17	0.14	0.12
Point mutation	1	10	10	10	10	-	25	25	25	25	-	0.06	0.06	0.06	0.06	-
Deletion	1	6	6	6	6	-	-	-	-	-	-	0.08	0.08	0.08	0.08	-

Table 2.1. Descriptive statistics of patient age and allele size. Technical replicates were randomly eliminated to generate one measurement per patient blood draw. Smears were reduced to a single value represented by the first quartile between the low and high end of a smear. This single-value representation was added as a discrete allele to all other alleles in a sample to calculate the mean, median, and standard deviation (SD) of CGG-repeat size. Minimum and maximum CGG repeat size included the low and high end of a smear.

Model	Function	p-value $CGG_{unmethylated}$ slope	p-value mRNA slope
1	$0.0565*mRNA + 0.0019*Age - 0.0002*CGG_{unmethylated}$	0.0596	$9.3x10^{-6}$
2	$0.0474*mRNA + 0.0027*Age + 0.0000289*CGG_{methylated} - 0.0523*$Fraction Methylated	NA	0.0090
3	$0.0420*mRNA + 0.0018*Age - 0.0002*CGG_{unmethylated} + 0.0000014*CGG_{methylated} -$ 0.0686*Fraction Methylated	0.0513	0.0181

Table 2.2 FMR1 mRNA statistics for nested linear mixed-effects modeling. Three nested linear mixed-effects models were generated to assess factors contributing to FMRP levels. In all three models, FMR1 mRNA positively (positive slope) and significantly (p-value < 0.05) contributed to FMRP levels.

Figures

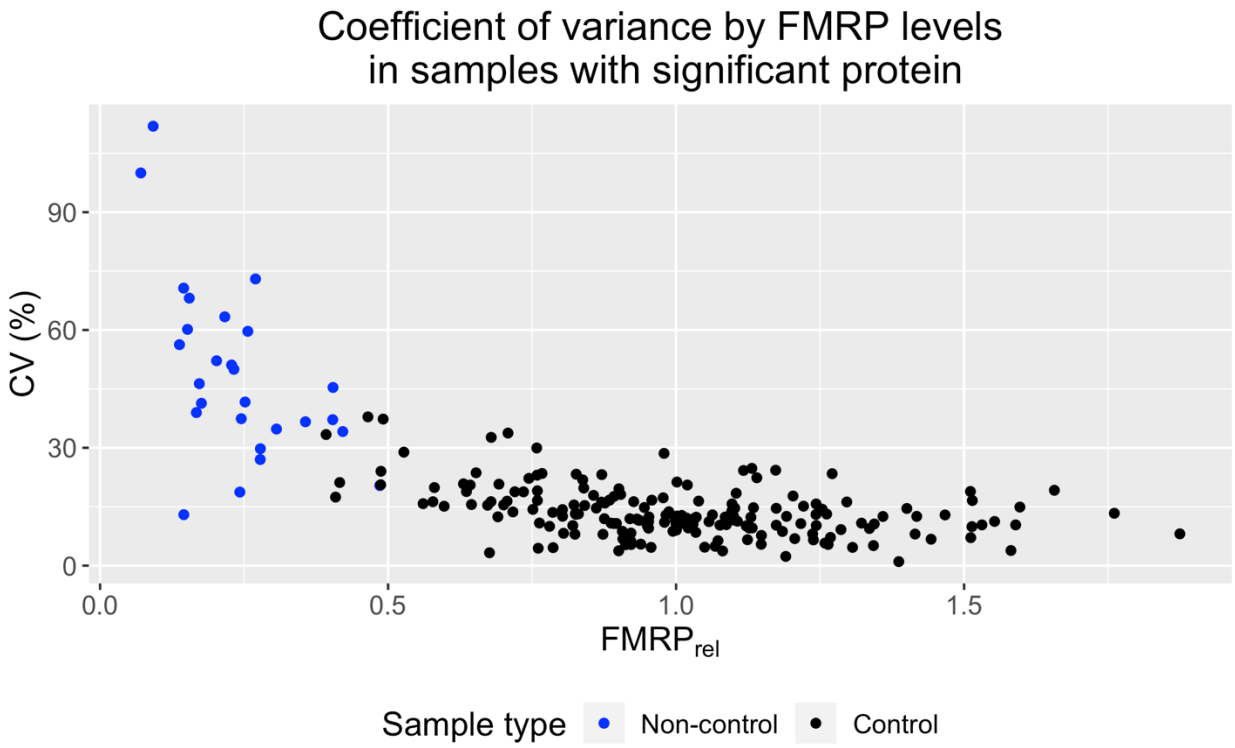


Fig. 2.1. Coefficient of variance (CV) by FMRP level. Relative FMRP was plotted against CV (%) for all samples with significant FMRP by one-sample t-test on corrected FRET ratios (See Methods).

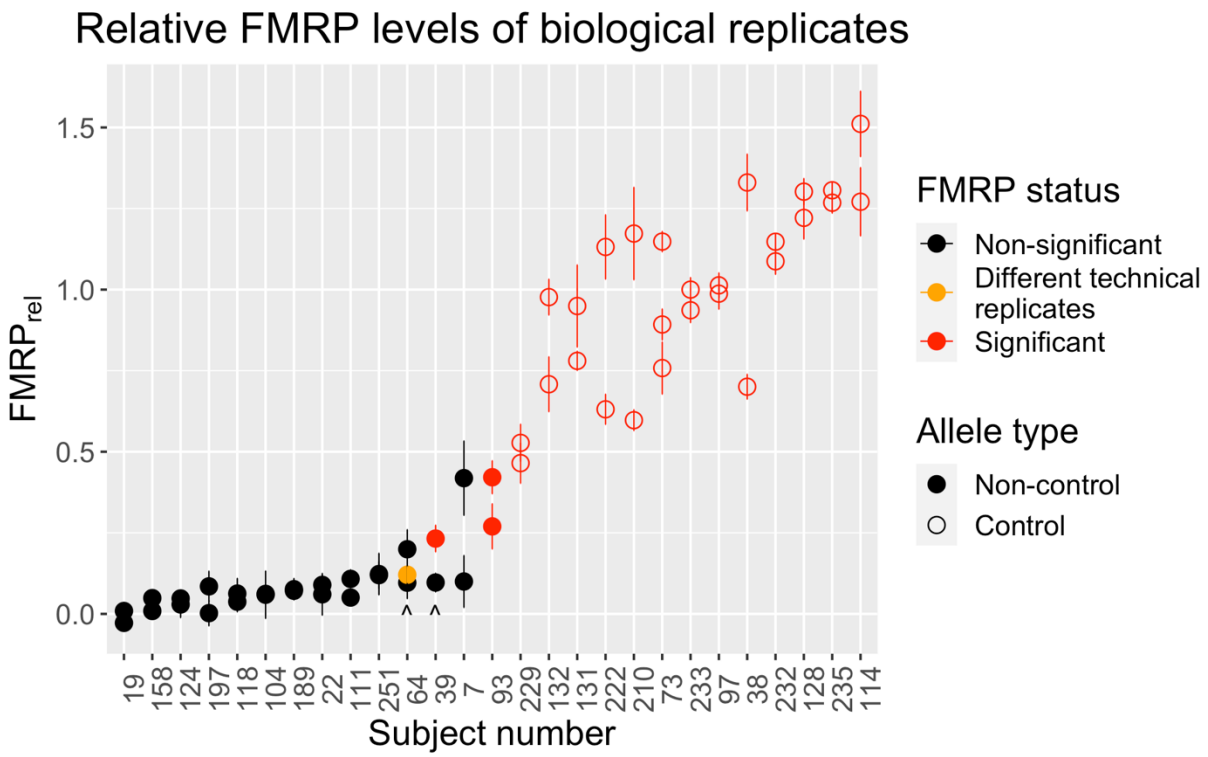


Fig. 2.2. Deviation in relative FMRP among biological replicates. FMRP (relative to the mean of samples with control alleles) for biological replicates was plotted by an individual's unique subject number for this study (n = 27 individuals with biological replicates). Subject number was arranged by mean relative FMRP (mean ± SEM). Biological replicates were defined as samples from the same individual, but from separate blood draws. When biological replicates include technical replicates, the mean of the technical replicates was used. Two individuals (7.4%) with non-control alleles had relative FMRP that differed in significance between their biological replicates (^). Both occurred at the transition in ability to significantly detect FMRP, indicating that the samples likely have FMRP, but at levels difficult to detect by significance testing. "Different technical replicates" is defined as one technical replicate with significant for FMRP and the other with non-significant FMRP.

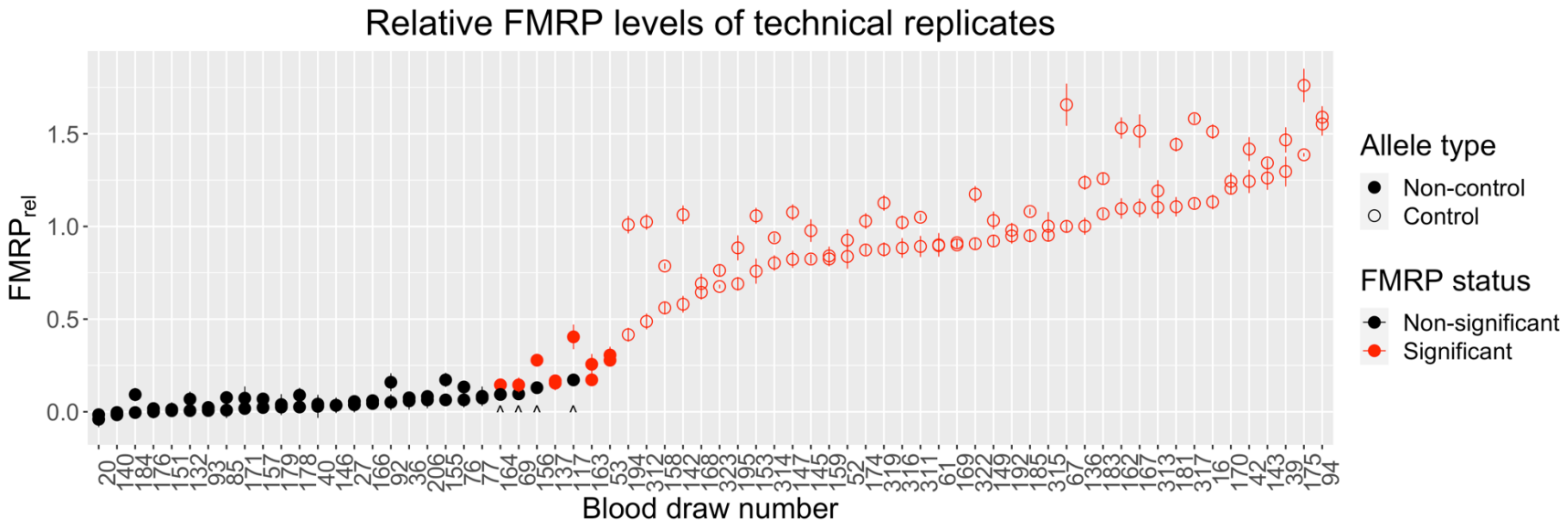


Fig. 2.3. Deviation in relative FMRP among technical replicates. Relative FMRP for technical replicates was plotted by an individual's unique blood draw number for this study (n = 68 individuals with technical replicates). Blood draw number was arranged by minimum relative FMRP (mean ± SEM). Technical replicates were defined as samples from the same individual and the same blood draw. Four individuals (5.9%) with non-control alleles had relative FMRP that differed in significance between their technical replicates (^). All occurred at the transition in ability to significantly detect FMRP, indicating that the samples likely have FMRP, but at levels difficult to detect by significance testing.

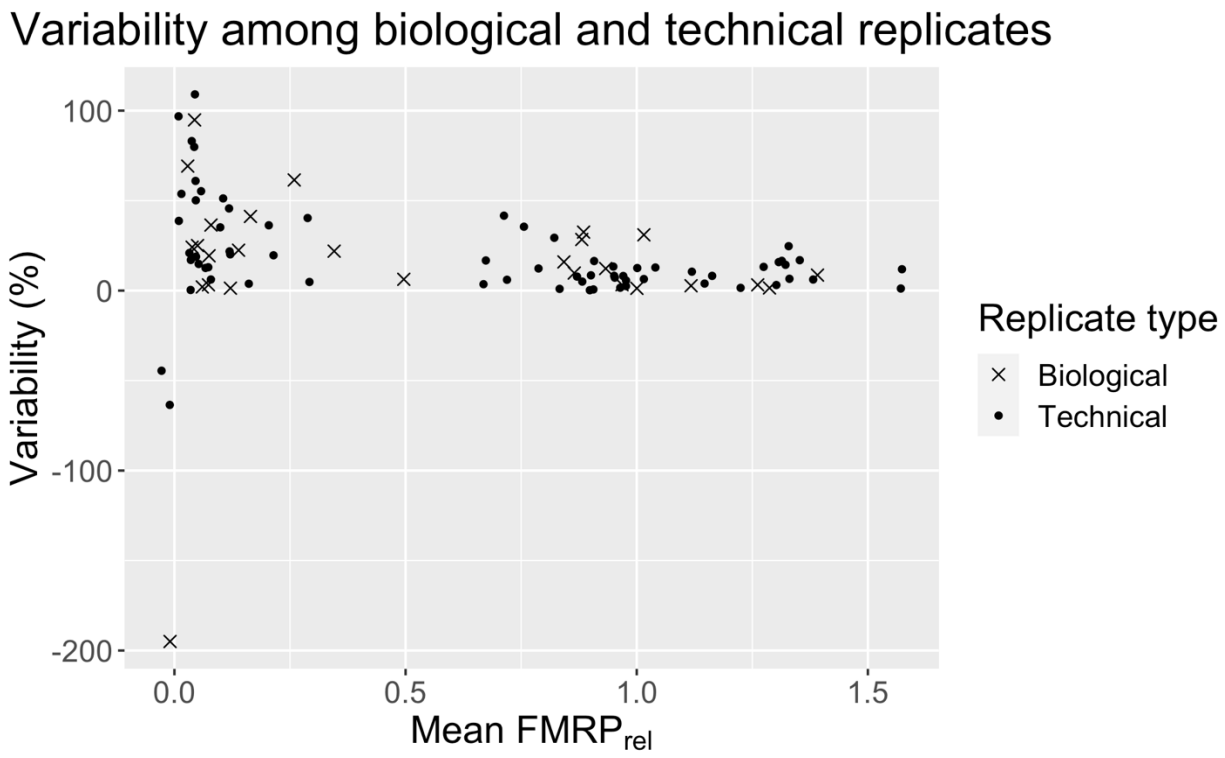


Fig. 2.4. Variance among biological and technical replicates. Standard error of the mean (SEM) was calculated for biological and technical replicates, respectively. Variability was calculated as $\text{Variability}\ (\%) = \frac{\text{SEM}}{\text{mean}}\ \text{x}\ 100$, and plotted against the mean relative FMRP of the replicates. SEM, rather than CV, was used to account for unequal sample size among replicate groups (n = 2-4).

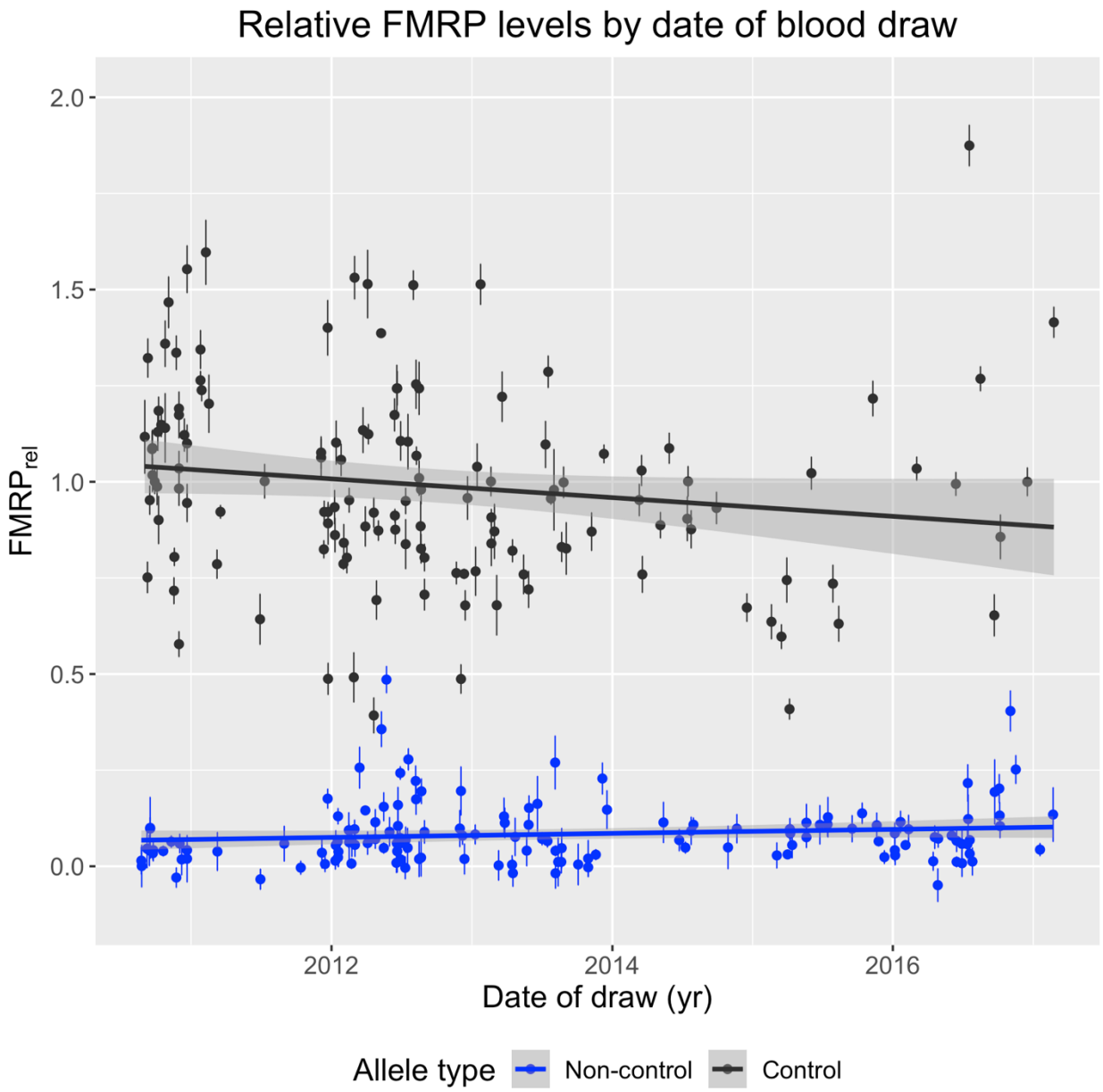


Fig. 2.5. Relative FMRP levels by date of draw. Biological (same individual, different blood draw) and technical (same individual, same blood draw) replicates were removed at random to generate one sample per individual with known date of blood draw. Relative FMRP (mean ± SEM) from PBMCs was plotted against the date of blood draw and a linear regression was fitted showing independence of FMRP on date of draw and thus length of time stored in liquid nitrogen. Control: n = 135, Estimate/slope = -0.000067 relative FMRP units/day, p-value = 0.072. Non-control: n = 143, Estimate/slope = 0.000014 relative FMRP units/day, p-value = 0.14.

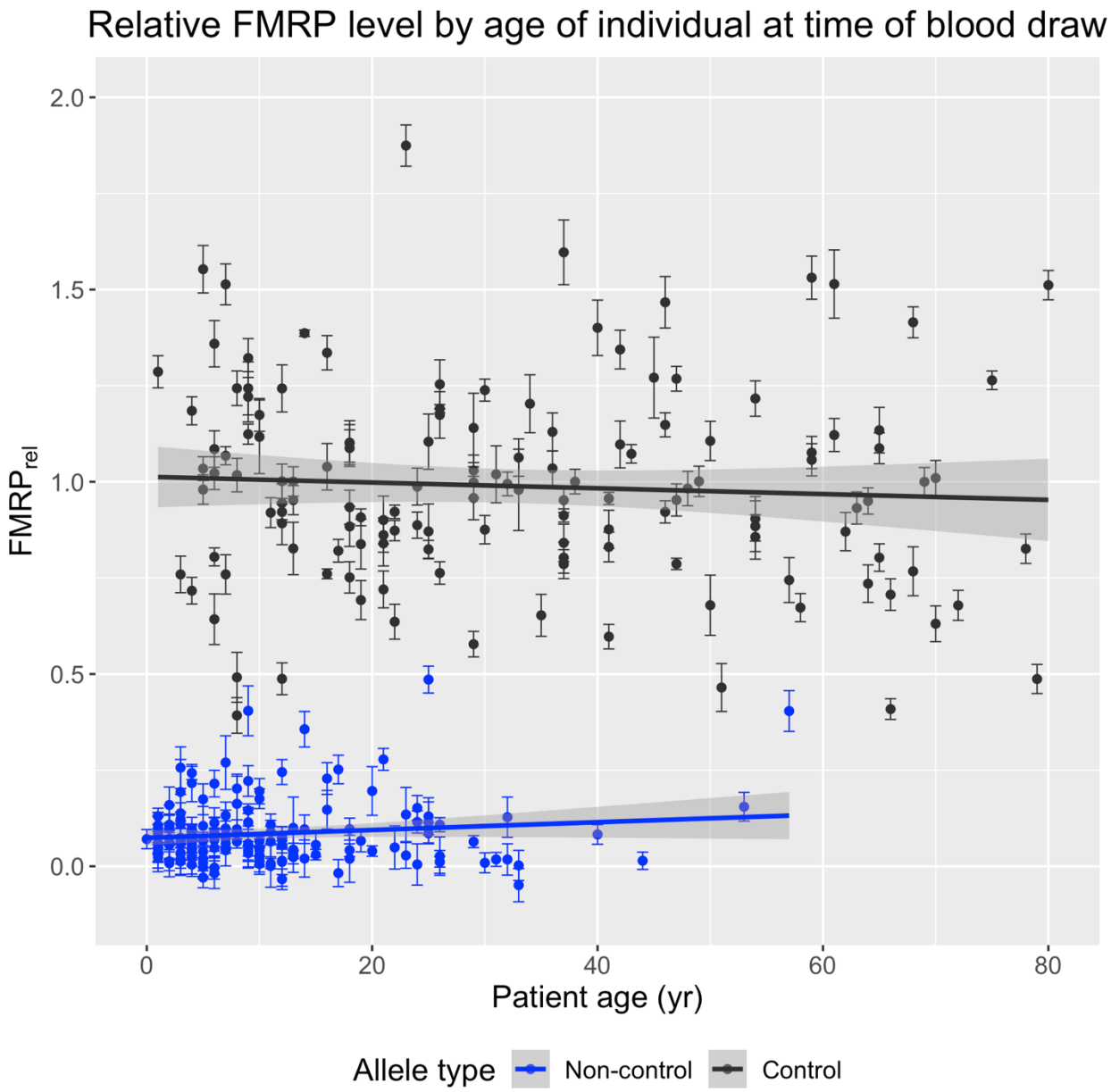


Fig. 2.6. Relative FMRP levels by age at time of blood draw. Biological (same individual, different blood draw) and technical (same individual, same blood draw) replicates were removed at random to generate one sample per individual. Relative FMRP (mean ± SEM) from PBMCs was plotted against age of individual in years at the time of blood draw and a linear regression was fitted showing independence of FMRP on age. Control: n = 138, Estimate/slope = -0.00075 FMRP units/day, p-value = 0.47. Non-control: n = 155, Estimate/slope = 0.0010 FMRP units/day, p-value = 0.13.

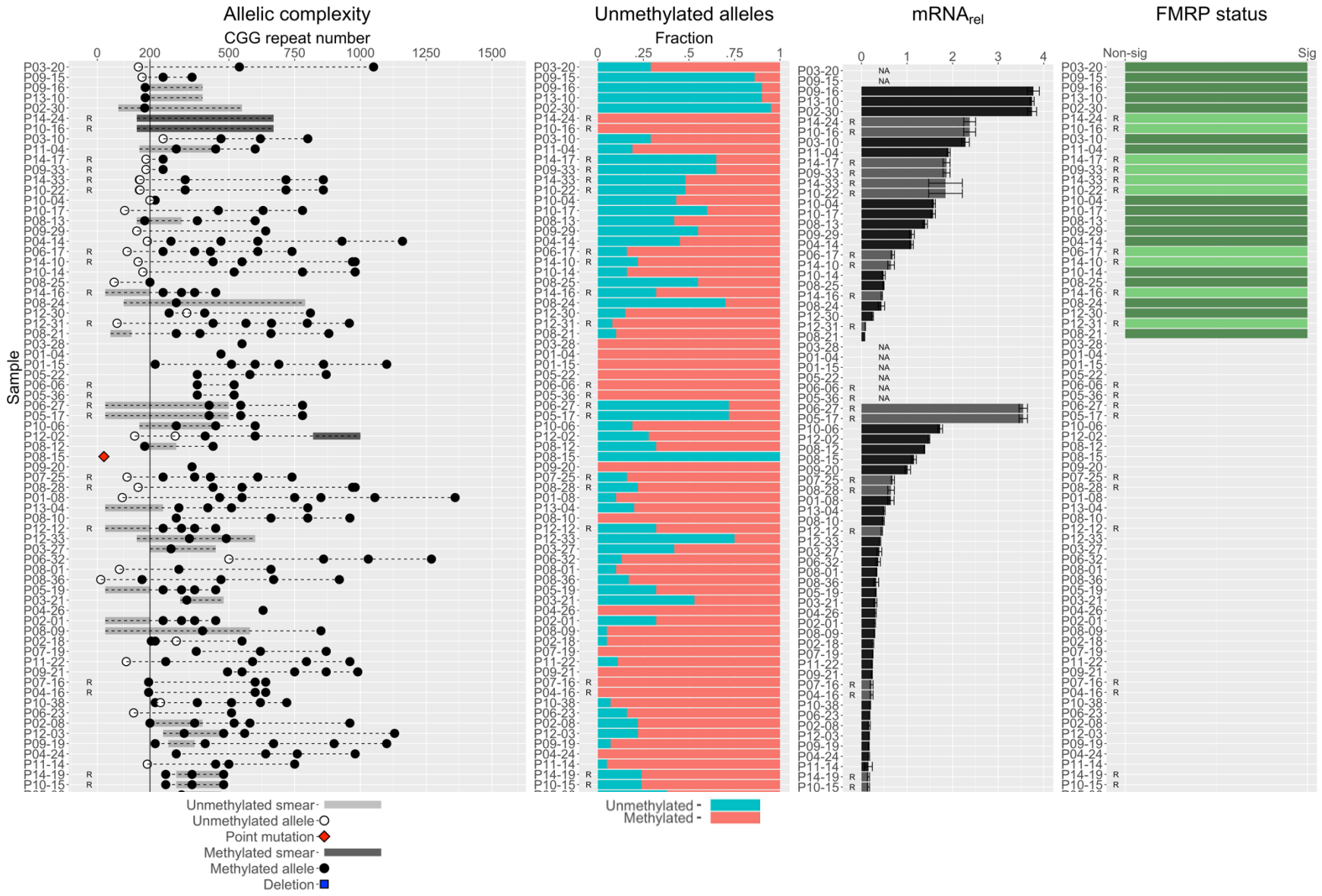


Fig. 2.7. Genetic and molecular characteristics of non-control samples. Continued on next page.

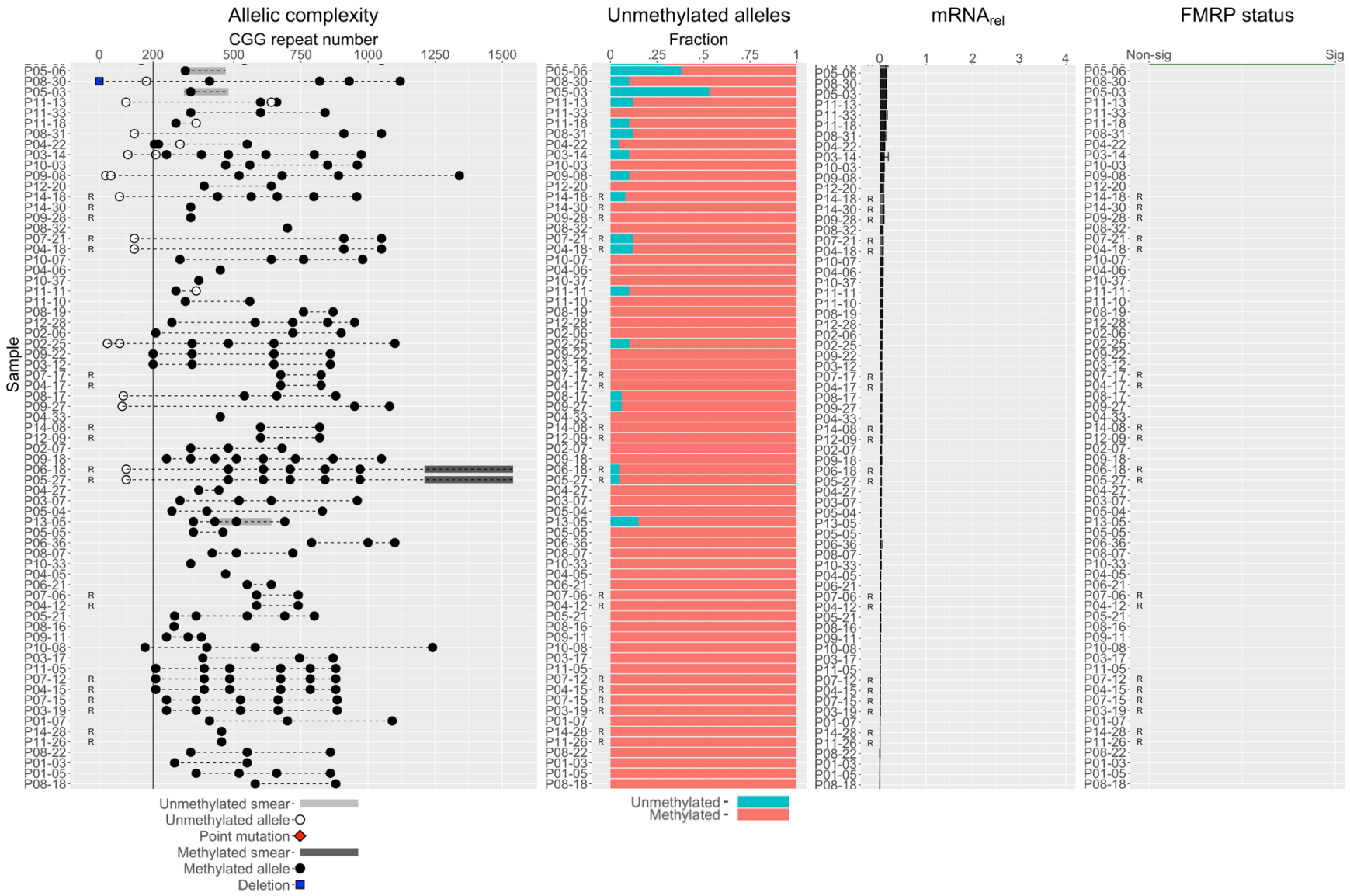
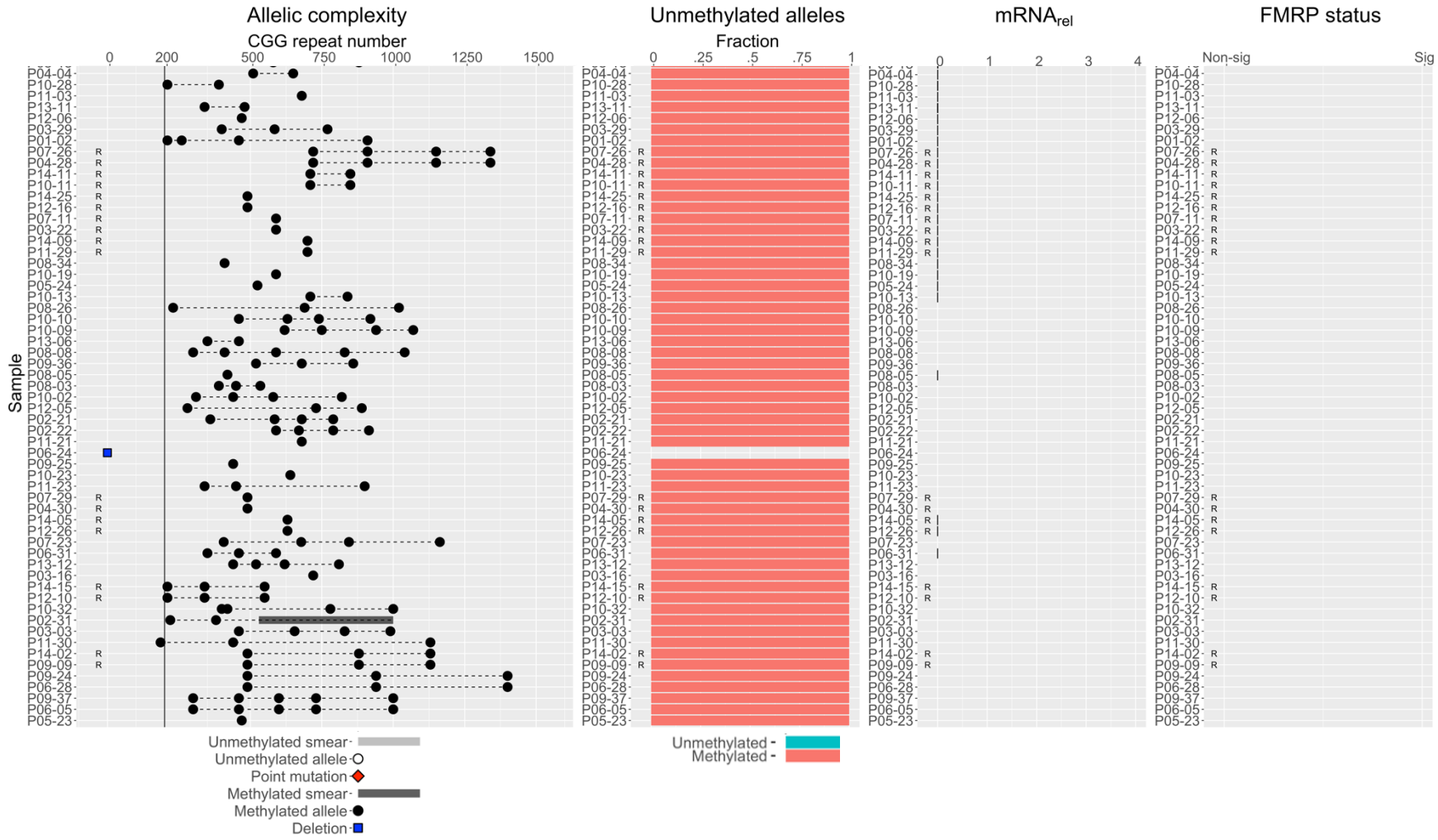


Fig. 2.7. Genetic and molecular characteristics of non-control samples. Continued on next page.



2.7. Genetic and molecular characteristics of non-control samples. Continued on next page.

Fig. 2.7. Genetic and molecular characteristics of non-control samples. All observations of each patient were included. Samples that had a technical replicate were denoted by the letter "R" and were shown in a lighter hue. Samples were arranged first by FMRP significance then by mRNA level. Sample P08-15 contains a control allele with a point mutation. It produces mRNA, but its protein is not detected by the Cisbio FRET assay. Sample P06-24 is a deletion sample. *FMR1* is not present. Therefore, no methylation analysis was performed. *Allelic complexity*. All alleles identified via capillary electrophoresis were plotted by CGG-repeat size and connected via a dashed line for each sample. Black circles: methylated alleles; white circles: unmethylated alleles; dark gray highlights: methylated smear range; light gray highlights: unmethylated smear range. *Unmethylated alleles*. Teal: fraction of unmethylated alleles; salmon: fraction of methylated alleles in a sample. $mRNA_{rel}$. Relative mRNA (mean ± SE) was determined via RT-qPCR and normalized to the mean of samples with control alleles. Samples whose mRNA was not evaluated are denoted by "NA." *FMRP status.* One-sample T-tests were performed on FRET ratios corrected to lysis buffer and deletion controls. A sample had significant FMRP when $p < 0.05$ for the highest concentration assayed. Only 27 samples with extended CGG repeats had significant levels of FMRP. All were mosaic for methylation and/or allele class.

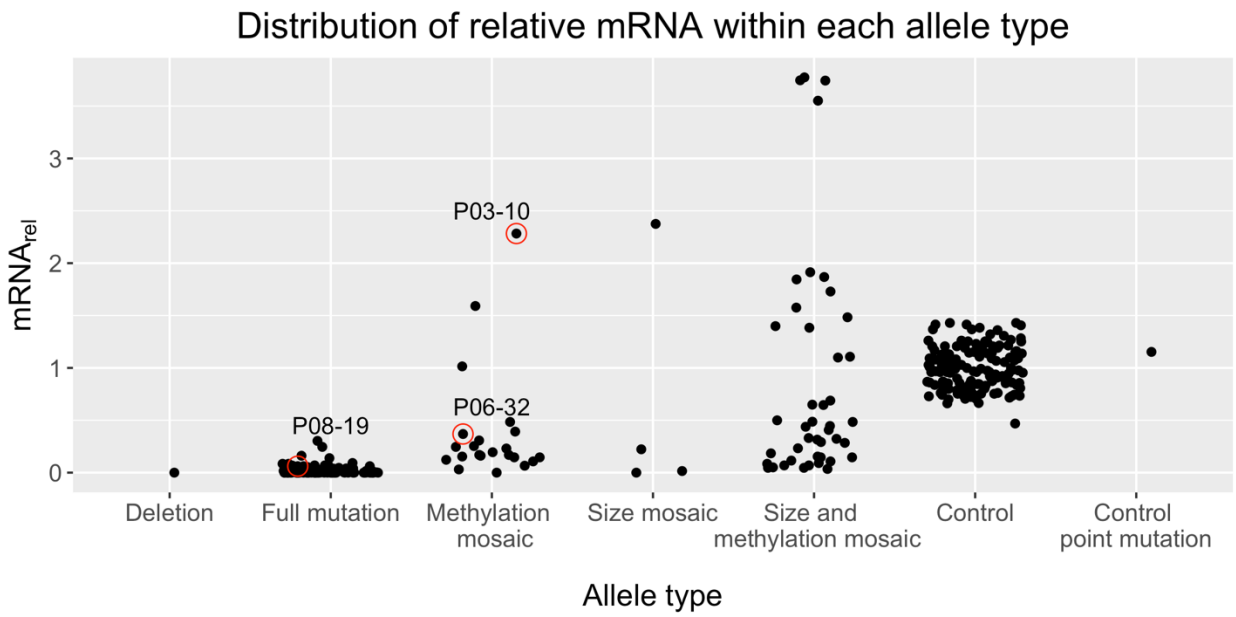


Fig. 2.8. Distribution of relative *FMR1* mRNA within each allele type. Technical replicates were randomly removed to keep one observation per patient blood draw. Deletion: *FMR1* not present; Full mutation: only full mutation methylated alleles; Methylation mosaics: some methylated and some unmethylated full mutation alleles in same individual; Size mosaic: some alleles under 200 CGG repeats and some full mutation alleles; Size and methylation mosaics: some alleles smaller than 200 CGG repeats and some alleles of varying methylation status above 200 CGG repeats; Control: control alleles; Control point mutation: control allele with a point mutation that prevents detection by Cisbio FRET assay.

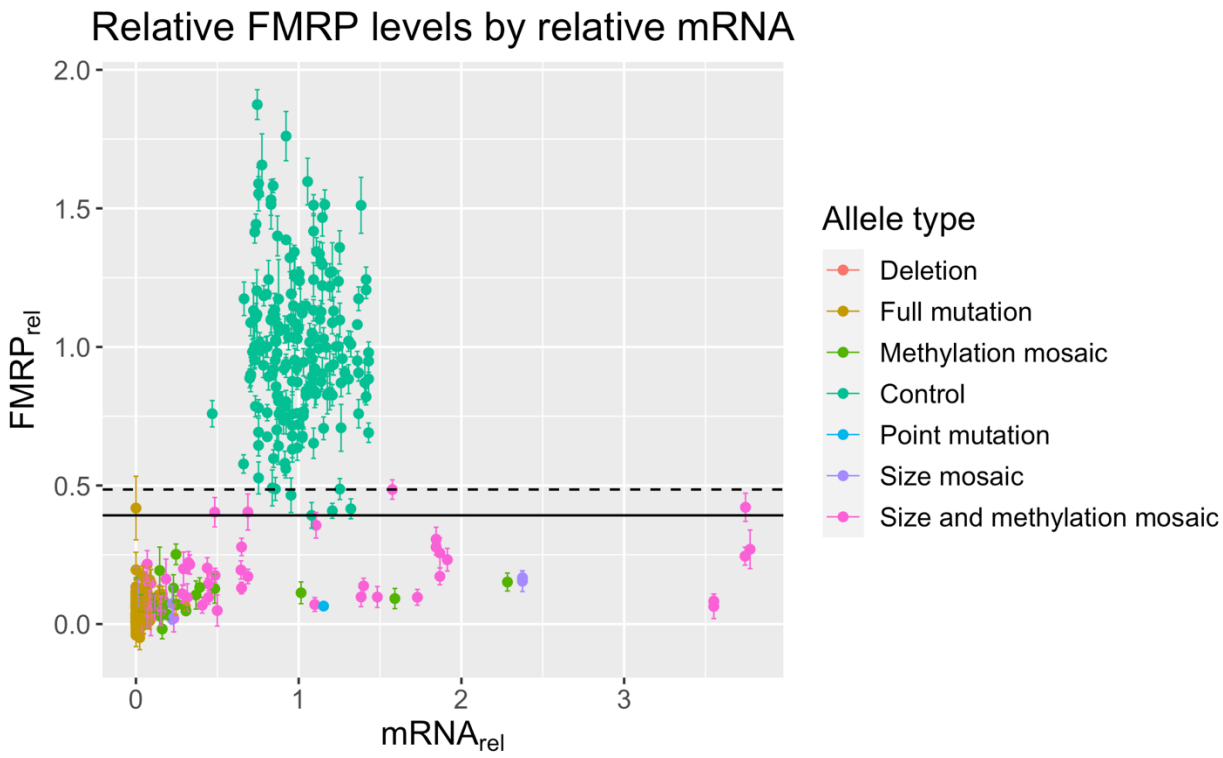


Fig. 2.9. FMRP levels by mRNA level. Relative *FMR1* mRNA was plotted against relative FMRP (mean ± SE) for all allele types, without controlling for subject age, CGG size, or CGG methylation status. Relative levels are normalized to that of the mean for control samples. All samples from each subject were included. Relative FMRP levels for samples with extended CGG repeats rarely approach those of control samples, regardless of mRNA level. Dashed line: maximum FMRP level of non-control samples. Solid line: minimum FMRP level of control samples.

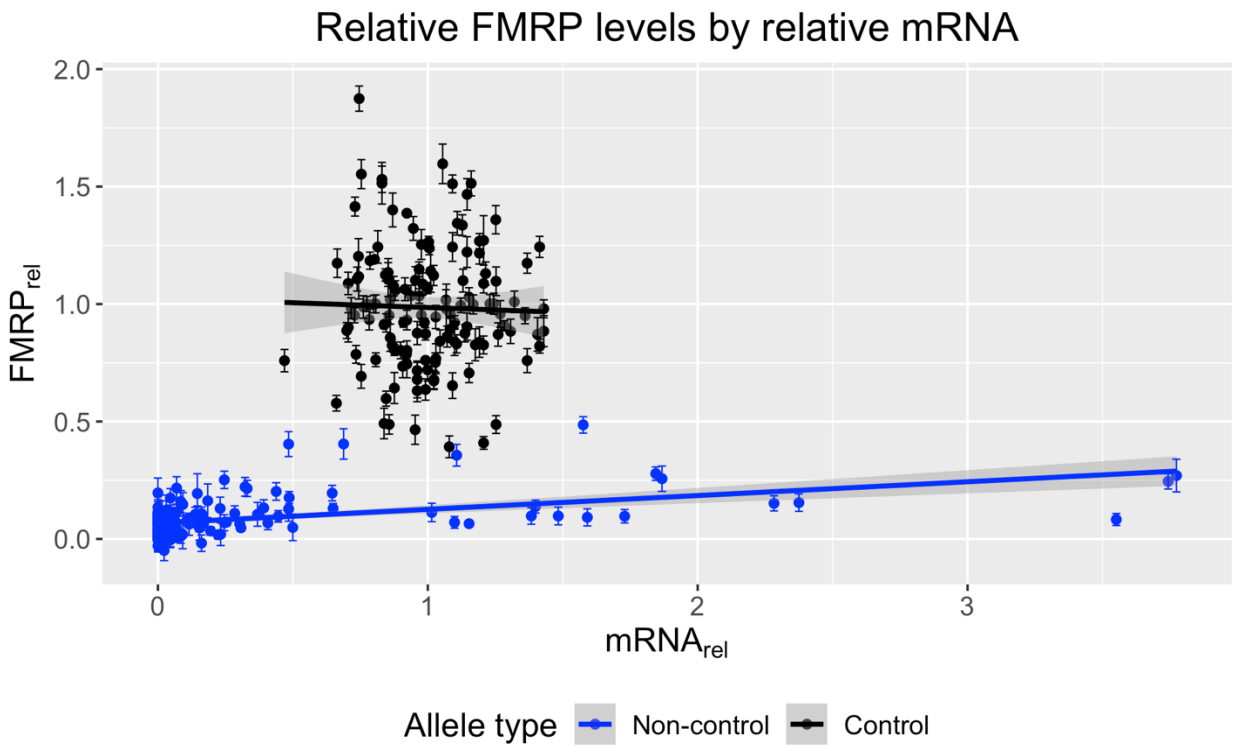


Fig. 2.10. Relative FMRP positively associates with relative *FMR1* mRNA for individuals with non-control alleles. Biological and technical replicates were removed at random to generate one measurement per individual. Relative mRNA (mean ± SE) from PBMCs was plotted against relative FMRP (mean ± SE) and a linear regression was fit. No association between *FMR1* mRNA and FMRP was found for individuals with control alleles. However, *FMR1* mRNA significantly associated with FMRP in patients with non-control alleles. Control: n = 134, Estimate/slope = -0.041, p-value = 0.727. Non-control: n = 148, Estimate/slope = 0.059, p-value = 1.21×10^{-9}.

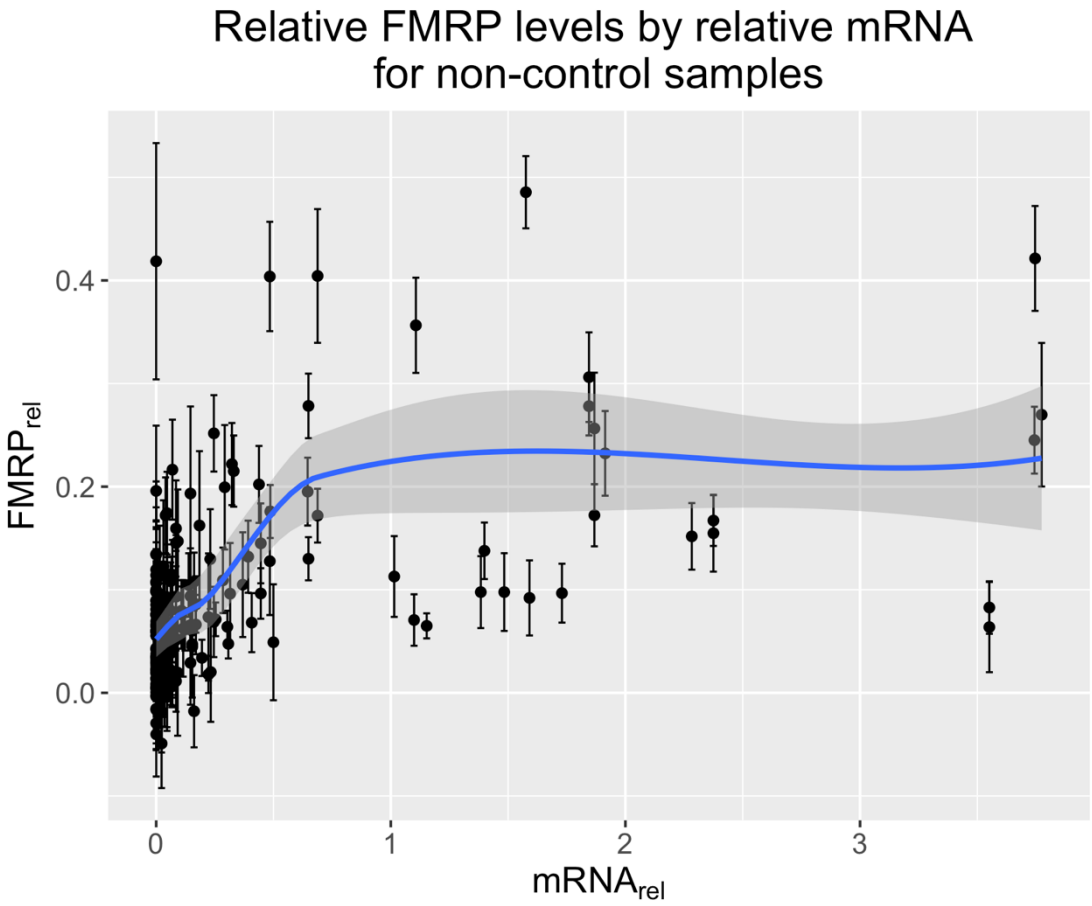


Fig. 2.11. Relative mRNA levels by relative FMRP levels for non-control samples. Relative *FMR1* mRNA was plotted against relative FMRP (mean ± SE) for all samples with non-control alleles with available mRNA data (n = 155), without controlling for subject age, CGG size, or CGG methylation status. Relative levels are normalized to that of the mean for control samples. All samples from each subject were included. A loess regression was used to visualize data trends: increasing mRNA from 0 to ~0.5 leads to increasing FMRP. However, FMRP plateaus at ~0.25 despite increasing mRNA levels past ~0.5.

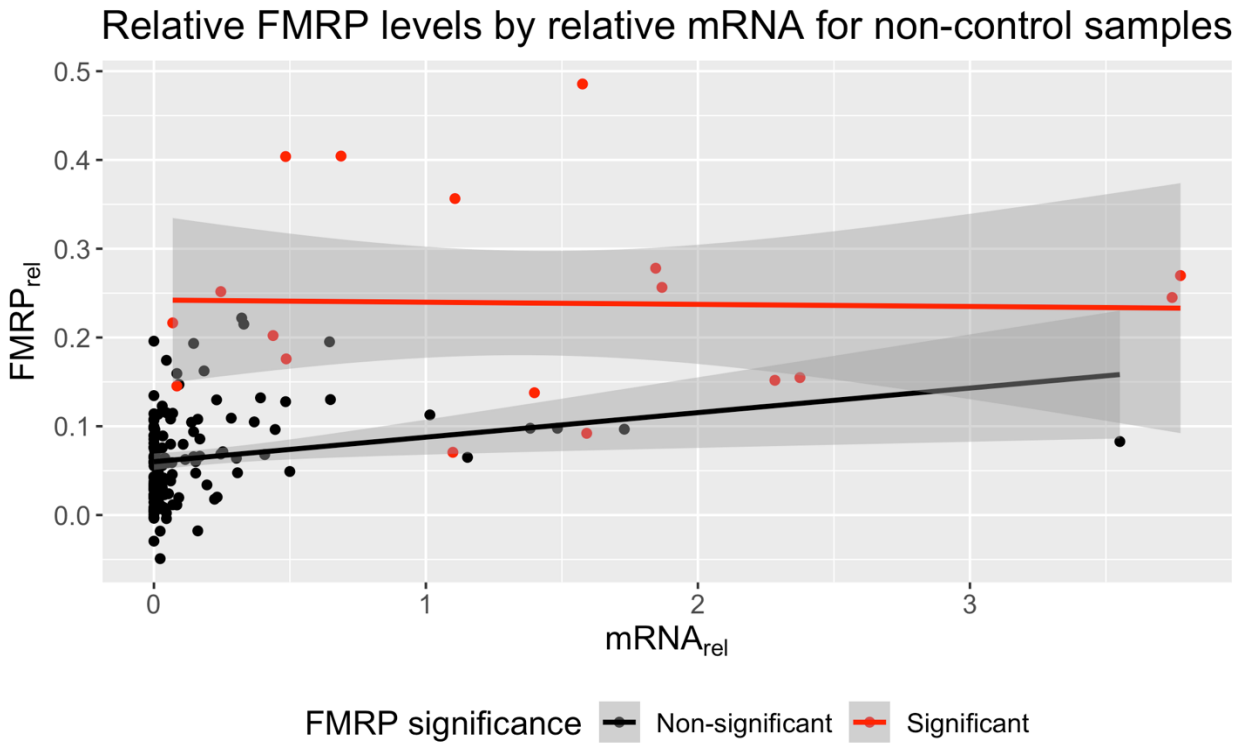


Fig. 2.12. The positive association between relative *FMR1* mRNA and relative FMRP in non-control samples is driven by those with non-significant FMRP. Biological and technical replicates were removed at random to generate one measurement per subject. Relative mRNA (mean ± SE) from PBMCs was plotted against relative FMRP (mean ± SE) for non-control samples. A linear regression was fit for FMRP(+) (significant) and FMRP(-) (non-significant) samples, respectively. No association between *FMR1* mRNA and FMRP was found for FMRP(+) individuals. However, *FMR1* mRNA significantly associated with FMRP in FMRP(-) individuals, suggesting the presence of FMRP below the level of detection via one-sample T-tests. FMRP(-): n = 131, Estimate/slope = 0.028, p-value = 0.011. FMRP(+): n = 18, Estimate/slope = -0.0024, p-value = 0.93.

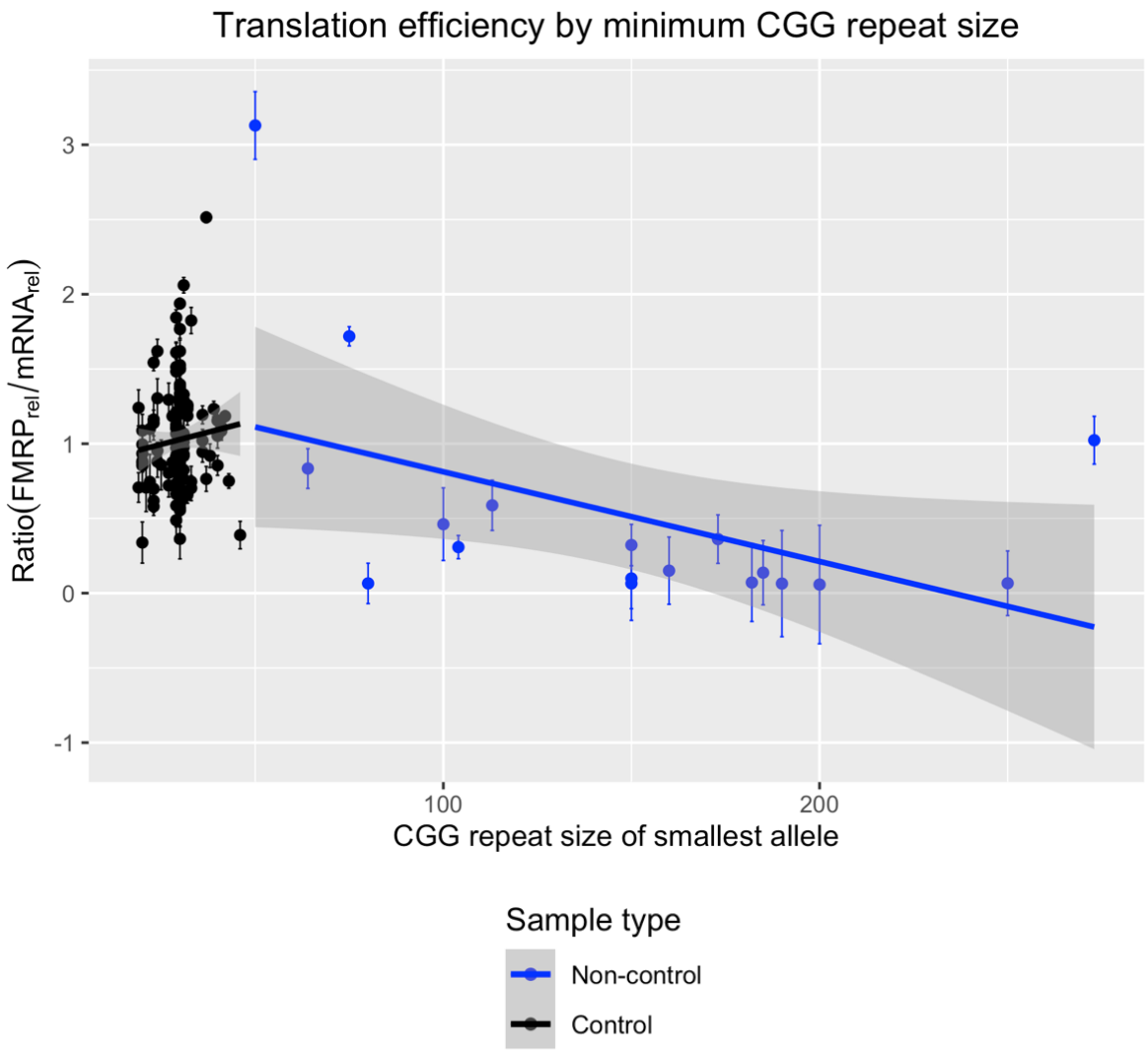


Fig. 2.13. Translation efficiency by smallest CGG repeat size for samples with significant FMRP. Biological and technical replicates were removed at random to generate one measurement per individual with significant levels of FMRP. The ratio of relative FMRP to relative mRNA was used as a measure of translation efficiency and plotted against the smallest CGG repeat of a sample, regardless of methylation status. A linear regression was fit for control and non-control samples, separately. Translation efficiency was independent of CGG repeat size for control alleles but showed a significantly negative correlation with repeat size for samples with non-control alleles. Control: n = 134, Estimate/slope = 0.0065 relative efficiency units per CGG, p-value = 0.31. Non-control: n = 18, Estimate/slope = -0.0060 relative efficiency units per CGG, p-value = 0.045.

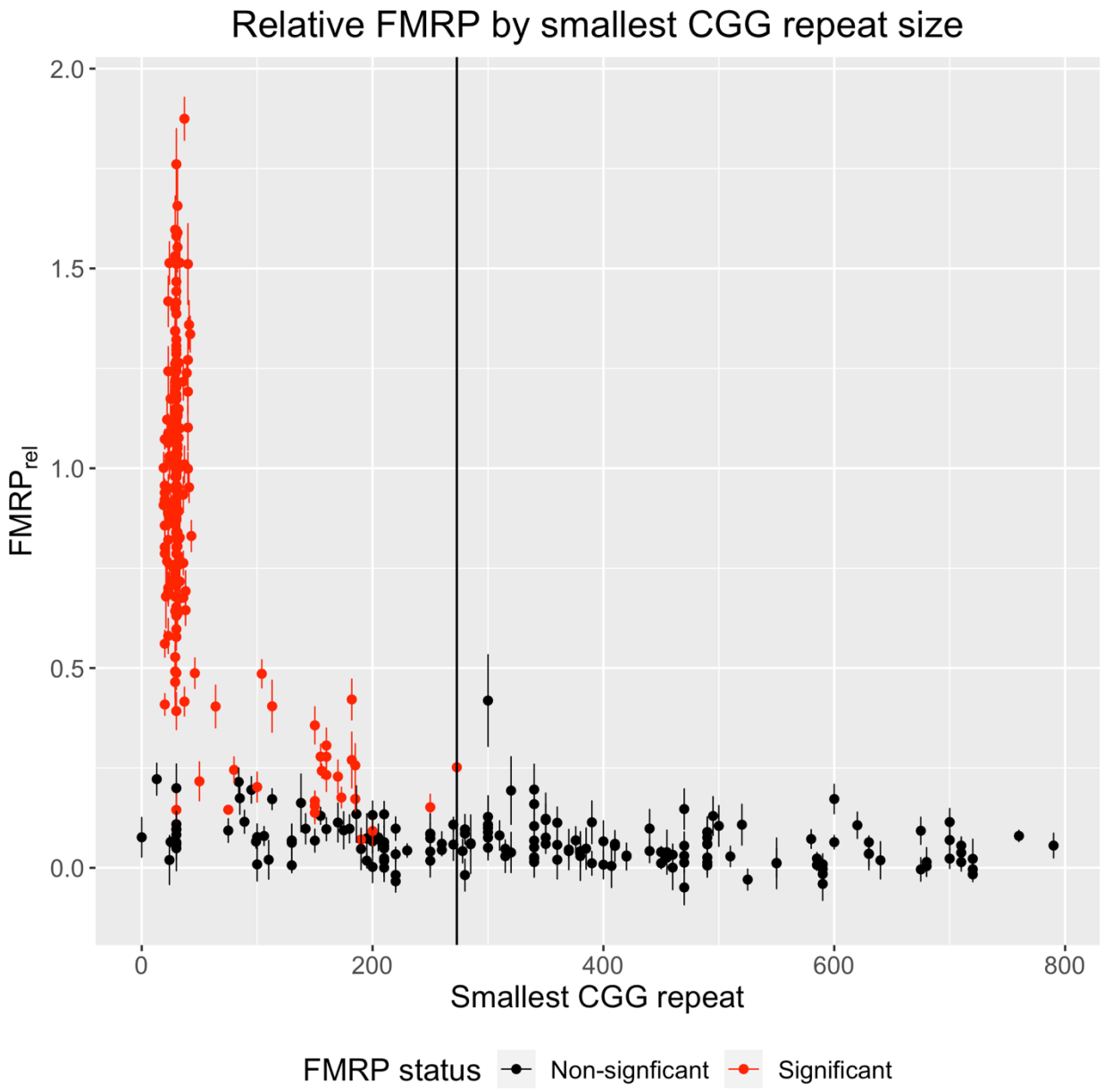


Fig. 2.14. Relative FMRP by smallest CGG repeat size. Relative FMRP (mean ± SE) was plotted against the CGG repeat size of the smallest allele of each sample, regardless of methylation status. Samples were color coded by FMRP status. No sample with a minimum allele size above 273 CGGs produced significant FMRP.

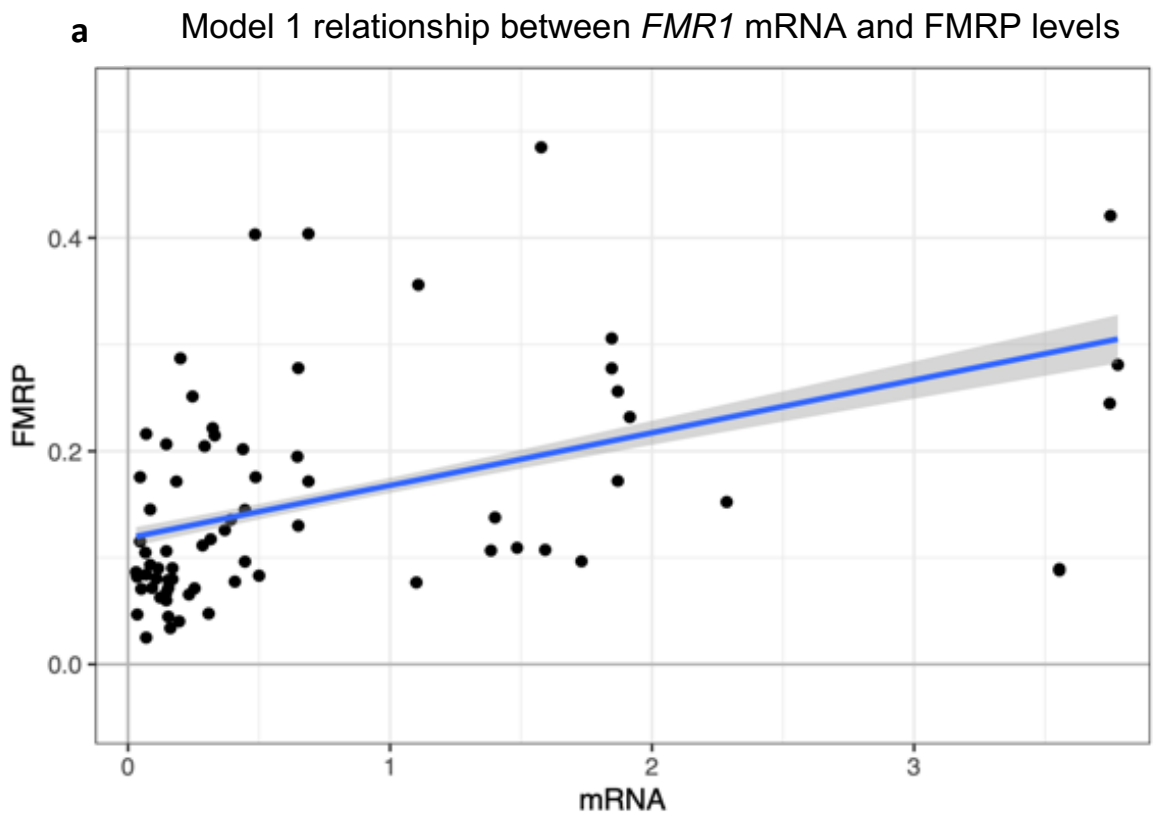


Fig. 2.15. Nested linear mixed-effects models showed relative FMRP was significantly dependent on relative *FMR1* mRNA for individuals with non-control alleles. Continued on next page.

b Model 2 relationship between *FMR1* mRNA and FMRP levels

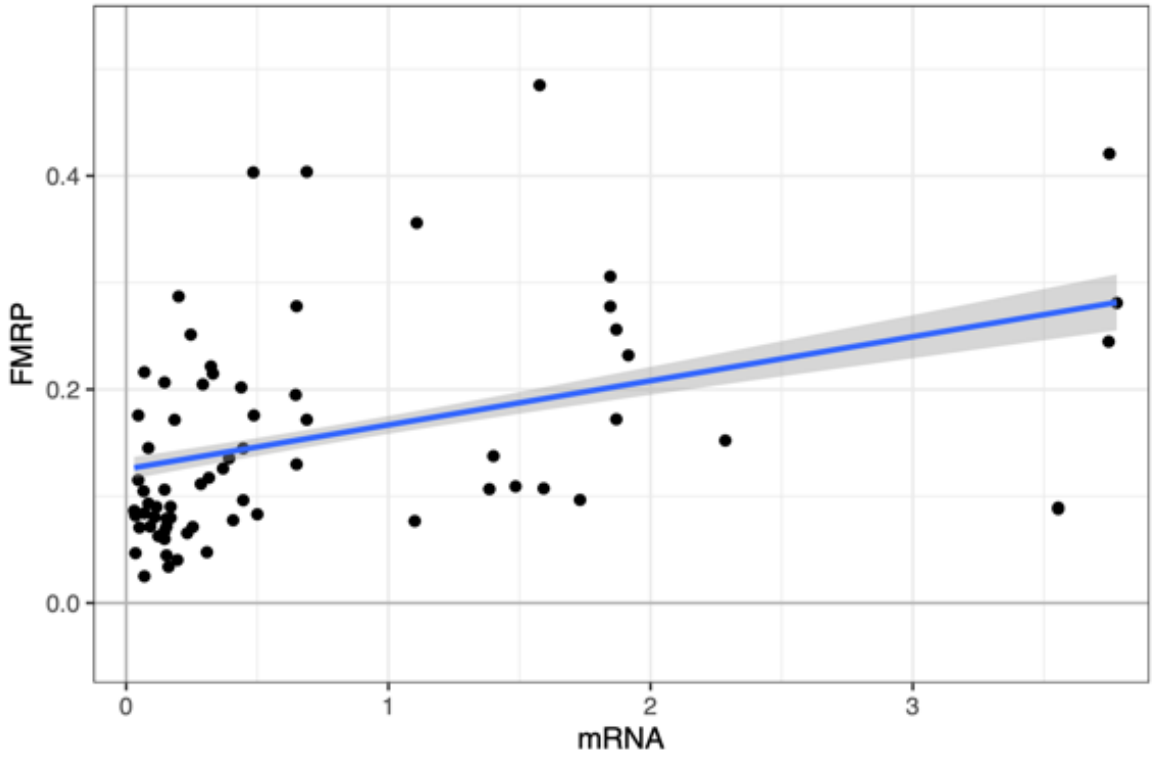


Fig. 2.15. Nested linear mixed-effects models showed relative FMRP was significantly dependent on relative *FMR1* mRNA for individuals with non-control alleles. Continued on next page.

c Model 3 relationship between *FMR1* mRNA and FMRP levels

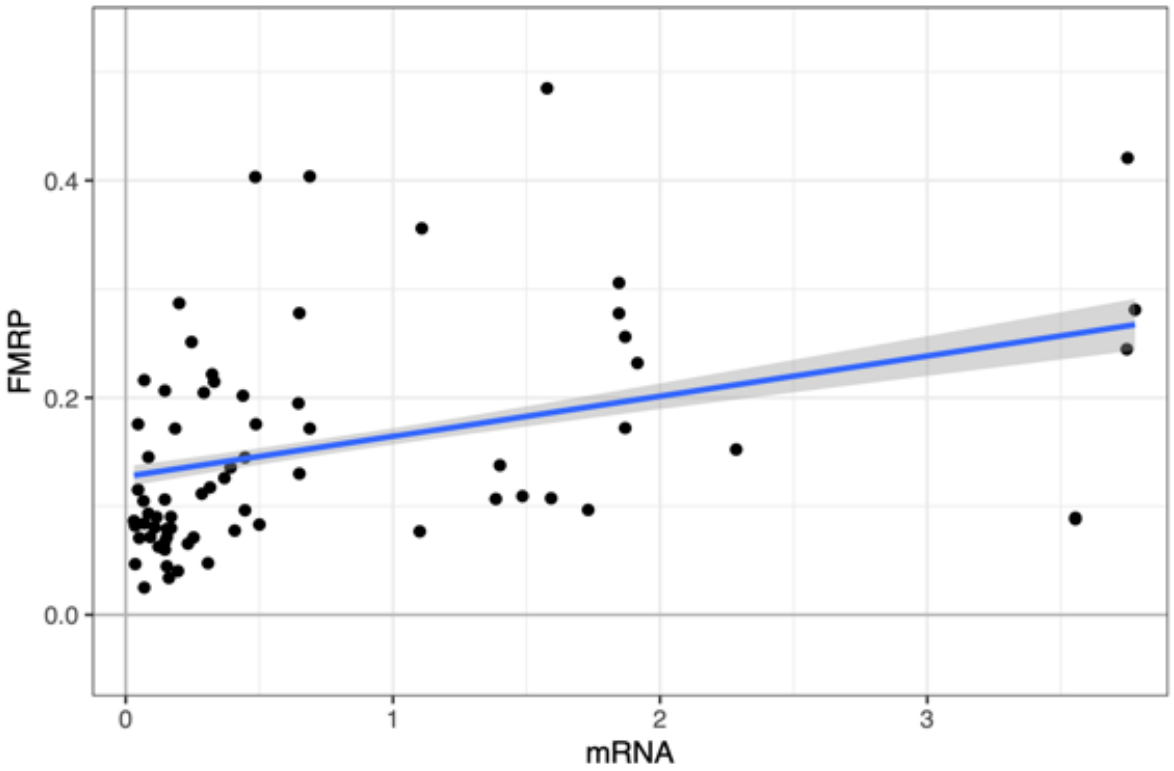


Fig. 2.15. Nested linear mixed-effects models showed relative FMRP was significantly dependent on relative *FMR1* mRNA for individuals with non-control alleles. Relative mRNA was plotted against relative FMRP for 72 observations from 54 individuals with extended CGG-repeats and both methylated and unmethylated alleles. Nested mixed-effects modeling was performed, and the predicted relationship between FMRP and *FMR1* mRNA was plotted over the data for three models (See Table 2.2), controlling for age and **a)** median unmethylated CGG repeat size **b)** median methylated CGG repeat size, and the fraction of methylated alleles **c)** both unmethylated and methylated median CGG repeat size, and the fraction of methylated alleles. Non-mRNA fixed effects were set to the mean value for age, CGG-repeat, and fraction methylated, where applicable. All models showed a similar significant ($p < 0.02$ for all) and positive (slope > 0) effect of *FMR1* mRNA on FMRP levels.

References

Adayev, T., LaFauci, G., Dobkin, C., Caggana, M., Wiley, V., Field, M., Wotton, T., Kascsak, R., Nolin, S. L., Glicksman, A., Hosmer, N., & Ted Brown, W. (2014). Fragile X protein in newborn dried blood spots. *BMC Medical Genetics*, *15*(1), 1–8. https://doi.org/10.1186/s12881-014-0119-0

Adayev, T., Lafauci, G., Xu, W., Dobkin, C., Kascsak, R., Brown, W. T., & Goodman, J. H. (2021). Development of a quantitative fmrp assay for mouse tissue applications. *Genes*, *12*(10), 1516. https://doi.org/10.3390/genes12101516

Banerjee, P., Schoenfeld, B. P., Bell, A. J., Choi, C. H., Bradley, M. P., Hinchey, P., Kollaros, M., Park, J. H., Mcbride, S. M. J., & Dockendorff, T. C. (2010). *Short-and Long-Term Memory Are Modulated by Multiple Isoforms of the Fragile X Mental Retardation Protein*. https://doi.org/10.1523/JNEUROSCI.6369-09.2010

Bidinosti, M., Shimshek, D. R., Mollenhauer, B., Marcellin, D., Schweizer, T., Lotz, G. P., Schlossmacher, M. G., & Weiss, A. (2012). Novel One-step Immunoassays to Quantify α-Synuclein: APPLICATIONS FOR BIOMARKER DEVELOPMENT AND HIGH-THROUGHPUT SCREENING. *Journal of Biological Chemistry*, *287*(40), 33691–33705. https://doi.org/10.1074/JBC.M112.379792

Boggs, A. E., Schmitt, L. M., McLane, R. D., Adayev, T., LaFauci, G., Horn, P. S., Dominick, K. C., Gross, C., & Erickson, C. A. (2022). Optimization, validation and initial clinical implications of a Luminex-based immunoassay for the quantification of Fragile X Protein from dried blood spots. *Scientific Reports*, *12*(1), 1–12. https://doi.org/10.1038/s41598-022-09633-8

Brackett, D. M., Qing, F., Amieux, P. S., Sellers, D. L., Horner, P. J., & Morris, D. R. (2013). Fmr1 Transcript Isoforms: Association with Polyribosomes; Regional and Developmental Expression in Mouse Brain. *PLOS ONE*, *8*(3), e58296. https://doi.org/10.1371/JOURNAL.PONE.0058296

Brouwer, J. R., Mientjes, E. J., Bakker, C. E., Nieuwenhuizen, I. M., Severijnen, L. A., Van der Linde, H. C., Nelson, D. L., Oostra, B. A., & Willemsen, R. (2007). Elevated Fmr1 mRNA levels and reduced protein expression in a mouse model with an unmethylated Fragile X full mutation. *Experimental Cell Research*, *313*(2), 244–253. https://doi.org/10.1016/j.yexcr.2006.10.002

Budimirovic, D. B., Schlageter, A., Filipovic-Sadic, S., Protic, D. D., Bram, E., Mahone, E. M., Nicholson, K., Culp, K., Javanmardi, K., Kemppainen, J., Hadd, A., Sharp, K., Adayev, T., Lafauci, G., Dobkin, C., Zhou, L., Brown, W. T., Berry-Kravis, E., Kaufmann, W. E., & Latham, G. J. (2020). A genotype-phenotype study of high-resolution FMR1 nucleic acid and protein analyses in fragile X patients with neurobehavioral assessments. *Brain Sciences*, *10*(10), 1–25. https://doi.org/10.3390/brainsci10100694

Cabal-Herrera, A. M., Tassanakijpanich, N., Salcedo-Arellano, M. J., & Hagerman, R. J. (2020). Fragile X-Associated Tremor/Ataxia Syndrome (FXTAS): Pathophysiology and Clinical Implications. *International Journal of Molecular Sciences*, *21*(12). https://doi.org/10.3390/ijms21124391

Carroll, R., Shaw, M., Arvio, M., Gardner, A., Kumar, R., Hodgson, B., Heron, S., McKenzie, F., Järvelä, I., & Gecz, J. (2020). EXOME REPORT: Two novel intragenic variants in the FMR1 gene in patients with suspect clinical diagnosis of Fragile X syndrome and no CGG repeat expansion. *European Journal of Medical Genetics*, *63*(10), 104010. https://doi.org/10.1016/j.ejmg.2020.104010

De Vries, B. B. A., Halley, D. J. J., Oostra, B. A., & Niermeijer, M. F. (1998). The fragile X syndrome. *Journal of Medical Genetics*, *35*(7), 579–589. https://doi.org/10.1192/bjp.160.1.24

De Vries, B. B. A., Jansen, C. C., Verheij, C., Willemsen, R., Van Hemel, J. O., Van Den Ouweland, A. M. W., Niermeijer, M. F., Oostra, B. A., & Halley, D. J. J. (1996). Variable FMR1 gene methylation of large expansions leads to variable phenotype in three males from one fragile X family. *Journal of Medical Genetics*, *33*(12), 1007–1010. https://doi.org/10.1136/jmg.33.12.1007

De Vries, B. B. A., Severijnen, L. A., Jacobs, A., Olmer, R., Halley, D. J. J., Oostra, B. A., & Willemsen, R. (2003). FMRP expression studies in blood and hair roots in a fragile X family with methylation mosaics. *Journal of Medical Genetics*, *40*(7), 535–539. https://doi.org/10.1136/jmg.40.7.535

Devys, D., Lutz, Y., Rouyer, N., Bellocq, J. P., & Mandel, J. L. (1993). The FMR-1 protein is cytoplasmic, most abundant in neurons and appears normal in carriers of a fragile X premutation. *Nature Genetics*, *4*(4), 335–340. https://doi.org/10.1038/ng0893-335

Fatemi, S. H., & Folsom, T. D. (2011). The role of fragile X mental retardation protein in major mental disorders. In *Neuropharmacology* (Vol. 60, Issues 7–8, pp. 1221–1226). https://doi.org/10.1016/j.neuropharm.2010.11.011

Feng, Y., Zhang, F., Lokey, L. K., Chastain, J. L., Lakkis, L., Eberhart, D., & Warren, S. T. (1995). Translational suppression by trinucleotide repeat expansion at FMR1. *Science*, *268*(5211), 731–734. https://doi.org/10.1126/science.7732383

Filipovic-Sadic, S., Sah, S., Chen, L., Krosting, J., Sekinger, E., Zhang, W., Hagerman, P. J., Stenzel, T. T., Hadd, A. G., Latham, G. J., & Tassone, F. (2010). A novel FMR1 PCR method for the routine detection of low abundance expanded alleles and full mutations in fragile X syndrome. *Clinical Chemistry*, *56*(3), 399–408. https://doi.org/10.1373/clinchem.2009.136101

Fink, D. A., Nelson, L. M., Pyeritz, R., Johnson, J., Sherman, S. L., Cohen, Y., & Elizur, S. E. (2018). Fragile X Associated Primary Ovarian Insufficiency (FXPOI): Case Report and Literature Review. *Frontiers in Genetics*, *9*. https://doi.org/10.3389/FGENE.2018.00529

Genç, B., Müller-Hartmann, H., Zeschnigk, M., Deissler, H., Schmitz, B., Majewski, F., Von Gontard, A., & Doerfle, W. (2000). Methylation mosaicism of 5'-(CGG)(n)-3' repeats in fragile X, premutation and normal individuals. *Nucleic Acids Research*, *28*(10), 2141–2152. https://doi.org/10.1093/nar/28.10.2141

Graef, J. D., Wu, H., Ng, C., Sun, C., Villegas, V., Qadir, D., Jesseman, K., Warren, S. T., Jaenisch, R., Cacace, A., & Wallace, O. (2020). Partial FMRP expression is sufficient to normalize neuronal hyperactivity in Fragile X neurons. *European Journal of Neuroscience*, *51*(10), 2143–2157. https://doi.org/10.1111/ejn.14660

Hagerman, R. J., & Hagerman, P. J. (2021). FXTAS: Pathophysiology and Management. *Current Opinion in Neurology*, *34*(4), 541. https://doi.org/10.1097/WCO.0000000000000954

Hagerman, R. J., Hull, C. E., Safanda, J. F., Carpenter, I., Staley, L. W., O'Connor, R. A., Seydel, C., Mazzocco, M. M. M., Snow, K., Thibodeau, S. N., Kuhl, D., Nelson, D. L., Caskey, C. T., & Taylor, A. K. (1994). High functioning fragile X males: Demonstration of an unmethylated fully expanded FMR-1 mutation associated with protein expression. *American Journal of Medical Genetics*, *51*(4), 298–308. https://doi.org/10.1002/ajmg.1320510404

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*. https://doi.org/10.3389/FPSYT.2018.00564

Iwahashi, C., Tassone, F., Hagerman, R. J., Yasui, D., Parrott, G., Nguyen, D., Mayeur, G., & Hagerman, P. J. (2009). A quantitative ELISA assay for the fragile X mental retardation 1 protein. *Journal of Molecular Diagnostics*, *11*(4), 281–289. https://doi.org/10.2353/jmoldx.2009.080118

Jacquemont, S., Pacini, L., Jønch, A. E., Cencelli, G., Rozenberg, I., He, Y., D'Andrea, L.,

Pedini, G., Eldeeb, M., Willemsen, R., Gasparini, F., Tassone, F., Hagerman, R., Gomez-Mancilla, B., & Bagni, C. (2018). Protein synthesis levels are increased in a subset of individuals with fragile X syndrome. *Human Molecular Genetics*, *27*(12), 2039–2051. https://doi.org/10.1093/hmg/ddy099

Jiraanont, P., Kumar, M., Tang, H. T., Espinal, G., Hagerman, P. J., Hagerman, R. J., Chutabhakdikul, N., & Tassone, F. (2017). Size and methylation mosaicism in males with Fragile X syndrome. *Expert Review of Molecular Diagnostics*, *17*(11), 1023–1032. https://doi.org/10.1080/14737159.2017.1377612

Kenneson, A., Zhang, F., Hagedorn, C. H., & Warren, S. T. (2001). Reduced FMRP and increased FMR1 transcription is proportionally associated with CGG repeat number in intermediate-length and premutation carriers. *Human Molecular Genetics*, *10*(14), 1449–1454. https://doi.org/10.1093/hmg/10.14.1449

Kim, K., Hessl, D., Randol, J. L., Espinal, G. M., Schneider, A., Protic, D., Aydin, E. Y., Hagerman, R. J., & Hagerman, P. J. (2019). Association between IQ and FMR1 protein (FMRP) across the spectrum of CGG repeat expansions. *PLoS ONE*, *14*(12). https://doi.org/10.1371/journal.pone.0226811

Kumari, D., Gazy, I., & Usdin, K. (2019). Pharmacological reactivation of the silenced FMR1 gene as a targeted therapeutic approach for fragile X syndrome. In *Brain Sciences* (Vol. 9, Issue 2, p. 39). Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/brainsci9020039

Lafauci, G., Adayev, T., Kascsak, R., Kascsak, R., Nolin, S., Mehta, P., Brown, W. T., & Dobkin, C. (2013). Fragile X screening by quantification of FMRP in dried blood spots by a luminex immunoassay. *Journal of Molecular Diagnostics*, *15*(4), 508–517. https://doi.org/10.1016/j.jmoldx.2013.02.006

Lafauci, G., Adayev, T., Kascsak, R., & Ted Brown, W. (2016). Detection and quantification of the fragile X mental retardation protein1(FMRP). *Genes*, *7*(12), 1–16. https://doi.org/10.3390/genes7120121

Lessard, M., Chouiali, A., Drouin, R., Sébire, G., & Corbin, F. (2012). Quantitative measurement of FMRP in blood platelets as a new screening test for fragile X syndrome. *Clinical Genetics*, *82*(5), 472–477. https://doi.org/10.1111/j.1399-0004.2011.01798.x

Loesch, D. Z., Huggins, R. M., Bui, Q. M., Taylor, A. K., Pratt, C., Epstein, J., & Hagerman, R. J. (2003). Effect of fragile X status categories and FMRP deficits on cognitive profiles estimated by robust pedigree analysis. *American Journal of Medical Genetics*, *122 A*(1), 13–23. https://doi.org/10.1002/ajmg.a.20214

Loesch, D. Z., Huggins, R. M., & Hagerman, R. J. (2004). Phenotypic Variation and FMRP Levels in Fragile X. In *Mental Retardation and Developmental Disabilities Research Reviews* (Vol. 10, Issue 1, pp. 31–41). https://doi.org/10.1002/mrdd.20006

Loesch, D. Z., Sherwell, S., Kinsella, G., Tassone, F., Taylor, A., Amor, D., Sung, S., & Evans, A. (2012). Fragile X-associated tremor/ataxia phenotype in a male carrier of unmethylated full mutation in the FMR1 gene. *Clinical Genetics*, *82*(1), 88–92. https://doi.org/10.1111/j.1399-0004.2011.01675.x

Ludwig, A. L., Espinal, G. M., Pretto, D. I., Jamal, A. L., Arque, G., Tassone, F., Berman, R. F., & Hagerman, P. J. (2014). CNS expression of murine fragile X protein (FMRP) as a function of CGG-repeat size. *Human Molecular Genetics*, *23*(12), 3228–3238. https://doi.org/10.1093/hmg/ddu032

Ludwig, A. L., Hershey, J. W., & Hagerman, P. J. (2011). Initiation of translation of the FMR1 mRNA occurs predominantly through 5′end-dependent ribosomal scanning. *J. Mol. Biol*, *407*(1), 21–34. https://doi.org/10.1016/j.jmb.2011.01.006

McConkie-Rosell, A., Lachiewicz, A. M., Spiridigliozzi, G. A., Tarleton, J., Schoenwald, S., Phelan, M. C., Goonewardena, P., Ding, X., & Brown, W. T. (1993). Evidence that methylation of the FMR-1 locus is responsible for variable phenotypic expression of the

fragile X syndrome. *American Journal of Human Genetics*, *53*(4), 800–809.
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
Menon, V., Leroux, J., White, C. D., & Reiss, A. L. (2004). Frontostriatal deficits in fragile X syndrome: Relation to FMR1 gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(10), 3615–3620. https://doi.org/10.1073/PNAS.0304544101/SUPPL_FILE/04544FIG6.PDF
Nolin, S. L., Glicksman, A., Houck, G. E., Brown, W. T., & Dobkin, C. S. (1994). Mosaicism in fragile X affected males. *American Journal of Medical Genetics*, *51*(4), 509–512. https://doi.org/10.1002/ajmg.1320510444
Nolin, S. L., Glicksman, A., Tortora, N., Allen, E., Macpherson, J., Mila, M., Vianna-Morgante, A. M., Sherman, S. L., Dobkin, C., Latham, G. J., & Hadd, A. G. (2019). Expansions and contractions of the FMR1 CGG repeat in 5,508 transmissions of normal, intermediate, and premutation alleles. *American Journal of Medical Genetics, Part A*, *179*(7), 1148–1156. https://doi.org/10.1002/ajmg.a.61165
Pan, Q., Shai, O., Lee, L. J., Frey, B. J., & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*, *40*(12), 1413–1415. https://doi.org/10.1038/ng.259
Peprah, E., He, W., Allen, E., Oliver, T., Boyne, A., & Sherman, S. L. (2009). Examination of FMR1 transcript and protein levels among 74 premutation carriers. *Journal of Human Genetics 2010 55:1*, *55*(1), 66–68. https://doi.org/10.1038/jhg.2009.121
Pieretti, M., Zhang, F., Fu, Y. H., Warren, S. T., Oostra, B. A., Caskey, C. T., & Nelson, D. L. (1991). Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell*, *66*(4), 817–822. https://doi.org/10.1016/0092-8674(91)90125-I
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
Pretto, D. I., Hunsaker, M. R., Cunningham, C. L., Greco, C. M., Hagerman, R. J., Noctor, S. C., Hall, D. A., Hagerman, P. J., & Tassone, F. (2013). Intranuclear inclusions in a fragile X mosaic male. *Translational Neurodegeneration*, *2*(1), 1–9. https://doi.org/10.1186/2047-9158-2-10
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
Pretto, D. I., 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). Clinical and molecular implications of mosaicism in FMR1 full mutations. *Frontiers in Genetics*, *5*, 318. https://doi.org/10.3389/fgene.2014.00318
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*, *8*(12), 1482–1488.
Richter, J. D., & Zhao, X. (2021). The molecular biology of FMRP: new insights into fragile X syndrome. In *Nature Reviews Neuroscience* (Vol. 22, Issue 4, pp. 209–222). Nature Publishing Group. https://doi.org/10.1038/s41583-021-00432-0
Robertson, E. E., Hall, D. A., McAsey, A. R., & O'Keefe, J. A. (2016). Fragile X-associated Tremor/Ataxia Syndrome: Phenotypic comparisons with other Movement Disorders. *The Clinical Neuropsychologist*, *30*(6), 849. https://doi.org/10.1080/13854046.2016.1202239

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*, *11*(10), 1780. https://doi.org/10.3390/diagnostics11101780

Santa María, L., Pugin, A., Alliende, M. A., Aliaga, S., Curotto, B., Aravena, T., Tang, H. T., Mendoza-Morales, G., Hagerman, R., & Tassone, F. (2014). FXTAS in an unmethylated mosaic male with fragile X syndrome from Chile. *Clinical Genetics*, *86*(4), 378–382. https://doi.org/10.1111/cge.12278

Schmitt, L. M., Shaffer, R. C., Hessl, D., & Erickson, C. (2019). Executive Function in Fragile X Syndrome: A Systematic Review. *Brain Sciences 2019, Vol. 9, Page 15*, *9*(1), 15. https://doi.org/10.3390/BRAINSCI9010015

Schneider, A., Winarni, T. I., Cabal-Herrera, A. M., Bacalman, S., Gane, L., Hagerman, P., Tassone, F., & Hagerman, R. (2020). Elevated FMR1-mRNA and lowered FMRP – A double-hit mechanism for psychiatric features in men with FMR1 premutations. *Translational Psychiatry*, *10*(1), 205. https://doi.org/10.1038/s41398-020-00863-w

Schutzius, G., Bleckmann, D., Kapps-Fouthier, S., di Giorgio, F., Gerhartz, B., & Weiss, A. (2013). A quantitative homogeneous assay for fragile X mental retardation 1 protein. *Journal of Neurodevelopmental Disorders*, *5*(1), 1–9. https://doi.org/10.1186/1866-1955-5-8

Singh, K., Gaur, P., & Prasad, S. (2007). Fragile x mental retardation (Fmr-1) gene expression is down regulated in brain of mice during aging. *Molecular Biology Reports*, *34*(3), 173–181. https://doi.org/10.1007/s11033-006-9032-8

Singh, K., & Prasad, S. (2008). Differential expression of Fmr-1 mRNA and FMRP in female mice brain during aging. *Molecular Biology Reports*, *35*(4), 677–684. https://doi.org/10.1007/s11033-007-9140-0

Sittler, A., Devys, D., Weber, C., & Mandel, J. L. (1996). Alternative Splicing of Exon 14 Determines Nuclear or Cytoplasmic Localisation of FMR1 Protein Isoforms. *Human Molecular Genetics*, *5*(1), 95–102. https://doi.org/10.1093/HMG/5.1.95

Sitzmann, A. F., Hagelstrom, R. T., Tassone, F., Hagerman, R. J., & Butler, M. G. (2018). Rare FMR1 gene mutations causing fragile X syndrome: A review. *American Journal of Medical Genetics, Part A*, *176*(1), 11–18. https://doi.org/10.1002/ajmg.a.38504

Tabolacci, E., Nobile, V., Pucci, C., & Chiurazzi, P. (2022). Mechanisms of the FMR1 Repeat Instability: How Does the CGG Sequence Expand? *International Journal of Molecular Sciences*, *23*(10), 1–17. https://doi.org/10.3390/ijms23105425

Tassone, F., Hagerman, R. J., Chamberlain, W. D., & Hagerman, P. J. (2000). Transcription of the FMR1 gene in individuals with fragile X syndrome. *American Journal of Medical Genetics - Seminars in Medical Genetics*, *97*(3), 195–203. https://doi.org/10.1002/1096-8628(200023)97:3<195::AID-AJMG1037>3.0.CO;2-R

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. https://doi.org/10.1002/(SICI)1096-8628(19990528)84:3<250::AID-AJMG17>3.0.CO;2-4

Tassone, F., Hagerman, R. J., Loesch, D. Z., Lachiewicz, A., Taylor, A. K., & Hagerman, P. J. (2000). Fragile X males with unmethylated, full mutation trinucleotide repeat expansions have elevated levels of FMR1 messenger RNA. *American Journal of Medical Genetics*, *94*(3), 232–236. https://doi.org/10.1002/1096-8628(20000918)94:3<232::AID-AJMG9>3.0.CO;2-H

Tassone, F., Hagerman, R. J., Taylor, A. K., Gane, L. W., Godfrey, T. E., & Hagerman, P. J. (2000). Elevated levels of FMR1 mRNA 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

Tassone, F., Hagerman, R. J., Taylor, A. K., & Hagerman, P. J. (2001). A majority of fragile X males with methylated, full mutation alleles have significant levels of FMR1 messenger

RNA. *Journal of Medical Genetics*, *38*(7), 453–456. https://doi.org/10.1136/jmg.38.7.453
Tassone, F., Hagerman, R. J., Taylor, A. K., Mills, J. B., Harris, S. W., Gane, L. W., & Hagerman, P. (2000). Clinical involvement and protein expression in individuals with the FMR1 premutation. *American Journal of Medical Genetics*, *91*(2), 144–152. https://doi.org/10.1002/(SICI)1096-8628(20000313)91:2<144::AID-AJMG14>3.0.CO;2-V
Tassone, F., Longshore, J., Zunich, J., Steinbach, P., Salat, U., & Taylor, A. K. (1999). Tissue-specific methylation differences in a fragile X premutation carrier. *Clinical Genetics*, *55*(5), 346–351. https://doi.org/10.1034/j.1399-0004.1999.550508.x
Tassone, F., Pan, R., Amiri, K., Taylor, A. K., & Hagerman, P. J. (2008). 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. *Journal of Molecular Diagnostics*, *10*(1), 43–49. https://doi.org/10.2353/jmoldx.2008.070073
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. https://doi.org/10.1002/(SICI)1096-8628(19990528)84:3<233::AID-AJMG14>3.0.CO;2-6
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.bbagrm.2017.08.007
Usdin, K., Hayward, B. E., Kumari, D., Lokanga, R. A., Sciascia, N., & Zhao, X. N. (2014). Repeat-mediated genetic and epigenetic changes at the FMR1 locus in the Fragile X-related disorders. In *Frontiers in Genetics* (Vol. 5, Issue JUL, p. 226). Frontiers Research Foundation. https://doi.org/10.3389/fgene.2014.00226
Verheij, C., Bakker, C. E., de Graaff, E., Keulemans, J., Willemsen, R., Verkerk, A. J. M. H., Galjaard, H., Reuser, A. J. J., Hoogeveen, A. T., & Oostra, B. A. (1993). Characterization and localization of the FMR-1 gene product. *Nature*, *363*(12), 722–724.
Verkerk, A. J. M. H., De Graaff, E., De Boulle, K., Eichler, E. E., Konecki, D. S., Reyniers, E., Manca, A., Poustka, A., Willems, P. J., Nelson, D. L., & Oostra, B. A. (1993). Alternative splicing in the fragile X gene FMR1. *Human Molecular Genetics*, *2*(4), 399–404. http://hmg.oxfordjournals.org/
Weisman-Shomer, P., Cohen, E., & Fry, M. (2000). Interruption of the fragile X syndrome expanded sequence d(CGG)(n) by interspersed d(AGG) trinucleotides diminishes the formation and stability of d(CGG)(n) tetrahelical structures. *Nucleic Acids Research*, *28*(7), 1535–1541. https://doi.org/10.1093/nar/28.7.1535
Willemsen, R., Mohkamsing, S., Vries, B. De, Devys, D., Ouweland, A. Van Den, Mandel, J. L., Galjaard, H., & Oostra, B. (1995). Rapid antibody test for fragile X syndrome. *The Lancet*, *345*(8958), 1147–1148.
Willemsen, R., Smits, A., Mohkamsing, S., Van Beerendonk, H., De Haan, A., De Vries, B., Van Den Ouweland, A., Sistermans, E., Galjaard, H., & Oostra, B. A. (1997). Rapid antibody test for diagnosing fragile X syndrome: A validation of the technique. *Human Genetics*, *99*(3), 308–311. https://doi.org/10.1007/s004390050363
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
Zachary, A. A., Lucas, D. P., Detrick, B., & Leffell, M. S. (2009). Naturally occurring interference in Luminex® assays for HLA-specific antibodies: Characteristics and resolution. *Human Immunology*, *70*(7), 496–501. https://doi.org/10.1016/j.humimm.2009.04.001
Zafarullah, M., Tang, H. T., Durbin-Johnson, B., Fourie, E., Hessl, D., Rivera, S. M., & Tassone, F. (2020). FMR1 locus isoforms: potential biomarker candidates in fragile X-associated

tremor/ataxia syndrome (FXTAS). *Scientific Reports*, *10*(1), 11099. https://doi.org/10.1038/s41598-020-67946-y

Zhao, X. N., & Usdin, K. (2016). UPS and downs: Mechanisms of repeat instability in the fragile X-related disorders. In *Genes* (Vol. 7, Issue 9). MDPI AG. https://doi.org/10.3390/genes7090070

Zumwalt, M., Ludwig, A., Hagerman, P. J., & Dieckmann, T. (2007). Secondary structure and dynamics of the r(CGG) repeat in the mRNA of the fragile X mental retardation 1 (FMR1) gene. *RNA Biology*, *4*(2), 93–100. http://www.ncbi.nlm.nih.gov/pubmed/17962727

Chapter 3

Conclusion

Jamie L. Randol

The genetic puzzle continues

The *FMR1*-associated disorders arising from the altered expression of the expanded CGG-repeat allele represent, collectively, a fascinating and enigmatic molecular puzzle. The gene is either epigenetically silenced, leading to the neurodevelopmental disorder, fragile X syndrome (FXS); or becomes overexpressed, resulting in both reproductive (FXPOI) and neurodegenerative (FXTAS) disorders, depending only on the size of the repeat expansion. However, neither the trigger for silencing nor for overexpression are understood at present (Yang et al., 2016) (Hagerman & Hagerman, 2021), (Hagerman et al., 2018), (Fink et al., 2018) (Pieretti et al., 1991; Richter & Zhao, 2021).

The study of the FMR1-associated disorders is complicated by the complex nature of the factors governing its expression. For example, *FMR1* DNA and mRNA are known to form R-loops and higher-order structures, such as hairpins and G-quadruplexes (Zhao & Usdin, 2021). The gene also suffers from both mitotic and meiotic repeat size instability (Nolin et al., 2019; Zhao & Usdin, 2016). Moreover, it is prone to methylation mosaicism in full mutation carriers (Genç et al., 2000); and it is unclear how methylation couples with CGG-repeat length to affect gene expression. For instance, some methylated alleles still produce mRNA (see chapter 2 of this dissertation). To add to this complexity, expanded repeats foster a shift in transcriptional and translations start sites to upstream loci and can use both canonical and non-canonical polyadenylation signals (Beilina et al., 2004; Tassone et al., 2011; Todd et al., 2013). Furthermore, there are many alternative splicing isoforms that can affect expression and function of FMRP (Banerjee et al., 2010; Sittler et al., 1996; Verkerk et al., 1993).

Clearly, there is a lot to unravel. This dissertation sought to elucidate some of the factors contributing to this complexity by exploring features of *FMR1* gene expression in individuals with FXTAS and FXS, specifically. FXTAS derives from toxic transcription of the expanded CGG repeat that is often elevated with a paradoxical reduction in FMRP (Hagerman & Hagerman, 2016; Tassone, Hagerman, Taylor, Gane, et al., 2000; Tassone, Hagerman, Taylor, Mills, et al., 2000). However, why there is altered gene expression is not well understood despite over two decades of research (Hagerman et al., 2001). One of three prominent models of FXTAS pathogenesis proposes that co-transcriptional R-loop formation in the presence of the expanded repeat triggers an unresolved DNA damage response (DDR) at the *FMR1* locus, which ultimately leads to cellular toxicity (Hagerman & Hagerman, 2016; Loomis et al., 2014). To investigate this, I quantified *FMR1* R-loop abundance in control and premutation carriers with and without FXTAS. However, no association between steady-state R-loop frequency and CGG repeat size, mRNA level, or FXTAS were found. Therefore, these data suggest that any FXTAS-associated DDR is not materially triggered by R-loop frequency. It is yet unclear whether other R-loop features may play a role in the FXTAS DDR. However, a DDR from mitochondrial dysfunction – and how the expanded CGG repeat would drive such dysfunction – remains an important avenue of future research (Robin et al., 2017).

FXS typically results from loss of the synaptically-relevant FMRP protein due to promoter methylation and consequent gene silencing (Pieretti et al., 1991; Richter & Zhao, 2021). Data from Chapter 2 of this dissertation support findings from previous studies and spotlight the complexity of the locus in terms of both size and methylation mosaicism and their effect on FMRP production. We found that most individuals (83%) with FXS have complex genotypes. Only 17% of FXS patients in the study possessed a single, methylated full mutation allele. Furthermore, significant levels of FMRP were only detected in mosaics. However, non-significant levels of FMRP were successfully modeled by methylation status and repeat size, implying that FMRP was present below the significance threshold and that a more sensitive

assay could uncover more about the interplay of these molecular features. Importantly, the molecular mechanisms by which altered *FMR1* expression leads to FXS and FXTAS remain important areas of investigation.

References

Banerjee, P., Schoenfeld, B. P., Bell, A. J., Choi, C. H., Bradley, M. P., Hinchey, P., Kollaros, M., Park, J. H., Mcbride, S. M. J., & Dockendorff, T. C. (2010). *Short-and Long-Term Memory Are Modulated by Multiple Isoforms of the Fragile X Mental Retardation Protein*. https://doi.org/10.1523/JNEUROSCI.6369-09.2010

Beilina, A., Tassone, F., Schwartz, P. H., Sahota, P., & Hagerman, P. J. (2004). Redistribution of transcription start sites within the FMR1 promoter region with expansion of the downstream CGG-repeat element. *Human Molecular Genetics*, *13*(5), 543–549. https://doi.org/10.1093/hmg/ddh053

Fink, D. A., Nelson, L. M., Pyeritz, R., Johnson, J., Sherman, S. L., Cohen, Y., & Elizur, S. E. (2018). Fragile X Associated Primary Ovarian Insufficiency (FXPOI): Case Report and Literature Review. *Frontiers in Genetics*, *9*. https://doi.org/10.3389/FGENE.2018.00529

Genç, B., Müller-Hartmann, H., Zeschnigk, M., Deissler, H., Schmitz, B., Majewski, F., Von Gontard, A., & Doerfle, W. (2000). Methylation mosaicism of 5'-(CGG)(n)-3' repeats in fragile X, premutation and normal individuals. *Nucleic Acids Research*, *28*(10), 2141–2152. https://doi.org/10.1093/nar/28.10.2141

Hagerman, R. J., & Hagerman, P. J. (2016). Fragile X-associated tremor/ataxia syndrome — features, mechanisms and management. *Nature Reviews Neurology*, *12*(7), 403–412. https://doi.org/10.1038/nrneurol.2016.82

Hagerman, R. J., & Hagerman, P. J. (2021). FXTAS: Pathophysiology and Management. *Current Opinion in Neurology*, *34*(4), 541. https://doi.org/10.1097/WCO.0000000000000954

Hagerman, R. J., Leehey, M., Heinrichs, W., Tassone, F., Wilson, R., Hills, J., Grigsby, J., Gage, B., & Hagerman, P. J. (2001). Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology*, *57*(1), 127–130. https://doi.org/10.1212/WNL.57.1.127

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*. https://doi.org/10.3389/FPSYT.2018.00564

Loomis, E. W., Sanz, L. A., Chédin, F., & Hagerman, P. J. (2014). Transcription-Associated R-Loop Formation across the Human FMR1 CGG-Repeat Region. *PLoS Genetics*, *10*(4). https://doi.org/10.1371/journal.pgen.1004294

Nolin, S. L., Glicksman, A., Tortora, N., Allen, E., Macpherson, J., Mila, M., Vianna-Morgante, A. M., Sherman, S. L., Dobkin, C., Latham, G. J., & Hadd, A. G. (2019). Expansions and contractions of the FMR1 CGG repeat in 5,508 transmissions of normal, intermediate, and premutation alleles. *American Journal of Medical Genetics, Part A*, *179*(7), 1148–1156. https://doi.org/10.1002/ajmg.a.61165

Pieretti, M., Zhang, F., Fu, Y. H., Warren, S. T., Oostra, B. A., Caskey, C. T., & Nelson, D. L. (1991). Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell*, *66*(4), 817–822. https://doi.org/10.1016/0092-8674(91)90125-I

Richter, J. D., & Zhao, X. (2021). The molecular biology of FMRP: new insights into fragile X syndrome. In *Nature Reviews Neuroscience* (Vol. 22, Issue 4, pp. 209–222). Nature Publishing Group. https://doi.org/10.1038/s41583-021-00432-0

Robin, G., López, J. R., Espinal, G. M., Hulsizer, S., Hagerman, P. J., & Pessah, I. N. (2017). Calcium dysregulation and Cdk5-ATM pathway involved in a mouse model of fragile X-associated tremor/ataxia syndrome. *Human Molecular Genetics*, *26*(14), 2649–2666. https://doi.org/10.1093/hmg/ddx148

Sittler, A., Devys, D., Weber, C., & Mandel, J. L. (1996). Alternative Splicing of Exon 14 Determines Nuclear or Cytoplasmic Localisation of FMR1 Protein Isoforms. *Human Molecular Genetics*, *5*(1), 95–102. https://doi.org/10.1093/HMG/5.1.95

Tassone, F., De Rubeis, S., Carosi, C., La Fata, G., Serpa, G., Raske, C., Willemsen, R., Hagerman, P. J., & Bagni, C. (2011). Differential usage of transcriptional start sites and polyadenylation sites in FMR1 premutation alleles. *Nucleic Acids Research*, *39*(14), 6172–6185. https://doi.org/10.1093/nar/gkr100

Tassone, F., Hagerman, R. J., Taylor, A. K., Gane, L. W., Godfrey, T. E., & Hagerman, P. J. (2000). Elevated levels of FMR1 mRNA 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

Tassone, F., Hagerman, R. J., Taylor, A. K., Mills, J. B., Harris, S. W., Gane, L. W., & Hagerman, P. (2000). Clinical involvement and protein expression in individuals with the FMR1 premutation. *American Journal of Medical Genetics*, *91*(2), 144–152. https://doi.org/10.1002/(SICI)1096-8628(20000313)91:2<144::AID-AJMG14>3.0.CO;2-V

Todd, P. K., Oh, S. Y., Krans, A., He, F., Sellier, C., Frazer, M., Renoux, A. J., Chen, K. chun, Scaglione, K. M., Basrur, V., Elenitoba-Johnson, K., Vonsattel, J. P., Louis, E. D., Sutton, M. A., Taylor, J. P., Mills, R. E., Charlet-Berguerand, N., & Paulson, H. L. (2013). CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron*. https://doi.org/10.1016/j.neuron.2013.03.026

Verkerk, A. J. M. H., De Graaff, E., De Boulle, K., Eichler, E. E., Konecki, D. S., Reyniers, E., Manca, A., Poustka, A., Willems, P. J., Nelson, D. L., & Oostra, B. A. (1993). Alternative splicing in the fragile X gene FMR1. *Human Molecular Genetics*, *2*(4), 399–404. http://hmg.oxfordjournals.org/

Yang, W., Fan, C., Chen, L., Cui, Z., Bai, Y., & Lan, F. (2016). Pathological Effects of the FMR1 CGG-Repeat Polymorphism (5-55 Repeat Numbers): Systematic Review and Meta-Analysis. *Tohoku J. Exp. Med*, *239*(1), 57–66. https://doi.org/10.1620/tjem.239.57

Zhao, X. N., & Usdin, K. (2016). UPS and downs: Mechanisms of repeat instability in the fragile X-related disorders. In *Genes* (Vol. 7, Issue 9). MDPI AG. https://doi.org/10.3390/genes7090070

Zhao, X. N., & Usdin, K. (2021). (Dys)function follows form: Nucleic acid structure, repeat expansion, and disease pathology in FMR1 disorders. *International Journal of Molecular Sciences*, *22*(17), 1–15. https://doi.org/10.3390/ijms22179167