Title

Structure and Alternative Splicing of the Antisense FMR1 (ASFMR1) Gene

Journal

Molecular Neurobiology, 60(4)

ISSN

0893-7648

Authors

Zafarullah, Marwa

Li, Jie

Tseng, Elizabeth

et al.

Publication Date

2023-04-01

DOI

10.1007/s12035-022-03176-9

Peer reviewed

HHS Public Access

Author manuscript

Mol Neurobiol. Author manuscript; available in PMC 2023 August 28.

Published in final edited form as:

Mol Neurobiol. 2023 April ; 60(4): 2051–2061. doi:10.1007/s12035-022-03176-9.

Structure and Alternative Splicing of the Antisense *FMR1* (*ASFMR1*) Gene

Marwa Zafarullah[1], **Jie Li**[2], **Elizabeth Tseng**[3], **Flora Tassone**[1,4]

[1]Department of Biochemistry and Molecular Medicine, University of California Davis, School of Medicine, Sacramento, CA 95817, USA

[2]Bioinformatics Core, Genome Center, University of California Davis, Davis, CA 95616, USA

[3]Pacific Biosciences, Inc, Menlo Park, CA 94025, USA

[4]MIND Institute, University of California Davis Medical Center, Sacramento, CA 95817, USA

Abstract

Fragile X-associated tremor/ataxia syndrome (FXTAS) is a neurodegenerative disorder caused by an expansion of 55–200 CGG repeats (premutation) in the 5'-UTR of the *FMR1* gene. Bidirectional transcription at *FMR1* locus has been demonstrated and specific alternative splicing of the Antisense *FMR1* (*ASFMR1*) gene has been proposed to have a contributing role in the pathogenesis of FXTAS. The structure of *ASFMR1* gene is still uncharacterized and it is currently unknown how many isoforms of the gene are expressed and at what level in premutation carriers (PM) and if they may contribute to the premutation pathology. In this study, we characterized the *ASFMR1* gene structure and the transcriptional landscape by using PacBio SMRT sequencing with target enrichment (IDT customized probe panel). We identified 45 *ASFMR1* isoforms ranging in sizes from 523 bp to 6 Kb, spanning approximately 59 kb of genomic DNA. Multiplexing and sequencing of six human brain samples from PM samples and normal control

✉Flora Tassone, ftassone@ucdavis.edu.

Author Contribution Designed Research.
Conception: F. Tassone
Organization: M. Zafarullah, and F. Tassone
Execution: M. Zafarullah, F. Tassone
Analyzed Data
Design: F. Tassone, E. Tseng
Execution: J.LI, E. Tseng,
Review and Critique: M. Zafarullah, J. Li, E. Tseng, F. Tassone
Manuscript Preparation
Writing the first draft: M. Zafarullah
Review and Critique: M. Zafarullah, J. Li, E. Tseng, F. Tassone
All authors read and approved the final manuscript.

Competing Interests M.Z.: No disclosures to report.
J.L.: No disclosures to report.
E.T: E.T. is an employee of Pacific Biosciences.
F.T. has received the funding from Azrieli Foundation and Zynerba for studies in Fragile X syndrome.

Ethics Approval All samples were collected under approved Institutional Review Board (IRB) protocols and all methods were carried out in accordance with the relevant guidelines and regulations.

Consent to Participate Informed consent was obtained from all individual participants included in the study.

Consent for Publication Not applicable.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12035-022-03176-9.

(HC) were carried out on the PacBio Sequel platform. We validated the presence of these isoforms by qRT-PCR and Sanger sequencing and characterized the acceptor and donor splicing site consensus sequences. Consistent with previous studies conducted in other tissue types, we found a high expression of *ASFMR1* isoform *Iso131bp* in brain samples of PM as compared to HC, while no differences in expression levels were observed for the newly identified isoforms *IsoAS1* and *IsoAS2*. We investigated the role of the splicing regulatory protein Sam68 which we did not observe in the alternative splicing of the *ASFMR1* gene. Our study provides a useful insight into the structure of *ASFMR1* gene and transcriptional landscape along with the expression pattern of various newly identified novel isoforms and on their potential role in premutation pathology.

Keywords

Fragile X-associated Tremor/ataxia Syndrome; *ASFMR1* ; *FMR1* ; Alternative Splicing; Isoforms; Long Read Sequencing

Introduction

Fragile X and associated disorders are a group of diseases caused by a trinucleotide CGG repeat expansion in the 5’ UTR of the *FMR1* gene. While a repeat expansion greater than 200 CGG leads to Fragile X Syndrome (FXS), the most common form of inherited intellectual disability, a repeat expansion between 55 and 200 CGG repeats (known as premutation) confers the risk of developing Fragile X-associated tremor/ataxia syndrome (FXTAS), a late-onset neurodegenerative disorder characterized by intention tremor, gait ataxia, autonomic dysfunction, and Parkinsonism [1, 2]. The proposed molecular mechanisms of FXTAS pathogenesis include the sequestration of CGG binding proteins including the splicing regulatory protein Sam68 [3] by the *FMR1* mRNA [4] which elevated expression leads to RNA toxicity, production of toxic FMRPolyG, and the chronic activation of DNA damage response [5].

The transcriptional and posttranscriptional events control the gene expression levels by the production of multiple distinct mRNA isoforms. These isoforms increase protein diversity and play an essential role in brain development [6] and cell survival [7]. The alteration in these mechanisms results in dysregulation of gene expression and is found to be contributed to the development of various neurological disorders including autism spectrum disorder (ASD) [8], Alzheimer’s disease (AD) [9], and others.

The dysregulation of *FMR1* gene expression has been implicated in the pathogenesis of FXTAS, where an increased level of many isoforms was observed in PM [10–12]. Bidirectional transcription has been observed at the *FMR1* locus through comprehensive transcriptome analyses (*FMR 4* [13], *FMR 5,* and *FMR6*) [14]. Importantly, a unique antisense *FMR1* (*ASFMR1*) gene overlapping the CGG repeat region of the *FMR1* gene was identified. Its high expression level in brain samples of PM and no expression in those with a full mutation endorse a potential contribution of the *ASFMR1* to the clinical phenotype of FXTAS and FXS [15]. The *ASFMR1* is known to exhibit a premutation-specific alternative spliced isoform, the *Iso131bp,* which is mainly expressed in PM compared to HC [10,

16–18]. However, no studies had been conducted to characterize the full alternative splicing repertoire of the *ASFMR1* gene.

Over the past decade, with the ability to generate a thousand kilobases with high accuracy the long-read, single-molecule DNA sequencing technologies have emerged as a powerful tool in genomics. An amplicon-based approach had been used in identifying complex alternative splicing at the *FMR1* gene [11] but it was found to be limited by the PCR primer design. Using the targeted enrichment, such as the Integrated DNA Technologies (IDT) specific probes, a highly accurate full-length transcript reads that can provide a comprehensive isoform view of low expressive genes like *ASFMR1* can be obtained [19, 20].

In this study, we characterized the structure of the *ASFMR1* gene and identified novel splicing isoforms by using the IDT customized probe panel with long-read sequencing. We demonstrate that targeted capture is a cost-effective way of jointly characterizing the structure of the lowly expressed *ASFMR1* transcript and its transcriptional landscape.

Materials and Methods

Tissue Samples

Frozen postmortem human cerebellum tissue from three PM cases diagnosed with FXTAS (mean CGG repeat number = 87; range 86–88 CGG repeats), obtained from the Medical Investigation of Neurodevelopmental Disorders Institute Brain Repository at the University of California at Davis in Sacramento, CA. Three-age matched HC (mean CGG repeat number = 30; range 20–41 CGG repeats) brain tissues were obtained from the Brain and Tissue Bank for Developmental Disorders of the National Institute of Child Health and Human Development at the University of Maryland in Baltimore, MD (Table 1). All samples were collected under approved Institutional Review Board (IRB) protocols and all methods were carried out in accordance with the relevant guidelines and regulations.

qRT-PCR analysis was carried out from total RNA isolated from 10 frozen postmortems human cerebellum tissue derived from PM cases diagnosed with FXTAS (mean CGG repeat number = 88; range 64–113 CGG repeats), and from five age-matched HC. To study the potential effect of Sam68 on the alternative splicing of the *ASFMR1* transcript, fibroblast cell lines from three PM cases, (mean CGG repeat number = 118; range 78–170 CGG repeats) and three age-matched HC (mean CGG repeat number = 25; range 22–30 CGG repeats) were established at the MIND Institute.

Cell Culture and Reagents

The following method for human fibroblast cell culturing was originally described in our previous report [21]. For the siRNA transfection to knock down the Sam68, an established human fibroblast cell line was placed in treated tissue culture T-25 flask (Corning Life Science, USA) with a modified Fibroblast Growth Medium [1:1 solution of Gibco AmnioMAX-C100 supplemented with 15% AmnioMAX-C100 Supplement (Invitrogen, Carlsbad, CA, USA) and RPMI-1640 Basal Medium supplemented with 10% fetal bovine serum (Invitrogen), 1X primocin (Invitrogen), 1% non-essential amino acids (Invitrogen),

and 1:250 fungizone (J R Scientific, Woodland, CA, USA)] with media exchange every 5 days and allowed to reach 90% confluence prior to splitting. Fibroblast cultures were passaged no more than 3 times prior to transfection.

RNA Extractions

Total RNA was isolated from brain samples and from fibroblast cell cultures in a clean and RNA-designated area by using TRIzol reagent (Invitrogen, Carlsbad, CA). Concentration was measured by spectrophotometric analysis. For the PacBio library preparation (see below) the quality of the RNA samples and lack of significant degradation was confirmed by measurements of the RNA Integrity Number (RIN, Table 1) utilizing an Agilent 2100 Bioanalyzer system.

CGG Repeat Sizing and Methylation Status

CGG repeat allele size and methylation status were assessed using a combination of Southern Blot analysis and PCR amplification. For Southern blot analysis, 5–10 μg of isolated genomic DNA from brain tissue was digested with EcoRI and NruI, run on an agarose gel, transferred on a nylon membrane, and hybridized with the *FMR1*-specific dig-labeled StB12.3 as detailed in [22]. PCR analysis was performed using *FMR1* specific primers (AmplideX PCR/CE, Asuragen, Inc.); amplicons were visualized by capillary electrophoresis, as previously reported, and analyzed using Gene Mapper software [23].

Library Preparation and Sequencing

cDNA Capture Using IDT Xgen® Lockdown® Probes and Single-Molecule Isoform-Sequencing—One microgram of total RNA per reaction was reverse transcribed using the Clontech SMARTer cDNA synthesis kit and six sample-specific barcoded oligos dT (16mer barcode sequences, Table 1). Three reverse transcription (RT) reactions were processed in parallel for each sample. PCR optimization was used to determine the optimal amplification cycle number for the downstream large-scale PCR reactions. A single primer (primer IIA from the Clontech SMARTer kit 5′ AAG CAG TGG TAT CAA CGC AGA GTA C 3′) was used for all PCR reactions post-RT. Large-scale PCR products were purified separately with 1X AMPure PB beads, and the bioanalyzer was used for QC. An equimolar pool of 6-plex barcoded cDNA library (5 μg total) was input into the probe-based capture with a custom-designed *ASFMR1* gene panel. Due to the low gene expression, five captures were performed with 5 ug input and 16 h incubation followed by 11 cycles of post-capture amplification. A SMRTBell library was constructed using 135 ng of captured and re-amplified cDNA. One SMRT Cell 1 M (20 h movie) was sequenced on the PacBio Sequel platform using 2.0 chemistry.

Clustering and Determining Haplotypes for CT-Rich Region—Subsequences aligned to chrX (hg38) were extracted for each sample. After inspecting the size distribution of these subsequences, they were clustered by size and sequence similarity using a combination of python and MUSCLE [[24] Edgar, 2004], and a consensus sequence was generated independently for each cluster.

Isoform Analysis—Sequencing of the barcoded cDNA data was obtained on one SMRT Cell 1 M on the PacBio Sequel system using 2.0 chemistry. Bioinformatics analysis was done using the IsoSeq3 application in the PacBio SMRT Analysis v6.0.0 to obtain high-quality, full-length isoform sequences.

Identification of Alternative Donor and Acceptor Sites—High-quality full-length transcript sequences from the IsoSeq data were mapped to the hg38 genome and collapsed into unique transcripts using the cDNA_Cupcake (https://github.com/Magdoll/cDNA_Cupcake) scripts, then classified using the SQANTI3 (https://github.com/ConesaLab/SQANTI3) tool against the Gencode v33 annotation.

RT-PCR and Sanger Sequencing

To validate the existence of novel *ASFMR1* transcripts, 3 ug of total RNA derived from the PM postmortem cerebellum included in this study was reverse transcribed using SuperScript IV (Thermofisher, Waltham, MA, USA) following manufacturers' recommendations. Primers were designed by using the software Primer3 and further checked in the In-silico PCR tool of the UCSC Genome Browser to confirm that they were specific for the region to be amplified. The following primers and probes were designed to sequence first four splicing isoforms including a* (forward primer 5'- TGA CAG GTC GCA CTG CCT -3', reverse primer 5'- CGC CCA AAA TCT GGT GAG AGA -3'), b* (forward primer 5'- GAG GGC CAG AAC GCC CAT TTC -3', reverse primer 5'- ACA GTG TCT GGC ACA CGA TAG A -3'), c* (forward primer 5'- GAG GGC CAG AAC GCC CAT TTC -3', reverse primer 5'- CCA TCT AGC AAC AGT GTC AAA ACA T -3'), and d* (forward primer 5'- TGA CAG GTC GCA CTG CCT -3', reverse primer 5'- TTT GAT GGG CTC ACC AGT AGA C -3'). The amplicons, which sizes were 210 bp, 57 bp, 323 bp, and 187 bp respectively, were isolated on a 4% metaphor agarose gel and the appropriate size bands were excised from the gel and purified with the Qiagen gel purification kit. Isolated DNA was Sanger sequenced following standard procedures.

mRNA Expression Levels

mRNA expression levels were measured by qRT-PCR. cDNA synthesis and real-time PCRs (qRT-PCR) using both Assays-On-Demand from Applied Biosystems (Applied Biosystems, Foster City, CA, United States) and custom-designed TaqMan primers and probe assays were as previously described[[25]] Custom-designed primers and probes were used to quantify the *ASFMR1* gene and *ASFMR1 131 bp* splice isoform [15]. The following primers and probes were designed to quantify the expression levels of two of the novel *ASFMR1* transcripts identified: *IsoAS1* (forward primer 5'-GGC CAG AAC GCC CAT TTC-3', reverse primer 5'- TCT GGT GAG AGA GGG CGT AGA-3') and *IsoAS2* (forward primer 5'- CAT GGA TTG AGT CGG GAT CTC −3', reverse primer 5'- GAT GCT CCG TCA ACT TCT CAG TT-3'). For Sam68 the Assays-On-Demand from Applied Biosystems (Applied Biosystems, Foster City, CA, United States) was used. Data were analyzed using the Sigma plot and the graphs were generated using PRISM software.

Sam68 Knockdown

In vitro transfection was performed in human fibroblast cell lines derived from three PM and three age-matched HC which stably expresses Sam68. siRNAs were obtained as ready-annealed, purified duplexes from Dharmacon Research (Lafayette, CO, USA). Cells were plated 24 h prior to the experiment into 6-well plates at 100,000 cells/well. Transfection was carried out using Lipofectamine™ RNAiMAX according to the instructions of the manufacturer for forward transfection. For lipofection treated cells, -ve cells (cells treated with RNAiMAX without siRNA) and untreated cells (cells without RNAiMAX) from both HC and PM were grown under the same condition. After 4 h incubation, the transfection medium was replaced by a fresh culture medium, incubated for 24 h, and lysed with TRIzol. Total RNA was extracted, and qRT-PCR was performed. The data was analyzed, and the graphs were generated using PRISM software.

Results

In this study, we used the customized probe-based capture by using the PacBio long-read sequencing platform to characterize the structure of the *ASFMR1* gene and to identify novel *ASFMR1* isoforms from 6 human brain samples derived from PM and from HC (Fig. 1). Using the expression levels for FMR1 and ASFMR1 in GTEx database (https://gtexportal.org/home/ the GTEx Portal on 09/28/2022) for brain tissues, the highest expression levels are about 80 and 0.5 TPM. The captured transcripts in the current study present 3524.7–18,579.5 and 383.0–1189.6 TMP, for *FMR1* and *ASFMR1* respectively. Therefore, the enrichment is significant, at 44–232 and 766–2379-fold for *FMR1* and *ASFMR1* respectively. Bioinformatics analysis identified the *ASFMR1* gene consists of 13 exons (Table 2) that spanned across the 59 kb of the genomic region, in which 45 isoforms were generated by the usage of new splicing sites, exon skipping, and intron retention. Several isoforms were expressed at a low level and detected only in PM as compared to HC.

Identification of Novel Splicing Variants of the ASFMR1 Gene

We obtained 497,240 full-length non-chimeric (FLNC) reads from PM brain samples and 474,711 FLNC reads for samples derived from age and gender-matched HC. To identify unique transcripts, each sample (from pooled tissues but with barcode information recorded) was run through the PacBio customized bioinformatics pipeline to acquire high-quality consensus isoform sequences, which were then mapped to the hg38 reference genome. After final filtration 16,105 unique transcripts were identified in the PM group, of which, 135 (0.7%) mapped to the *FMR1* locus. Of the 16,394 unique transcripts identified in the HC group, 164 (1%) mapped to the *FMR1* locus (Table 3). Among these transcripts, we identified a total of 45 unique *ASFMR1* isoforms ranging in size from 523 bp to 6 Kb (including the exon 1, 2, 3, and 4) (Fig. 2a, Supplementary Table 1) and spanning approximately 59 kb of the genomic region. The overlap between *ASFMR1* and *FMR1* gene as reported by [15] was covered with customized design primers and Sanger sequencing (Fig. 2b). Out of 45 unique isoforms, 19 were more highly expressed and detected only in PM, while 17 were more highly expressed and detected only in HC, and nine were common in both PM and HC.

We validated the existence of four novel isoforms by qRT-PCR and Sanger sequencing: the *IsoAS1* and *IsoAS2 isoforms* are shown in (Fig. 3a, b), while 2 additional ones, missing two different portions of the gene, were also confirmed by Sanger sequencing and are shown in (Fig. 3c, d). All identified isoforms have canonical splice sites (GT-AG or GC-AG) (Table 2) supported by multiple full-length reads. Most of the isoforms have four exons, differing only in the use of upstream and downstream regions.

ASFMR1 Isoforms Expressed in Human Brain Tissues

To determine the expression level of several *ASFMR1* isoforms, including the previously identified *Iso131bp* [15] and selected newly identified isoforms *IsoAS1* and *IsoAS2* along with the *ASFMR1* and *FMR1* genes, qRT–PCR from ten PM and five HC postmortem cerebellum brain samples was performed. Interestingly, we found higher expression of the *ASFMR1* ($p = 0.019$; Fig. 4a) and of the *Iso131bp* ($p = 0.021$; Fig. 4b) transcripts in PM as compared to HC. However, we did not observe a significant difference in the expression levels of either the *FMR1* gene (Fig. 4c) or the *IsoAS1* (Fig. 4d) and *IsoAS2* (Fig. 4e) isoforms between the two groups. This expression profile suggests a potential role for *ASFMR1* and the *ASFMR1 Iso131bp* mRNAs, in the pathogenesis of FXTAS.

Sam68 Does Not Affect Alternative Splicing of the *ASFMR1* Gene

To evaluate the role of Sam68 in the alternative splicing of the *ASFMR1* gene, we transfected fibroblast cell lines established from three PM and three HC individuals followed by quantitative expression analysis via qRT-PCR. A significant decrease in the expression levels of the Sam68 in transfected HC cells compared to the lipofectamine treated (CL; $p = 0.027$; Fig. 5a) and untreated group (CU; $p = 0.013$; Fig. 5a) was observed. Consistently, we also observed a decreased Sam68 expression, indicating a successful knockdown of Sam68, in transfected cells from PM compared to the lipofectamine only treated and (PL; $p = 0.001$; Fig. 5a) to the untreated cells (PU; $p < 0.01$; Fig. 5a). No difference in the expression level of the *FMR1* gene (Fig. 5b) and of the *ASFMR1* isoforms including *Iso131bp* (Fig. 5c), *IsoAS1* (Fig. 5d), and *IsoAS2* (Fig. 5e) was observed, suggesting that splicing alterations observed within the *ASFMR1* gene are not affected by the sequestration of Sam68.

Discussion

Antisense transcripts constitute an additional layer of regulation of gene expression. Thus, their characterization is relevant for the identification of candidates, particularly those associated with neurodegenerative disorders (Zucchelli et al., 2019-Mol Neurob.). In this study, we characterized the structure of the *ASFMR1* gene by using the targeted enrichment on multiplexed cDNA libraries with the PacBio long-read sequencing platform. Using the human brain tissues derived from the male carriers with FXTAS we identified novel *ASFMR1* isoforms, expressed only in PM but not detected in HC, and confirmed the presence and the expression of a subset of these transcripts via orthogonal methods. Our study shows that splicing regulation of the *ASFMR1* gene is varied meaning as compared to the *FMR1* gene.

Unstable repeat expansions cause more than 30 neurodegenerative and neuromuscular diseases of which the bidirectional transcription has been identified in at least 14 of these disorders [26]. The first repeat expansion was discovered in the *FMR1* gene in 1991 [27], which leads to the development of FXS. Two decades later, the presence of bidirectional transcription and the *ASFMR1* was reported [15]; however, the structure of the *FMR1* gene is well defined, and there is little knowledge available regarding the structure of the *ASFMR1* gene. Using targeted capture with long-read sequencing, we characterized the structure of the *ASFMR1* gene and demonstrated the presence of 13 exons spanning over 59 kb of genomic DNA. Our goal was to understand the expression and the role of the *ASFMR1* in the pathogenesis of FXTAS.

To date, a total of 45 *FMR1* mRNA isoforms have been identified, some of which are observed only in PM as compared to HC individuals [11]. Similarly, it was reported that [15] also the *ASFMR1* gene goes under alternative splicing, and, that of the five isoforms only the isoform *Iso131bp* overlapped with the *FMR1* gene and was differentially expressed in PM, as confirmed later in other studies [10, 16–18]. In this study, we identified 45 novel *ASFMR1* transcripts, 19 of which were detected only in PM, 17 only in HC, and nine both in groups. However, it is possible that additional isoforms may present, in either the PM or the HC group or in both, at very low level of expression and undetectable by this approach. We further validated the presence of these isoforms using qRT-PCR and sequencing analysis. Since all our identified isoforms showed canonical splice sites (GT-AG or GC-AG) and the obtained sequences did not overlap with the *FMR1* gene, we covered the gap with customized designed primers and Sanger sequencing. The characterization of these newly identified isoforms in this study is providing a better understanding of the *ASFMR1* gene regulation, as an imbalance in their expression could lead to an altered functional diversity with neurotoxic consequences. Besides, it will also help in elucidating the mechanism(s) by which “toxic gain of function” of the *ASFMR1* mRNA may play a role in the development of FXTAS and potentially for the development of pharmacological target treatment for this disorder.

The expression profile of the *FMR1* gene shows its relatively high abundance in human tissues such as the brain, kidney, ovary, testes, and thyroid to barely detectable levels in skeletal muscle [15, 28]. Moreover, [11] the transcriptional *FMR1* isoforms distribution pattern in peripheral blood mononuclear cells, fibroblasts, and brain tissue samples did not show any tissue-specific pattern [11]. The *ASFMR1* gene is fully expressed in all the tissues, with relatively high expression in the brain and kidney and barely detectable in the heart, placenta, prostate, skeletal muscle, thyroid, and trachea [15]. Although we did not observe any significant difference in the expression levels of the newly identified *IsoAS1* and *IsoAS2* isoforms, we confirmed a significantly high expression of the *ASFMR1* gene as well as of the *ASFMR1 Iso131bp* in the brain tissues of PM as compared to HC suggesting their potential role in the pathogenesis of FXTAS.

Sam68 is a splicing regulator encoded by the *KHDRBS1* gene and is involved in various aspects of RNA metabolism, from transcription to alternative splicing and then to nuclear export [29]. It is also involved in many signaling pathways linked to cell response to stimuli, and cell cycle transitions [30]. The expanded CGG repeats in the *FMR1* gene in PM result

in the 2–8 folds increase in the level of *FMR1* mRNA [4] which is one of the molecular mechanisms leading to the development of the FXTAS. Previously, it was reported that mRNA containing the extended CGG repeats form a large intranuclear RNA aggregated that sequesters many essential proteins including the splicing factor Sam68 which results in the loss of its splicing-regulatory function [3]. We showed that the alternative splicing observed in the *ASFMR1* gene does not appear driven by the sequestering of the Sam68 and different regulatory factors are likely involved. As fibroblasts are likely not the best cell model to use, due to the lack of a robust evidence of RNA toxicity, further studies, in different cell models, are needed to understand the splicing mechanism involved in the regulation of *ASFMR1* that could contribute to the pathogenesis of FXTAS and of other related disorders.

Conclusion

Antisense genes within trinucleotide repeat expansion are associated with various neurodegenerative disorders and their altered expression could have a real impact on the clinical manifestation of neurodegeneration. Although huge efforts have been made to understand the complex mechanisms underlying these diseases, the contribution of these genes and repeat expanded RNAs remains elusive. Our study lays the groundwork for the characterization of the *ASFMR1* gene, and the identification of the unique isoforms by using the PacBio long-read targeted capture from human brain regions. This knowledge is highly valuable and applicable for a better understanding of the FXTAS and related disorders development. Further studies are needed to fully discover the biological and pathological significance of these novel isoforms to improve the success of therapeutic strategies for age-related neurodegenerative disorders such as FXTAS.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank the participants of the community-based studies who donated their time and sample for this study. This paper is dedicated to the memory of Matteo.

Funding

This work was supported by the NIH grants RO1HD036071 and RO1MH07804.

Data Availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Hagerman RJ, Leehey M, Heinrichs W et al. (2001) Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. Neurology 57:127–130 [PubMed: 11445641]
2. Mila M, Alvarez-Mora MI, Madrigal I, Rodriguez-Revenga L (2018) Fragile X syndrome: an overview and update of the *FMR1* gene. Clin Genet 93:197–205 [PubMed: 28617938]

3. Sellier C, Rau F, Liu Y et al. (2010) Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. EMBO J 29:1248–1261 [PubMed: 20186122]
4. Tassone F, Hagerman RJ, Taylor AK et al. (2000) Elevated levels of FMR1 mRNA in carrier males: a new mechanism of involvement in the Fragile-X syndrome. Am J Human Genet 66:6–15 [PubMed: 10631132]
5. Hagerman RJ, Hagerman P (2016) Fragile X-associated tremor/ataxia syndrome - features, mechanisms and management. Nat Rev Neurol 12:403–412 [PubMed: 27340021]
6. Su C-H, Dhananjaya D, Tarn W-Y (2018) Alternative splicing in neurogenesis and brain development. Front Mol Biosci 5:12 [PubMed: 29484299]
7. Paronetto MP, Passacantilli I, Sette C (2016) Alternative splicing and cell survival: from tissue homeostasis to disease. Cell Death Differ 23:1919–1929 [PubMed: 27689872]
8. Quesnel-Vallières M, Dargaei Z, Irimia M et al. (2016) Misregulation of an activity-dependent splicing network as a common mechanism underlying autism spectrum disorders. Mol Cell 64:1023–1034 [PubMed: 27984743]
9. Koch L (2018) Altered splicing in Alzheimer transcriptomes. Nat Rev Genet 19:738–739
10. Pretto DI, Eid JS, Yrigollen CM et al. (2015) Differential increases of specific FMR1 mRNA isoforms in premutation carriers. J Med Genet 52:42–52 [PubMed: 25358671]
11. Tseng E, Tang H-T, AlOlaby RR et al. (2017) Altered expression of the FMR1 splicing variants landscape in premutation carriers. Biochim Biophys Acta Gene Regul Mech 1860:1117–1126 [PubMed: 28888471]
12. Zafarullah M, Tang H-T, Durbin-Johnson B et al. (2020) FMR1 locus isoforms: potential biomarker candidates in fragile X-associated tremor/ataxia syndrome (FXTAS). Sci Rep 10:11099 [PubMed: 32632326]
13. Khalil AM, Faghihi MA, Modarresi F et al. (2008) A novel RNA transcript with antiapoptotic function is silenced in fragile X syndrome. PLoS One 3:e1486 [PubMed: 18213394]
14. Pastori C, Peschansky VJ, Barbouth D et al. (2014) Comprehensive analysis of the transcriptional landscape of the human FMR1 gene reveals two new long noncoding RNAs differentially expressed in Fragile X syndrome and Fragile X-associated tremor/ataxia syndrome. Hum Genet 133:59–67 [PubMed: 24005575]
15. Ladd PD, Smith LE, Rabaia NA et al. (2007) An antisense transcript spanning the CGG repeat region of FMR1 is upregulated in premutation carriers but silenced in full mutation individuals. Hum Mol Genet 16:3174–3187 [PubMed: 17921506]
16. Al Olaby RR, Tang H-T, Durbin-Johnson B et al. (2018) Assessment of molecular measures in non-FXTAS male premutation carriers. Front Genet 9:302 [PubMed: 30186307]
17. Vittal P, Pandya S, Sharp K et al. (2018) splice variant: a predictor of fragile X-associated tremor/ataxia syndrome. Neurol Genet 4:e246 [PubMed: 30065951]
18. Loesch DZ, Godler DE, Evans A et al. (2011) Evidence for the toxicity of bidirectional transcripts and mitochondrial dysfunction in blood associated with small CGG expansions in the FMR1 gene in patients with parkinsonism. Genet Med 13:392–399 [PubMed: 21270637]
19. Tseng E, Rowell WJ, Glenn O-C et al. (2019) The landscape of SNCA transcripts across synucleinopathies: new insights from long reads sequencing analysis. Front Genet 10:584 [PubMed: 31338105]
20. Sheynkman GM, Tuttle KS, Laval F et al. (2020) ORF Capture-Seq as a versatile method for targeted identification of full-length isoforms. Nat Commun 11:2326 [PubMed: 32393825]
21. Pretto DI, Mendoza-Morales G, Lo J et al. (2014) CGG allele size somatic mosaicism and methylation in FMR1 premutation alleles. J Med Genet 51:309–318 [PubMed: 24591415]
22. Tassone F, Pan R, Amiri K et al. (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. J Mol Diagn 10:43–49 [PubMed: 18165273]
23. Filipovic-Sadic S, Sah S, Chen L et al. (2010) A novel FMR1 PCR method for the routine detection of low abundance expanded alleles and full mutations in fragile X syndrome. Clin Chem 56:399–408 [PubMed: 20056738]
24. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797 [PubMed: 15034147]

25. Tassone F, Hagerman RJ, Chamberlain WD, Hagerman PJ (2000) Transcription of the FMR1 gene in individuals with fragile X syndrome. Am J Med Genet 97:195–203 [PubMed: 11449488]
26. Castro AF, Loureiro JR, Bessa J, Silveira I (2020) Antisense transcription across nucleotide repeat expansions in neurodegenerative and neuromuscular diseases: progress and mysteries. Genes 11:1418. 10.3390/genes11121418 [PubMed: 33261024]
27. Verkerk AJ, Pieretti M, Sutcliffe JS et al. (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell 65:905–914 [PubMed: 1710175]
28. Hinds HL, Ashley CT, Sutcliffe JS et al. (1993) Tissue specific expression of FMR-1 provides evidence for a functional role in fragile X syndrome. Nat Genet 3:36–43 [PubMed: 8490651]
29. Sánchez-Jiménez F, Sánchez-Margalet V (2013) Role of Sam68 in post-transcriptional gene regulation. Int J Mol Sci 14:23402–23419 [PubMed: 24287914]
30. Frisone P, Pradella D, Di Matteo A et al. (2015) SAM68: signal transduction and RNA metabolism in human cancer. Biomed Res Int 2015:528954 [PubMed: 26273626]

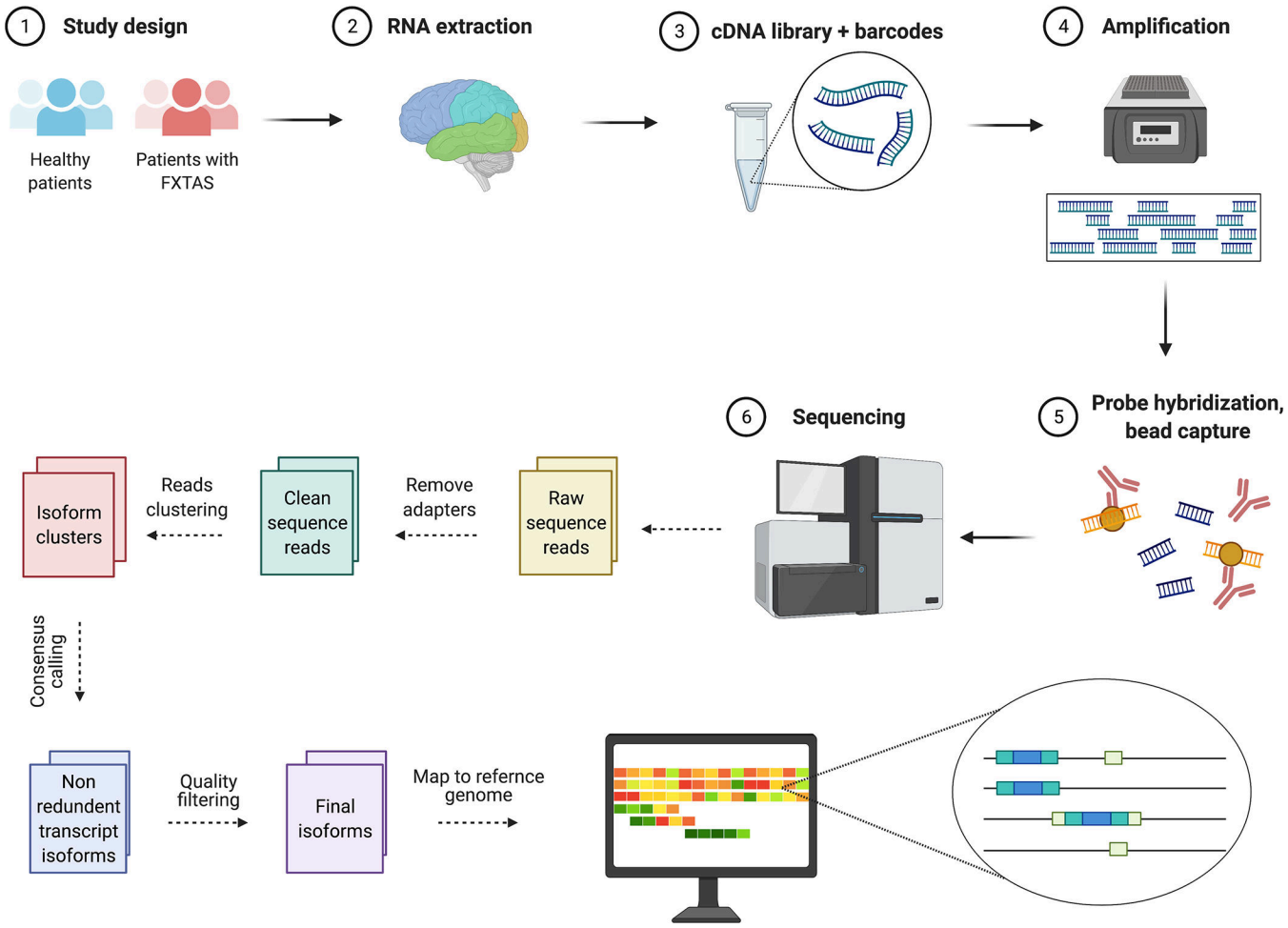
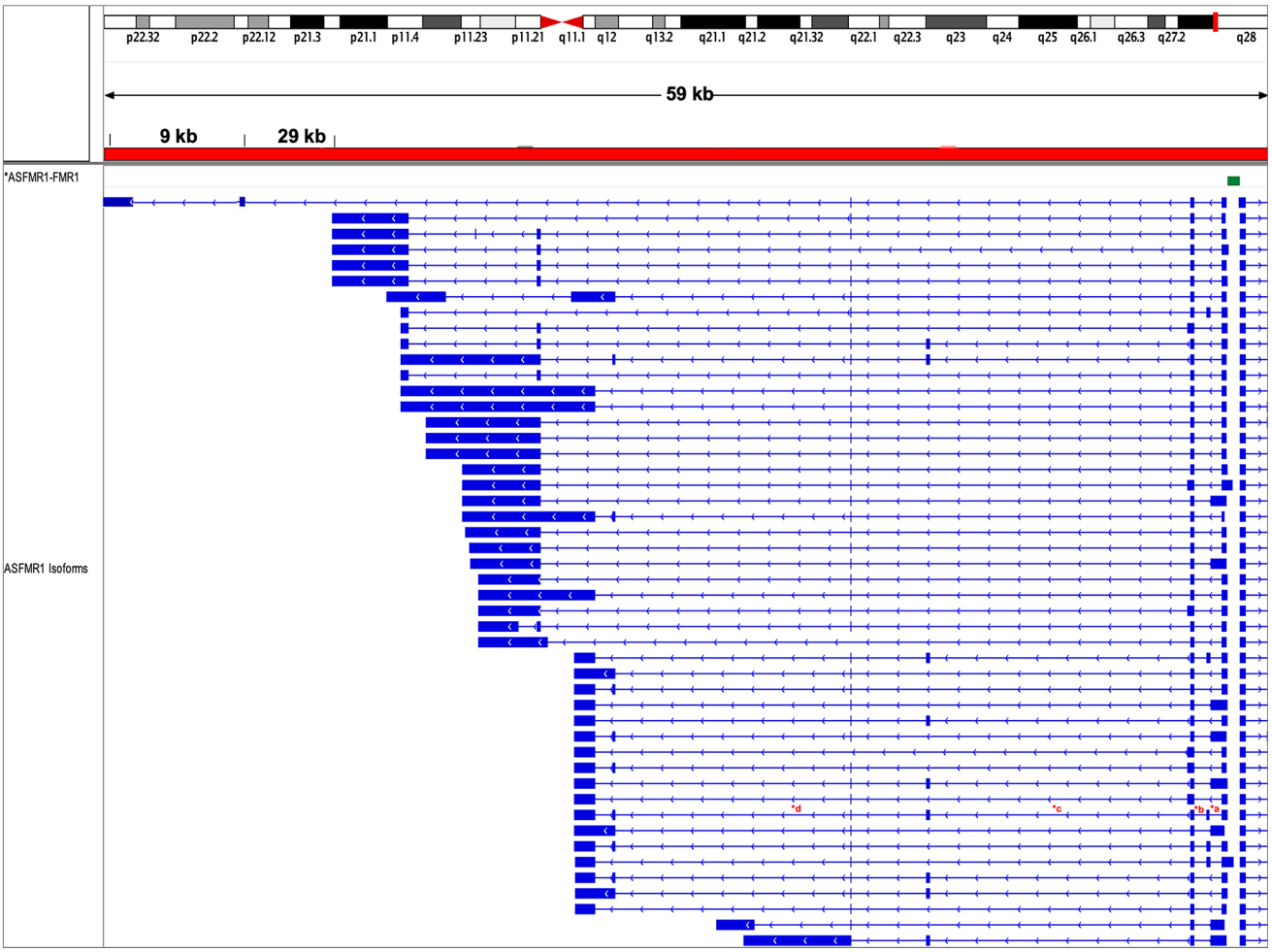


Fig. 1.

Schematic presentation of the study design. RNA was extracted from postmortem brain tissues of normal individuals and FXTAS patients. cDNA libraries were made using probe hybridization and sequenced on the PacBio Sequel system. Analysis was performed using PacBio software by following the bioinformatics pipeline started from the cleaning of raw sequence reads and end up at the final mapping of identified unique isoforms to the reference genome

a.



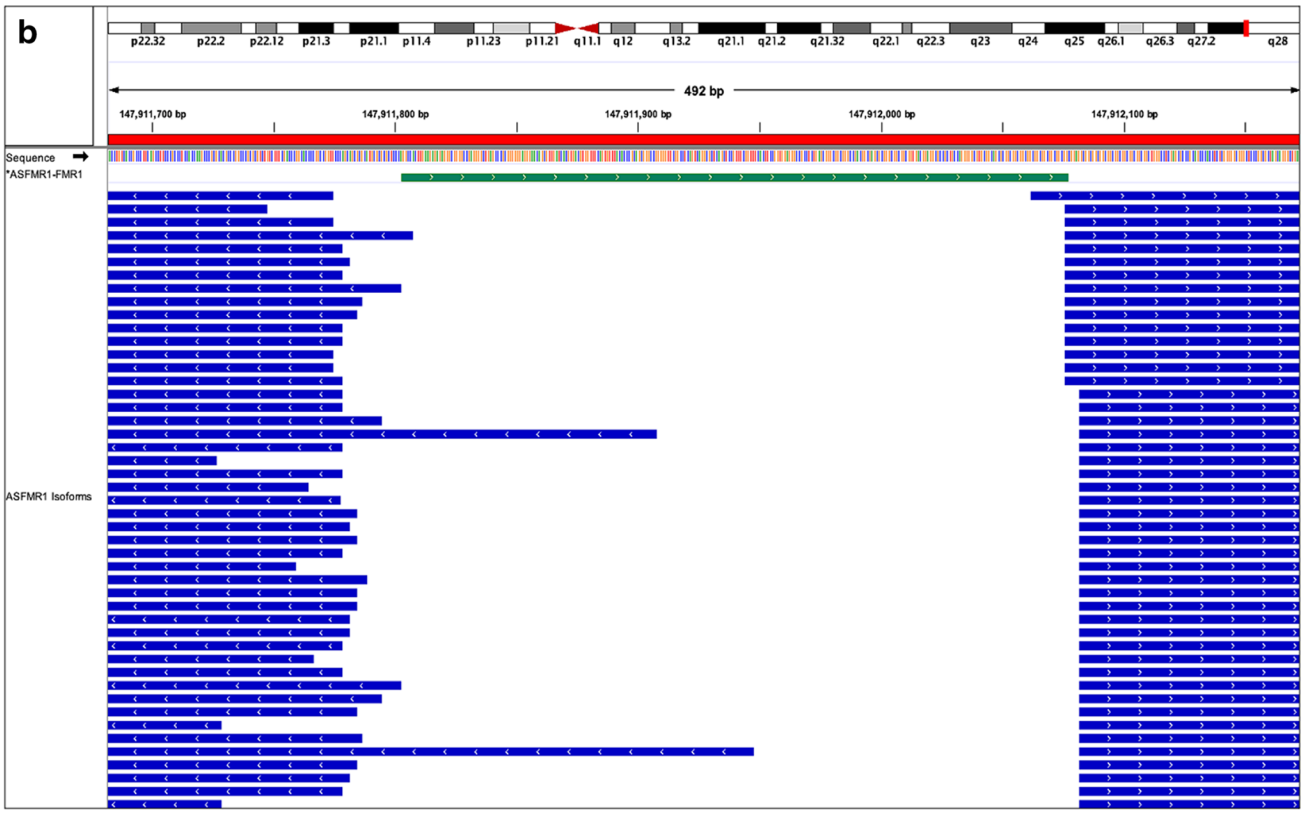


Fig. 2.

Characterization of the *ASFMR1* gene structure with the novel identified splice variants. Visualization of *ASFMR1* isoforms using Integrative Genomic Viewer (IGV) on chromosome X in human genome assembly GRCh38 (hg38). **a** The unique identified *ASFMR1* isoforms spanning 59 kb of genomic DNA are shown. The blue box represents the transcribed region for each given isoform and their size ranges between 532 and 6 kb. The green box represents the overlapping region between the *ASFMR1 and FMR1* genes (see zoom image in **b**). **b** A zoom image of the overlapping region of AS*FMR1-FMR1* genes obtained by Sanger sequencing is shown in green. The transcripts on the right side of the window correspond to the *FMR1* gene while the transcript on the left corresponds to the *ASFMR1* gene

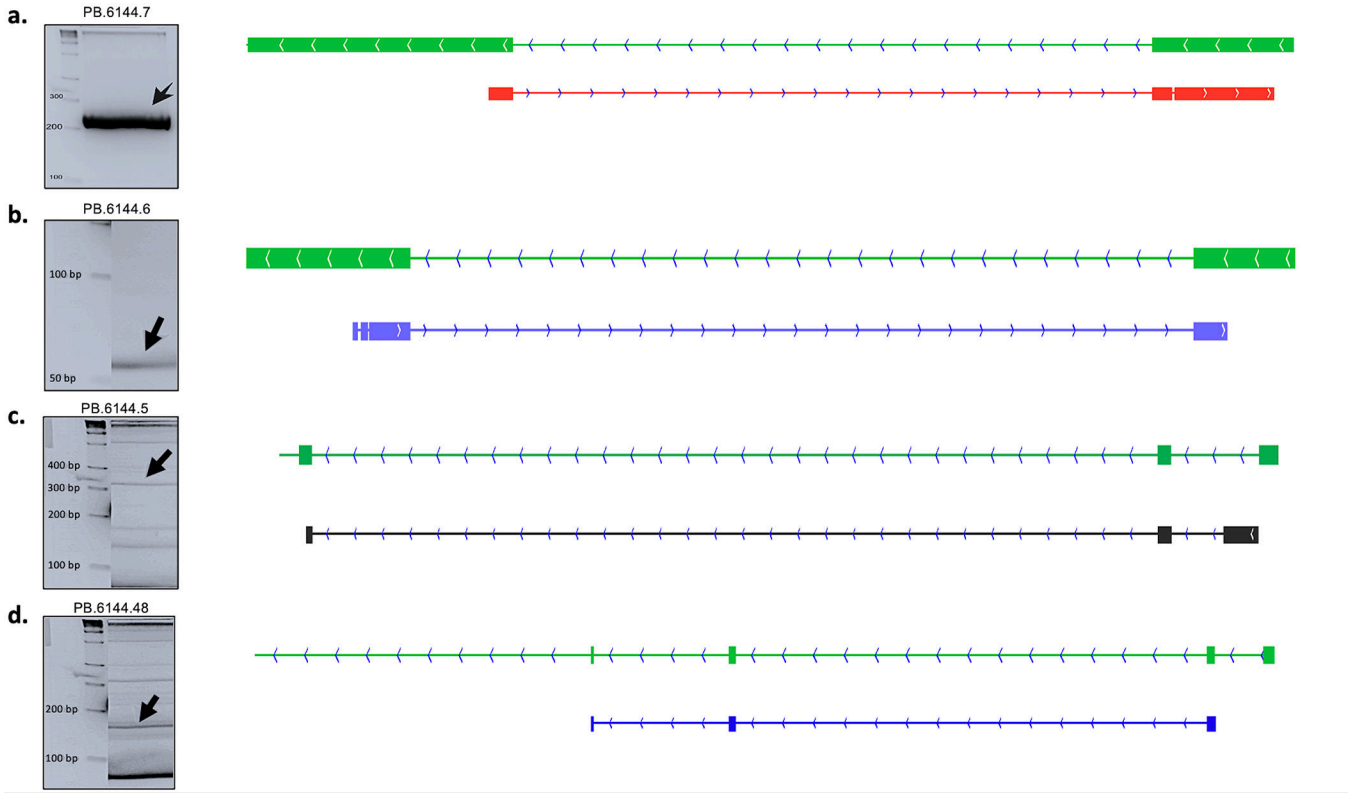


Fig. 3.

Validation of the *ASFMR1* mRNA isoforms by qRT-PCR. Validation of a number of novel splicing isoforms was performed by RT-PCR and Sanger sequencing. A subset of four splicing isoforms (**a**, **b**, **c**, **d**) is depicted in this figure. The left panel is representing the qRT-PCR product run on gel with the black arrow showing the band corresponding to the isoform under study, while the right panel is showing the Sanger sequencing blast result where the sequence color in green at top is representing the sequence from the genome browser and the one in the bottom (red, blue, and black) is the sequence from the Sanger sequencing. The location of the splicing events occurred in these isoforms is indicated in Fig. 2a (***a**, ***b**, ***c**, ***d**)

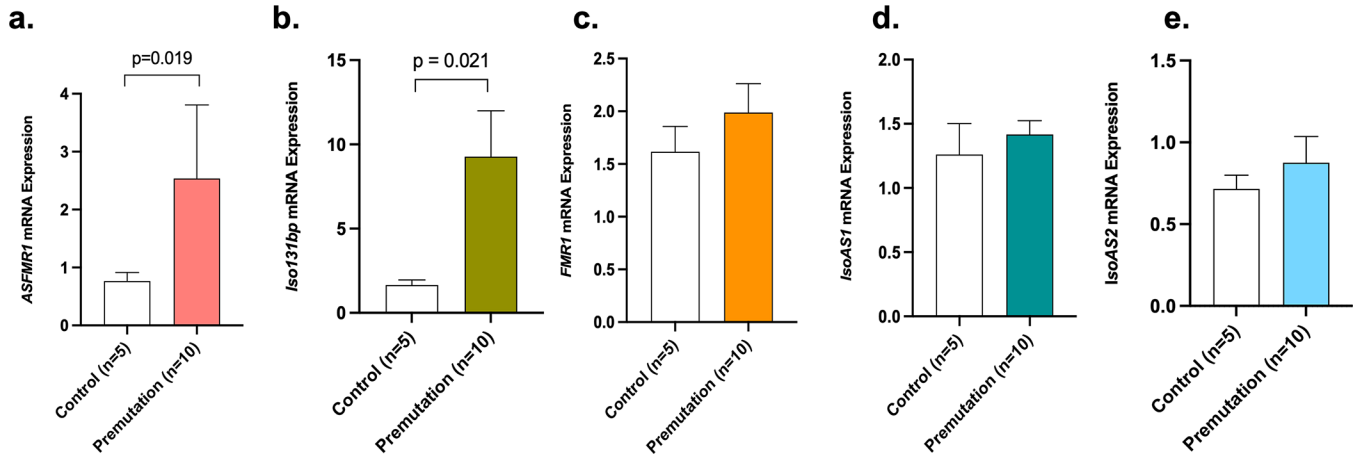


Fig. 4.

Expression levels of *ASFMR1* mRNA isoforms in PM and HC. Bar plots showing significant increased levels of *ASFMR1* mRNA ($p = 0.019$; (**a**) and, of the *Iso131bp* mRNA ($p = 0.021$; (**b**) only in the PM compared to HC. No significant expression between the two groups was observed for the *FMR1* gene (**c**) and for the two *ASFMR1* isoforms *IsoAS1* (**d**) and *IsoAS2* (**e**)

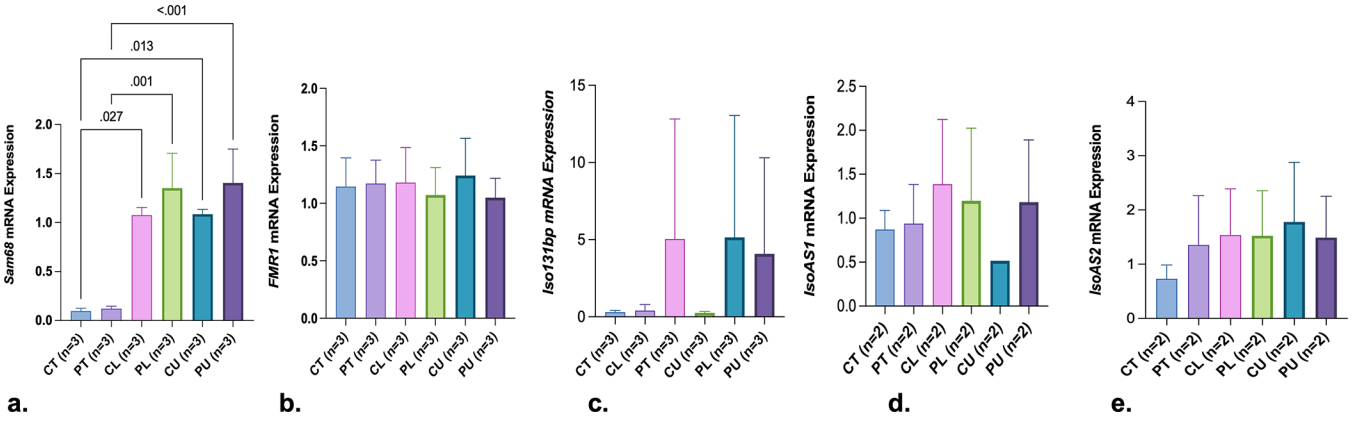


Fig. 5.

Sam68 does not affect the alternate splicing of *ASFMR1*. **a** Bar plots showing the knockdown of the Sam68 in fibroblast transfected cells from Healthy Controls (CT) and premutation (PT) compared to the lipofectamine treated in absence of siRNA [CL; PL] and compared to the untreated group in absence of both lipofectamine and siRNA [CU; PU]. No difference in the expression levels of *FMR1* gene (**b**), and of the *ASFMR1* isoforms including Iso*131bp* (**c**), *IsoAS1* (**d**) and *IsoAS2* (**e**) was observed between these groups of controls and premutation

Table 1:

Demographics of participants and sequence of specific barcoded oligos

Samples	CGG repeat	Gender	Category	AGE at Death	PMI	RIN	Sequence
N1	30	M	HC	61	24	8.2	AAGCAGTGGTATCAACGCAGAGTACCACATATCAGAGTGCGTTTTTTTTTTTTTTTTTTTTTTT
N2	20	M	HC	40	23	7.6	AAGCAGTGGTATCAACGCAGAGTACACACACAGACTGTGAGTTTTTTTTTTTTTTTTTTTTTTT
N3	41	M	HC	58	17	8.5	AAGCAGTGGTATCAACGCAGAGTACACACATCTCGTGAGAGTTTTTTTTTTTTTTTTTTTTTTT
P1	86	M	PM	79	7	8.2	AAGCAGTGGTATCAACGCAGAGTACCACGCACACACGCGCGTTTTTTTTTTTTTTTTTTTTTTT
P2	88	M	PM	55	NA	7.5	AAGCAGTGGTATCAACGCAGAGTACCACTCGACTCTCGCGTTTTTTTTTTTTTTTTTTTTTTTT
P3	87	M	PM	75	5	7.7	AAGCAGTGGTATCAACGCAGAGTACCATATATATCAGCTGTTTTTTTTTTTTTTTTTTTTTTTT

Table 2:

hg38 coordinates of *ASFMR1* exons

hg 38 Coordinates	chrX:Start	chrX:End	Size
exon 1	147896132	147896647	515
exon 2	147897050	147897131	81
exon 3	147902756	147902787	31
exon 4	147904576	147904671	95
exon 5	147910893	147910995	102
exon 6	147911281	147911349	68
exon 7	147911642	147911784	142

Table 3:

Reads from PacBio long read sequencing platform

Groups	Subjects	CCS	FLNC	FL %	# of unique transcripts	# of on-target unique transcripts	# of FL attributed to unique transcripts	# of FL attributed to on-target transcripts	On-Target FL%	Splice_site	Canonical
Premutation	3	577008	497240	86%	16105	135	286431	1891	0.7%	GTAG	canonical
Normal	3	541658	474711	88%	16394	164	280075	2728	1.0%	GTAG	canonical