ARTICLE | Article in Press

The *FMR1* isoform landscape reveals new spliced variants in women's blood and ovarian tissues

Marina Luz Ingravidi · Cecilia S. Fernández · Eloisa Caviglia · Liliana Dain · Laura Kamenetzky · Ianina Ferder

Received: 19 August 2026 / Accepted: 29 October 2026
© The Author(s) 2026

Abstract

FMR1 encodes fragile X messenger ribonucleoprotein 1 (FMRP), an RNA-binding protein involved in post-transcriptional regulation and highly expressed in ovarian tissue. Although extensive alternative splicing of *FMR1* has been described, the full-set transcript landscape in ovarian cells remains incompletely characterized. Here, we used Oxford Nanopore Technologies long-read amplicon sequencing to characterize full-length *FMR1* transcripts spanning exons 1–17 in human blood and ovarian granulosa cells. We identified 30 high-confidence isoforms accounting for 95% of amplicon-captured *FMR1* expression. The predominant transcripts lacked exon 12 and retained exon 14, consistent with previous reports. Long-read sequencing also resolved previously uncharacterized splicing events, including infrequent splice-site usage in exons 15 and 17, exon 11 skipping, and partial retention of introns 2 and 7. Several transcripts preserve the open reading frame, whereas others introduce premature termination codons and may undergo nonsense-mediated decay or produce truncated FMRP isoforms. Notably, the *FMR1* expression profile was conserved between blood and granulosa cells, indicating that blood may serve as an accessible proxy tissue for studying *FMR1* isoform expression. These findings expand the known *FMR1* transcript repertoire and demonstrate the value of long-read sequencing for resolving complex splicing, providing a basis for investigating these isoforms' role in ovarian function.

Keywords *FMR1* mRNA isoforms · Human granulosa cells · Female reproduction · Long-read transcript sequencing · Alternative splicing

Introduction

The Fragile X Messenger Ribonucleoprotein 1 (former Fragile X Mental Retardation 1) gene (*FMR1*) consists of 17 exons spanning 38 kb of the Xq27.3 chromosome and codes for the protein Synaptic Functional Regulator *FMR1* (formerly Fragile X Mental Retardation protein, FMRP)[1,2]. *FMR1* is implicated in several disorders, depending on the number of CGG repeats in its 5'UTR region. In the general population, normal alleles have 29–30 repeats[3,4], while expansions of the CGG repeats over the normal range are classified as permutation (PM) when the allele has 55–200 repeats or full mutation (FM), with expansions greater than 200 repeats[5]. The full

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.



https://doi.org/10.1038/s41598-026-78955-8

Published online: 10 November 2026

scientific reports

mutation is responsible for the Fragile X Syndrome, the most commonly inherited cause of intellectual disability and the primary monogenic cause of autism[6], while the premutation is associated with three other *FMR1*-associated disorders: Fragile X-associated Tremor/Ataxia Syndrome (FXTAS), Fragile X-associated Primary Ovarian Insufficiency (FXPOI) and the Fragile X-associated neuropsychiatric disorders (FXAND)[5,7].

Alternative splicing of *FMR1* generates multiple isoforms, many of which encode distinct protein isoforms. The isoforms are generated mainly through exon skipping of exons 12 and 14 and from the use of alternative splice sites in exons 15 and 17[2,8,9]. Exclusion of exon 12 produces an *FMR1* isoform with a shorter KH2 RNA-binding domain, whereas exclusion of exon 14 removes the nuclear translocation signal and introduces a frameshift leading to a premature termination codon (PTC). Moreover, three and two alternative splice sites have been described for exons 15 and 17, respectively. The usage of the second acceptor site in exon 15 eliminates an important phosphorylation site, while alternative splicing in the third acceptor site retains this deletion and also removes amino acids involved in the methylation of FMRP[10,11]. The functional significance of the splice acceptor site within exon 17, which results in the loss of 17 amino acids, remains unclear, although this region has been proposed to be important for the localization of nuclear FMRP isoforms within Cajal bodies[12]. Splice variants lacking exons 3 and 11 have also been described[9,13], although the implications of these splicing events are still unknown. Finally, Yang et al. reported the inclusion of a 140-bp sequence from intron 9 as an alternative exon in human peripheral blood[14].

The advent of long-read high-throughput sequencing technologies led to the identification of 49 additional isoforms in various human tissues, including muscle, brain, heart, and testes, of which 30 appeared to be expressed only in premutation carriers[9,13]. However, these studies were limited to analyzing alternative splicing only between exons 9 and 17. Therefore, to our knowledge, no studies to date have described the expression of full-length *FMR1* isoforms in any tissue or species.

Among the disorders associated with the *FMR1* premutation, our research focuses on Fragile X-associated primary ovarian insufficiency (FXPOI), with particular emphasis on characterizing *FMR1* splicing isoforms in the ovary. FXPOI is characterized by impaired ovarian function, reduced fertility, and the onset of primary ovarian insufficiency before the age of 40. Approximately 20% of women carrying the *FMR1* premutation develop FXPOI, representing an approximately 20-fold higher risk than that observed in the general population[15,16]. Despite its clinical impact, FXPOI remains the least studied disorder associated with the *FMR1* repeated expansion.

Previously, we reported the expression of multiple *Fmr1* splicing isoforms during rat folliculogenesis, identified through standard RT-PCR amplification and Sanger sequencing of regions containing alternative splicing sites[17]. In the present work, we characterized the repertoire of *FMR1* variants within the amplicon-defined transcript window spanning exons 1 to 17, in human blood and ovarian granulosa cells using long-read sequencing. We aim to evaluate whether isoform expression in blood is representative of that in the ovary, specifically in granulosa cells, as blood tissue is much more accessible and easier to obtain than ovarian tissue. Granulosa cells play a central role in folliculogenesis and ovarian function and represent one of the most relevant cell types for investigating the molecular mechanisms underlying Fragile X-associated primary ovarian insufficiency (FXPOI). However, the diversity and full set of *FMR1* isoforms expressed in these cells has not previously been investigated.

Long Read sequencing, such as Oxford Nanopore Technologies (ONT), can generate reads from cDNA over 30 kb[18] enabling the sequencing of entire transcript isoforms in a single read and allowing comprehensive analysis of isoform expression across different tissues. In particular, amplicon sequencing, a targeted approach, involves PCR amplification of specific genomic or transcript regions followed by direct sequencing, allowing high-resolution characterization of isoform diversity. With this in mind, we employed ONT amplicon sequencing to characterize *FMR1* isoforms in blood and ovarian granulosa cells obtained from oocyte donor women. We identified 30 high-confidence isoforms accounting for 95% of amplicon-captured *FMR1* expression. The expression profile was conserved between blood and granulosa cells, indicating that blood may serve as an accessible proxy tissue for studying *FMR1* isoform expression.

Materials and methods

Ethical approval

This study was approved by the CEGYR Research Ethics Committee of the Reproductive Medicine and Genetics Clinic CEGYR, Buenos Aires, Argentina (protocol #6.6.25, Rev. 1). All procedures performed in this study were in accordance with the 1964 Declaration of Helsinki and its later amendments, or comparable ethical standards. Written informed consent was obtained from all participants recruited for this study.

Subjects

Peripheral blood and follicular fluid samples were obtained from five women (aged 25–32) enrolled in the oocyte donation program at the Fertility Centre CEGYR (Buenos Aires, Argentina). All participants had CGG repeat length in the *FMR1* gene within the normal range.

Two peripheral blood samples were obtained from each participant, one before (blood 1, B1) and one after ovarian hyperstimulation (blood 2, B2). The inclusion of the post-stimulation blood sample (B2) was motivated by the fact that granulosa cells were collected following the hyperstimulation protocol, allowing us to assess the potential impact of hormonal treatment on *FMR1* isoform expression. The hyperstimulation process consisted of the administration of corifollitropin alfa (Elonva 150 µg) for 7 days, followed by follitropin alfa (GONAL-F 150 IU) + MENOPUR 150 IU + triptorelin (2 GONAPEPTYL) + Progesterone 200 mg. Remnant follicular fluid was collected at the time of oocyte retrieval in sterile 50 mL tubes.

Leukocyte purification from peripheral blood samples

A total of 5–6 mL of peripheral blood was transferred to a 50 mL conical tube, and RBC lysis buffer was added to a final volume of 45 mL. After incubation for 10 min at room temperature, the sample was centrifuged at 600 × g for 10 min to pellet the cells. The supernatant was discarded, and the pellet was resuspended in 1 mL of RBC lysis buffer and transferred to a 1.5 mL microcentrifuge tube. Following a 5 min incubation at room temperature, the cells were centrifuged at 600 x g for 2 min. The supernatant was removed, and the pellet was washed once with sterile PBS (600 x g 2 min). Finally, the cells were resuspended in 700 µL of TRIzol reagent and stored at −80 °C until RNA extraction.

Isolation of granulosa cells from follicular fluid

A total of 50–100 mL of follicular fluid was centrifuged at 400 × g for 10 min. The resulting pellet, containing granulosa cells (GC), was resuspended in 1 mL of sterile DMEM: F12 medium supplemented with antibiotics and centrifuged again at 180 × g for 5 min. The supernatant was discarded, and the pellet was resuspended in 1 mL of DMEM: F12. The cell suspension was carefully layered over 3 mL of LymphoSep separation medium (MP Biomedicals) in a 15 mL conical tube[19]. After centrifugation at 600 × g for 30 min, GCs were collected from the interface, resuspended in 5 mL of sterile RBC lysis buffer, and incubated for 5 min at room temperature. The cells were then centrifuged at 180 × g for 5 min, and the resulting pellet was resuspended in 1 mL of DMEM: F12 medium. The GCs were plated in culture dishes containing DMEM: F12 + 10% FBS, where GCs adhered to the surface, while lymphocytes remained non-adherent. After 48 h, non-adherent cells were discarded by removing the culture medium and washing with PBS. The attached GCs were collected using TRIzol reagent and stored at −80 °C until RNA extraction. All follicular fluid samples were processed immediately after collection on the same day.

RNA isolation

Total RNA was extracted according to the manufacturer's instructions from TRIzol-stored leukocytes purified from peripheral blood samples and from the GCs isolated from the follicular fluid as described in the previous section. RNA integrity was evaluated by agarose gel electrophoresis and quantification was performed using a NanoDrop spectrophotometer (Thermo Fisher Scientific).

RT-PCR

Complementary DNA (cDNA) was synthesized from 1 µg of total RNA using oligo(dT) primers and Maxima H Minus Reverse Transcriptase (Thermo Fisher Scientific) in a final reaction volume of 20 µL. Technical replicates were generated for two samples (B2_E and GC_E) by splitting each RNA sample into two aliquots, each of which underwent independent cDNA synthesis and library preparation (Ea and Eb, respectively). To amplify all potential *FMR1* transcript isoforms spanning exon 1 through 17, we used a forward primer 5′-GAAGATGGAGGAGC TGGTGG-3 located in exon 1 and a reverse primer 5′-ACATTTAAGCAAGTTAGCGCCT-3' located in exon 17; (see Supplementary Fig. 1 for the reference sequence ENST00000370475 and primers positions). PCR amplification was performed with the Expand Long Template PCR System (Roche) and the following cycling conditions: initial denaturation at 94 °C for 2 min; 10 cycles of 94 °C denaturation for 10 s, 60 °C annealing for 30 s, 68 °C elongation for 5 min; followed by 25 cycles of 94 °C denaturation for 15 s, 60 °C annealing for 30 s, 68 °C elongation for 5 min adding 20 s elongation for each successive cycle and a final elongation of 7 min at 68 °C.

Library preparation for ONT amplicon-sequencing

A total of 15 biological samples from five different women were used for library preparation. Each sample corresponded to *FMR1* amplicons generated by PCR using cDNA synthesized from GC and peripheral blood samples. To serve as technical controls, an additional cDNA synthesis reaction was performed for two of the samples (one B2 and one GC), resulting in a total of 17 samples included in the sequencing run.

After amplification, each sample was purified using 1X Agencourt AMPure XP beads (Beckman Coulter) and quantified using the Qubit dsDNA High Sensitivity Assay (Thermo Fisher Scientific). Sample purity was evaluated using a NanoDrop spectrophotometer™ (Thermo Fisher Scientific). Library preparation, multiplexing, and sequencing were performed using the Native Barcoding Kit 24 V14 (SQK-NBD114.24, Oxford Nanopore Technologies, UK) following the manufacturer's recommendations for amplicon sequencing. Briefly, amplicon DNA was subjected to DNA end-repair and dA-tailing using the NEBNext Ultra II End Repair/dA-Tailing Module (New England Biolabs). For samples with lower PCR yields, the maximum available mass was utilized (ranging from 27.5 to 379.8 ng). Native barcodes were then ligated to each sample. Following barcode ligation, the 17 samples were pooled and sequencing adapters were subsequently ligated to the barcoded samples. The final library was quantified, and 50 fmol was loaded onto a PromethION R10.4.1 Flow Cell (Oxford Nanopore Technologies) and sequenced on a PromethION P2 Solo device for approximately 25 h using MinKNOW software v.24.06.10. Basecalling (super-accurate basecalling model, SUP; minimum Q-score = 10), demultiplexing, and barcode and adapter trimming were performed post-run using Dorado (v.0.9.6).

FMR1 isoform identification

To assess sequencing quality, read length distributions were visualized using NanoComp (v.1.24.0). No minimum read length threshold was applied for downstream analysis; reads of all lengths were retained to avoid discarding possible genuine short isoforms. Quality filtering was limited to the minimum Q-score threshold of 10 applied during basecalling (Dorado v.0.9.6). SeqKit (v2.6.1) was used to generate summary statistics of the read sets. Retained reads were aligned to the human reference genome (GRCh38.p14, Ensembl release 115) using

Minimap2 (v2.24-r1122) with splice-aware parameters. The resulting coordinate-sorted SAM files were converted to BAM format, sorted, and indexed using SAMtools (v1.18), followed by conversion to BED12 format with BEDtools (v.2.31.1). Primary genome alignment processing, isoform identification, and quantification were performed using FLAIR (v2.0.0)[20], enabling the detection of splicing combinations resulting in new isoforms - including unannotated exon skipping and alternative exon usage events - alongside canonical transcripts. Reads were first corrected against the reference annotation using the FLAIR correct module, which takes as input the BED12-formatted aligned reads and the Ensembl GTF annotation and adjusts the genomic coordinates of splice sites in each read to match the nearest annotated splice site within a 15 bp window; reads with splice sites that cannot be reconciled with the annotation are excluded[20]. Corrected reads from individual samples were then assembled into high-confidence transcript models and quantified using FLAIR collapse (v2.0.0). This step incorporated raw long-reads in FASTQ format to evaluate isoform support, applying --stringent and --check_splice criteria to ensure high-accuracy splice site definition. Finally, high-confidence isoforms identified across individual barcodes were merged into a unified non-redundant transcriptome using an updated version of the software FLAIR (v2.2.0), specifically the combine module, to collapse transcript models with identical exon-junction chains into a single representative isoform per dataset.

Reproducibility and ROC curve analysis

To evaluate the consistency of isoform detection between technical replicates, a Receiver Operating Characteristic (ROC) curve analysis was performed using the pROC package in R. For this analysis, the "presence" of an isoform within a sample was defined using a heuristic minimum abundance threshold of >5 reads. For datasets with available technical replicates, one replicate (Ea) was established as the baseline reference by applying this presence threshold to binarize the data into present (1) or absent (0) states. The abundance scores from the corresponding second technical replicate (Eb) were then utilized as the predictor variable. The Area Under the Curve (AUC) was calculated to quantify the predictive accuracy and validate the reproducibility of isoform detection between technical replicates. To complement the analysis, Spearman rank correlation was additionally performed on CPM-normalized isoform abundance values across all 123 high-confidence isoforms, using the cor.test function in R (method = "spearman"). The number of isoforms detected (CPM>0) in each replicate individually and in both replicates simultaneously was also recorded as an additional measure of detection consistency.

FMR1 isoform high-quality filtering design

To retain robust and biologically relevant isoforms, a multi-step filtering strategy was implemented. First, to validate the structural continuity of the identified transcript models, we assessed whether reads contained the complete sequence of the expected amplicon (spanning exons 1 and 17) for each isoform. A read was classified as "complete" if its spliced alignment blocks - after merging gaps of ≤ 10 bp attributable to ONT-derived indels - completely covered (100%) all internal exons of the assigned isoform, as defined by their genomic coordinates in the combined GTF annotation and the PCR amplicon design. Partial coverage of the terminal exons was permitted to account for primer binding site positioning inherent to the PCR amplification strategy (see Supplementary Fig. 1 for primers coordinates). Second, to ensure expression reliability and reproducibility, each isoform was required to have more than five assigned reads and be detected in at least $n-1$ samples (barcodes) within its respective tissue group. Additionally, for isoforms in the B2 and GC groups, this minimum abundance threshold (>5 reads) had to be met in both of their respective technical replicates. To identify potential technical biases, a principal component analysis (PCA) of the initially filtered dataset was performed using variance-stabilized transformed (VST) counts. Unsupervised clustering was conducted using the K-means algorithm, specifying $k=3$ centers corresponding to the three biological groups (B1, B2, and GC). Following the identification and exclusion of outlier samples based on the 95% confidence ellipses, the $n-1$ presence criteria were recalculated and reapplied to the remaining dataset to obtain the final list of high-confidence isoforms.

Differential expression analysis

Differential expression analysis was performed at the isoform level using the limma-voom framework, with duplicateCorrelation function which allows paired designed[21–23]. The isoform-level count matrix, containing the previously filtered high-confidence isoforms was normalized using trimmed mean of M-values (TMM). Pairwise comparisons were conducted for B1 versus GC, B2 versus GC, and B1 versus B2. P-values were adjusted for multiple testing using the Benjamini-Hochberg procedure to control the false discovery rate. Differentially expressed isoforms (DEIs) were defined as those with adjusted p-value < 0.05 and $|\log2FC| > 1$.

FMR1 isoform expression profile analysis

To assess putative differences in the number of isoforms retained in the 95% abundance thresholds of the different tissues analyzed (e.g. a difference in expression profile vs. a technical artifact of the threshold criterion), we performed a Spearman rank correlation analysis across all 123 high-confidence isoforms. CPM-normalized expression values were averaged across samples within each tissue group (B1, B2, and GC), and isoforms were ranked by decreasing abundance within each tissue. Spearman's rank correlation coefficients were calculated between tissue pairs using the cor.test function in R (method = "spearman"). Additionally, the cumulative abundance distribution of all 123 isoforms was examined per tissue to identify isoforms present in the top 95% of one tissue but not another, and their abundance rank in GC was recorded.

Protein sequence prediction from transcript isoform sequences

Open reading frames (ORFs) and corresponding protein sequences for the identified transcript isoforms were predicted using TransDecoder2 (v1.1.0). To ensure biological relevance and accurate reading frame assignment, downstream analysis was restricted to predicted ORFs initiating from the canonical *FMR1* start codon (ATG). To determine the structural impact of alternative splicing, the predicted amino acid sequences were aligned against the canonical FMRP protein (UniProt Q06787), encoded by ENST00000370475 using ClustalW. This alignment allowed for the precise identification of structural variations, including frameshifts and the introduction of premature termination codons (PTCs).

Following sequence alignment, the susceptibility of each isoform to nonsense-mediated mRNA decay (NMD) was evaluated by applying established canonical rulesets[24–26]. Only transcript isoforms that were not previously described and were predicted to escape NMD - and therefore capable of producing stable, modified, or truncated proteins - were selected for further potential functional characterization.

Conserved protein domains within the NMD-escaping predicted isoforms were systematically annotated using the InterProScan web service (https://www.ebi.ac.uk/interpro/search/sequence/), with domain identification explicitly based on the Pfam database. Furthermore, specific functional regions of interest were manually curated and mapped onto the sequences according to previously published literature (NLS:[27], NES:[28], RGG:[29], NoLS:[30]). This comprehensive annotation strategy provided the structural framework for the final visual representation of the FMRP protein variants.

Results

In this study we employed ONT long-read sequencing to characterize the expression of *FMR1* isoforms spanning exons 1 to 17 from blood and ovarian granulosa cells obtained from healthy women. Figure 1 summarizes the applied study workflow. Following cDNA amplicon sequencing, a total of 48 gigabases and 14.6 million reads were generated across 17 samples, with a mean read length of 3,638 bp and a mean N50 of 3,948 bp, consistent with the expected amplicon sizes. After basecalling quality filtering (minimum Q-score = 10), an average

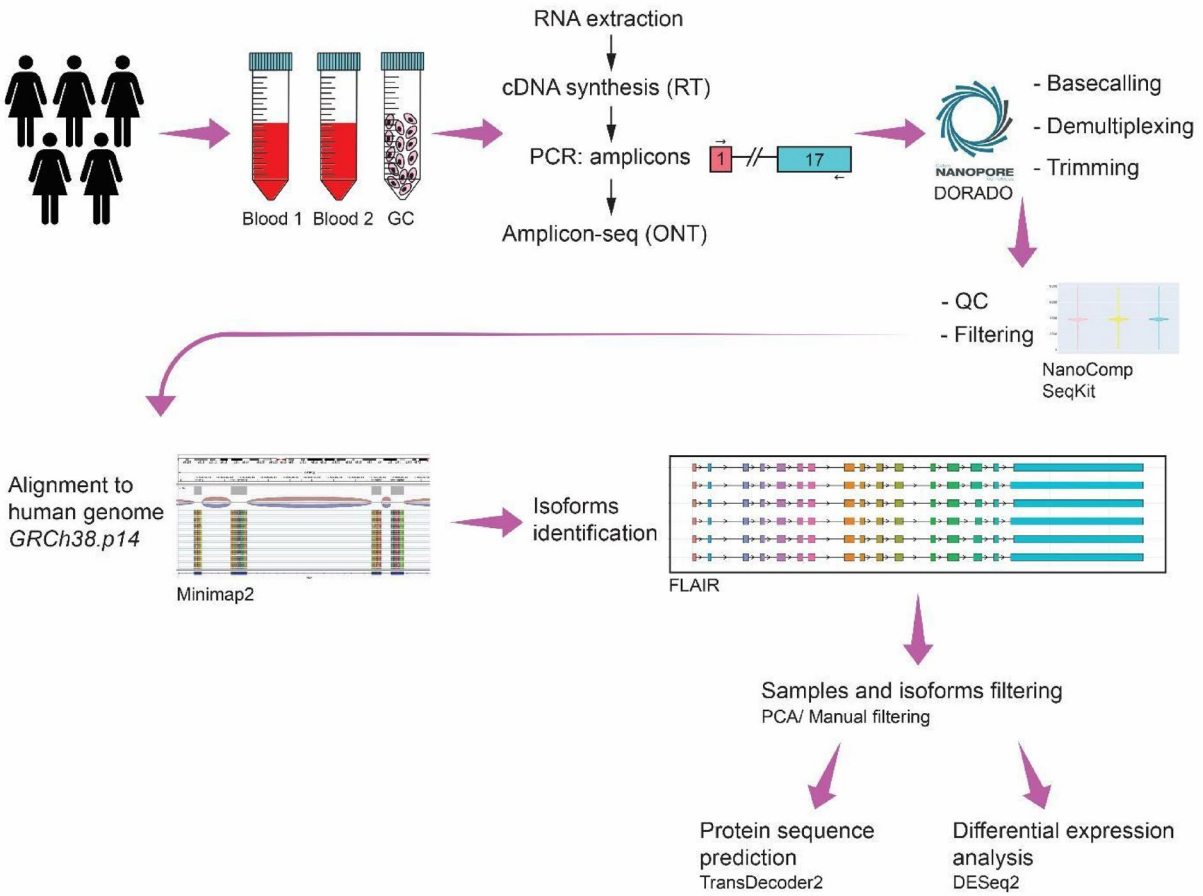


Fig. 1 Experimental workflow diagram. Two blood samples were collected from women enrolled in an oocyte donation program: one before hyperstimulation treatment (Blood 1, B1; $n=5$) and the second post-stimulation (Blood 2, B2; $n=5$). Additionally, remnant ovarian granulosa cells (GCs) were isolated from follicular fluid obtained during oocyte retrieval (GCs; $n=5$). For one participant (woman E), technical replicates were generated for both B2 and GC samples by splitting the RNA into two independent library preparations (B2_Ea/B2_Eb and GC_Ea/GC_Eb), resulting in an augmented n for these tissues (B2 = 6, GC = 6). RNA was extracted from all samples, reverse transcribed, and the complete *FMR1* cDNA was PCR-amplified and sequenced following ONT protocols. Three samples were excluded as outliers following PCA-based quality control (final barcodes retained: B1 = 5, B2 = 5, GC = 4). Various bioinformatic tools were used for quality control, alignment, isoform identification, principal component analysis (PCA), differential expression analysis and mRNA translation.

of 481,603 reads per sample was retained for downstream analysis (Supplementary Table 1). Alignment to the human reference genome (GRCh38.p14, Ensembl release 115) yielded a 99.99% mapping rate, with approximately 86% of these reads assigned to an *FMR1* isoform. Principal component analysis (PCA) identified three outlier samples (barcodes 02, 03, and 09), which were excluded from downstream analysis (Supplementary Fig. 2, Panel C). For the remaining samples, *FMR1* mapping efficiency was robust (80–95%, Supplementary Fig. 2, Panel D), and technical replicates demonstrated high consistency in isoform detection (ROC AUC of 0.89 and 0.83 for GC and B2, respectively; Supplementary Fig. 2, Panel A). To further characterize reproducibility, Spearman rank correlation between technical replicates showed strong concordance in isoform abundance profiles for both B2 (rho = 0.767; $p<0.001$) and GC (rho = 0.826; $p<0.001$). At the detection level, 110 of 123 isoforms were consistently detected in both B2 replicates (115 and 116 isoforms respectively), and 95 of 123 in both GC replicates (97 and 102 isoforms respectively), confirming that isoform detection and quantification were highly consistent across independent library preparations from the same biological sample (Supplementary Fig. 2 panel B).

After filtering, a total of 123 high-confidence *FMR1* isoforms were identified (Supplementary Table 2). These transcript models showed strong structural support, with an average of 78% of assigned reads classified as complete indicating full length coverage of the long transcript amplicon analyzed in this work (see Supplementary Fig. 1 for primer design and coordinates). In addition, to further evaluate whether expression profiles differed between tissues, Spearman rank correlation was calculated across all 123 high-confidence isoforms using CPM-normalized values. Isoform abundance rankings were strongly correlated between blood and GC (B1 vs. GC: rho = 0.83; B2 vs. GC: rho = 0.77; both $p < 0.001$), suggesting broad conservation of the *FMR1* isoform expression profile across tissues within the amplicon transcript analyzed in this work. The complete abundance distribution across all 123 isoforms per tissue is shown in Supplementary Table 5.

To focus on the most biologically relevant transcripts, we restricted our analysis to the subset of isoforms that cumulatively accounted for 95% of the *FMR1* amplicons abundance across all three tissues (Fig. 2). This criterion retained 26 isoforms in blood samples (B1 and B2) and 17 in granulosa cells (GC), yielding 30 distinct isoforms overall. The 30 isoform-set has a mean of 93% of the reads with 100% coverage (Supplementary Table 8). Although fewer isoforms were retained within the 95% abundance threshold in GC (17 versus 26 in B1 and B2), this difference does not appear to reflect an absence of those transcripts in granulosa cells or a difference in sequencing depth. Mean per-sample raw read counts and reads mapped to *FMR1* were comparable across tissue groups after removing the outliers (raw reads per tissue: GC: ~601,000; B1: ~543,000; B2: ~588,000; reads mapped to *FMR1* per tissue: GC: ~543,000; B1: ~488,000; B2: ~542,000), ruling out sequencing depth as a confounding factor.

Of the 13 isoforms present in the top 95% of blood but not GC, all were detectable in GC with CPM > 0: five ranked immediately below the 95% threshold (positions 18–24, mean CPM 1,568–3,448) and the remaining eight between positions 31–79. Within this highly expressed subset, structural confidence was even greater, with full-length read support averaging 93.9% (range: 86.4–96.2%; Fig. 3).

Among these 30 transcripts, the 10 most highly expressed isoforms were consistently dominant across all samples, despite minor variations in relative abundance (Fig. 2). The two predominant isoforms in every sample were isoform 1 (ENST00000218200) and isoform 2 (ENST00000687593), which together accounted for approximately 40% of total *FMR1* isoform expression. Both isoforms lack exon 12, include exon 14, and utilize the first splice acceptor site in exon 15; they differ only in their use of alternative acceptor sites in exon 17. The next two most highly expressed isoforms were isoform 6 (ENST00000440235) and isoform 5 (ENST00000691214), collectively accounting for 12–15% of total expression. Both lack exon 12, include exon 14, and use the third acceptor site in exon 15, differing only in their splice site selection in exon 17. The following four isoforms - isoform 9 (ENST00000370477), isoform 18 (ENST00000691111), isoform 7 (ENST00000690137), and isoform 13 (ENST00000370475) - together contributed 13–14.5% of total isoform expression. All four include exon 14; however, isoforms 7 and 13 additionally include exon 12. These isoforms further differ in their use of splice sites in exons 15 and 17. Finally, the two least abundant isoforms among the top ten, each representing ~ 1.5-2% of total expression, varied slightly between samples and included isoforms 3, 12 (ENST00000370470), 19, 88, and 108 (ENST00000943062) (Fig. 2A). Isoforms 3, 19, and 88 have not been previously annotated and therefore lack corresponding Ensembl transcript (ENST) identifiers. Notably, isoforms 88, 108 and 167 harbor an infrequent splice acceptor site in exon 15 (Supplementary Fig. 4) representing a fourth acceptor site located 15 base pairs downstream of the canonical first site, in addition to the three widely described ones[9,17,31]. The remaining isoforms within the 95% expression threshold (Figs. 2 and 3) included several previously unannotated isoforms: 4, 8, 10, 20, 23, 26, 32, 36, 58, 82, 86, and 167. Isoforms 8, 14 (ENST00000685491), 36, and 86 exhibit an infrequent splice acceptor site in exon 17 that results in the exclusion of 5 base pairs downstream of the canonical first acceptor site (Supplementary Fig. 4). This alternative site represents a third acceptor site, in addition to the two widely described[9,12,17]. Isoforms 82 and 58 contain a 137-bp sequence between exons 2 and 3 consistent with partial intron retention, while isoform 32 includes a 92-bp fragment derived from the intron between exons 7 and 8. Finally, isoform 15 (ENST00000943063) is the only isoform among the 30 identified that lacks exon 11. This isoform has been previously reported[13], although the functional consequences of this splicing event remain unclear.

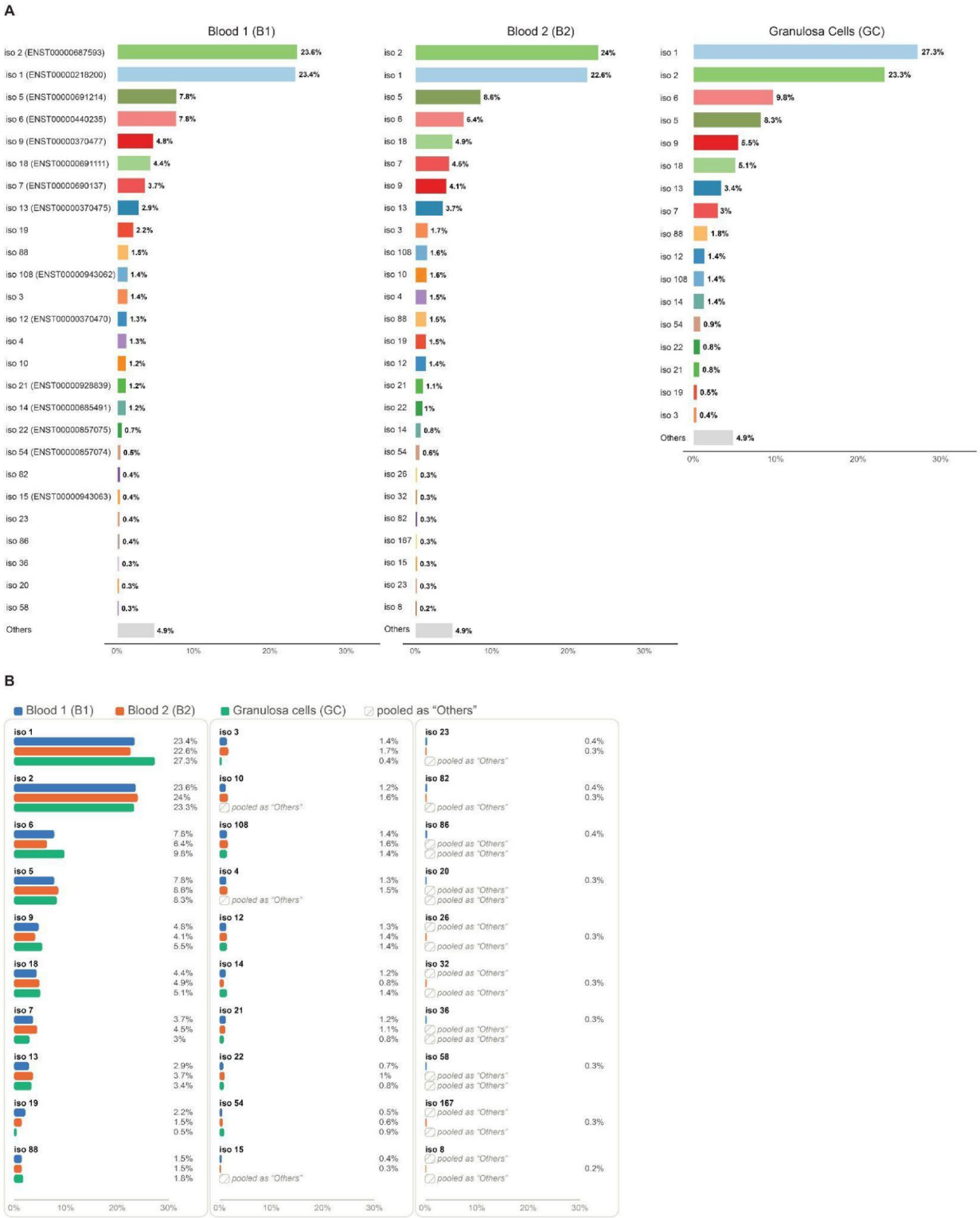


Fig. 2 Distribution of *FMR1* transcript isoforms in blood and granulosa cells. **(A)** The relative contribution of each *FMR1* isoform to total *FMR1* expression is shown for Blood 1 (B1), Blood 2 (B2), and granulosa cells (GC). Values represent the percentage of total *FMR1* reads assigned to each isoform. **(B)** The same data are displayed as a grouped bar chart to allow direct comparison of B1, B2, and GC values for each isoform. Isoforms comprising the 95% of the amplicon-captured expression are displayed individually; all remaining low-abundance isoforms are pooled as "Others". Raw and normalized counts are shown in Supplementary Tables 2 and 3, respectively.

https://doi.org/10.1038/s41598-026-78955-8

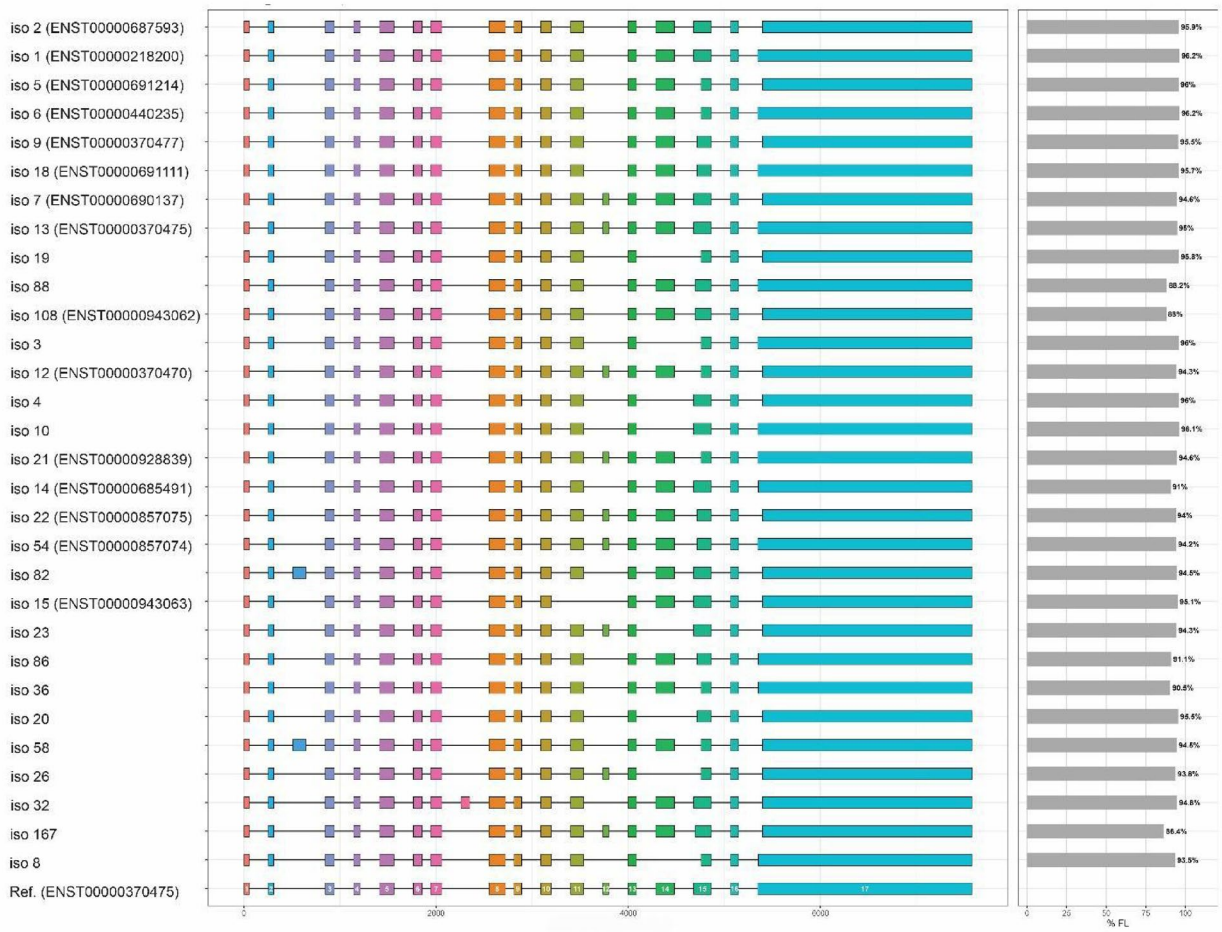


Fig. 3 Structure of the 30 isoforms accounting for 95% of amplicon-captured expression across blood and granulosa cells. Left panel: Isoforms accounting for the 95% of amplicon-captured expression were combined into a non-redundant set of 30 transcripts. Each row corresponds to a distinct isoform, with exons represented by colored boxes and introns by connecting lines. Differences in exon length reflect the use of alternative splice sites within exons. The canonical reference transcript (ENST00000370475) is included for comparison. Right panel: Bar plot displaying the percentage of full-length (FL) or "complete" reads spanning exons 1 to 17 supporting each corresponding transcript.

In addition, we performed differential expression analysis on the high confidence isoforms. We compared isoform expression across samples and found that there was no differential expression in all 123 isoforms (adj p-value > 0.05, |log2FC| < 1) (Supplementary Fig. 3, Supplementary Table 6). Overall, these results indicate that *FMR1* isoform expression is conserved among the tissue types studied with no significant changes observed (Fig. 4).

To further investigate the potential functional implications of the sequence variation observed in the newly described *FMR1* isoforms, we analyzed the putative amino acid sequences encoded by each isoform (Fig. 5). After in silico translation and protein domain analysis, we observed that splicing-induced frameshifts may lead to the recurrent loss or truncation of C-terminal of three regulatory domains: nuclear export signals (NES); RNA-binding domain (RGG); Nucleolar localization signal (NoLS) in isoform 3, 8, 19 and 26 (Supplementary Table 7). Furthermore, premature stop codons indicate that 7 isoforms (isoforms 4, 10, 20, 23, 32, 58 and 82) could be degraded by NMD process.



Fig. 4 Widely and infrequently used alternative splicing acceptor sites in *FMR1* exons 15 and 17. Exons are represented by colored boxes and introns by connecting lines. Canonical splice acceptor sites are shown in black and infrequent sites in orange. Numbers indicate the different splice sites. The length and sequence of the regions excluded by the use of alternative splice sites are indicated below each diagram.

Discussion

The *FMR1* gene encodes fragile X messenger ribonucleoprotein (FMRP), a multidomain RNA-binding protein that contains an N-terminal region involved in protein–protein interactions and three KH domains that contribute to the regulation of mRNA translation[32,33]. FMRP shuttles between the nucleus and cytoplasm through its nuclear localization signal (NLS) and nuclear export signal (NES), although it is predominantly cytoplasmic. Its activity is further regulated by post-translational modifications, including phosphorylation and arginine methylation, which influence its stability, ubiquitination and RNA-binding properties[34]. *FMR1* expression has been documented in ovarian tissue in both humans and rodents, with expression detected in oocytes, granulosa cells and theca cells throughout follicular development. These observations, together with evidence linking *FMR1* expression to granulosa-cell function and follicular development, support a potential role for FMRP in ovarian function and oocyte maturation[17,27,35–38].

The extensive repertoire of alternatively spliced *FMR1* transcripts reported across tissues and species suggests substantial transcriptomic complexity. However, the detection of a transcript does not necessarily imply the production of a functional protein. Alternative splicing events that alter the open reading frame (ORF) may introduce premature termination codons (PTCs), potentially resulting in nonsense-mediated mRNA decay (NMD) or the production of truncated proteins. Thus, characterizing the transcript landscape is an important first step towards understanding the functional complexity of *FMR1* expression.

In the present study, we used long-read amplicon sequencing based on Oxford Nanopore Technologies (ONT) to provide, to our knowledge, the first comprehensive characterization of continuous reads of *FMR1* transcripts, in human peripheral blood and ovarian granulosa cells. By sequencing transcripts spanning exons 1–17, this approach enabled the direct resolution of exon combinations and alternative splice-site usage within individual transcripts, overcoming an important limitation of short-read sequencing, in which transcript reconstruction often requires computational inference from multiple short reads. We identified 30 reproducibly detected isoforms accounting for 95% of amplicon-captured *FMR1* expression, including 26 isoforms in blood and 17 in granulosa cells. Our results showed that the lower isoform count within the GC 95% threshold reflects a more concentrated abundance distribution rather than true isoform absence, and that isoform ranking profiles were strongly correlated across tissues. Nevertheless, this conservation claim should be interpreted with caution, as GCs were analyzed after 48 h of primary culture following hormonal hyperstimulation. This culture period was chosen to enrich the sample with GCs by eliminating residual blood cells that could interfere with the analysis following follicular fluid purification. Even so, potential alterations in isoform expression induced by the in vitro

https://doi.org/10.1038/s41598-026-78955-8

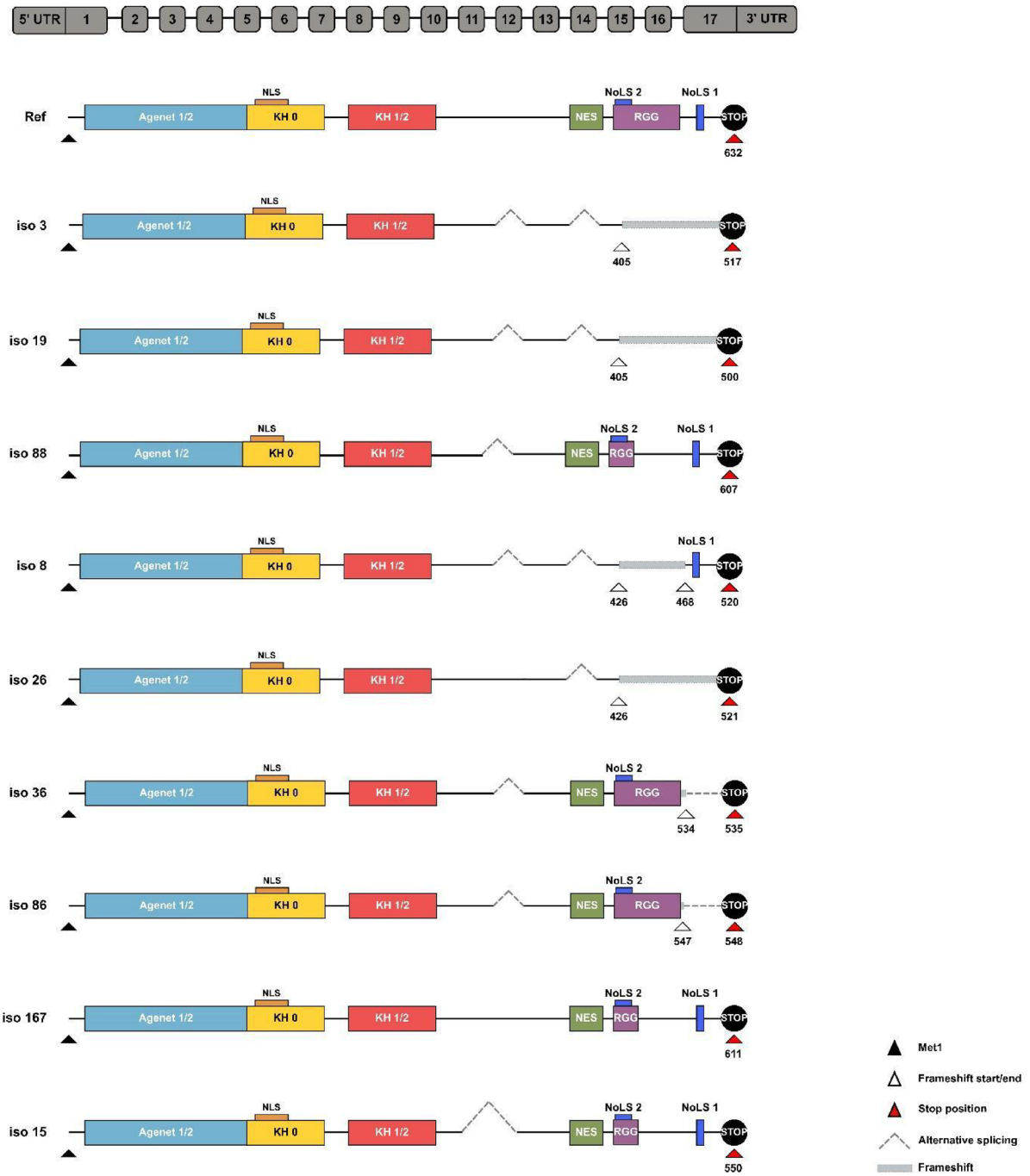


Fig. 5 Predicted functional domain architecture of novel FMRP isoforms predicted to escape the NMD pathway. The top section provides a schematic representation of exons 1 through 17 of the transcripts. Below, the reference isoform (Ref, 632 aa) displays the canonical arrangement of N-terminal domains (Agenet 1 and 2, KH 0, 1 and 2) and C-terminal motifs, including localization signals (NLS, NES, NoLS 1 and 2) and the RNA-binding domain (RGG). Novel isoforms (3, 19, 88, 8, 26, 167) and previously described isoform 15 (ENST00000943063) exhibit C-terminal reconfigurations driven by alternative splicing events (dashed lines). Textured gray bars indicate frameshift regions, delimited by white triangles (frameshift start/end). Start (Met1) and termination (Stop) codons are denoted by black and red triangles, respectively, indicating the final predicted amino acid length for each protein.

culture conditions cannot be entirely excluded, as well as possible residual lymphocyte contamination in the GC sample. It is important to note that, even though we observed no differential expression in all the high confidence expressed isoforms, there are limitations associated with applying this analysis to a single-gene amplicon dataset.

Across samples, the two most abundant transcripts were isoforms 1 (ENST00000218200) and 2 (ENST00000687593). Both lack exon 12 and retain exon 14, consistent with previous observations that transcripts lacking exon 12 are among the predominant *FMR1* isoforms across tissues and species. The frequent inclusion of exon 14 is also consistent with previous reports and is of particular interest because exon 14 encodes the NES of FMRP, providing a potential molecular basis for the predominantly cytoplasmic localization of the protein[9,17,31,35,39–41]. Together, these findings indicate that the major *FMR1* transcripts expressed in blood and granulosa cells are broadly consistent with the canonical isoform repertoire described in other tissues. Beyond these predominant transcripts, we identified substantial variation in the use of alternative splice sites and in the inclusion or exclusion of exons 12 and 14. Alternative splicing within exon 15 has previously been proposed to generate FMRP variants with different post-translational modification profiles and potentially distinct RNA-binding or regulatory properties[41]. In contrast, exclusion of exon 14 alters the ORF and introduces a PTC, potentially rendering these transcripts susceptible to NMD depending on the splice acceptor site used in exon 15. However, whether such transcripts are efficiently degraded by NMD remains unresolved. In HEK293T cells, knockdown of UPF1, a core component of the NMD pathway, did not increase the abundance of *FMR1* transcripts lacking exon 14[42]. Notably, FMRP has itself been reported to interact with UPF1 with NMD hyperactivation observed in FMR1-deficient cells[43]. Given that FMRP can also bind *FMR1* mRNA[44], these observations suggest a putative feedback mechanism linking *FMR1* alternative splicing and NMD. Functional studies will therefore be necessary to determine whether *FMR1* transcripts predicted to be NMD-sensitive are indeed targeted for degradation through this pathway.

An important finding of this study was the identification of transcripts that use infrequent annotated splice acceptor sites. In particular, several isoforms used an alternative acceptor site in exon 15 that removes 15 bp while preserving the ORF, potentially generating an FMRP variant lacking five amino acids. In contrast, use of an infrequent acceptor site in exon 17 removes only 5 bp, shifting the ORF and introducing a PTC. Since this PTC occurs downstream of the final exon–exon junction, the transcript may be predicted to escape canonical NMD and could potentially be translated into a truncated FMRP isoform. However, the relationship between PTC position and NMD sensitivity is more complex than the canonical exon-junction rule, and transcripts containing PTCs located far upstream of the poly(A) tail may also undergo degradation[45,46].

We want to emphasize that one of the important findings in our work is the characterization of the *FMR1* isoform landscape that arises from the use of frequent and infrequent splicing sites generating new splicing variants that have not been described before. Importantly, while our approach enables the description of novel combinations of previously annotated splice sites, it does not identify new splice sites. Moreover, the possibility that these infrequently used splicing sites may represent sequencing or read assembly errors in our experiments seems unlikely, as they were observed in more than one isoform and were found consistently across multiple samples. Nevertheless, experimental validation with more independent RNA-sequencing approaches, junction-specific RT-PCR followed by Sanger sequencing across independent women, or confirmation on an independent sequencing platform, would address this.

We also identified several previously unannotated *FMR1* isoforms. These included novel combinations of known alternative splicing events, exon 11 exclusion, and intron retention. Isoform 15 (ENST00000943063), which lacks exon 11, was detected among the reproducibly expressed transcripts. Although exon 11 skipping has previously been reported[13], the complete transcript structure and its functional consequences have not been established. Notably, exon 11 exclusion preserves the ORF and would result in the deletion of 45 amino acids. The absence of known functional domains within this region makes the consequences of this deletion difficult to predict at the protein level, and further work is needed to determine whether this transcript produces a stable protein with altered function.

Three isoforms containing intron-retention events were also identified. These events introduce PTCs and could therefore generate NMD-sensitive transcripts. Of particular interest, a 137-bp sequence between exons 2 and 3 was retained in two independent isoforms. The recurrence of this event raises the possibility that it represents a regulated splicing event rather than stochastic splicing noise. Intron retention is increasingly recognized as a mechanism contributing to transcript regulation and proteome diversity across tissues and species[47–49]. Nevertheless, distinguishing functional intron retention from aberrant or incompletely processed transcripts remains challenging, particularly when transcripts are expressed at low abundance. This distinction is particularly relevant when interpreting novel *FMR1* isoforms. NMD does not exclusively eliminate aberrant transcripts; naturally occurring alternative splicing (AS) can generate NMD-sensitive transcripts that contribute to physiological regulation of gene expression. Such events have been described as regulatory AS-NMD[50]. The mere presence of a PTC-containing *FMR1* transcript therefore cannot establish that the transcript is biologically irrelevant.

An additional consideration is the reference annotation against which the transcripts identified in this study were mapped. The Ensembl annotation incorporates evidence generated using different transcriptomic approaches, and much of the historical transcript annotation has relied on short-read sequencing and computational transcript reconstruction. Short-read data are highly informative for identifying expressed splice junctions but have limited capacity to resolve complete transcript structures when multiple alternative splicing events occur within the same molecule. Consequently, some *FMR1* isoforms identified by long-read sequencing may not be represented in current annotations, while annotated transcripts may not necessarily correspond to experimentally validated, full-length molecules. Our results therefore highlight the complementary value of long-read sequencing for refining transcript annotations and resolving complex alternative splicing patterns. However, it is worth mentioning that the repertoire described herein is restricted to isoforms within the amplicon-defined transcript spanning exons 1 through 17. Consequently, any isoforms with alternative 5' or 3' ends lacking either primer-binding region were not detected.

An important question is whether all identified isoforms are physiologically relevant. At present, our data cannot establish this. Transcript abundance provides evidence of expression but does not demonstrate protein production, cellular localization or biological activity. Conversely, low-abundance transcripts should not necessarily be considered biologically irrelevant. An isoform expressed at low abundance in bulk tissue could be highly expressed in a specific cell population, developmental stage or physiological condition. In addition, several low-abundance isoforms may have similar functional consequences and therefore contribute cumulatively to the *FMR1* transcript or protein repertoire. For this reason, the novel isoforms identified here provide hypotheses for future functional studies rather than definitive evidence of distinct FMRP functions.

The identification of previously unannotated *FMR1* transcripts in granulosa cells is particularly relevant to ovarian biology. FMRP has been implicated in granulosa-cell function, follicular development and ovarian germ-cell maturation, while alterations in *FMR1* expression and premutation-associated *FMR1* biology are relevant to fragile X-associated primary ovarian insufficiency (FXPOI)[17,36–38]. The present findings suggest that *FMR1* regulation in the ovary may involve not only changes in total transcript abundance but also changes in the relative expression of individual transcript isoforms.

Although the number of individuals analyzed in this study is limited, our results provide a first comprehensive characterization of the diversity of alternative splicing combinations among *FMR1* isoforms encompassing exons 1–17. However, whether the expression of specific *FMR1* isoforms is associated with follicular development or contributes to the development of FXPOI remains unknown. The observed correlations between *FMR1* isoform expression in GCs and peripheral blood should also be interpreted with caution, given the limited sample size. Further studies involving larger cohorts will be necessary to validate these findings and to determine whether specific isoform expression patterns are associated with FXPOI.

Our results also indicate that *FMR1* isoform expression is conserved across the tissue types examined, with no significant differences observed. However, the amplicon-seq strategy used in this work may mask potential differences in expression levels, as absolute expression of *FMR1* isoforms is better evaluated through a whole transcriptomic tissue assay. Even so, peripheral blood may represent a readily accessible tissue for future

non-invasive studies of *FMR1* isoforms as potential biomarkers of FXPOI development in premutation carriers. Overall, the long-read sequencing strategy established here provides a framework for investigating *FMR1* transcript diversity in clinically relevant tissues.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-78955-8.

Acknowledgements The authors wish to thank Sol Ruiz, Lucila Viappiani, Ezequiel Petrillo and Agustín Baricalla for their valuable insights and helpful discussions regarding the bioinformatic analysis. Most importantly, we express our deepest appreciation to the women who donated oocytes for their vital contribution of blood and granulosa cells to this work.

Author contributions MLI RNA extraction, QC analysis, library construction, sequencing, raw data interpretation, transcriptomic analysis, protein analysis, drafting of the paper; CSF women recruitment, sample collection; EC sample collection; LD design and concept of the study, discussing results, writing of the final manuscript; LK design and concept of the study, sequencing, raw data interpretation, discussing results, writing of the final manuscript; IF design and concept of the study, GCs culture, RNA extraction, QC analysis, raw data interpretation, discussing results, writing of the final manuscript. All of the authors read and approved the final manuscript. LK and IF: These authors contributed equally as corresponding authors.

Funding This study was supported by the National Agency for Science and Technology Promotion (ANPCyT), Argentina, project PICT 2021 018 and the National Scientific and Technical Research Council (CONICET), project PIP 2023–2025 GI −91CO.

Data availability GitHub: https://github.com/maru-luz/FMR1-isoform-detection.gitBioProject: http://www.ncbi.nlm.nih.gov/bioproject/1509465, ID PRJNA1509465.

Declarations

Competing interests The authors declare no competing interests.

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

References

1. Eichler, E. E., Richards, S., Gibbs, R. A. & Nelson, D. L. Fine structure of the human FMR1 gene. *Hum. Mol. Genet.* **2**, 1147–1153 (1993).
2. Verkerk, A. J. et al. Alternative splicing in the fragile X gene FMR1. *Hum. Mol. Genet.* **2**, 399–404 (1993).
3. Fernandez-Carvajal, I. et al. Screening for expanded alleles of the FMR1 gene in blood spots from newborn males in a Spanish population. *J. Mol. Diagn.* **11**, 324–329 (2009).
4. Jung, S. & Richter, J. D. Trinucleotide repeat expansion and RNA dysregulation in fragile X syndrome: emerging therapeutic approaches. *RNA* **31**, 307–313 (2025).
5. Gillett, D. A., Tigro, H., Wang, Y. & Suo, Z. FMR1 disorders: Basics of biology and therapeutics in development. *Cells* **13**, 2100 (2024).
6. Souski, H., Benmakhlouf, Y. & Mechita, M. B. Clinical, genetic and molecular divergences between full mutation and premutation in fragile X syndrome: A systematic review. *Int. J. Dev. Neurosci.* **85**, e70054 (2025).
7. Dijkstra, A. A. et al. Neuropathology of FMR1-premutation carriers presenting with dementia and neuropsychiatric symptoms. *Brain Commun.* **3**, fcab007 (2021).
8. Ashley, C. T. et al. Human and murine FMR-1: alternative splicing and translational initiation downstream of the CGG-repeat. *Nat. Genet.* **4**, 244–251 (1993).

9. Pretto, D. I. et al. Differential increases of specific FMR1 mRNA isoforms in premutation carriers. *J. Med. Genet.* **52**, 42–52 (2015).
10. Dolzhanskaya, N., Merz, G. & Denman, R. B. Alternative splicing modulates protein arginine methyltransferase-dependent methylation of fragile X syndrome mental retardation protein†. *Biochemistry* **45** 10385–10393 https://doi.org/10.1021/bi0525019 (2006).
11. Ceman, S. et al. Phosphorylation influences the translation state of FMRP-associated polyribosomes. *Hum. Mol. Genet.* **12**, 3295–3305 (2003).
12. Dury, A. Y. et al. Nuclear fragile X mental retardation protein is localized to Cajal bodies. *PLoS Genet.* **9**, e1003890 (2013).
13. Tseng, E., Tang, H. T., AlOlaby, R. R., Hickey, L. & Tassone, F. Altered expression of the FMR1 splicing variants landscape in premutation carriers. *Biochim. Biophys. Acta Gene Regul. Mech.* **1860**, 1117–1126 (2017).
14. Yang, W. J. et al. Further identification of a 140 bp sequence from amid intron 9 of human FMR1 gene as a new exon. *BMC Genet.* **21**, 63 (2020).
15. Alvarez-Mora, M. I. et al. Evaluation of FMR4, FMR5 and FMR6 expression levels as non-invasive biomarkers for the diagnosis of fragile X-associated primary ovarian insufficiency (FXPOI). *J. Clin. Med.* **11**, 2186 (2022).
16. Fink, D. A. et al. Fragile X associated primary ovarian insufficiency (FXPOI): Case report and literature review. *Front. Genet.* **9**, 529 (2018).
17. Ferder, I. C. et al. Expression and characterisation of Fmr1 splice variants during folliculogenesis in the rat. *Reprod. Fertil. Dev.* **34**, 1034–1042 (2022).
18. Chen, Y. et al. A systematic benchmark of Nanopore long-read RNA sequencing for transcript-level analysis in human cell lines. *Nat. Methods*. **22**, 801–812 (2025).
19. Zırh, S. et al. A new isolation and culture method for granulosa cells. *Cell. Tissue Bank.* **22**, 719–726 (2021).
20. Tang, A. D. et al. Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia reveals downregulation of retained introns. *Nat. Commun.* **11**, 1438 (2020).
21. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
22. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
23. Smyth, G. K., Ritchie, M., Thorne, N. & Wettenhall, J. Wei Shi and Yifang Hu. Linear Models for Microarray and RNA-Seq Data User's Guide. https://www.bioconductor.org/packages/devel/bioc/vignettes/limma/inst/doc/usersguide.pdf
24. Lindeboom, R. G. H., Supek, F. & Lehner, B. The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nat. Genet.* **48**, 1112–1118 (2016).
25. Kurosaki, T., Popp, M. W. & Maquat, L. E. Publisher Correction: Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat. Rev. Mol. Cell. Biol.* **20**, 384 (2019).
26. Nagy, E. & Maquat, L. E. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem. Sci.* **23**, 198–199 (1998).
27. Bächner, D. et al. Enhanced expression of the murine FMR1 gene during germ cell proliferation suggests a special function in both the male and the female gonad. *Hum. Mol. Genet.* **2**, 2043–2050 (1993).
28. Eberhart, D. E., Malter, H. E., Feng, Y. & Warren, S. T. The fragile X mental retardation protein is a ribonucleoprotein containing both nuclear localization and nuclear export signals. *Hum. Mol. Genet.* **5**, 1083–1091 (1996).
29. Vasilyev, N. et al. Crystal structure reveals specific recognition of a G-quadruplex RNA by a β-turn in the RGG motif of FMRP. *Proc. Natl. Acad. Sci. U S A*. **112**, E5391–E5400 (2015).
30. Taha, M. S. et al. Subcellular fractionation and localization studies reveal a direct interaction of the fragile X mental retardation protein (FMRP) with nucleolin. *PLoS One*. **9**, e91465 (2014).
31. Brackett, D. M. et al. FMR1 transcript isoforms: association with polyribosomes; regional and developmental expression in mouse brain. *PLoS One*. **8**, e58296 (2013).
32. Catalano, F. et al. Unraveling in vitro phase separation and aggregation properties of the structured region of FMRP and the impact of Fragile X syndrome-linked mutations. *FEBS J.* **293**, 6155–6175 (2026).
33. Troilo, F. et al. The elusive folding mechanism of KH1 domain of Fragile X messenger ribonucleoprotein 1. *Arch. Biochem. Biophys.* **783**, 110909 (2026).
34. Richter, J. D. & Zhao, X. The molecular biology of FMRP: new insights into fragile X syndrome. *Nat. Rev. Neurosci.* **22**, 209–222 (2021).
35. Ferder, I. et al. Expression of fragile X mental retardation protein and Fmr1 mRNA during folliculogenesis in the rat. *Reproduction* **145**, 335–343 (2013).
36. Nguyen, X. P. et al. Regulation of bone morphogenetic protein receptor type II expression by FMR1/fragile X mental retardation protein in human granulosa cells in the context of poor ovarian response. *Int. J. Mol. Sci.* **25**, 10643 (2024).
37. Rehnitz, J. et al. FMR1 expression in human granulosa cells increases with exon 1 CGG repeat length depending on ovarian reserve. *Reprod. Biol. Endocrinol.* **16**, 65 (2018).

38. Rosario, R. et al. FMRP associates with cytoplasmic granules at the onset of meiosis in the human oocyte. *PLoS One*. **11**, e0163987 (2016).
39. Devys, D., Lutz, Y., Rouyer, N., Bellocq, J. P. & Mandel, J. L. The FMR-1 protein is cytoplasmic, most abundant in neurons and appears normal in carriers of a fragile X premutation. *Nat. Genet.* **4**, 335–340 (1993).
40. Fu, X. et al. Alternatively spliced products lacking exon 12 dominate the expression of fragile X mental retardation 1 gene in human tissues. *Mol. Med. Rep.* **12**, 1957–1962 (2015).
41. Xie, W., Dolzhanskaya, N., LaFauci, G., Dobkin, C. & Denman, R. B. Tissue and developmental regulation of fragile X mental retardation 1 exon 12 and 15 isoforms. *Neurobiol. Dis.* **35**, 52–62 (2009).
42. Corrêa-Velloso, J. C. et al. Fmr1 exon 14 skipping in late embryonic development of the rat forebrain. *BMC Neurosci.* **23**, 32 (2022).
43. Kurosaki, T. et al. Loss of the fragile X syndrome protein FMRP results in misregulation of nonsense-mediated mRNA decay. *Nat. Cell. Biol.* **23**, 40–48 (2021).
44. Edwards, M. & Joseph, S. The fragile X proteins differentially regulate translation of reporter mRNAs with G-quadruplex structures. *J. Mol. Biol.* **434**, 167396 (2022).
45. Brogna, S. & Wen, J. Nonsense-mediated mRNA decay (NMD) mechanisms. *Nat. Struct. Mol. Biol.* **16**, 107–113 (2009).
46. Wu, C., Roy, B., He, F., Yan, K. & Jacobson, A. Poly(A)-binding protein regulates the efficiency of translation termination. *Cell. Rep.* **33**, 108399 (2020).
47. Monteuuis, G., Wong, J. J. L., Bailey, C. G., Schmitz, U. & Rasko, J. E. J. The changing paradigm of intron retention: regulation, ramifications and recipes. *Nucleic Acids Res.* **47**, 11497–11513 (2019).
48. Porras-Tobias, A. L., Caldera, A. & Castro-Piedras, I. Intron Retention: A reemerging paradigm in RNA biology and post-transcriptional gene regulation. *Genes (Basel)*. **16**, 986 (2025).
49. Wong, J. J. L. & Schmitz, U. Intron retention: importance, challenges, and opportunities. *Trends Genet.* **38**, 789–792 (2022).
50. Yap, K., Ramakumar, S., Hamid, F. & Makeyev, E. V. Regulation of eukaryotic gene expression through functional coupling between alternative splicing and nonsense-mediated decay. *J. Mol. Biol.* **438**, 169764 (2026).

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

Authors and Affiliations

Marina Luz Ingravidi[1,2] · Cecilia S. Fernández[3] · Eloisa Caviglia[3] · Liliana Dain[1,2] · Laura Kamenetzky[1,2] · Ianina Ferder[1,2]

✉ Laura Kamenetzky
lauka@fbmc.fcen.uba.ar

✉ Ianina Ferder
iferder@fbmc.fcen.uba.ar

[1] Instituto de Biociencias, Biotecnología y Biología Traslacional (iB3) - FCEN - UBA, Intendente Güiraldes 2620 (C1428EG), Buenos Aires, Argentina

[2] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Godoy Cruz 2290 (C1425FQB), Buenos Aires, Argentina

[3] Novagen- CEGYR, Buenos Aires, Argentina