frontiers in Genetics

BRIEF RESEARCH REPORT
published: 13 November 2020
doi: 10.3389/fgene.2020.557147

Development and Validation of a Mathematical Model to Predict the Complexity of *FMR1* Allele Combinations

*Bárbara Rodrigues[1,2], Emídio Vale-Fernandes[2,3], Nuno Maia[1,2], Flávia Santos[1,2], Isabel Marques[1,2], Rosário Santos[1,2], António J. A. Nogueira[4] and Paula Jorge[1,2]**

[1] Molecular Genetics Unit, Centro de Genética Médica Dr. Jacinto Magalhães (CGMJM), Centro Hospitalar Universitário do Porto (CHUP), Porto, Portugal, [2] Unit for Multidisciplinary Research in Biomedicine (UMIB), Institute of Biomedical Sciences Abel Salazar (ICBAS), University of Porto, Porto, Portugal, [3] Centre for Medically Assisted Procreation/Public Gamete Bank, Centro Materno-Infantil do Norte Dr. Albino Aroso (CMIN), Centro Hospitalar Universitário do Porto (CHUP), Porto, Portugal, [4] Center for Environmental and Marine Studies (CESAM), Department of Biology, University of Aveiro, Aveiro, Portugal

OPEN ACCESS

Edited by:
Laia Rodriguez-Revenga, Hospital Clínic de Barcelona, Spain

Reviewed by:
María Isabel Alvarez-Mora, University Hospital October 12, Spain
Carolyn M. Yrigollen, University of California, Davis, United States

***Correspondence:**
Paula Jorge
paulajorge.cgm@chporto.min-saude.pt

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

Received: *29 April 2020*
Accepted: *13 October 2020*
Published: *13 November 2020*

Citation:
Rodrigues B, Vale-Fernandes E, Maia N, Santos F, Marques I, Santos R, Nogueira AJA and Jorge P (2020) Development and Validation of a Mathematical Model to Predict the Complexity of FMR1 Allele Combinations. Front. Genet. 11:557147. doi: 10.3389/fgene.2020.557147

The polymorphic trinucleotide repetitive region in the *FMR1* gene 5′UTR contains AGG interspersions, particularly in normal-sized alleles (CGG < 45). In this range repetitive stretches are typically interrupted once or twice, although alleles without or with three or more AGG interspersions can also be observed. AGG interspersions together with the total length of the repetitive region confer stability and hinder expansion to pathogenic ranges: either premutation (55 < CGG < 200) or full mutation (CGG > 200). The AGG interspersions have long been identified as one of the most important features of *FMR1* repeat stability, being particularly important to determine expansion risk estimates in female premutation carriers. We sought to compute the combined AGG interspersion numbers and patterns, aiming to define *FMR1* repetitive tract complexity combinations. A mathematical model, the first to compute this cumulative effect, was developed and validated using data from 131 young and healthy females. Plotting of their allelic complexity enabled the identification of two statistically distinct groups – *equivalent* and *dissimilar* allelic combinations. The outcome, a numerical parameter designated *allelic score*, depicts the repeat substructure of each allele, measuring the allelic complexity of the *FMR1* gene including the AGGs burden, thus allowing new behavioral scrutiny of normal-sized alleles in females.

Keywords: *FMR1* gene, CGG repeats, AGG interspersion pattern, modeling allelic complexity, *allelic score*

INTRODUCTION

The fragile X-related disorders result from the expansion of a CGG-repeat tract in the 5′ untranslated region of the *FMR1* gene (Xq27.3), coding for the fragile X mental retardation protein (FMRP), an RNA-binding protein that regulates expression of several genes (Man et al., 2017). Depending on the number of CGG repeats, *FMR1* alleles can be categorized into four classes: normal (CGG < 45), intermediate or "gray zone" (45 < CGG < 54), premutation

(55 < CGG < 200), and full mutation (CGG > 200) (Biancalana et al., 2015). Premutations causing *FMR1* mRNA overexpression and reduced FMRP synthesis, underly both fragile X-associated tremor/ataxia syndrome (FXTAS, OMIM #300623) and fragile X-associated primary ovarian insufficiency (FXPOI, OMIM #311360). The full mutation alleles undergo hypermethylation, leading to gene silencing and absence of FMRP, causing fragile X syndrome (FXS, OMIM #300624), the most common heritable cause of intellectual disability (Man et al., 2017). Due to the repeat tract instability, above a threshold expansions and contractions can be observed both in the germline and in the somatic cells. Some rare contraction events can originate mosaicism with mutated and normal alleles in clinically typical fragile-X phenotypes (Maia et al., 2016). In the normal population, the vast majority of the alleles contain one or more AGG interspersions within the repetitive tract, usually at every 9 or 10 CGG repeat intervals, being highly stable. In higher repeat ranges, the number of AGGs tends to be progressively smaller as the size of the repetitive tract increases (Yrigollen et al., 2012; Mila et al., 2018; Manor et al., 2019). The AGG interspersions together with the repetitive region's total length confer stability and hinder expansion to pathogenic size-ranges (Latham et al., 2014; Domniz et al., 2018; McGinty and Mirkin, 2018). In premutation female carriers, the risk of having a child with FXS depends on both the repeat length and AGG interspersions (Ardui et al., 2018). The incidence of normal pure alleles (without interspersions) is low and their origin as well as the phenotypic impact in females, are still debatable. It has been proposed that "low zone" alleles, variably determined to be CGG $\leq$ 26 or CGG $\leq$ 23, are associated with different phenotypic outcomes (Mailick et al., 2014; Gleicher et al., 2015; Rehnitz et al., 2018). Some studies show that they are associated with decreased ovarian reserve and fertility issues, due to a mechanism not yet elucidated, possibly different from that involved in premutated alleles (Gleicher et al., 2015; Wang et al., 2017), although such negative effects were not corroborated by others (Spitzer et al., 2012; Ruth et al., 2016). These contradictory assumptions require further studies to elucidate the clinical impact of "low zone" alleles.

Few studies focus on AGG interspersion patterns to assess allele stability, within the normal range. Given the importance of understanding the cumulative effect of the CGG repeat tract length and its AGG interspersions, we developed a mathematical model that considers these patterns and produces a functional model predicting the complexity of allele combinations (*allelic score*).

MATERIALS AND METHODS

Study Population

Young and potentially fertile females were recruited among candidates for oocyte donation at the Portuguese Public Gamete Bank, Centro Materno-Infantil do Norte Dr. Albino Aroso (CMIN), Centro Hospitalar Universitário do Porto (CHUP). The donor population, originating from the entire national territory, includes actively recruited students from major Portuguese universities, with a wide range of nationalities. Around 10% of the donor candidates were of foreign nationality, 95% were Caucasian and about 30% of those who donated at our center lived outside Porto (Galvão et al., 2017). Two independent cohorts were used for development (cohort 1) and for validation (cohort 2) studies. Cohort 1, n = 50, mean age 25.4 $\pm$ 3.93 years (range 18–33), recruited between 2016 and 2017. Cohort 2, n = 81, mean age 26.5 $\pm$ 3.86 years (range 19–33), collected between 2018 and 2019. All participants provided written informed consent, and this project was approved by the Hospital's Ethics Committee (2018.231/201-DEFI/200-CES).

FMR1 Repeat Region Substructure Profile

Sizing of *FMR1* alleles had been previously obtained as part of the routine oocyte donor's protocol, on blood samples. Categorizing the respective genotype followed the ACMG/EMQN guidelines: normal (CGG < 45), intermediate or "gray zone" (45 < CGG < 54), premutation (55 < CGG < 200), and full mutation (CGG > 200) (Monaghan et al., 2013; Biancalana et al., 2015). AGG interspersion pattern was determined by Triplet Repeat Primed-PCR using FRAXA PCR kit LabGscan™ (Diagnostica Longwood, Zaragoza, Spain), according to the manufacturer's instructions. This method allowed the confirmation of the total repeat length and the characterization of the CGG/AGG substructure. Thirteen samples with different patterns were additionally verified by Sanger sequencing to confirm the previously determined CGG/AGG pattern.

Statistical Analysis

Hierarchical Cluster Analysis using euclidean distance as a metric to evaluate similarity was used in statistical software SPSS® version 26 (IBM developer, 2019: SPSS Statistics version 26 – Armonk, New York, United States). Linear regression of the linearized form of an exponential model [i.e., regression of ln(score 2) against score 1] was used to obtain a functional model to relate the complexity of both alleles in each sample. The analysis of covariance (ANCOVA), as outlined by Zar (2010), was used to compare the regression models, and derive common regression lines, with *allelic scores* as variables [i.e., score 1 and ln(score 2)]. All statistical tests were carried out for a significance level of 0.05.

Determination of X-Chromosome Inactivation Pattern and *FMR1* Methylation Status

X-chromosome inactivation (XCI) pattern was determined by the human androgen-receptor assay (HUMARA), resorting to the CAG trinucleotide repeat located in the first exon and two methylation-sensitive endonuclease sites located upstream of the *AR* gene (Allen et al., 1992). The percentage of allele activity was determined using the peak heights, and normalized to the corresponding undigested allele peak

height. The *FMR1* methylation status was determined using AmplideX® mPCR *FMR1* kit (Asuragen, Inc., Austin, TX, United States), according to the manufacturer's instructions. The mPCR assay determines both the number of CGG repeats and the percentage promoter methylation of each *FMR1* allele.

RESULTS

A similar *FMR1* CGG size distribution was obtained in both cohorts with normal alleles, ranging from 15 to 40 CGG in cohort 1 and from 15 to 44 CGG in cohort 2 (n = 127, 97%) and intermediate genotypes, one allele with 48 CGG in cohort 1 and three alleles with 45 CGG in cohort 2 (n = 4, 3%) (**Tables 1**, **2**). Homozygosity was observed in eleven samples (22%, cohort 1), of which nine shared the same CGG/AGG substructure, and in seventeen samples (21%, cohort 2), of which thirteen shared the same AGG pattern. In line with previous publications, the vast majority of the alleles (93%) showed one or two AGGs, 5% were pure (4, cohort 1 and 9, cohort 2) and the remaining 2% showed three AGG interspersions. The most common structure, $(CGG)_{10}AGG(CGG)_9AGG(CGG)_9$, was identified in 29 (29%, cohort 1) and 40 alleles (25%, cohort 2). Similar to other worldwide populations, a highly polymorphic CGG/AGG substructure was observed: forty-one and fifty-five unique patterns were identified in cohorts 1 and 2, respectively (**Tables 1**, **2**; Yrigollen et al., 2014).

Development of the Mathematical Model

A mathematical model was developed to integrate the AGG interspersion number and pattern and the total repeat length, reflecting the CGG/AGG substructure. The result score, named *allelic score*, was calculated separately for each allele as follows:

$$Allelic\ score = \left(\sum_{i=1}^{n} R_i \times\ 4^{i-1}\right) + \left(R_{n+1} \times\ 4^{n}\right)$$

where,

R_i: number of CGG repeats before the first AGG interspersion of order i;

i: CGG repeat order number;

n: total number of AGG interspersions;

R_{n+1}: number of CGG repeats after the last AGG interspersion.

Base-4 numeral system was used to ensure that the *allelic score* is unique to each of the AGG interspersion patterns and sufficiently spaced.

For the purpose of addressing allelic complexity, two different aspects of the allelic structure are considered: number of AGG interspersions and number of CGG repeats between interspersions. Higher relevance is given to the number of interspersions as, for alleles with identical number of CGG repeats, higher number of AGG interspersions is usually linked with allelic stability (Maia et al., 2016; Manor et al., 2019). As example, an allele with two AGGs shows an *allelic score* of 193 whereas an allele with a similar length but only one AGG has an *allelic score* of 59.

$$Allelic\ score\left[(CGG)_9\,AGG\,(CGG)_{10}\,AGG\,(CGG)_9\right] =$$
$$[(9 \times 4^{1-1}) + (10 \times 4^{2-1})] + (9 \times 4^2) = 193$$
$$Allelic\ score\left[(CGG)_{10}\,AGG\,(CGG)_{19}\right] =$$
$$(19 \times 4^{1-1}) + (10 \times 4^1) = 59$$

This mathematical model is protected with a national patent (reference – 115244) and international patent application submitted on december 6, 2019 (reference – PCT/IB2019/060520).

Application and Validation of the Mathematical Model

Allelic scores ranged from 15 to 825 (cohort 1) and 15 to 828 (cohort 2), with most samples scoring below 220 (95.4%) and six with a score in the order of 800, due to the presence of three AGG interspersions (**Tables 1**, **2**). Scores under 220 either represent zero, one or two AGG interspersions; above two AGG interspersions, the *allelic score* grows exponentially. An exploratory cluster analysis identified four major clusters, with observations within each quadrant separated in both axes by an *allelic score* of 150 (**Supplementary Figures 1**, **2**). Similar behaviors were observed among the two quadrants where *allelic scores* were both lower than 150 or both higher than 150, and the other two where alleles show low and high *allelic score*, allowing the definition of two groups. The *equivalent* group contains samples where both alleles show a similar complexity, and the *dissimilar* group with samples where alleles show a different complexity. These groups include samples with three AGGs as the behavior of their alleles fits that of other samples in the same quadrant (**Supplementary Figures 1**, **2**). In both groups, an exponential model was used to describe the correlation between the *allelic score* of each allele. Significant correlations were found: cohort 1 – *equivalent* group: r = 0.8092; df = 24; $p < 0.0001$ and *dissimilar* group: $r = -0.7067$; df = 22; $p < 0.0001$ (**Supplementary Figure 3**). To validate the mathematical models and their reproducibility, a covariance (ANCOVA) analysis was used to compare the models calculated for cohort 1 and the same models computed using cohort 2 data (*equivalent* group: r = 0.8603; df = 43; $p < 0.0001$ and *dissimilar* group: $r = -0.8716$; df = 33; $p < 0.0001$) (**Supplementary Figure 4**). There was no statistically significant difference between cohort 1 (development cohort) and cohort 2 (validation cohort) with respect to the *equivalent* and *dissimilar* group's models, as demonstrated by the coincident regression lines (**Supplementary Figure 5**). A more robust model including all observations (both cohorts) was derived: *equivalent* group – $F_{(2,68)} = 1.8048$; $p = 0.1723$: ln(score 2) = 3.6452 + 0.0088 × score 1 and *dissimilar* group – $F_{(2,55)} = 0.9574$; $p = 0.3902$: ln(score 2) = 5.6944 – 0.0065 × score 1.

Seven samples from each group (cohort 2) were tested for XCI pattern (**Supplementary Table 1**). Interestingly, in a sample

TABLE 1 | Cohort 1 data used to calculate the *allelic scores*, and identify the two groups, *equivalent* (white background) and *dissimilar* (gray background).

Allele 1			Allele 2		
CGG/AGG Pattern	**Repeat length**	***Allelic score***	**CGG/AGG Pattern**	**Repeat length**	***Allelic score***
(CGG)$_{8}$AGG(CGG)$_{9}$	18[$]	**41**	(CGG)$_{23}$AGG(CGG)$_{9}$	33	**101**
(CGG)$_{10}$AGG(CGG)$_{9}$	20[$]	**49**	(CGG)$_{10}$AGG(CGG)$_{9}$	20[$]	**49**
(CGG)$_{9}$AGG(CGG)$_{8}$AGG(CGG)$_{9}$	20[$]	**185**	(CGG)$_{10}$AGG(CGG)$_{8}$AGG(CGG)$_{9}$	29	**201**
(CGG)$_{10}$AGG(CGG)$_{9}$	20[$]	**49**	(CGG)$_{20}$AGG(CGG)$_{9}$	30	**89**
(CGG)$_{10}$AGG(CGG)$_{9}$	20[$]	**49**	(CGG)$_{13}$AGG(CGG)$_{9}$	23[$]	**61**
(CGG)$_{9}$AGG(CGG)$_{11}$	21[$]	**47**	(CGG)$_{12}$AGG(CGG)$_{16}$	29	**64**
(CGG)$_{10}$AGG(CGG)$_{11}$	22[$]	**51**	(CGG)$_{13}$AGG(CGG)$_{16}$	30	**68**
(CGG)$_{9}$AGG(CGG)$_{13}$	23[$]	**49**	(CGG)$_{12}$AGG(CGG)$_{25}$	38	**73**
(CGG)$_{9}$AGG(CGG)$_{15}$	25[$]	**51**	(CGG)$_{10}$AGG(CGG)$_{19}$	30	**59**
(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	29	**189**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	29	**189**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{10}$	31	**206**
(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	29	**189**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	29	**189**	(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	29	**189**
(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	29	**189**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	29	**189**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{9}$AGG(CGG)$_{10}$AGG(CGG)$_{9}$	30	**193**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{14}$	35	**210**
(CGG)$_{9}$AGG(CGG)$_{10}$AGG(CGG)$_{9}$	30	**193**	(CGG)$_{10}$AGG(CGG)$_{11}$AGG(CGG)$_{9}$	32	**213**
(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{10}$	30	**190**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{8}$AGG(CGG)$_{9}$	39	**825**
(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{9}$AGG(CGG)$_{10}$AGG(CGG)$_{9}$	30	**193**	(CGG)$_{10}$AGG(CGG)$_{11}$AGG(CGG)$_{9}$	32	**213**
(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{12}$	32	**192**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{18}$	39	**214**
(CGG)$_{8}$AGG(CGG)$_{9}$AGG(CGG)$_{21}$	40	**185**	(CGG)$_{9}$AGG(CGG)$_{8}$AGG(CGG)$_{29}$	48[#]	**205**
(CGG)$_{15}$	15[$]	**15**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{7}$AGG(CGG)$_{9}$	17[$]	**37**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{10}$AGG(CGG)$_{9}$	20[$]	**49**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{10}$AGG(CGG)$_{9}$	20[$]	**49**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{10}$	31	**206**
(CGG)10AGG(CGG)9	20[$]	**49**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{10}$AGG(CGG)$_{9}$	20[$]	**49**	(CGG)$_{10}$AGG(CGG)$_{10}$AGG(CGG)$_{10}$	32	**210**
(CGG)$_{9}$AGG(CGG)$_{10}$	20[$]	**46**	(CGG)$_{10}$AGG(CGG)$_{8}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	39	**813**
(CGG)$_{9}$AGG(CGG)$_{10}$	20[$]	**46**	(CGG)$_{10}$AGG(CGG)$_{8}$AGG(CGG)$_{9}$	29	**201**
(CGG)$_{9}$AGG(CGG)$_{10}$	20[$]	**46**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{10}$AGG(CGG)$_{9}$	20[$]	**49**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{8}$AGG(CGG)$_{9}$	39	**825**
(CGG)$_{9}$AGG(CGG)$_{10}$	20[$]	**46**	(CGG)$_{10}$AGG(CGG)$_{8}$AGG(CGG)$_{7}$AGG(CGG)$_{9}$	37	**805**
(CGG)$_{9}$AGG(CGG)$_{10}$	20[$]	**46**	(CGG)$_{9}$AGG(CGG)$_{12}$AGG(CGG)$_{9}$	32	**201**
(CGG)$_{25}$	25[$]	**25**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{10}$	31	**206**
(CGG)$_{15}$AGG(CGG)$_{9}$	25[$]	**69**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{10}$	31	**206**
(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	29	**189**	(CGG)$_{9}$AGG(CGG)$_{29}$	39	**65**
(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	29	**189**	(CGG)$_{11}$AGG(CGG)$_{20}$	32	**64**
(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**	(CGG)$_{10}$AGG(CGG)$_{22}$	33	**62**
(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**	(CGG)$_{10}$AGG(CGG)$_{20}$	31	**60**
(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**	(CGG)$_{10}$AGG(CGG)$_{22}$	33	**62**
(CGG)$_{30}$	30	**30**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**	(CGG)$_{10}$AGG(CGG)$_{20}$	31	**60**
(CGG)$_{30}$	30	**30**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**

Homoallelism for CGG-repeat length (black background) and homozygosity for both CGG-repeat length and AGG pattern (allelic score in green background).
[#] *intermediate size.*
[$] *normal "low zone" alleles (see section "Discussion").*

TABLE 2 | Cohort 2 data used to calculate the *allelic scores*, and identify the two groups, *equivalent* (white background) and *dissimilar* (gray background).

Allele 1			Allele 2		
CGG/AGG Pattern	**Repeat length**	***Allelic score***	**CGG/AGG Pattern**	**Repeat length**	***Allelic score***
(CGG)$_{15}$	15[$]	**15**	(CGG)$_{10}$AGG(CGG)$_{9}$	20[$]	**49**
(CGG)$_{18}$	18[$]	**18**	(CGG)$_{10}$AGG(CGG)$_{9}$	20[$]	**49**
(CGG)$_{10}$AGG(CGG)$_{9}$	20[$]	**49**	(CGG)$_{25}$AGG(CGG)$_{9}$	35	**109**
(CGG)$_{10}$AGG(CGG)$_{9}$	20[$]	**49**	(CGG)$_{10}$AGG(CGG)$_{9}$	20[$]	**49**
(CGG)$_{10}$AGG(CGG)$_{9}$	20[$]	**49**	(CGG)$_{20}$	20[$]	**20**
(CGG)$_{10}$AGG(CGG)$_{9}$	20[$]	**49**	(CGG)$_{10}$AGG(CGG)$_{19}$	30	**59**
(CGG)$_{11}$AGG(CGG)$_{9}$	21[$]	**53**	(CGG)$_{12}$AGG(CGG)$_{10}$	23[$]	**58**
(CGG)$_{9}$AGG(CGG)$_{13}$	23[$]	**49**	(CGG)$_{9}$AGG(CGG)$_{19}$	29	**55**
(CGG)$_{9}$AGG(CGG)$_{13}$	23[$]	**49**	(CGG)$_{12}$AGG(CGG)$_{32}$	45[#]	**80**
(CGG)$_{13}$AGG(CGG)$_{9}$	23[$]	**61**	(CGG)$_{24}$	24[$]	**24**
(CGG)$_{10}$AGG(CGG)$_{13}$	24[$]	**53**	(CGG)$_{13}$AGG(CGG)$_{16}$	30	**68**
(CGG)$_{16}$AGG(CGG)$_{9}$	26[$]	**73**	(CGG)$_{29}$	29	**29**
(CGG)$_{9}$AGG(CGG)$_{18}$	28	**54**	(CGG)$_{9}$AGG(CGG)$_{28}$	38	**64**
(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	29	**189**	(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	29	**189**
(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	29	**189**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{10}$AGG(CGG)$_{8}$AGG(CGG)$_{9}$	29	**201**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{8}$	29	**204**
(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	29	**189**	(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	29	**189**
(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	29	**189**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	29	**189**	(CGG)$_{9}$AGG(CGG)$_{12}$AGG(CGG)$_{9}$	32	**201**
(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	29	**189**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{10}$	30	**190**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{16}$	37	**212**
(CGG)$_{9}$AGG(CGG)$_{10}$AGG(CGG)$_{9}$	30	**193**	(CGG)$_{10}$AGG(CGG)$_{11}$AGG(CGG)$_{9}$	32	**213**
(CGG)$_{9}$AGG(CGG)$_{10}$AGG(CGG)$_{9}$	30	**193**	(CGG)$_{10}$AGG(CGG)$_{11}$AGG(CGG)$_{9}$	32	**213**
(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{10}$	30	**190**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{8}$	39	**828**
(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**	(CGG)$_{10}$AGG(CGG)$_{10}$AGG(CGG)$_{10}$	32	**210**
(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{9}$AGG(CGG)$_{10}$AGG(CGG)$_{9}$	30	**193**	(CGG)$_{10}$AGG(CGG)$_{14}$AGG(CGG)$_{9}$	35	**225**
(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{10}$	30	**190**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{20}$	41	**216**
(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{9}$AGG(CGG)$_{10}$AGG(CGG)$_{9}$	30	**193**	(CGG)$_{10}$AGG(CGG)$_{11}$AGG(CGG)$_{9}$	32	**213**
(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{10}$	30	**190**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{19}$	40	**215**
(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{10}$	30	**190**	(CGG)$_{9}$AGG(CGG)$_{11}$AGG(CGG)$_{9}$	31	**197**
(CGG)$_{9}$AGG(CGG)$_{10}$AGG(CGG)$_{9}$	30	**193**	(CGG)$_{10}$AGG(CGG)$_{11}$AGG(CGG)$_{9}$	32	**213**
(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**	(CGG)$_{10}$AGG(CGG)$_{10}$AGG(CGG)$_{9}$	31	**209**
(CGG)$_{9}$AGG(CGG)$_{10}$AGG(CGG)$_{9}$	30	**193**	(CGG)$_{10}$AGG(CGG)$_{11}$AGG(CGG)$_{9}$	32	**213**
(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**	(CGG)$_{10}$AGG(CGG)$_{12}$AGG(CGG)$_{9}$	33	**217**
(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{10}$	31	**206**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{10}$	31	**206**
(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{11}$	31	**191**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{17}$	38	**213**
(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{19}$	39	**199**	(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$AGG(CGG)$_{15}$	45[#]	**771**
(CGG)$_{10}$AGG(CGG)$_{5}$	16[$]	**45**	(CGG)$_{10}$AGG(CGG$_{9}$AGG(CGG)$_{10}$	31	**206**
(CGG)$_{10}$AGG(CGG$_{9}$	20[$]	**49**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**
(CGG)$_{9}$AGG(CGG)$_{10}$	20[$]	**46**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{8}$	29	**204**
(CGG)$_{10}$AGG(CGG)$_{9}$	20[$]	**49**	(CGG)$_{10}$AGG(CGG)$_{9}$AGG(CGG)$_{9}$	30	**205**

(Continued)

TABLE 2 | Continued

Allele 1			Allele 2		
CGG/AGG Pattern	**Repeat length**	***Allelic score***	**CGG/AGG Pattern**	**Repeat length**	***Allelic score***
$(CGG)_{10}AGG(CGG)_9$	20[§]	**49**	$(CGG)_{10}AGG(CGG)_{10}AGG(CGG)_{10}$	32	**210**
$(CGG)_{10}AGG(CGG)_9$	20[§]	**49**	$(CGG)_{10}AGG(CGG)_9AGG(CGG)_9$	30	**205**
$(CGG)_9AGG(CGG)_{10}$	20[§]	**46**	$(CGG)_{10}AGG(CGG)_{11}AGG(CGG)_9$	32	**213**
$(CGG)_9AGG(CGG)_{10}$	20[§]	**46**	$(CGG)_{10}AGG(CGG)_8AGG(CGG)_9$	29	**201**
$(CGG)_{20}$	20[§]	**20**	$(CGG)_9AGG(CGG)_9AGG(CGG)_9$	29	**189**
$(CGG)_{10}AGG(CGG)_9$	20[§]	**49**	$(CGG)_{10}AGG(CGG)_9AGG(CGG)_9$	30	**205**
$(CGG)_9AGG(CGG)_{10}$	20[§]	**46**	$(CGG)_{10}AGG(CGG)_9AGG(CGG)_9$	30	**205**
$(CGG)_{10}AGG(CGG)_9$	20[§]	**49**	$(CGG)_{10}AGG(CGG)_9AGG(CGG)_9$	30	**205**
$(CGG)_9AGG(CGG)_{10}$	20[§]	**46**	$(CGG)_{10}AGG(CGG)_8AGG(CGG)_9$	29	**201**
$(CGG)_{10}AGG(CGG)_9$	20[§]	**49**	$(CGG)_{10}AGG(CGG)_9AGG(CGG)_9$	30	**205**
$(CGG)_{10}AGG(CGG)_{11}$	22[§]	**51**	$(CGG)_{12}AGG(CGG)_7AGG(CGG)_9$	30	**229**
$(CGG)_{13}AGG(CGG)_9$	23[§]	**61**	$(CGG)_{10}AGG(CGG)_9AGG(CGG)_9$	30	**205**
$(CGG)_{12}AGG(CGG)_{10}$	23[§]	**58**	$(CGG)_9AGG(CGG)_{10}AGG(CGG)_9$	30	**193**
$(CGG)_{13}AGG(CGG)_9$	23[§]	**61**	$(CGG)_{10}AGG(CGG)_9AGG(CGG)_9$	30	**205**
$(CGG)_{13}AGG(CGG)_9$	23[§]	**61**	$(CGG)_{10}AGG(CGG)_9AGG(CGG)_9$	30	**205**
$(CGG)_{13}AGG(CGG)_9$	23[§]	**61**	$(CGG)_{10}AGG(CGG)_9AGG(CGG)_9$	30	**205**
$(CGG)_{27}$	27	**27**	$(CGG)_{10}AGG(CGG)_9AGG(CGG)_9$	30	**205**
$(CGG)_9AGG(CGG)_9AGG(CGG)_9$	29	**189**	$(CGG)_{29}AGG(CGG)_9$	39	**125**
$(CGG)_9AGG(CGG)_9AGG(CGG)_9$	29	**189**	$(CGG)_{38}$	38	**38**
$(CGG)_9AGG(CGG)_{19}$	29	**55**	$(CGG)_{10}AGG(CGG)_9AGG(CGG)_9$	30	**205**
$(CGG)_9AGG(CGG)_9AGG(CGG)_9$	29	**189**	$(CGG)_{10}AGG(CGG)_{20}$	31	**60**
$(CGG)_9AGG(CGG)_{19}$	29	**55**	$(CGG)_{10}AGG(CGG)_9AGG(CGG)_{10}$	31	**206**
$(CGG)_{10}AGG(CGG)_9AGG(CGG)_9$	30	**205**	$(CGG)_{22}AGG(CGG)_9$	32	**97**
$(CGG)_{10}AGG(CGG)_{19}$	30	**59**	$(CGG)_{10}AGG(CGG)_9AGG(CGG)_{10}$	31	**206**
$(CGG)_9AGG(CGG)_{10}AGG(CGG)_9$	30	**193**	$(CGG)_{10}AGG(CGG)_{19}$	30	**59**
$(CGG)_{30}$	30	**30**	$(CGG)_{10}AGG(CGG)_9AGG(CGG)_{10}$	31	**206**
$(CGG)_{10}AGG(CGG)_9AGG(CGG)_9$	30	**205**	$(CGG)_{33}AGG(CGG)_9$	43	**141**
$(CGG)_{10}AGG(CGG)_9AGG(CGG)_9$	30	**205**	$(CGG)_{10}AGG(CGG)_{19}$	30	**59**
$(CGG)_{10}AGG(CGG)_{20}$	31	**60**	$(CGG)_{10}AGG(CGG)_9AGG(CGG)_{23}$	44	**219**
$(CGG)_9AGG(CGG)_{21}$	31	**57**	$(CGG)_{10}AGG(CGG)_8AGG(CGG)_{25}$	45[#]	**217**
$(CGG)_{10}AGG(CGG)_{20}$	31	**60**	$(CGG)_{10}AGG(CGG)9AGG(CGG)_{10}$	31	**206**

Homoallelism for CGG-repeat length (black background) and homozygosity for both CGG-repeat length and AGG pattern (allelic score in green background).
[#]intermediate size. [§]normal "low zone" alleles (see section "Discussion").

belonging to the *dissimilar* group, *FMR1* mPCR showed extreme skewing (85%) toward the smallest "low zone" allele.

DISCUSSION

Our study focused on developing a tool to score and evaluate the complexity of the *FMR1* gene repetitive tract structure. To this end, a mathematical model was designed that computes the *FMR1* gene CGG repeat length, as well as the AGG interspersion number and pattern. The output, a number designated *allelic score*, deciphers a functional model to predict the complexity of allele combinations. Two cohorts of young, healthy, and potentially fertile females were used independently for development and validation studies. The fact that two statistically significant groups, *equivalent* and *dissimilar*, were identified in both cohorts, justified the pooling of data. Furthermore, the identification of two groups shows the model's ability to compare the complexity of the two alleles. Interestingly, the *dissimilar* group is enriched with "low zone" heterozygous samples (herein defined as CGG $\leq$ 26). It has been proposed that these "low zone" alleles may exert negative effects, although controversial (Spitzer et al., 2012; Mailick et al., 2014; Gleicher et al., 2015; Ruth et al., 2016). Another study claims that normal *FMR1* repeat length outside 26 > CGG > 34 concur with a higher XCI skew, a putative mechanism underlying the ovarian reserve impairment (as assessed by AMH), particularly in infertile older females (Barad et al., 2017). Moreover, the AGG "protective" effect toward a decreased risk of ovarian malfunction was observed in females carrying premutated alleles with two or more interspersions (Lekovich et al., 2018). According to our model, these alleles would show a high *allelic score*, which seems to suggest a correlation between the allelic complexity and a protective effect. Replication of these results is still required

using larger control and patient cohorts. Nonetheless, with this mathematical model developed to calculate the *FMR1 allelic score*, further research can now be undertaken with a different perspective in terms of *FMR1* characterization.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committe of Centro Hospitalar Universitário do Porto. Written informed consent from the participants was obtained in accordance with the national legislation (lei 12/2005) and the institutional requirements. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

PJ conceived and designed the study together with BR. AN developed the mathematical model and performed the statistical analysis with BR who also carried out laboratory work, analyzed the data, and drafted the manuscript. FS performed methylation/inactivation studies. EV-F, NM, IM, and RS provided critical feedback, helped conduct the research, and contributed toward the manuscript. All authors discussed the final results and critically reviewed the manuscript.

FUNDING

This work was supported by national funds: FCT/MCTES (Fundação para a Ciência e a Tecnologia) – Project Reference SFRH/BD/136398/2018 to BR, UMIB (Unidade Multidisciplinar de Investigação Biomédica) – Reference UIDP/00215/2020 and UIDB/00215/2020, DEFI (Departamento de Ensino, Formação e Investigação) – Reference 2015-DEFI/145/12, and CESAM (Centro de Estudos do Ambiente e do Mar) – Reference UIDP/50017/2020 and UIDB/50017/2020.

ACKNOWLEDGMENTS

We gratefully acknowledge the Center for Medically Assisted Procreation/Public Gamete Bank, Centro Materno-Infantil do Norte Dr. Albino Aroso (CMIN), Centro Hospitalar Universitário do Porto (CHUP). Without the invaluable help of all collaborators (potential donors, clinicians, nurses, and embryologists), our work would not have been possible. Special thanks to Isabel Sousa Pereira for recruiting cohort 1 participants. We are also grateful for important feedback from international reviewers.

REFERENCES

Allen, R. C., Zoghbi, H. Y., Annemarie, B., Moseley, H. M. R., and Belmont, J. W. (1992). Methylation of *Hpall* and *Hhal* sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. *Am. J. Hum. Genet.* 51, 1229–1212. doi: 10.1158/1538-7445.am2014-ct404

Ardui, S., Race, V., de Ravel, T., Van Esch, H., Devriendt, K., and Matthijs, G. (2018). Detecting AGG interruptions in females with a *FMR1* premutation by long-read single-molecule sequencing: a 1 year clinical experience. *Front. Genet.* 9:150. doi: 10.3389/fgene.2018.00150

Barad, D. H., Darmon, S., Weghofer, A., Latham, G. J., Wang, Q., Kushnir, V. A., et al. (2017). Association of skewed X-chromosome inactivation with *FMR1* CGG repeat length and anti-mullerian hormone levels: a cohort study. *Reprod. Biol. Endocrinol.* 15:34. doi: 10.1186/s12958-017-0250-9

Biancalana, V., Glaeser, D., McQuaid, S., and Steinbach, P. (2015). EMQN best practice guidelines for the molecular genetic testing and reporting of fragile X syndrome and other fragile X-associated disorders. *Eur. J. Hum. Genet.* 23, 417–425. doi: 10.1038/ejhg.2014.185

Domniz, N., Ries-Levavi, L., Cohen, Y., Marom-Haham, L., Berkenstadt, M., Pras, E., et al. (2018). Absence of AGG interruptions is a risk factor for full mutation expansion among israeli fmr1 premutation carriers. *Front. Genet.* 9:606. doi: 10.3389/fgene.2018.00606

Galvão, A., Vale-Fernandes, E., Pereira, I. S., Fraga, S., Lourenço, C., Morgado, A., et al. (2017). "Applicants for oocyte donors from the Portuguese public gamete bank: who are they?," in *CMIN SUMMIT 17 - Inovações e Controvérsias na Saúde da Mulher e da Criança* (Porto).

Gleicher, N., Yu, Y., Himaya, E., Barad, D. H., Weghofer, A., Wu, Y., et al. (2015). Early decline in functional ovarian reserve in young women with low (CGGn <26) *FMR1* gene alleles. *Transl. Res.* 166, 502–507. doi: 10.1016/j.trsl.2015.06.014

Latham, G. J., Coppinger, J., Hadd, A. G., and Nolin, S. L. (2014). The role of AGG interruptions in fragile X repeat expansions: a twenty-year perspective. *Front. Genet.* 5:244. doi: 10.3389/fgene.2014.00244

Lekovich, J., Man, L., Xu, K., Canon, C., Lilienthal, D., Stewart, J. D., et al. (2018). CGG repeat length and AGG interruptions as indicators of fragile X–associated diminished ovarian reserve. *Genet. Med.* 20, 957–964. doi: 10.1038/gim.2017.220

Maia, N., Loureiro, J. R., Oliveira, B., Marques, I., Santos, R., Jorge, P., et al. (2016). Contraction of fully expanded *FMR1* alleles to the normal range: predisposing haplotype or rare events? *J. Hum. Genet.* 62, 1–7. doi: 10.1038/jhg.2016.122

Mailick, M. R., Hong, J., Rathouz, P., Baker, M. W., Greenberg, J. S., Smith, L., et al. (2014). Low-normal *FMR1* CGG repeat length: phenotypic associations. *Front. Genet.* 5:309. doi: 10.3389/fgene.2014.00309

Man, L., Lekovich, J., Rosenwaks, Z., and Gerhardt, J. (2017). Fragile X-associated diminished ovarian reserve and primary ovarian insufficiency from molecular mechanisms to clinical manifestations. *Front. Mol. Neurosci.* 10:290. doi: 10.3389/fnmol.2017.00290

Manor, E., Gonen, R., Sarussi, B., Keidar-Friedman, D., Kumar, J., Tang, H. T., et al. (2019). The role of AGG interruptions in the *FMR1* gene stability: a survey in ethnic groups with low and high rate of consanguinity. *Mol. Genet. Genomic Med.* 7, 1–14. doi: 10.1002/mgg3.946

McGinty, R. J., and Mirkin, S. M. (2018). Cis- and trans-modifiers of repeat expansions: blending model systems with human genetics. *Trends Genet.* 34, 448–465. doi: 10.1016/j.tig.2018.02.005

Mila, M., Alvarez-Mora, M. I., Madrigal, I., and Rodriguez-Revenga, L. (2018). Fragile X syndrome: an overview and update of the *FMR1* gene. *Clin. Genet.* 93, 197–205. doi: 10.1111/cge.13075

Monaghan, K. G., Lyon, E., and Spector, E. B. (2013). ACMG standards and guidelines for fragile X testing: a revision to the disease-specific supplements to the standards and guidelines for clinical genetics laboratories of the American

College of medical genetics and genomics. *Genet. Med.* 15, 575–586. doi: 10.1038/gim.2013.61

Rehnitz, J., Alcoba, D. D., Brum, I. S., Dietrich, J. E., Youness, B., Hinderhofer, K., et al. (2018). *FMR1* expression in human granulosa cells increases with exon 1 CGG repeat length depending on ovarian reserve. *Reprod. Biol. Endocrinol.* 16:65 doi: 10.1186/s12958-018-0383-5

Ruth, K. S., Bennett, C. E., Schoemaker, M. J., Weedon, M. N., Swerdlow, A. J., and Murray, A. (2016). Length of *FMR1* repeat alleles within the normal range does not substantially affect the risk of early menopause. *Hum. Reprod.* 31, 2396–2403. doi: 10.1093/humrep/dew204

Spitzer, L. T., Johnstone, E. B., Huddleston, H. G., Cedars, L. M., Davis, G., and Fujimoto, V. (2012). *FMR1* repeats and ovarian reserve: CGG repeat number does not influence antral follicle count. *J. Fertil. Vitr.* 02, 10–13. doi: 10.4172/2165-7491.1000105

Wang, Q., Kushnir, V. A., Darmon, S., Barad, D. H., Wu, Y., Zhang, L., et al. (2017). Reduced RNA expression of the *FMR1* gene in women with low (CGGn<26) repeats. *Fertil. Steril.* 108:e143. doi: 10.1016/j.fertnstert.2017.07.432

Yrigollen, C. M., Durbin-Johnson, B., Gane, L., Nelson, D. L., Hagerman, R., Hagerman, P. J., et al. (2012). AGG interruptions within the maternal *FMR1* gene reduce the risk of offspring with fragile X syndrome. *Genet. Med.* 14, 729–736. doi: 10.1038/gim.2012.34

Yrigollen, C. M., Sweha, S., Durbin-Johnson, B., Zhou, L., Berry-Kravis, E., Fernandez-Carvajal, I., et al. (2014). Distribution of AGG interruption patterns within nine world populations. *Intractable Rare Dis. Res.* 3, 153–161. doi: 10.5582/irdr.2014.01028

Zar, J. H. (2010). *Bioestatistical Analysis*, 5th Edn. Upper Saddle River, NJ: Prentice Hall, doi: 10.1007/978-1-4939-2917-7

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.