



DATA NOTE

REVISED **The FRAXA and FRAXE allele repeat size of boys from the Avon Longitudinal Study of Parents and Children (ALSPAC)**
[version 2; peer review: 2 approved]

Rosie Clark ¹, Steven Gregory ¹, Susan Ring ^{1,2}, Patricia Jacobs³, Sarah Ennis ³, Anna Murray ⁴, Genette Ellis ¹, Jean Golding ¹, Kate Northstone ¹, Marcus Pembrey¹

¹Population Health Sciences, Bristol Medical School, University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, UK

²MRC Integrative Epidemiology Unit, University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, UK

³Human Genetics, Faculty of Medicine, University of Southampton, University Road, Southampton, SO17 1BJ, UK

⁴Medical School, University of Exeter, Barrack Road, Exeter, EX2 5DW, UK

v2 **First published:** 02 Aug 2019, 4:116 (<https://doi.org/10.12688/wellcomeopenres.15342.1>)
Latest published: 27 Jan 2020, 4:116 (<https://doi.org/10.12688/wellcomeopenres.15342.2>)

Abstract

The FRAXA and FRAXE alleles of the FMR1 and FMR2 genes located on the X chromosome contain varying numbers of trinucleotide repeats. Large numbers of repeats at FRAXA (full mutations) manifest as Fragile X syndrome, associated with mental impairment that affects males more severely. In this paper, we present the dataset of frequencies of FRAXA and FRAXE repeat size extracted from DNA samples collected from boys enrolled in the Avon Longitudinal Study of Parents and Children (ALSPAC). DNA data were extracted from samples collected in ALSPAC clinics from several types of samples: cord blood, venepuncture blood taken at 43 months, 61 months, seven years or nine years. The DNA was amplified at FRAXA and FRAXE using fluorescent PCR in the Wessex Regional Genetics Laboratory, Salisbury District Hospital. The mean repeat size for FRAXA is 28.92 (S.D. 5.44), the median 30 and the range 8 to 68. There were particularly high numbers of boys with repeat sizes of 20 (10.67%) and 23 (7.35%). The mean repeat size for FRAXE is 17.41 (S.D. 3.94), with median of 16 and range of 0 to 61. There is a relatively high degree of variation of the FRAXA repeat size particularly and we suggest the extensive data available from the ALSPAC study opens up areas of research into understanding phenotypes associated with relatively unexplored repeat sizes. This could be particularly interesting for the lower repeat sizes occurring with high frequency at FRAXA in this population. As the data can be linked to exposures and phenotypes, it will provide a resource for researchers worldwide.

Keywords

ALSPAC, longitudinal cohort, FRAXA, FRAXE, Triplet repeats

Open Peer Review

Reviewer Status

	Invited Reviewers	
	1	2
version 2 (revision) 27 Jan 2020	 report	 report
version 1 02 Aug 2019	 report	 report

- Randi J. Hagerman** , University of California, Davis, Sacramento, USA
- Timing Liu** , National University of Singapore, Singapore, Singapore
University of Cambridge, Cambridge, UK
- Samuel S. Chong** , National University of Singapore, Singapore, Singapore
National University Health System, Singapore, Singapore

Any reports and responses or comments on the article can be found at the end of the article.



This article is included in the [Avon Longitudinal](#)

[Study of Parents and Children \(ALSPAC\)](#) gateway.

Corresponding author: Rosie Clark (r.clark@bristol.ac.uk)

Author roles: **Clark R:** Writing – Original Draft Preparation, Writing – Review & Editing; **Gregory S:** Formal Analysis, Writing – Review & Editing; **Ring S:** Data Curation, Writing – Review & Editing; **Jacobs P:** Conceptualization, Funding Acquisition, Investigation, Supervision, Writing – Review & Editing; **Ennis S:** Investigation, Writing – Review & Editing; **Murray A:** Investigation, Writing – Review & Editing; **Ellis G:** Formal Analysis, Writing – Review & Editing; **Golding J:** Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Northstone K:** Formal Analysis, Writing – Review & Editing; **Pembrey M:** Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the Wellcome Trust [102215], the ALSPAC core programme grant, and a Programme Grant awarded to Patricia Jacobs. The final cleaning and writing of this paper was undertaken using funds from the John Templeton Foundation [60828]. The UK Medical Research Council and Wellcome [102215] and the University of Bristol currently provide core support for ALSPAC. This publication is the work of the authors; Rosie Clark and Jean Golding will serve as guarantors for the contents of this paper. A comprehensive list of grants funding is available on the ALSPAC website (<http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf>). *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

Copyright: © 2020 Clark R *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Clark R, Gregory S, Ring S *et al.* **The FRAXA and FRAXE allele repeat size of boys from the Avon Longitudinal Study of Parents and Children (ALSPAC) [version 2; peer review: 2 approved]** Wellcome Open Research 2020, 4:116 (<https://doi.org/10.12688/wellcomeopenres.15342.2>)

First published: 02 Aug 2019, 4:116 (<https://doi.org/10.12688/wellcomeopenres.15342.1>)

REVISED Amendments from Version 1

We have added information into the paper about the pattern of data across ethnicities.

We have added further information about differences in allele classification criteria in the strengths and limitations section, including giving the percentages referring to the classification of 41 to 54 in the intermediate allele class. Given the differences in allele classifications, we have provided supplementary material of a table giving all the individual genotype information (frequency of each count of repeats at FRAXA and FRAXE). This will allow researchers to classify or split the data in whichever way they prefer.

Any further responses from the reviewers can be found at the end of the article

Introduction

The mutation associated with the Fragile X syndrome is within a CGG trinucleotide repeat array in the FMR1 gene, identified in 1991 (Pieretti *et al.*, 1991). The gene is located in the Xq27.3 region of the long arm of the X chromosome (Lubs, 1969; Sutherland, 1979; Sutherland, 1977). The array contains a variable number of repeats, and each repeat size is regarded in genetic terms as an FMR1 allele (often referred to as a FRAXA allele) (reviewed in Pembrey *et al.*, 2001). Similarly, FRAXE syndrome is caused by the expansion of a GCC repeat in the FMR2 gene (Knight *et al.*, 1993). FRAXA allele classes are categorised by the number of repeats and defined as full mutation (>200), premutation (61–200), intermediate (41–60), common (11–40) and minimal (<11) (Murray *et al.*, 1996). FRAXE allele classes are similarly defined, but are slightly different for intermediate (31–60) and common (11–31) classes.

Full mutations of FRAXA (> 200 CGG repeats with silencing of the FMR1 gene) are more common than those of FRAXE (Knight *et al.*, 1996); in the literature, the population prevalence of full FRAXA mutations in males was estimated to be between 1 in 4000 and 1 in 5000, though acknowledging biases due to informed consent (Chandrasekara *et al.*, 2017; Youings *et al.*, 2000). For females the estimation is lower, at around 1 in 8000 (Crawford *et al.*, 1999). FRAXE full mutation prevalence in males has been estimated at 1 in 23,400 (Youings *et al.*, 2000).

Individuals with a full mutation inherit it from a female who carries a full mutation or premutation (Pembrey *et al.*, 2001). Around a third of female gene carriers are intellectually impaired, usually less severely than males (Pembrey *et al.*, 2001). Most males with Fragile X syndrome (full mutation) have a severe to moderate degree of mental impairment (Pembrey *et al.*, 2001). The full mutation type for FRAXE syndrome has been found to result in much milder impairment than the FRAXA phenotype (Crawford *et al.*, 1999).

In females, there is a well-established phenotypic effect of FRAXA premutation sized alleles of an increased risk of premature ovarian failure (Chandrasekara *et al.*, 2017; Macpherson *et al.*, 2003; Youings *et al.*, 2000). There is also some evidence of an increased incidence of late-onset neurological disorder characterised by tremor and ataxia (FXTAS) in premutation males, with around 40% at risk, and, to a lesser extent, females, at ~8% (Chandrasekara *et al.*, 2017; Hagerman *et al.*, 2001; Hagerman *et al.*, 2004). Several studies have suggested that

intermediate alleles at both FRAXE and FRAXA may have a role in cognitive impairment in boys, more reliably for FRAXA (Murray *et al.*, 1996; Youings *et al.*, 2000). However, this finding has not been replicated in some other multinational studies (Youings *et al.*, 2000). Specifically, a study investigating FRAXA intermediate alleles in ALSPAC data even found a significant deficiency of special educational needs (SEN) cases in boys with FRAXA intermediate expansions (Ennis *et al.*, 2006).

An issue within the literature is that the definitions of the boundaries of classes of repeat numbers are sometimes different across studies (e.g. changes over time due to knowledge base or differing clinical definitions), thus making it harder to compare frequencies (see Cornish *et al.*, 2008; Pembrey *et al.*, 2001; Youings *et al.*, 2000). The general variation of frequencies of repeat sizes is often overlooked, as are phenotypic associations with relatively low repeat sizes (particularly differences within allele classes). Additionally, some studies have found a correlation between the length of repeats and degree of phenotype, even within allele classes (Cornish *et al.*, 2008).

Here we present a dataset of repeat size frequencies at FRAXA and FRAXE of boys from the ALSPAC study, which could provide an important resource for researchers internationally.

Materials and methods**Ethical approval and consent**

Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees (Birmingham, 2018). Informed consent for the use of data via questionnaires and clinics was obtained from participants following the recommendations of the ALSPAC Ethics and Law Committee at the time. Questionnaires were completed by parents in their own homes and the return of a questionnaire to the study offices was interpreted as giving tacit consent to the study. Full details of the approvals obtained are available from the ethics pages of the study website (<http://www.bristol.ac.uk/alspac/researchers/research-ethics/>). Study members have the right to withdraw their consent for elements of the study or from the study entirely at any time. Consent for biological samples was collected in accordance with the Human Tissue Act (2004).

The ALSPAC sample

The ALSPAC study was designed to include all pregnant mothers with an expected date of delivery between 1st April 1991 and 31st December 1992 living in a specific area in the South West Regional Health Authority of England (Boyd *et al.*, 2013; Fraser *et al.*, 2013). The number of enrolled pregnant mothers was 14541, which resulted in 13988 infants who survived to one year. When the oldest children were approximately seven years of age, an attempt was made to bolster the initial sample with eligible cases who had failed to join the study originally. Including these additional children results in 14,901 infants who survived to one year. Data collection largely consisted of self-completion questionnaires (principally based on psychosocial factors, physical environments and health) filled in by mothers, partners, children and teachers. The study methodology (Golding *et al.*, 2001), enrolment and response rates are given in detail on the study website (<http://bristol.ac.uk/alspac/index.html>). Please note that the study website contains

details of all the data that is available through a fully searchable data dictionary and variable search tool (<http://www.bristol.ac.uk/alspac/researchers/our-data/>). The data presented here on maternal age, parity, parental education levels and socio-economic status were obtained during pregnancy from maternal self-completion questionnaires.

Sample collection

The DNA of children in the study was gathered by the ALSPAC study team; samples were collected from 5275 males in total (Ennis *et al.*, 2006). Permission for DNA extraction was obtained as written consent from the study mothers during pregnancy (for cord blood), as well as in ALSPAC clinics for venepuncture blood taken at 43 months, 61 months, seven years and nine years. Some samples were extracted from buccal wash when the child did not consent to venepuncture. The ALSPAC study team double coded all samples for anonymity.

DNA amplification and analysis

Cord blood samples were collected in heparin at birth and stored at -70°C for five to eight years before DNA extraction (Jones *et al.*, 2000). Blood samples from children at 43 or 61 months were stored for one month to two years before extraction (Jones *et al.*, 2000). Samples from children at seven and nine years were stored at -20°C for up to three weeks before DNA extraction (Jones *et al.*, 2000). The samples were provided to the laboratory as 250ng aliquots in 96 well plates with eight wells on each plate left empty for laboratory control, consisting of DNA with known CGG repeat number and water controls (Ennis *et al.*, 2006). When there was both a cord blood and clinic sample were available for the same boy, the clinic sample was used to maximise genotyping efficiency, as heparin can inhibit PCR, and to minimise maternal contamination issues (Ennis *et al.*, 2006). The DNA was amplified using fluorescent PCR (involving fluorescently labelled oligonucleotide primers) across the CGG FRAXA repeat and GCC FRAXE repeat. This took place in the Wessex Regional Genetics Laboratory.

Two multiplex PCR reactions were carried out, the details of which are given elsewhere (Jones *et al.*, 2000; Murray *et al.*, 1996; Youings *et al.*, 2000). For both multiplex reactions, products were denatured and separated on a polyacrylamide urea gel and then run on an ABI 373A Stretch machine for eight hours at 40 W, 2500 V and 35 mA (Murray *et al.*, 1996). The gel data were analysed on 672 GENESCAN software (ABI/Perkin Elmer) and then imported into GENOTYPER software (ABI/Perkin Elmer) to assign alleles (Murray *et al.*, 1996).

Data processing

Genotype data were returned to the ALSPAC Study (University of Bristol). The anonymised dataset is in SPSS format and analysis was carried out using a combination of SPSS (version 11) statistical software and SAS (SAS Institute Inc., Cary, N.C.). The dataset includes: the ALSPAC pregnancy identifier (anonymised), whether the pregnancy was singleton/twins/triplets, the category of FRAXA/E repeat size, and the number of repeats at FRAXA/E.

Dataset

In all, there were 5087 children for whom the number of repeats (RPTs) at FRAXA and 5070 at FRAXE were assessed. Of

these, 192 were additional cases recruited at seven years old. The proportion of the 7676 boys in the study population for whom RPT data were obtained was 66.3% for FRAXA and 66.1% for FRAXE. Those who were missed included: those for whom permission had not been given and thus DNA was not available (2401); those where the assay had not worked (164), including those with high levels of premutations and the full mutation; and additionally, where there was not enough DNA, or samples had a heterozygous genotype indicating the possibility of two X chromosomes or maternal contamination (27) (Ennis *et al.*, 2006). As PCR is not possible for amplification of large premutations/full mutations, these individuals are not included in this dataset.

Table 1 shows the percentage of boys for whom there is FRAXA repeat data within each category of sociodemographic

Table 1. Proportion (and percentage) of boys for whom the number of FRAXA repeats (RPTs) have been obtained by sociodemographic attributes of their parents. The education level ranges from less than O-level/GCSE (A) to degree level qualification (E). The social class categories are as previously defined ('Standard occupation classification', 1990).

Variable	Proportion with FRAXA RPTs	% with FRAXA RPTs
Maternal education		
A (lowest)	755 /1322	57.1
B	418 /641	65.2
C	1577 /2237	70.5
D	1139 /1439	79.2
E (highest)	618 /812	76.1
Paternal education		
A (lowest)	974 /1612	60.4
B	343 /515	66.6
C	950 /1334	71.2
D	1129 /1579	71.5
E (highest)	863 /1143	75.5
Social class (father's Occupation)		
I (Higher professional)	479 /630	76.0
II	1373 /1916	71.7
IIINm	436 /619	70.4
IIIM	1187 /1779	66.7
IV	361 /535	67.5
V (unskilled)	110 /163	67.5
Maternal age		
<20	151 /327	46.2
20–24	778 /1378	56.5
25–29	1892 /2813	67.3
30–34	1397 /1998	69.9
35+	547 /766	71.4
Parity of the mother		
0	1997 /2988	66.8
1	1596 /2368	67.4
2	683 /989	69.1
3+	255 /404	63.1

attributes. Among the parents of boys in the ALSPAC study, FRAXA repeat information was slightly more likely to be available when the education levels achieved by the mother and the father were higher, the father’s occupation was of a higher social class, and the mothers were older. However, there was no difference in availability by the number of pregnancies the mother had had where the baby reached viability (parity). The patterns in regard to FRAXE were similar.

The error rate for FRAXA and FRAXE genotypes (i.e. CGG repeat number difference >1) were 0.52% and 0.65% respectively, calculated by comparing duplicate samples. Duplicates that were not matched and samples which failed to amplify were categorised as failed (Ennis *et al.*, 2006).

FRAXA

The mean number of repeats (RPTs) at FRAXA is 28.92 (S.D. 5.44), with a median of 30 and a range of 8 to 68. Using the arbitrary definitions in the literature (Murray *et al.*, 1996), only three of the boys would be categorised as having a premutation

(defined as 61-200 RPTs) and 168 as intermediary (41-60 RPTs) (Table 2). In all, approximately 11% have an RPT >32. The frequency distribution also has a substantially high number of individuals with repeat numbers of 20 (10.67%) and 23 (7.35%) disrupting the normal distribution around the mean (Figure 1).

Table 2. The frequency of boys in each class of FRAXA repeats. This is also expressed as a percentage of the number of boys for which the number of repeats (RPT) data is available.

Class of FRAXA repeat	Frequency	Percentage
Minimal (<11)	<5	0.06
Common (11-40)	4,913	96.58
Intermediate (41-60)	168	3.3
Premutation (61-200) unmethylated	<5	0.06
Total	5,087	100

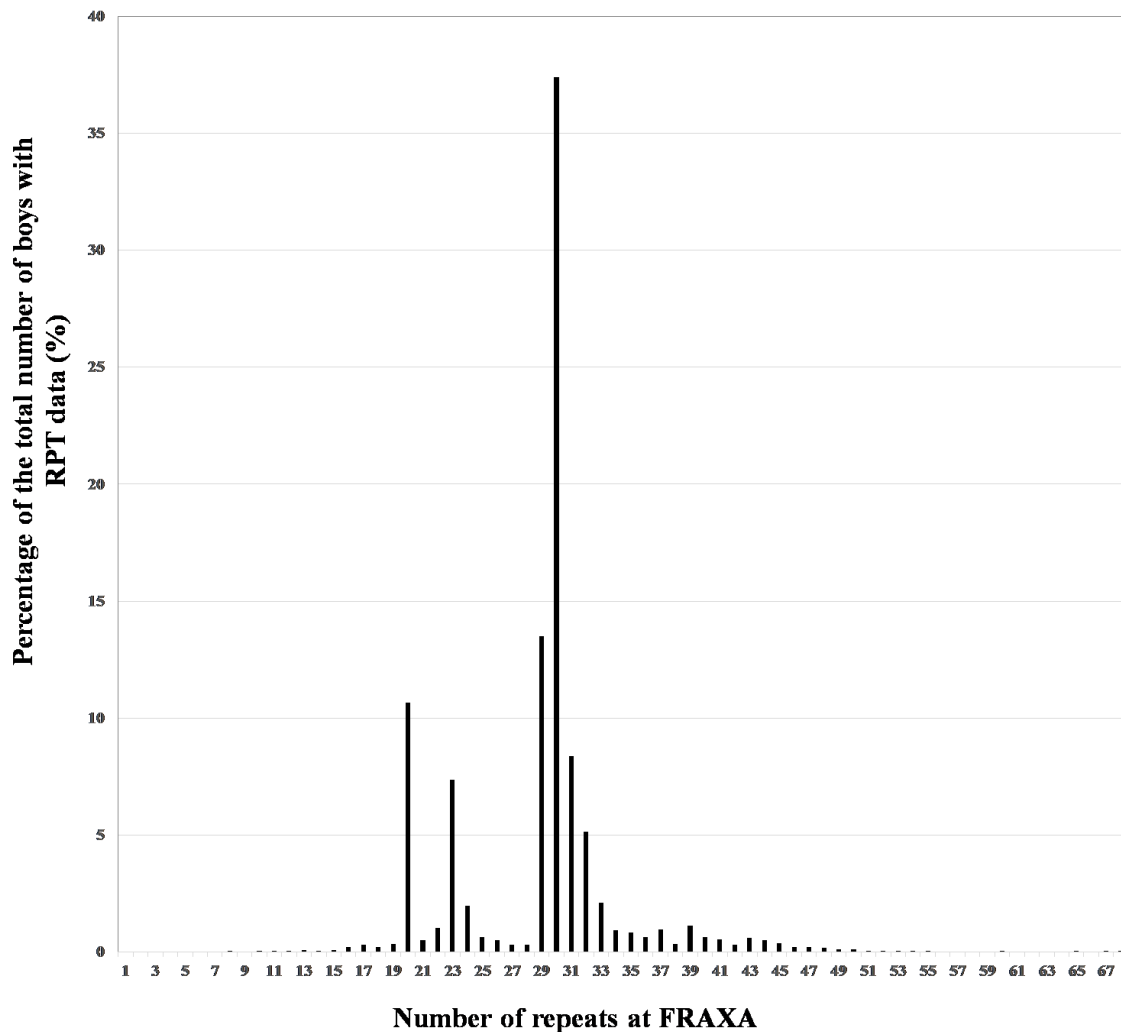


Figure 1. Distribution of FRAXA repeats. Graph of the distribution of number of repeats at FRAXA, for all boys for which valid number of repeats (RPT) data was obtained.

FRAXE

The mean number of RPTs at FRAXE is 17.41 (S.D. 3.94), with a median of 16 and a range of 0 to 61. Using the definitions in the literature (Murray *et al.*, 1996), one of the boys would be classed as having a premutation (Table 3). There is less variation in the FRAXE RPTs compared to the FRAXA RPTs (Figure 2), but a suggestion of a bimodal distribution. The pattern of data is unchanged across ethnicities for both FRAXE and FRAXA (5% non-white).

Strengths and limitations of the dataset

One of the strengths of this data is that it includes a very large number of individuals and we present the general variation

Table 3. The frequency of boys in each class of FRAXE repeats. In the third column, this is expressed as a percentage of the number of boys for which the number of repeats (RPT) data is available.

Class of FRAXE repeat	Frequency	Percentage
Minimal (<11)	43	0.85
Common (11-30)	4983	98.28
Intermediate (31-60)	43	0.85
Premutation (61-200) unmethylated	<5	0.02
Total	5070	100

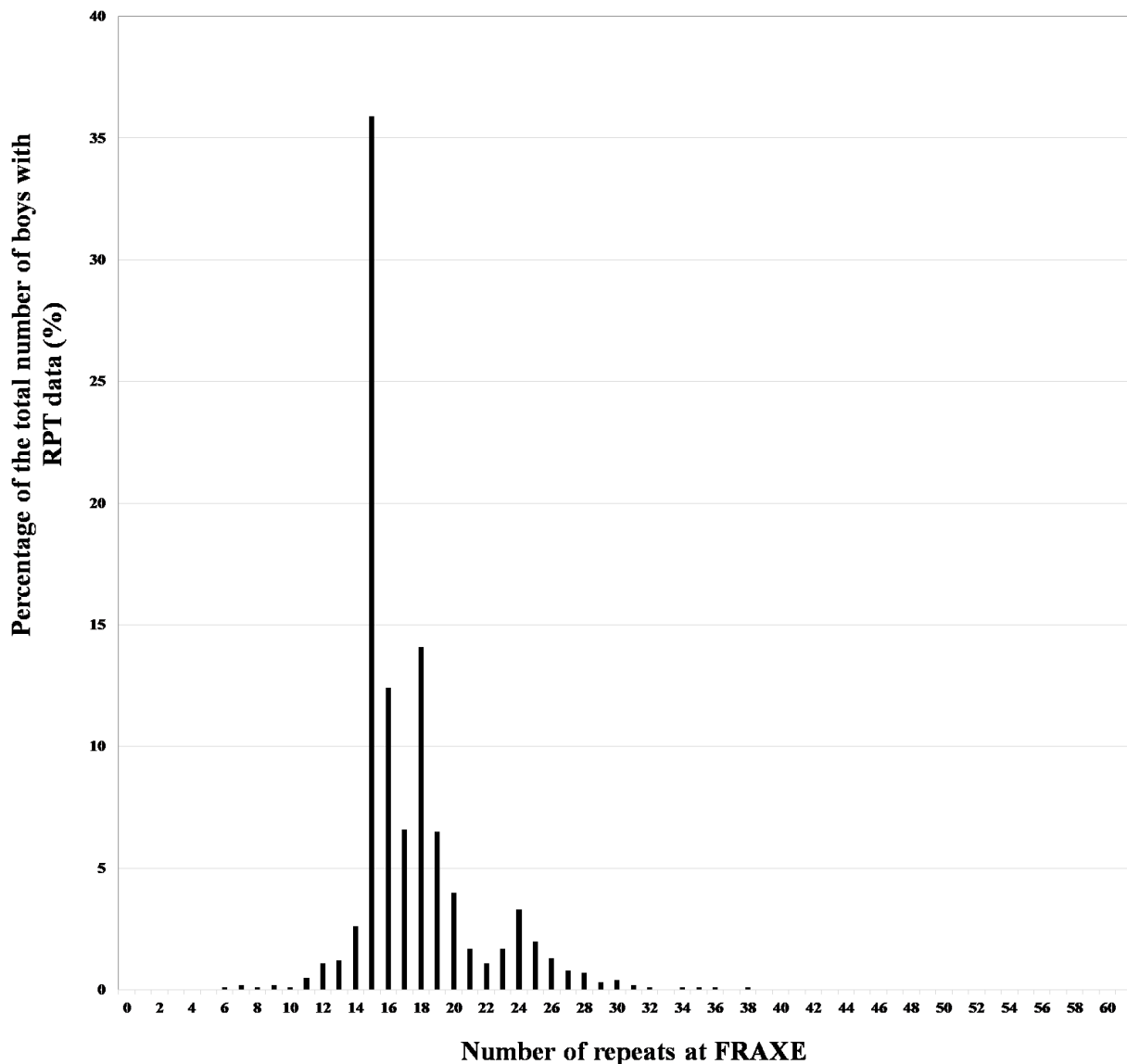


Figure 2. Distribution of FRAXE repeats. Graph of the distribution of number of repeats at FRAXE, for all boys for which valid number of repeats (RPT) data was obtained.

across RPT sizes, which is sometimes overlooked in the literature. Additionally, given the extensive information gathered about children in the ALSPAC study, there is potential to look at the RPT size variation to ascertain if there are any correlations with other variables. This could open up interesting avenues of research into the phenotypes associated with different RPT sizes, particularly for the allele categories rarely studied in the literature so far (such as ‘minimal’ and ‘common’) and for the relatively unexpected peak seen in the FRAXA data around 20/23 RPTs (Figure 1). Very few studies have looked at phenotypes associated with very low numbers of repeats at FRAXA or FRAXE. A few smaller studies did find association of alleles of less than 26 repeats at FRAXA with increased risk of developing fertility problems, behavioural problems and/or having children with developmental disability or psychiatric illness; this has not been replicated by larger studies and as such remains controversial (Kraan *et al.*, 2018).

There was some bias in that the boys without RPT data were more likely to have had mothers who were younger, less well educated and of lower social class (Table 1). This needs to be taken into account in further studies, including subsequent correlations with variables from the ALSPAC data. Another limitation is that ascertaining repeat sizes of large pre-mutation or full mutation alleles was not possible and these data are not represented in the dataset. Additionally, as discussed in the introduction the boundaries of classes of repeat numbers vary across studies, and particularly for FRAXA there is another commonly used distinction (intermediate as 41-54 repeats and pre-mutation as 55-200 repeats). However, if we classify the boundaries this way, then the intermediate percentage only decreases a very small amount to 3.24 and pre-mutation increases to 0.12. Although differing boundaries limits comparison across studies, we have included a full table of genotype data in Supplementary material (DOI: [10.6084/m9.figshare.11695440](https://doi.org/10.6084/m9.figshare.11695440)) such that researchers can classify whichever way they prefer.

Data availability

Underlying data

ALSPAC data access is through a system of managed open access. The steps below highlight how to apply for access to the

data included in this paper and all other ALSPAC data. Please read the ALSPAC access policy (http://www.bristol.ac.uk/media-library/sites/alspac/documents/researchers/data-access/ALSPAC_Access_Policy.pdf), which describes the process of accessing the data and biological samples in detail and outlines the costs associated with doing so.

1. You may also find it useful to browse our fully searchable research proposals database (<https://proposals.epi.bristol.ac.uk/>), which lists all research projects that have been approved since April 2011.
2. Please submit your research proposal (<https://proposals.epi.bristol.ac.uk/>) for consideration by the ALSPAC Executive Committee using the online process. You will receive a response within 10 working days to advise you whether your proposal has been approved.

If you have any questions about accessing data, please email: alspac-data@bristol.ac.uk (data) or bbl-info@bristol.ac.uk (samples).

The ALSPAC data management plan (<http://www.bristol.ac.uk/media-library/sites/alspac/documents/researchers/data-access/alspac-data-management-plan.pdf>) describes in detail the policy regarding data sharing, which is through a system of managed open access.

Extended data

The supplementary material for this paper is a table of the genotype data: the frequencies of numbers of repeats at FRAXA and FRAXE. This can be accessed at DOI: [10.6084/m9.figshare.11695440](https://doi.org/10.6084/m9.figshare.11695440).

Acknowledgements

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses.

References

- Birmingham K: **Pioneering ethics in a longitudinal study.** Policy Press, 2018; 136. [Publisher Full Text](#)
- Boyd A, Golding J, Macleod J, *et al.*: **Cohort Profile: the ‘children of the 90s’—the index offspring of the Avon Longitudinal Study of Parents and Children.** *Int J Epidemiol.* 2013; 42(1): 111–127. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chandrasekara B, Wijesundera S, Chong SS, *et al.*: **Prevalence of Fragile X Syndrome among children receiving special education and carrier states in first degree relatives.** *Ceylon Med J.* 2017; 62(2): 92–96. [PubMed Abstract](#) | [Publisher Full Text](#)
- Cornish KM, Li L, Kogan CS, *et al.*: **Age-dependent cognitive changes in carriers of the fragile X syndrome.** *Cortex.* 2008; 44(6): 628–636. [PubMed Abstract](#) | [Publisher Full Text](#)
- Crawford DC, Meadows KL, Newman JL, *et al.*: **Prevalence and phenotype consequence of FRAXA and FRAXE alleles in a large, ethnically diverse, special education-needs population.** *Am J Hum Genet.* 1999; 64(2): 495–507. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ennis S, Murray A, Youings S, *et al.*: **An investigation of FRAXA intermediate allele phenotype in a longitudinal sample.** *Ann Hum Genet.* 2006; 70(Pt 2): 170–180. [PubMed Abstract](#) | [Publisher Full Text](#)
- Fraser A, Macdonald-Wallis C, Tilling K, *et al.*: **Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort.** *Int J Epidemiol.* 2013; 42(1): 97–110. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Golding J, Pembrey M, Jones R, *et al.*: **ALSPAC—the Avon Longitudinal Study of Parents and Children. I. Study methodology.** *Paediatr Perinat Epidemiol.* 2001; 15(1): 74–87. [PubMed Abstract](#) | [Publisher Full Text](#)

Hagerman RJ, Leavitt BR, Farzin F, *et al.*: **Fragile-X-associated tremor/ataxia syndrome (FXTAS) in females with the FMR1 premutation.** *Am J Hum Genet.* 2004; **74**(5): 1051–1056.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Hagerman RJ, Leehay M, Heinrichs W, *et al.*: **Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X.** *Neurology.* 2001; **57**(1): 127–130.

[PubMed Abstract](#) | [Publisher Full Text](#)

Jones RW, Ring S, Tyfield L, *et al.*: **A new human genetic resource: a DNA bank established as part of the Avon longitudinal study of pregnancy and childhood (ALSPAC).** *Eur J Hum Genet.* 2000; **8**(9): 653–60.

[PubMed Abstract](#) | [Publisher Full Text](#)

Knight SJ, Flannery AV, Hirst MC, *et al.*: **Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation.** *Cell.* 1993; **74**(1): 127–134.

[PubMed Abstract](#) | [Publisher Full Text](#)

Knight SJ, Ritchie RJ, Chakrabarti L, *et al.*: **A study of FRAXE in mentally retarded individuals referred for fragile X syndrome (FRAXA) testing in the United Kingdom.** *Am J Hum Genet.* 1996; **58**(5): 906–13.

[PubMed Abstract](#) | [Free Full Text](#)

Kraan CM, Bui QM, Field M, *et al.*: **FMR1 allele size distribution in 35,000 males and females: a comparison of developmental delay and general population cohorts.** *Genet Med.* 2018; **20**(12): 1627–1634.

[PubMed Abstract](#) | [Publisher Full Text](#)

Lubs HA: **A marker X chromosome.** *Am J Hum Genet.* 1969; **21**(3): 231–44.

[PubMed Abstract](#) | [Free Full Text](#)

Macpherson J, Waghorn A, Hammans S, *et al.*: **Observation of an excess of fragile-X premutations in a population of males referred with spinocerebellar**

ataxia. *Hum Genet.* 2003; **112**(5–6): 619–620.

[PubMed Abstract](#) | [Publisher Full Text](#)

Murray A, Youings S, Dennis N, *et al.*: **Population screening at the FRAXA and FRAXE loci: molecular analyses of boys with learning difficulties and their mothers.** *Hum Mol Genet.* 1996; **5**(6): 727–735.

[PubMed Abstract](#) | [Publisher Full Text](#)

Office of Population Censuses and Surveys: **Standard occupation classification.** London: HMS, 1990.

[Reference Source](#)

Pembrey ME, Barnicoat AJ, Carmichael B, *et al.*: **An assessment of screening strategies for fragile X syndrome in the UK.** *Health Technol Assess.* 2001; **5**(7): 1–95.

[PubMed Abstract](#) | [Publisher Full Text](#)

Pieretti M, Zhang FP, Fu YH, *et al.*: **Absence of expression of the FMR-1 gene in fragile X syndrome.** *Cell.* 1991; **66**(4): 817–822.

[PubMed Abstract](#) | [Publisher Full Text](#)

Sutherland GR: **Fragile sites on human chromosomes: demonstration of their dependence on the type of tissue culture medium.** *Science.* 1977; **197**(4300): 265–266.

[PubMed Abstract](#) | [Publisher Full Text](#)

Sutherland GR: **Heritable fragile sites on human chromosomes. III. Detection of fra(X)(q27) in males with X-linked mental retardation and in their female relatives.** *Hum Genet.* 1979; **53**(1): 23–27.

[PubMed Abstract](#) | [Publisher Full Text](#)

Youings SA, Murray A, Dennis N, *et al.*: **FRAXA and FRAXE: the results of a five year survey.** *J Med Genet.* 2000; **37**(6): 415–421.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 2

Reviewer Report 20 March 2020

<https://doi.org/10.21956/wellcomeopenres.17209.r37691>


© 2020 Chong S et al. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Timing Liu 

¹ Department of Pediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

² Faculty of Biology, University of Cambridge, Cambridge, UK

Samuel S. Chong 

¹ Department of Pediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

² Khoo Teck Puat - National University Children's Medical Institute, National University Health System, Singapore, Singapore

We thank the authors for indicating the percentage of non-white individuals in the study, and for the table of detected allele frequencies. We understand the authors' reluctance to re-analyze the 164 samples for whom no alleles could be detected, given that the original analyses were probably generated more than two decades ago. However, this does mean that frequencies are available only for FRAXA alleles of not more than 68 repeats, and only FRAXE alleles of 61 or less repeats. As such, the reported frequencies of normal and intermediate FRAXA and FRAXE allele spectra can be accepted without question, but the premutation allele spectra and frequencies are incomplete in this study. On a separate note, we previously wrongly assumed that the relatively high percentage of the 20 and 23 repeat FRAXA alleles was novel. It has come to our attention, however, that there have been studies reporting relatively high frequencies of 20 and/or 23 repeats in certain European groups^{1,2}.

References

1. Larsen LA, Vuust J, Nystad M, Evseeva I, et al.: Analysis of FMR1 (CGG)(n) alleles and DXS548-FRAXAC1 haplotypes in three European circumpolar populations: traces of genetic relationship with Asia. *Eur J Hum Genet.* 2001; **9** (9): 724-7 [PubMed Abstract](#) | [Publisher Full Text](#)
2. Mingroni-Netto RC, Angeli CB, Auricchio MT, Leal-Mesquita ER, et al.: Distribution of CGG repeats and FRAXAC1/DXS548 alleles in South American populations. *Am J Med Genet.* 2002; **111** (3): 243-52 [PubMed Abstract](#) | [Publisher Full Text](#)

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Molecular genetics of repeat expansion disorders.

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 03 March 2020

<https://doi.org/10.21956/wellcomeopenres.17209.r37690>

© 2020 Hagerman R. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Randi J. Hagerman 

Medical Investigation of Neurodevelopmental Disorders (MIND) Institute, University of California, Davis, Sacramento, CA, USA

I thank the authors for adding the table of the allele sizes through the CGG repeats that they were able to detect, but I am still concerned about the 164 patients with a high end premutation or full mutation that they were not able to detect. Therefore their data cannot be used for prevalence figures for the premutation ranges or higher. If they are able to use the newer PCR techniques that can detect high end premutation carriers, then their data will be useful for prevalence studies.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: fragile X research

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 16 October 2019

<https://doi.org/10.21956/wellcomeopenres.16752.r36650>

© 2019 Chong S et al. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Timing Liu 

¹ Department of Pediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

² Faculty of Biology, University of Cambridge, Cambridge, UK

Samuel S. Chong 

¹ Department of Pediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

² Khoo Teck Puat - National University Children's Medical Institute, National University Health System, Singapore, Singapore

Clark *et al.* present some nice data on the distribution of triplet repeat sizes at the FRAXA and FRAXE fragile sites in a large population sample from the ALSPAC study.

It would be good for the authors to provide information about the break-down of the repeat size distribution by ethnic groups, given the UK population's ethnic heterogeneity, unless all the samples are of single origin Caucasian ethnicity. This point is pertinent given that the high frequency of 20- and 23-repeat FRAXA alleles in the ALSPAC cohort has not been previously reported elsewhere.

The authors have suggested that future studies could look at phenotypes associated with very low numbers of repeats at FRAXA or FRAXE, and certainly the 20- and 23-repeat FRAXA alleles would be interesting study candidates. Besides the repeat size, repeat structure (e.g. repeat interruptions) might also be worth a look with regard to transmission stability and cognitive phenotype variability/expressivity.

Another point that needs addressing is the fate of the 164 samples that were excluded from further analysis due to standard PCR amplification failures. The authors acknowledged that some of these samples likely include high-level pre-mutation and full mutation samples. It is a pity that there were no follow-up analyses of these samples. Judging from the slab gel instruments used, the PCR genotyping was likely performed in the 1990s and Southern blot re-analysis then may not have been an option if DNA quantity was limited. However, the authors ought to now consider re-analysing these few samples using the newer and more powerful TP-PCR methods capable of sizing all premutations and detecting all full mutations. This would render a more complete picture of the FRAXA and FRAXE repeat size spectrum and allele distribution and frequencies in the study cohort. The paper's title, and the statement "a dataset of repeat size frequencies at FRAXA and FRAXE of boys from the ALSPAC study", would then also be more justified.

Given the different repeat size classifications adopted in different studies, it would be useful if the individual genotype information be made available in the supplementary information. This will enable readers to re-plot and re-classify allele sizes according to their preferred standard.

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Partly

Are sufficient details of methods and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Molecular genetics of repeat expansion disorders.

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.

Author Response 22 Jan 2020

Rosie Clark, University of Bristol, Oakfield House, Oakfield Grove, Bristol, UK

We thank the reviewer for their constructive comments.

In response to the point regarding phenotypes associated with low numbers of repeats, while we agree that this would be worth looking at, the aim of this paper is to highlight the data that are available so that other researchers can investigate different research questions. This is what we understand to be the purpose of a Data Note paper.

Regarding the comment about PCR amplification, this is a good point but these methods are beyond our abilities and facilities currently. However, data from the ALSPAC biobank is available to researchers as outlined in our access policy.

All other comments have been addressed by changes to the text and new supplementary material provided.

Competing Interests: No competing interests were disclosed.

Reviewer Report 12 August 2019

<https://doi.org/10.21956/wellcomeopenres.16752.r36130>

© 2019 Hagerman R. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Randi J. Hagerman

Medical Investigation of Neurodevelopmental Disorders (MIND) Institute, University of California, Davis, Sacramento, CA, USA

Clark *et al.* describes the alleles that were obtained from a remarkable sample, the ALSPAC study which followed the offspring of women who delivered between April 1, 1991 and December 31, 1992. PCR testing was done for FRAXA and FRAXE alleles and reported here. Only boys were studied and only about 66% had successful results. The reasons for a low uptake included a lack of consent, which is understandable, but there were 164 samples where the assay did not work either because it was a high premutation or a full mutation or not enough DNA or a heterogeneous genotype with possibly 2 Xs or maternal contamination. They need to clarify the numbers that had a large premutation or a full mutation to help with the prevalence figures and if possible use Southern Blot testing or newer PCR techniques to clarify who had a full mutation. For the data that they did obtain there was a large number with a lower repeat in the FRAXA group, specifically 10.67% with 20 repeats and 7.35% with 23 repeats which is unusual compared to previous studies. It seems that the authors have missed a great opportunity to

correlate these low repeat numbers with the cognitive data that seems to be available at multiple time points in the ALSPAC study. The authors should do this if possible.

It is difficult to compare their prevalence results to previous studies because the standard definition of the premutation as 55 to 200 repeats is not used. Instead the premutation is defined as 61-200 and the intermediate is 41-60. Most other studies will use 41 to 54 repeats for the intermediate or gray zone. It would be helpful for them to define these categories in the introduction and use the standard definitions.

Under ALSPAC sample first paragraph "enrollment" is misspelled.

Overall this study demonstrates a high level of low CGG repeats in the FRAXA analysis but they have the opportunity to correlate this with the clinical data including cognitive testing and they need to do this.

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Partly

Are sufficient details of methods and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: fragile X research

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 22 Jan 2020

Rosie Clark, University of Bristol, Oakfield House, Oakfield Grove, Bristol, UK

We thank the reviewer for their constructive comments. Regarding their first point about numbers of individuals with a full mutation, these methods are beyond our abilities and facilities currently, but data from the ALSPAC biobank is available to researchers for further use as outlined in our access policy.

In response to the reviewer's second point about correlating repeat numbers with cognitive data, we agree that these points would be worth looking at but the aim of this Data Note paper is to highlight the data that are available so that other researchers can investigate different research questions.

Enrolment is the English spelling so we have not corrected this.

Other comments have been addressed as amendments to the text and supplementary material.

Competing Interests: No competing interests were disclosed.

