FMRI CGG Repeats: Reference Levels and Race–Ethnic Variation in Women With Normal Fertility (Study of Women’s Health Across the Nation)

Lisa M. Pastore, PhD¹, Ani Manichaikul, PhD^{2,3},
Xin Q. Wang, MS³, and Joel S. Finkelstein, MD⁴

Reproductive Sciences
2016, Vol. 23(9) 1225–1233
© The Author(s) 2016
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1933719116632927
rs.sagepub.com

Abstract

FMRI premutation carriers (55–199 CGG repeats), and potentially women with high normal (35–44) or low normal (<28) CGG repeats, are at risk of premature ovarian aging. The scarcity of population data on CGG repeats <45 CGG, and variation in race–ethnicity, makes it difficult to determine true associations. DNA was analyzed for *FMRI* CGG repeat lengths from 803 women (386 caucasians, 219 African Americans, 102 Japanese, and 96 Chinese) from the US-based Study of Women’s Health Across the Nation (SWAN). Participants had ≥ 1 menses in the 3 months before enrollment, ≥ 1 pregnancy, no history of infertility or hormonal therapy, and menopause ≥ 46 years. Statistical analyses used Fisher exact tests. Among these women with normal reproductive histories, significant *FMRI* repeat length differences were found across race–ethnicity for both the longer ($P = .0002$) and the shorter ($P < .0001$) alleles. The trinucleotide length variance was greater for non-Asian than Asian women ($P < .0001$), despite identical median values. Our data indicate that short allele lengths <25 CGG on one or both alleles are more common in non-Asian than Asian women. We confirm the minor allele in the 35 to 39 CGG range among Asians as reported previously. Only 2 (0.3%) premutation carriers were identified. These data demonstrate that *FMRI* distributions do vary by race–ethnicity, even within the “normal” range. This study indicates the need to control for race–ethnicity in *FMRI* ovarian aging research and provides race–ethnic population data for females separated by allele.

Keywords

race–ethnicity, genetic variation, *FMRI*, diminished ovarian reserve, women, primary ovarian insufficiency, reference population

Introduction

Specific trinucleotide repeat lengths in the *FMRI* gene (HGNC:3775) are associated with ovarian dysfunction. Women with premutation level repeats in this gene (55–199 CGG) are at increased risk of premature ovarian failure,^{1–3} alternatively termed primary ovarian insufficiency. Some reports have suggested that women with 35 to 44 CGGs,^{4,5} 35 to 54 CGGs,⁶ and <28 CGG repeat lengths⁷ may be associated with less severe ovarian dysfunction, often manifest as diminished ovarian reserve. These findings are less consistent, as reviewed by Pastore and Johnson.⁸ The associations between ovarian dysfunction and the *FMRI* CGG repeat lengths shorter than 55 CGG are speculative, however, because there are few reports that examined the distribution of *FMRI* CGG repeats in the normal range among fertile women and whether the distribution varies by ethnicity.

Population data are required to discriminate the normal range of *FMRI* CGG repeat lengths from the abnormal range. In particular, population data on the CGG repeat length in women with normal reproductive histories are required to

discriminate whether repeat lengths in women with early ovarian aging differ from the norm. Population data stratified by the higher and lower alleles in females are also important, as some researchers have raised the possibility that the lower allele confers increased risk of early ovarian aging⁷ and most of the research on the association between *FMRI* and ovarian aging has focused exclusively on the higher allele.

¹ Department of Obstetrics, Gynecology and Reproductive Medicine, Stony Brook Medicine, Stony Brook, NY, USA

² Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA

³ Division of Biostatistics and Epidemiology, Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA

⁴ Department of Medicine, Massachusetts General Hospital, Boston, MA, USA

Corresponding Author:

Lisa M. Pastore, Department of Obstetrics, Gynecology and Reproductive Medicine, Stony Brook University, Stony Brook, NY 11794-8091, USA.
Email: pastorestudies@gmail.com

The Study of Women's Health Across the Nation (SWAN) is a longitudinal multirace, multiethnic, multisite cohort study of a random sample of premenopausal or early perimenopausal women. The SWAN study collected data on a wide variety of physiologic, epidemiologic, and psychosocial changes, including a detailed characterization of the menopausal transition and, from some participants, a DNA sample. We assessed the distribution of *FMRI* CGG repeat lengths in SWAN to provide a robust estimate of the distribution of CGG repeats in a well-characterized population of women with normal reproductive histories.

Existing Population Data in the Range of <CGG₄₅ *FMRI* Repeat Lengths

The distribution of *FMRI* CGG repeat lengths less than 45 in the general population has been assessed in 14 population-based studies (Table 1). These reports include cohorts from Asia, Europe, and North America. The modal repeat length is 29 and 30 CGGs in most of these reports, with the exception of 2 Japanese studies (modal value was 27 or 28 CGG repeats),^{14,19} a study of Canadian newborns (modal value was 28),¹⁶ and a study of Israeli men (modal value of 27).¹² Five of these articles reported separate results for females.^{9-11,16,19} Nine of the 10 studies that included females reported the total allelic distribution,^{9,11,15-21} and none of those publications reported the *FMRI* distribution of the higher allele separate from the lower allele. The recent article by Voorhuis et al¹⁰ did report separate results for the higher and lower alleles among females from the general population, but their population included women with early menopause at ages 40 to 45 and infertility.

Existing Population Data on *FMRI* Differences by Race–Ethnicity

Using 8 general population studies of the distribution of *FMRI* CGG repeat lengths, Genereux and Laird reported that Asian and non-Asian populations followed similar distributional curves (their analysis was restricted to ≥ 40 CGG repeat lengths), but the curves were “almost completely nonoverlapping.”²² The Asian curve was left-shifted, which indicated a lower mean repeat length for intermediate and expanded alleles among Asians. The study by Otsuka et al¹⁹ might suggest potentially an overall lower distribution at repeats <40 CGG in Japanese populations, given their finding of a slightly lower modal value than reported elsewhere (Table 1). Huang et al compared their Chinese study population¹⁸ with caucasians from Crawford et al²³ and African Americans from Eichler et al²⁴ and found a statistically smaller number of CGG repeats among the Chinese ($\chi^2 P < .0001$). That same study also compared their Chinese data to reported Japanese¹⁹ and Indonesian²⁵ populations and found no differences in the CGG distributions.

The Present Study

This study aimed

- to provide population data on the *FMRI* CGG repeat distribution including details on the “normal” range of <45 repeats,
- to provide those data stratified by the higher and lower allele in addition to the total allelic distribution, and
- to examine racial–ethnic differences in the distributions.

The hypothesis was that the distribution of CGG repeats in the *FMRI* gene in women would vary by ethnicity, even within what is classified as the normal range (ie, <45 CGG repeats).^{26,27} The population was restricted to females with normal reproductive histories so as to serve as a comparative reference for research on early ovarian aging in women. This work is important for the interpretation of *FMRI* genetic results in studies of early ovarian aging, as these detailed data are lacking from the literature.

Methods

These data are from the SWAN study of middle-aged women from across the United States (www.swanstudy.org). The SWAN study is comprised of 2 stages: a cross-sectional telephone or in-home survey conducted between November 1995 and October 1997 and a longitudinal investigation as women age and experience the menopause transition. The SWAN enrollees were required to be premenopausal, not taking hormones, and between 42 to 52 years of age at enrollment. For further details on the SWAN study design, see Sowers et al.²⁸ Briefly, community-based samples of women were drawn from 7 geographic locations: Boston, Massachusetts; Chicago, Illinois; Detroit, Michigan; Los Angeles, California; Newark, New Jersey; Oakland, California; and Pittsburgh, Pennsylvania. Women self-identifying primarily with one or more of the following 5 racial/ethnic groups were interviewed: caucasian, African American, Chinese, Hispanic, and Japanese. Each site studied caucasians and one other ethnic group.

Participation in the SWAN Genetics Study

During years 6 and 7 of the study, materials, including buccal cells and whole blood, were collected to provide a source of DNA. Enrollees from the New Jersey site did not participate in the SWAN Genetics Study and that was the only site to enroll Hispanics; thus, DNA from Hispanic women are not available. This study was approved by the institutional review board (IRB) at each clinical site. In addition, each site obtained a certificate of confidentiality.

As described elsewhere,²⁹ 1538 women at the 6 SWAN sites provided whole blood samples with successful B-lymphocyte transformation. Currently, 1523 samples are available with approval to investigators outside the SWAN organization (<http://datawarehouse.swanrepository.com/aboutInventory.php>, as of March 6, 2014).

Table 1. Published General Population Studies Detailing the Normal *FMR1* CGG Repeat Distribution (<45 CGG).

	Author	Study Population Size; Country	Modal CGG Repeat Length	Was Analysis Conducted by Race/Ethnicity?	Analysis and/or Data on the Lower Allele?	Type of Study, Including If the Population Restricted on Reproductive Hx and/or Family Hx of FXS
Female only	Kim et al ⁹	5470; Korea	29, followed by 30	NA	No	Reproductive age women without a family history of FXS
	Voorhuis et al ¹⁰	3368; the Netherlands	31 on higher allele, 30 on lower allele	No	Yes	Cohort of women in a national breast cancer screening program restricted to those with natural menopause at age $\geq$ 40 years
	Weiss et al ¹¹	9329; Israel	30, followed by 29	Yes (Ashkenazi vs non-Ashkenazi Jewish)	No	Reproductive age women without a family history of FXS
Male only	Falick-Zaccai et al ¹²	40; Israel	28, followed by 27	No	NA	Unaffected Tunisian Jewish male control group
	Faradz et al ²⁵	1069; Indonesia	29, followed by 30	NA	NA	General population study
	Fernandez-Carvajal et al ¹³	5267; Spain	30, followed by 29	NA	NA	Newborn screening (blood spots)
Both females and males	Richards et al ¹⁴	142; Japan	28, followed by 29	NA	NA	Normal males though some were fathers of FXS offspring
	Crawford et al ¹⁵	3553; United States	30, followed by 29	Yes (caucasian vs AA)	No	Children only in this population
	Dawson et al ¹⁶	2000; Canada	28, followed by 27 and 29	No	No	Blood spots from consecutive newborns in 1 province, half male and half female
	Fu et al ¹⁷	492 alleles; United States + Canada + Europe data	29, followed by 30 and 28	Yes (no differences found)	No	Adults, with partial information of FXS family history
	Huang et al ¹⁸	1113; China	29, followed by 30	Yes (comparison to other reports)	No	Unaffected general population from mainland China
	Otsuka et al ¹⁹	837; Japan	27, followed by 26	NA	No	General population study (absence of major illness was confirmed by clinician)
	Strom et al ²⁰	119232; United States	30, followed by 29 and 31	No	No	General population study from a <i>FMR1</i> testing provider. All data are from people >14 years old
	Tzeng et al ²¹	200; Taiwan	29, followed by 30	NA	No	General population study of 300 chromosomes from 100 women and 100 men

Abbreviations: AA, African-American; FXS, fragile X syndrome; Hx, history; NA, not available.

Participation in This Fragile X Analysis

Eligibility for this *FMR1* analysis required women to (a) be premenopausal through age 45, (b) have a stored DNA sample, and (c) undergo menopause at the age of 46 or older. Women who had undergone either natural or surgical menopause were eligible for inclusion in the study, provided that they had still had periods after the age of 45. This definition has been used elsewhere to define a normal age at menopause.³⁰ In addition, women were excluded:

- if they had ever taken fertility medications,
- if they had ever had a period of 12 months when they could not become pregnant despite regular sexual activity without contraception,
- if they had never been pregnant, or
- if the answer to any of these 3 questions was missing.

This selection algorithm yielded 805 eligible participants. Two samples had insufficient DNA volume in the vial, yielding 803 samples available for this analysis. The DNA assays and the analysis of these coded data were approved by the University of Virginia (UVA) IRB (#11448).

Race–Ethnicity

Race–ethnicity was self-determined by respondents and was obtained by asking the following open-ended question: “How would you describe your primary racial or ethnic group?” This variable was available for all 803 women in this study.

FMR1 Assays

Molecular diagnostic testing was performed by the UVA Molecular Diagnostics Laboratory using capillary electrophoresis with peripheral venous blood samples on an automated ABI 3700 automated DNA sequencer. The PCR reaction contains 2 primer sets, one flanking the CGG repeat area in the 5′ untranslated region of *FMR1* and an internal control consisting of primers spanning a highly polymorphic region near the promotor of the androgen receptor gene.³¹ This assay approach yields highly robust counts of the CGG repeat length on the X chromosomes up to 100 repeats and accuracy of ± 1 CGG repeat.³² Samples that displayed only a single peak were presumed to be homozygous, as opposed to having a full mutation (>200 CGG) on 1 allele, given that these study participants represented a normal comparison population.

The assay results identified 4 women with 3 X alleles (17/30/40 caucasian, 21/29/43 Chinese, 21/29/43 caucasian, and 29/37/39 Chinese). This chromosomal change can be associated with an increased risk of learning and speech delays and occurs in one of every 1000 newborn girls (<http://ghr.nlm.nih.gov/condition/triple-x-syndrome>, accessed May 13, 2013). In a study of 1112 prenatal samples, 2 *FMR1* trisomy samples were described,³³ and 1 trisomy identified in another study of 735 newborn females,¹⁶ so this is a rare but not novel observation. We considered these to be incidental findings given no unusual phenotype in these women and were advised to ignore the

smallest of the 3 alleles in our statistical analysis (personal communication, UVA Molecular Diagnostics Lab, May 13, 2013). All CGG lengths for each of these 4 women and their race–ethnicity are presented for readers’ use.

Statistical Analysis

Because women have 2 X chromosomes, their *FMR1* results provide 2 numbers corresponding to the trinucleotide repeat length in each allele. Consistent with prior reports,^{7,10} the allele with the fewer number of CGG repeats was termed the “lower” allele or “allele 1,” and the allele with the greater number of CGG repeats was termed the “higher” allele or “allele 2.”

Standard descriptive statistics were calculated, including mean, standard deviation, median, and range. Participant characteristics were compared by 1-way analysis of variance (ANOVA) across race–ethnic groups for continuous variables and by χ^2 tests for categorical variables. Discrete categories for CGG repeat length were selected a priori to respond to prior reports of high normal^{4,5} and low normal⁷ repeats potentially being associated with early ovarian aging. The categorical distributions of CGG repeat lengths were compared across race–ethnic groups using a Fisher exact test with $\alpha = .05$. Quantitative comparisons of CGG repeat length, separate for alleles 1 and 2, were completed by 1-way ANOVA, with allele 2 counts analyzed after log transformation. A nonparametric Levene test for homogeneity of the variance across race–ethnic groups was performed using the aggregate alleles. The statistical analysis was conducted using SAS v9.3 (Cary, North Carolina), and the graphs were created with Stata v13 (StataCorp, Texas) and Excel.

Subsequent to the completion of the primary analysis, additional SWAN questionnaire data became available on whether the women had “visited a doctor for difficulty getting pregnant.” This question was posed to women through the reproductive history questionnaire, which was administered once in conjunction with the 13th annual SWAN follow-up. Among the 805 women in our study cohort, 24 had sought this medical advice. The primary analyses of race–ethnic differences among alleles 1 and 2 were repeated with elimination of these 24 women.

Results

The SWAN analytic cohort had an average age at enrollment of 47 years and an average age at menopause of 52 years (Table 2). The proportion of women who were married or cohabitating varied by race–ethnicity ($P < .0001$). Rates of ever having smoked varied by ethnicity ($P < .0001$), with this health history least reported by Chinese participants (7%). As reflective of our eligibility criteria, all women had been pregnant at least once, although not all of those pregnancies ended in a live birth. The proportion of nulliparous women in this cohort was approximately 7%.

Aggregate Allelic Frequencies

Homozygous CGG repeat lengths were common in all race–ethnic groups. Among caucasians, 43.8% were homozygous,

Table 2. Descriptive Statistics by Different Races in SWAN Study.

Variable	Caucasian (n = 386)	African American (n = 219)	Japanese (n = 102)	Chinese (n = 96)	All (n = 803)	P ^a
Age at enrollment						.3799
Mean (SD)	47.2 (2.64)	47.1 (2.57)	47.6 (2.33)	47.1 (2.50)	47.2 (2.57)	
Median	47.2	47.2	47.7	47.1	47.2	
Range	42-52	42-52	42-52	42-52	42-52	
Age at menopause						.5833
Mean (SD)	52.6 (2.48)	52.2 (2.37)	52.5 (2.40)	52.6 (2.28)	52.5 (2.41)	
Median	52.5	52.2	52.1	52.6	52.4	
Range (n)	46-58 (127)	47-58 (93)	46-57 (49)	48-56 (42)	46-58 (311)	
Marital status, %						<.0001 ^b
Single/never married	28 (7.4%)	33 (16.5%)	2 (2.0%)	1 (1.0%)	64 (8.2%)	
Married/living as married	276 (72.6%)	101 (50.5%)	90 (89.0%)	77 (80.2%)	544 (70.0%)	
Separated	8 (2.1%)	14 (7.0%)	2 (2.0%)	1 (1.0%)	25 (3.2%)	
Widowed	5 (1.3%)	13 (6.5%)	2 (2.0%)	4 (4.2%)	24 (3.1%)	
Divorced	63 (16.6%)	39 (19.5%)	5 (5.0%)	13 (13.5%)	120 (15.4%)	
Ever smoked, %	186 (48.2%)	103 (47.9%)	35 (34.3%)	7 (7.3%)	331 (41.4%)	<.0001
Parity						<.0001 ^c
Mean (SD)	2.16 (1.29)	2.71 (1.42)	2.09 (1.05)	2.10 (0.86)	2.29 (1.28)	
Median	2	3	2	2	2	
Range	0-9	0-9	0-5	0-5	0-9	
Mode	2	2	2	2	2	
Gravity						.0016 ^c
Mean (SD)	2.98 (1.66)	3.42 (1.68)	2.95 (1.59)	2.95 (1.31)	3.09 (1.63)	
Median	3	3	3	3	3	
Range	1-12	1-11	1-11	1-7	1-12	
Mode	2	3	2	2	2	

Abbreviations: ANOVA, analysis of variance; SD, standard deviation; SWAN, Study of Women's Health Across the Nation.

^a P value was from either ANOVA for a continuous variable or χ^2 test for a categorical variable.

^b Married/cohabitating versus the others.

^c P value was from nonparametric Kruskal-Wallis rank sum test.

and among African American women, 53.9% were homozygous, with 30 repeats on both alleles being the most common result for both groups. Among Japanese women, 63.7% were homozygous, and 29 repeats on both alleles were the most common result. Among Chinese women, 55.2% were homozygous, and 29/29 and 30/30 repeats were equally observed.

Including both alleles, the most prevalent CGG repeat length among caucasians was 30 CGG (33.0%), followed by 17.2% with 29 CGG and a minor allele at 20 CGG (4.9%). Among African American women, 34.7% had 30 repeats, followed by 25.3% with 29 CGG. Japanese and Chinese women in this cohort had modal values of 29 repeats (35.8% and 35.8%, respectively) and 30 repeats (34.3% and 34.4%, respectively). Chinese women also had a minor allele at 37 repeats (5.2%). The variance of the CGG lengths of the aggregate allelic distributions varied by race-ethnic group ($P < .0001$). The discrete total allelic distribution by race-ethnic group is graphed in Supplement Figure S1.

Higher FMR1 Allele (Allele 2)

The mean CGG repeat length of allele 2 did not vary by race-ethnicity ($P = .17$), with modal values of 29 and 30 across all race/ethnic groups (Table 3). The shortest repeat length among allele 2 in both Asian subgroups was 28 CGG. As shown in the

“detail” portion of Table 3, caucasian and African American women had a wider range of CGG repeat lengths on allele 2, both below 25 repeats and above 44 repeats, than the 2 Asian groups. As anticipated in this fertile female population, few premutation carriers were identified (only 2 carriers in this population: one caucasian with repeat lengths of 30 and 56 and one African American with repeat lengths of 29 and 63) for a prevalence of 0.25% (95% confidence interval [CI]: 0.07%-0.90%). The proportion of women with an intermediate repeat (45-54 CGGs) was 2.8% in caucasians, 2.7% among African Americans, 2.0% in Japanese, and 0% in Chinese, with an overall proportion of 2.4% (95% CI: 1.5%-3.7%; Table 3). Analysis of allele 2 for 4 discrete CGG repeat length categories (<35 CGG, 35-39 CGG, 40-44 CGG, and 45-54 CGG) indicated significant differences by ethnicity ($P = .0002$). For the discrete distribution of the higher allele by ethnic heritage, the reader is referred to Supplement Figure S2. Among the 2 Asian groups, there was a shoulder evident between 35 and 44 repeat lengths that was not apparent in caucasian and African American women (Figure S2).

Lower FMR1 Allele (Allele 1)

The mean CGG repeat length of allele 1 varied by race-ethnicity ($P < .0001$), even though the modal repeat length of allele 1

Table 3. The *FMRI* CGG Repeat Length by Different Race/Ethnicities in SWAN Study (Allele 2).

CGG Repeats	Caucasian (n = 386)	African American (n = 219)	Japanese (n = 102)	Chinese (n = 96)	All (n = 803)
Summary statistics					
Mean (SD)	31.90 (5.12)	31.29 (4.57)	31.42 (4.01)	32.39 (4.25)	31.73 (4.75)
Median	30	30	30	30	30
Range	19-56	20-63	28-52	28-44	19-63
Mode	30	30	29	29	30
Detailed distribution by 5 CGG repeat bands up through the premutation					
<15	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
15-19	2 (0.5%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (0.3%)
20-24	14 (3.6%)	5 (2.3%)	0 (0.0%)	0 (0.0%)	19 (2.4%)
25-29	74 (19.2%)	55 (25.1%)	39 (38.2%)	31 (32.3%)	199 (24.8%)
30-34	223 (57.8%)	134 (61.2%)	44 (43.1%)	37 (38.5%)	438 (54.6%)
35-39	31 (8.0%)	9 (4.1%)	14 (13.7%)	19 (19.8%)	73 (9.1%)
40-44	30 (7.8%)	9 (4.1%)	3 (2.9%)	9 (9.4%)	51 (6.4%)
45-49	8 (2.1%)	5 (2.3%)	1 (1.0%)	0 (0.0%)	14 (1.7%)
50-54	3 (0.8%)	1 (0.5%)	1 (1.0%)	0 (0.0%)	5 (0.6%)
>54	1 (0.3%)	1 (0.5%)	0 (0.0%)	0 (0.0%)	2 (0.3%)

Abbreviations: SD, standard deviation; SWAN, Study of Women's Health Across the Nation.

Table 4. The *FMRI* CGG Repeat Length by Different Race/Ethnicities in SWAN Study (Allele 1).

CGG Repeats	Caucasian (n = 386)	African American (n = 219)	Japanese (n = 102)	Chinese (n = 96)	All (n = 803)
Summary statistics					
Mean (SD)	27.08 (4.46)	27.87 (3.82)	29.29 (2.57)	29.16 (2.04)	27.82 (3.95)
Median	29	29	29	29	29
Range	5-43	8-38	19-37	21-37	5-43
Mode	30	30	30	30	30
Detailed distribution by 5 CGG repeat bands up through the premutation					
<15	2 (0.5%)	1 (0.5%)	0 (0.0%)	0 (0.0%)	3 (0.4%)
15-19	12 (3.1%)	4 (1.8%)	1 (1.0%)	0 (0.0%)	17 (2.1%)
20-24	110 (28.5%)	43 (19.6%)	5 (4.9%)	6 (6.3%)	164 (20.4%)
25-29	80 (20.7%)	73 (33.3%)	46 (45.1%)	43 (44.8%)	242 (30.1%)
30-34	180 (46.6%)	96 (43.8%)	46 (45.1%)	46 (47.9%)	368 (45.8%)
35-39	1 (0.3%)	2 (0.9%)	4 (3.9%)	1 (1.0%)	8 (1.0%)
40-44	1 (0.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.1%)
45-49	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
50-54	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
>54	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

Abbreviations: SD, standard deviation; SWAN, Study of Women's Health Across the Nation.

was 30 among each race-ethnic group (Table 4). There were no intermediate or premutation alleles in the lower allele of this fertile female population. The trinucleotide distribution of allele 1 was tightly clustered (90% of the women) between 25 and 34 CGG within both Asian groups. This is in contrast to caucasians and African Americans where 67.3% and 77.1%, respectively, of the women had between 25 and 34 CGG in allele 1. The CGG repeat length of allele 1 for 3 discrete repeat categories (<20, 20-24, ≥25) significantly varied by race-ethnicity ($P < .0001$). Both caucasian and African-American women were more likely to have fewer than 25 CGG in allele 1 than Japanese or Chinese women. For further detail on the CGG frequencies of allele 1 by ethnic heritage, see Supplemental Figure S3.

Sensitivity Analysis

The above analyses of race-ethnic differences among alleles 1 and 2 were repeated restricted to the women who reported that they had never seen an infertility specialist. Among that smaller data set (excluding 24 women), the results remained essentially unchanged (data not shown).

Discussion

In this study, differences in the *FMRI* CGG repeat length distributions were found by race-ethnicity on both allele 1 ($P < .0001$) and allele 2 ($P = .0002$) in this multiethnic, community-based sample of women with normal reproductive

histories. Although this overall finding was expected, other observations were not, including greater variation observed in the distribution of CGG repeats in African Americans and caucasians than in the Chinese and Japanese women ($P < .0001$). Importantly, we also found that *FMR1* distributions vary according to race-ethnicity even within the normal range of fewer than 45 CGG repeats. These findings have important implications for study design and the interpretation of research results, particularly highlighting the need to control for race-ethnicity when studying early ovarian aging diagnoses in women.

The median number of CGG repeats was identical in all 4 race-ethnic groups in allele 2 and allele 1 (30 and 29 CGG repeats, respectively). Despite that observation, the distributions differed significantly with greater variation of values in caucasians and African Americans than in Asians. The caucasian and African American women also had a greater likelihood of having fewer than 25 CGG repeats on allele 1. A minor allele (or shoulder) around 20 to 23 CGG has been reported previously among caucasians using a combined male plus female data set.¹⁵ Additionally, although the modal values were also nearly identical between the Asian and non-Asian populations (29 and 30 CGG repeats, respectively), there was clearly a minor allele around 35 to 39 CGG in the Asian women (Supplemental Figures S1 and S2), as observed in Indonesians,²⁵ Koreans,⁹ and Chinese populations.^{18,21} These results demonstrate that the use of summary statistics such as medians and modes hides potentially important differences in the distribution of the *FMR1* trinucleotide repeat and highlights the importance of examining a variety of parameters.

We cannot make a simple statement about race-ethnic differences in the distribution of *FMR1* repeat lengths among women with normal reproductive histories. Although others have reported a smaller mean CGG repeat length among pre-mutation carriers and a left-shifted distribution of 55 to 199 repeats in Asians compared with non-Asians,²² the Asian distribution of CGG repeats was not left-shifted in our population-based study of women when examining the entire distributional range. Our data would indicate that having short allele lengths <25 CGG on one or both alleles is much less likely in Asian women than in caucasians and African Americans. Smaller overall CGG repeat lengths in unaffected (ie, without evidence of fragile X syndrome [FXS]) individuals have been reported between a large Chinese study and both caucasians and African Americans ($P < .0001$);¹⁸ note that these prior analyses were not restricted to females and they did not comment on allele lengths below the modal value. Consistent with our findings among women with normal reproductive histories, a study of $n = 62$ oocyte donors found that none of the Asian donors had fewer than 26 CGG repeats, whereas 30% of caucasians and 50% of African Americans had <26 CGG (not significant).³⁴ It appears that heterogeneity exists within the *FMR1* trinucleotide repeat lengths by both race-ethnic group and by the range of repeat (eg, pre-mutation, normal range) in the general population of females. This information is important for the design

and analysis of future study cohorts and for the interpretation of research findings.

The observation of race-ethnic differences also has implications for the interpretation of laboratory reference ranges. As population-based studies continue to be published, meta-analyses may be in the best position to assess the value of creating race-ethnic specific laboratory reference ranges and/or reference ranges specific to a particular phenotype (such as a reference range for females of reproductive age for the assessment of fragile X-associated primary ovarian insufficiency).

The primary limitation of this study is that we cannot say with certainty that none of the cohort has a family history of FXS. Those data were not collected by the SWAN study, but the cohort would by definition exclude women with premature ovarian failure. Theoretically, inclusion of individuals with a family history of FXS would have a small upward bias in the CGG repeat lengths. Also, there was insufficient funding to measure potential variation in the repeat structure within different race/ethnic groups, such as AGG interruptions, which would have been informative given the reported differences in AGG interruption patterns across 9 countries³⁵ and ethnicity.¹¹ The primary strength of this study is the ability to examine race-ethnic genetic differences within a single study due to the SWAN recruitment design. Additional study strengths include the known normal fertility history, the large cohort size, and the community-based sample (so that founder effects are not influencing the distributions). The similarity of the proportion of pre-mutation (0.3%) and intermediate (2.3%) length alleles in this SWAN cohort to the literature^{36,37} provides support for the generalizability of our observations.

In an obstetrics and gynecology context, clinical concerns related to *FMR1* used to be restricted to the risk of FXS offspring, which focused on identifying pre-mutations (55-199 CGG) and full mutations (>200 CGG) among reproductive age women, and the transmission of repeat expansions over generations. In the past 10 years, reports have suggested that 35 to 44 CGG,^{4,5} 35 to 54 CGG,⁶ and <28 CGG repeat lengths⁷ may be associated with early ovarian aging. It has recently been reported that women who have both alleles with fewer than 23 repeats have an increased odds of having a child with a disability.³⁸ There are several articles with very large sample sizes with detailed *FMR1* CGG repeat lengths in the pre-mutation and/or full mutation range (eg, Berkenstadt et al reported the CGG repeat distribution above 49 CGG repeats in 40 000 women,³⁹ and Strom et al reported the higher allele-only distribution in 119 000 men and women²⁰), but few publications reported the distributions within the normal range in women and fewer still with data on allele 1 (Table 1). There is a need for population data below 45 CGG repeat lengths, such as the data included here, in order to interpret *FMR1* research results from female reproductive age women. Race-ethnic differences potentially can explain some of the ambiguity in the research results on the association of *FMR1* repeats with infertility.⁸ The race-ethnic specific population data in women separate for the higher and lower alleles, which have not previously been

published to our knowledge, can serve as a reference for both researchers and those interpreting clinical results.

As people are increasingly screened for genetic conditions, the predictive value of the *FMR1* repeat size will become more significant. Universal screening of pregnant women for *FMR1* premutations has been recommended,^{40,41} although not yet adopted. If universal screening is implemented, questions of future reproductive potential will need to be addressed. Thus, a clear association between the trinucleotide repeat length and the ovarian phenotypes, or a lack of an association, will need to be demonstrated to allow these individuals to make informed reproductive decisions. Normative, population-based studies such as these data from the SWAN cohort are needed in order to distinguish whether *FMR1* CGG repeats in phenotypic cases truly differ from population norms.

Acknowledgments

The authors thank James Bowden and Regina Seaner of the Molecular Diagnostics Lab at the University of Virginia for their *FMR1* processing/reporting of all the SWAN samples. They also thank the SWAN staff at Massachusetts General Hospital and the Repository at the University of Michigan for their assistance with the data sets/questionnaires and all the women who participated in SWAN.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Center for Child Health and Human Development at the National Institutes of Health (NIH, Grant R01HD068440 to L.M.P.). The Study of Women's Health Across the Nation (SWAN) has grant support from the National Institutes of Health (NIH), DHHS, through the National Institute on Aging (NIA), the National Institute of Nursing Research (NINR), and the NIH Office of Research on Women's Health (ORWH; Grants U01NR004061, U01AG012505, U01AG012535, U01AG012531, U01AG012539, U01AG012546, U01AG012553, U01AG012554, U01AG012495). The SWAN Repository is funded by NIH grant U01AG017719. This publication was also supported in part by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through UCSF-CTSI Grant Number UL1 RR024131. The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the NIA, NINR, ORWH, or the NIH.

Supplemental Material

The online supplemental figures and table are available at <http://rs.sagepub.com/supplemental>.

References

1. Allingham-Hawkins DJ, Babul-Hirji R, Chitayat D, et al. Fragile X premutation is a significant risk factor for premature ovarian failure: the International Collaborative POF in Fragile X study—preliminary data. *Am J Med Genet.* 1999;83(4):322-325.
2. Uzielli ML, Guarducci S, Lapi E, et al. Premature ovarian failure (POF) and fragile X premutation females: from POF to fragile X carrier identification, from fragile X carrier diagnosis to POF association data. *Am J Med Genet.* 1999;84(3):300-303.
3. Sullivan AK, Marcus M, Epstein MP, et al. Association of *FMR1* repeat size with ovarian dysfunction. *Hum Reprod.* 2005;20(2):402-412.
4. Streuli I, Fraisse T, Ibecheole V, Moix I, Morris MA, de Ziegler D. Intermediate and premutation *FMR1* alleles in women with occult primary ovarian insufficiency. *Fertil Steril.* 2009;92(2):464-470.
5. Pastore LM, Young SL, Baker VM, Karns LB, Williams CD, Silverman LM. Elevated prevalence of 35-44 *FMR1* trinucleotide repeats in women with diminished ovarian reserve. *Reprod Sci.* 2012;19(11):1226-1231.
6. Bretherick KL, Fluker MR, Robinson WP. *FMR1* repeat sizes in the gray zone and high end of the normal range are associated with premature ovarian failure. *Hum Genet.* 2005;117(4):376-382.
7. Gleicher N, Weghofer A, Oktay K, Barad D. Relevance of triple CGG repeats in the *FMR1* gene to ovarian reserve. *Reprod Biomed Online.* 2009;19(3):385-390.
8. Pastore LM, Johnson J. The *FMR1* gene, infertility and reproductive decision-making: a review. *Front Genet.* 2014;5:195.
9. Kim MJ, Kim do J, Kim SY, et al. Fragile X carrier screening in Korean women of reproductive age. *J Med Screen.* 2013;20(1):15-20.
10. Voorhuis M, Onland-Moret NC, Janse F, et al. The significance of fragile X mental retardation gene 1 CGG repeat sizes in the normal and intermediate range in women with primary ovarian insufficiency. *Hum Reprod.* 2014;29(7):1585-1593.
11. Weiss K, Orr-Urtreger A, Kaplan Ber I, et al. Ethnic effect on *FMR1* carrier rate and AGG repeat interruptions among Ashkenazi women. *Genet Med.* 2014;16(12):940-944.
12. Falik-Zaccai TC, Shachak E, Yalon M, et al. Predisposition to the fragile X syndrome in Jews of Tunisian descent is due to the absence of AGG interruptions on a rare Mediterranean haplotype. *Am J Hum Genet.* 1997;60(1):103-112.
13. Fernandez-Carvajal I, Walichiewicz P, Xiaosen X, Pan R, Hagerman PJ, Tassone F. Screening for expanded alleles of the *FMR1* gene in blood spots from newborn males in a Spanish population. *J Mol Diagn.* 2009;11(4):324-329.
14. Richards RI, Kondo I, Holman K, et al. Haplotype analysis at the *FRAXA* locus in the Japanese population. *Am J Med Genet.* 1994;51(4):412-416.
15. Crawford DC, Meadows KL, Newman JL, et al. Prevalence of the fragile X syndrome in African-Americans. *Am J Med Genet.* 2002;110(3):226-233.
16. Dawson AJ, Chodirker BN, Chudley AE. Frequency of *FMR1* premutations in a consecutive newborn population by PCR screening of Guthrie blood spots. *Biochem Mol Med.* 1995;56(1):63-69.
17. Fu Y-H, Kuhl DPA, Pizzuti A, et al. Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell.* 1991;67(6):1047-1058.
18. Huang W, Xia Q, Luo S, et al. Distribution of fragile X mental retardation 1 CGG repeat and flanking haplotypes in a large Chinese population. *Mol Genet Genomic Med.* 2015;3(3):172-181.

19. Otsuka S, Sakamoto Y, Siomi H, et al. Fragile X carrier screening and FMR1 allele distribution in the Japanese population. *Brain Dev.* 2010;32(2):110-114.
20. Strom CM, Crossley B, Redman JB, et al. Molecular testing for fragile X syndrome: lessons learned from 119,232 tests performed in a clinical laboratory. *Genet Med.* 2007;9(1):46-51.
21. Tzeng C, Cho W, Kuo P, Chen R. Pilot fragile X screening in normal population of Taiwan. *Diagn Mol Pathol.* 1999;8(3):152-156.
22. Genereux DP, Laird CD. Why do fragile X carrier frequencies differ between Asian and non-Asian populations? *Genes Genet Syst.* 2013;88(3):211-224.
23. Crawford DC, Zhang F, Wilson B, Warren ST, Sherman SL. Fragile X CGG repeat structures among African-Americans: identification of a novel factor responsible for repeat instability. *Hum Mol Genet.* 2000;9(12):1759-1769.
24. Eichler EE, Macpherson JN, Murray A, Jacobs PA, Chakravarti A, Nelson DL. Haplotype and interspersed analysis of the FMR1 CGG repeat identifies two different mutational pathways for the origin of the fragile X syndrome. *Hum Mol Genet.* 1996;5(3):319-330.
25. Faradz SM, Pattihla MZ, Leigh DA, et al. Genetic diversity at the FMR1 locus in the Indonesian population. *Ann Hum Genet.* 2000;64(pt 4):329-339.
26. Kronquist KE, Sherman SL, Spector EB. Clinical significance of tri-nucleotide repeats in fragile X testing: a clarification of American College of Medical Genetics guidelines. *Genet Med.* 2008;10(11):845-847.
27. Biancalana V, Glaeser D, McQuaid S, Steinbach P. 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.* 2015;23(4):417-425.
28. Sowers M, Crawford S, Sternfeld B, et al. SWAN: a multicenter, multiethnic, community-based cohort study of women and the menopausal transition. In: Lobo RA, Kelsey JL, Marcus R, eds. *Menopause: Biology and Pathobiology*. San Diego, CA: Academic Press; 2000:175-188.
29. Kardia SR, Chu J, Sowers MR. Characterizing variation in sex steroid hormone pathway genes in women of 4 races/ethnicities: the Study of Women's Health Across the Nation (SWAN). *Am J Med.* 2006;119(9 suppl 1):s3-s15.
30. Murray A, Schoemaker MJ, Bennett CE, et al. Population-based estimates of the prevalence of FMR1 expansion mutations in women with early menopause and primary ovarian insufficiency. *Genet Med.* 2014;16(1):19-24.
31. White BJ, Ayad M, Fraser A, et al. A 6-year experience demonstrates the utility of screening for both cytogenetic and FMR1 abnormalities in patients with mental retardation. *Genet Test.* 1999;3(3):291-296.
32. Larsen LA, Grønskov K, Nørgaard-Pedersen B, Brøndum-Nielsen K, Hasholt L, Vuust J. High-throughput analysis of Fragile X (CGG)_n alleles in the normal and premutation range by PCR amplification and automated capillary electrophoresis. *Hum Genet.* 1997;100(5-6):564-568.
33. Nolin SL, Glicksman A, Ding X, et al. Fragile X analysis of 1112 prenatal samples from 1991 to 2010. *Prenat Diagn.* 2011;31(10):925-931.
34. Gleicher N, Kim A, Weghofer A, Barad DH. Differences in ovarian aging patterns between races are associated with ovarian genotypes and sub-genotypes of the FMR1 gene. *Reprod Biol Endocrinol.* 2012;10:77.
35. Yrigollen CM, Sweha S, Durbin-Johnson B, et al. Distribution of AGG interruption patterns within nine world populations. *Intractable Rare Dis Res.* 2014;3(4):153-161.
36. Cronister A, Teicher J, Rohlfes EM, Donnenfeld A, Hallam S. Prevalence and instability of fragile X alleles: implications for offering fragile X prenatal diagnosis. *Obstet Gynecol.* 2008;111(3):596-601.
37. Maenner MJ, Baker MW, Broman KW, et al. FMR1 CGG expansions: prevalence and sex ratios. *Am J Med Genet B Neuropsychiatr Genet.* 2013;162b(5):466-473.
38. Mailick MR, Hong J, Rathouz P, et al. Low-normal FMR1 CGG repeat length: phenotypic associations. *Front Genet.* 2014;5:309.
39. Berkenstadt M, Ries-Levavi L, Cuckle H, Peleg L, Barkai G. Pre-conceptional and prenatal screening for fragile X syndrome: experience with 40,000 tests. *Prenat Diagn.* 2007;27(11):991-994.
40. Musci TJ, Caughey AB. Cost-effectiveness analysis of prenatal population-based fragile X carrier screening. *Am J Obstet Gynecol.* 2005;192(6):1905-1912; discussion 1912-1915.
41. Abrams L, Cronister A, Brown WT, et al. Newborn, carrier, and early childhood screening recommendations for fragile X. *Pediatrics.* 2012;130(6):1126-1135.