Identifying susceptibility genes for primary ovarian insufficiency on the high-risk genetic background of a fragile X premutation

Cristina E. Trevino, Ph.D.[a], **J. Christopher Rounds, Ph.D.**[b], **Krista Charen, M.S.**[a], **Lisa Shubeck, B.A.**[a], **Heather S. Hipp, M.D.**[c], **Jessica B. Spencer, M.D.**[c], **H. Richard Johnston, Ph.D.**[a], **Dave J. Cutler, Ph.D.**[a], **Michael E. Zwick, Ph.D.**[a,d], **Michael P. Epstein, Ph.D.**[a], **Anna Murray, Ph.D.**[e], **James N. Macpherson, Ph.D.**[f], **Montserrat Mila, Ph.D.**[g], **Laia Rodriguez-Revenga, Ph.D.**[g,h], **Elizabeth Berry-Kravis, M.D.**[i], **Deborah A. Hall, M.D.**[j], **Maureen A. Leehey, M.D.**[k], **Ying Liu, M.D.**[k], **Corrine Welt, M.D.**[l], **Stephen T. Warren, Ph.D.**[a,d,m], **Stephanie L. Sherman, Ph.D.**[a,d], **Peng Jin, Ph.D.**[a], **Emily G. Allen, Ph.D.**[a]

[a]Department of Human Genetics, Emory University, Atlanta, Georgia;

[b]Department of Cell Biology, Emory University, Atlanta, Georgia;

[c]Department of Gynecology and Obstetrics, Emory University, Atlanta, Georgia;

[d]Department of Pediatrics, Emory University, Atlanta, Georgia;

[e]University of Exeter Medical School, University of Exeter, Exeter, United Kingdom;

[f]Wessex Regional Genetics Laboratory, Salisbury District Hospital, Salisbury, United Kingdom;

[g]Biochemistry and Molecular Genetics Department, Hospital Clinic of Barcelona and Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain;

[h]CIBER of Rare Diseases (CIBERER), Instituto de Salud Carlos III, Spain;

[i]Departments of Pediatrics, Neurological Sciences, Biochemistry, Rush University Medical Center, Chicago, Illinois;

[j]Department of Neurological Sciences, Rush University, Chicago, Illinois;

[k]Department of Neurology, University of Colorado School of Medicine, Aurora, Colorado;

[l]Division of Endocrinology, Metabolism and Diabetes, University of Utah School of Medicine, Salt Lake City, Utah;

[m]Department of Biochemistry, Emory University, Atlanta, Georgia

Abstract

Objective: To identify modifying genes that explains the risk of fragile X-associated primary ovarian insufficiency (FXPOI).

Design: Gene-based, case/control association study, followed by a functional screen of highly ranked genes using a *Drosophila* model.

Setting: Participants were recruited from academic and clinical settings.

Patient(s): Women with a premutation (PM) who experienced FXPOI at the age of 35 years or younger (n = 63) and women with a PM who experienced menopause at the age of 50 years or older (n = 51) provided clinical information and a deoxyribonucleic acid sample for whole genome sequencing. The functional screen was on the basis of *Drosophila* TRiP lines.

Intervention(s): Clinical information and a DNA sample were collected for whole genome sequencing.

Main Outcome Measures: A polygenic risk score derived from common variants associated with natural age at menopause was calculated and associated with the risk of FXPOI. Genes associated with the risk of FXPOI were identified on the basis of the *P*-value from gene-based association test and an altered level of fecundity when knocked down in the *Drosophila* PM model.

Results: The polygenic risk score on the basis of common variants associated with natural age at menopause explained approximately 8% of the variance in the risk of FXPOI. Further, *SUMO1* and *KRR1* were identified as possible modifying genes associated with the risk of FXPOI on the basis of an untargeted gene analysis of rare variants.

Conclusions: In addition to the large genetic effect of a PM on ovarian function, the additive effects of common variants associated with natural age at menopause and the effect of rare modifying variants appear to play a role in FXPOI risk.

Fragile X-associated primary ovarian insufficiency (FXPOI [MIM: 311360]) is one of the disorders associated with the fragile X premutation (PM) size repeat expansions (55–200 unmethylated CGG repeats, PM) located in the 5′ untranslated region of the X-linked *FMR1* gene *(FMR1* [MIM: 309550]). On average, women in the general population experience natural menopause around the age of 51 years with 1% of women experiencing symptoms of menopause prematurely, a hallmark of primary ovarian insufficiency (POI) (1). Primary ovarian insufficiency is characterized by amenorrhea for at least 4–6 months before the age of 40 years in a setting of a high follicle-stimulating hormone level (2, 3). On average, women with a PM experience symptoms of menopause 5 years earlier than the general population, leading to a lifetime risk of approximately 20% for FXPOI (3–6). However, not all women with a PM suffer from POI. The identification of risk factors for FXPOI, and POI in general, can help predict the potential of a shortened reproductive window and provide possible interventions to help achieve family building plans and reduce the risk of untreated early hypogonadism.

Genetic factors that have been investigated in women with a PM to explain the incomplete penetrance of POI include PM CGG repeat structure (repeat length and AGG interruptions), skewing of X-chromosome inactivation, and genetic background. In women with a PM, repeat size is nonlinearly associated with FXPOI, with the greatest risk incurred at 80–100 repeats rather than with the largest PM alleles (4, 6–8). The AGG interrupt pattern among the CGG repeats within a PM allele does not appear to be associated with FXPOI risk (9). In addition, neither skewed X-chromosome inactivation nor the increased percentage of active X chromosomes harboring a PM has been associated with a higher risk of FXPOI (10–13). Two studies have provided indirect evidence for a genetic component being involved in explaining the risk of FXPOI. First, evidence for an additive genetic component, adjusting for repeat size, was identified in a large sample of PM carriers and noncarriers (14). Second, the average age of menopause among first-degree relatives of PM carriers was found to be associated with the risk of FXPOI (7). These findings suggest a significant polygenic component that explains the age of onset of FXPOI. Indeed, evidence for common genetic variants that explain, in part, the wide distribution of natural age at menopause comes from studies in the general population of women and natural age at menopause (15–17). For example, the large genome-wide association study (GWAS) of approximately 70,000 women presented by Day et al. (15) identified over 50 common variants associated with natural age of menopause.

With respect to the molecular consequences of carrying a PM, it is known that increased repeat size within the PM range is associated with increased transcription of *FMR1* mRNA, although fragile X mental retardation protein levels are the same or reduced (18–23). Unlike the full mutation (>200 methylated CGG repeats) where the *FMR1* gene is transcriptionally silenced leading to fragile X syndrome, the protein encoded by *FMR1* (fragile X mental retardation protein) is still produced by a PM allele (24). Much has been learned about potential PM-associated molecular mechanisms from fragile X-associated tremor/ataxia syndrome (FXTAS), the other well-established PM-associated disorder (25). For this neurodegenerative disorder, the toxic effect of the PM is found to be related to the long PM repeat tract in the *FMR1* mRNA. This repeat tract has the potential to form secondary structures, such as hairpins, that alter subsequent processes (26, 27). Evidence for at least two mechanisms has been identified. First, increased *FMR1* mRNA containing hairpin loops and other structures formed within the PM-size CGG repeats has been shown to sequester specific RNA-binding proteins, altering their normal functions (28–31). Second, repeat-associated non-ATG (RAN) translation, caused by translation machinery becoming stalled on a structure like the hairpins that form in CGG mRNA, produces small potentially toxic polypeptides, in this case alanine or glutamine polymers (32). Evidence for these two mechanisms playing a role in FXPOI, as well as in FXTAS, has been mounting (33, 34).

In addition, much has been gained from the *Drosophila* PM model with respect to mechanism. For example, Jin et al. (35) showed that the CGG repeat itself was sufficient to cause neuronal phenotypes associated with FXTAS [MIM: 300623]. In addition, this model clearly showed that specific CGG RNA-binding proteins, including *hnRNP A2/A1* [MIM: 600124], *CUGPB1* [MIM: 601074], *DROSHA* [MIM: 608828], *DGCR8* [MIM: 609030], *SAM68* [MIM: 602489], and *Pur-alpha* [MIM: 600473], alter neuronal function

via sequestration of these proteins (28–30, 34). On the basis of this, we hypothesize that these six RNA-binding proteins may be involved in ovarian dysfunction related to FXPOI.

This study aimed to identify modifying genes that explain the incomplete penetrance, or risk, of FXPOI. To perform this, we assessed the contribution of a polygenic component and common and rare variants.

MATERIALS AND METHODS

Overview of Approach

We tested the hypothesis that genetic modifiers interact with the *FMR1* PM or are additive to the PM to explain the incomplete penetrance of FXPOI. We garnered the power of whole genome sequencing (WGS), comparing genetic variants among women with a PM who experienced FXPOI or menopause at "extreme" ages. We used an "extreme phenotype" case/control approach vs. using family members (e.g., trios) to maximize the power, given the limitation of sample size because of the high cost of WGS. Although we could not identify whether variants were de novo, as could be performed with trios, we did not predict an excess of de novo deleterious variants of large effect. We chose the age of 35 years as an extreme lower limit for POI, which is 5 years earlier than the defined clinical diagnosis of the age of 40 years (cases, n = 63), and the age of 50 years as an extreme upper limit, which is 5 years older than the average for a PM carrier (controls, n = 51). For analysis of rare variation from the WGS data, we used gene-based approaches to aggregate information on such variation across a gene for inference and ranked top genes for screening. In addition to this untargeted approach, we used a candidate gene approach that focused on the six RNA-binding proteins known to bind to the *FMR1* PM mRNA. Highly ranked genes were then screened using *Drosophila* as a whole-organism reporter assay for evidence of a role in ovarian function.

Participants

Participants were identified through existing infrastructures that were established to recruit fragile X PM carriers, primarily through families who have a member diagnosed with fragile X syndrome. Recruitment and sample acquisition were coordinated through the National Fragile X Center at Emory University (Supplemental Methods and Materials). Once a participant provided consent, a blood or saliva sample was collected, and all women completed a standardized reproductive and health history questionnaire (Supplemental Methods). In addition, most women with a diagnosis an early-onset FXPOI, our reproductive endocrinologist (H.S.H.) conducted follow-up telephone interviews to review their reproductive history and capture the woman's diagnostic and treatment experience (36). Protocols and consent forms were approved by the Emory University Institutional Review Board, and informed consent was obtained from all participants.

For this study, all cases and controls comprised women who carried a PM, defined as an *FMR1* repeat allele with 55–200 unmethylated CGG repeats. Cases were further defined as unrelated PM carriers who had amenorrhea for at least 4–6 months at the age of 35 years because of FXPOI. Controls were unrelated PM carriers who went through natural

menopause or cessation of menses for 1 year at the age of 50 years. We excluded women whose age at menopause could have been affected by FXPOI-unrelated medical conditions, including chemotherapy or radiation therapy, gynecologic surgery (e.g., oophorectomy, hysterectomy, endometrial ablation), or an eating disorder.

Statistical and Bioinformatic Analysis

Whole genome sequencing was performed on 68 cases and 55 controls for this preliminary analysis by HudsonAlpha (Huntsville, AL). FASTQ files from paired-end WGS reads were mapped, and variants were called with PEMapper and PECaller (37), respectively. Variants were annotated using Bystro (http://bystro.io) (38). The mean coverage depth ± standard deviation (SD) of WGS was 30.783 ± 7.090 for samples, and the mean transition/ transversion ratio ± SD was 2.056 ± 0.008.

After standard quality control (QC) measures were conducted, we performed principal component analysis, using PLINK1.9 (39) to identify population stratification, a property that may lead to spurious associations. We identified a total of nine outlier samples for removal. The final dataset included 63 cases and 51 controls and 13,663,751 single nucleotide variants for analysis. As none of the principal components were significant after QC in the regression model nor changed the overall results as evaluated by the Q–Q plot, they were not included as covariates for parsimony (Supplemental Methods).

Common variant analysis.—Common single nucleotide polymorphisms (SNPs) were defined as having a minor allele frequency (MAF) > 0.05, as documented in gnomAD (40). A total of 3,055,728 such SNPs were identified and tested for association with FXPOI case/control status using logistic regression, adjusting for PM repeat and repeat size squared.

Rare variant analysis.—Rare variants were defined as having an MAF < 0.05 from gnomAD genomes data (40). Variants in which the reference allele was the minor allele were excluded. We used the optimal unified test implemented in SKAT-O (41) that maximizes the advantages of the gene-based burden tests and the sequence kernel association test (SKAT) (42). SKAT-O testing was performed using the SKAT package in R. Genes with the lowest P-value were evaluated as candidate genes. As expected, no genes reached Bonferroni correction for the untargeted approach (P<10e-6) in any of the rare variant tests given the small sample size. We prioritized those for further screening on the basis of having a P-value <.001, a fly ortholog, and literature references to ovarian phenotypes. For the candidate RNA-binding gene approach, statistical significance was evaluated at a Bonferroni correction of P<.008.

Polygenic risk score.—A polygenic component associated with the risk of FXPOI was assessed by combining the information from common genetic variants into a polygenic risk score (PRS). The discovery dataset used to calculate the PRS was on the basis of a large meta-analysis GWAS comprising 33 studies that included 69,360 women of European ancestry who experienced natural age at menopause. Natural age at menopause for this study was defined as age at last naturally occurring menstrual period followed by at least 12 consecutive months of amenorrhea starting between the ages of 40 and 60 years (15). All

studies used the full imputed set of HapMap Phase 2 autosomal SNPs, run with an additive model, including top principal components and study specific covariates. Single nucleotide polymorphisms with an MAF < 0.01 or with low imputation quality were excluded.

Our target dataset for which we calculated the PRS was the same as that used for the WGS studies—63 early-onset PM cases and 51 PM controls. The same standard QC measures described earlier were used before analyzing this dataset as well as removing the major histocompatibility complex region (Chr6: 25–34 Mb, hg19), a region of extended high linkage disequilibrium that can overly influence PRS results. The final target dataset included 724,760 total variants that overlap with the variants from the Day et al. (15) study.

PRSice-2 software (43) was used to measure the proportion of variance in FXPOI case–control status explained (measured by Nagelkerke's R^2) by the PRS, using associated SNPs on the basis of different P-value thresholds derived from the GWAS of Day et al. (15) (Supplemental Methods).

Drosophila Model Generation and Fecundity Analyses

Generation of a stable line expressing 90 CGG in the *Drosophila* germline.
—*Drosophila* with the PM repeat (90 CGG repeats) inserted on chromosome 2, as described by Jin et al. (35), were obtained from the laboratory of Dr. Peng Jin. The progeny of PM repeat flies and a germline-expressing *nanos>Gal4* line (Bloomington Stock #4442) were generated and crossed to a *Sp/CyO* stock to allow for capture of PM, *nanos>Gal4* recombinant chromosomes. Recombinant males were confirmed through polymerase chain reaction genotyping. Then, *nanos>Gal4,90CGG/Sp* males were crossed with a *Sp/CyO*, *tubulin>Gal80* stock to obtain a stable, balanced line *nanos>Gal4, 90CGG/ CyO,tubulin>Gal80*. On the basis of candidate gene selection guided by the human WGS rare variant analysis and from previously identified candidate RNA-binding proteins, *Drosophila* TRiP lines expressing RNA interference constructs against candidate genes were obtained from the Bloomington Drosophila Stock Center (Supplemental Table 1). Stocks carrying these RNAi constructs were then crossed with both germline-expressing *nanos>Gal4* alone and the *nanos>Gal4*, *90CGG* PM recombinant for fecundity experiments.

Fecundity testing.
—The level of fecundity was chosen as a reporter of ovarian dysfunction, as it is a widely used, relatively straightforward mid-throughput screen (44, 45) and eliminates aspects related to survival that would be associated with a phenotype such as the number of offspring. The detail protocol is provided in Supplemental Methods. Control stocks and the stable 90 CGG PM alone were both crossed with RNAi background stocks to establish baseline fecundity. Each candidate gene knockdown (KD) was compared with the baseline fecundity values established with controls crossed with Bloomington TRiP background lines (Bloomington Stocks #36303 or #36304). The initial screen was on the basis of three replicates, each including results from five flies per cage. To further examine top candidates from the initial screen, a follow-up screen was conducted with at least 10 replicates to increase sample size to ensure robust results. The outcome fecundity measure analyzed in subsequent regression models was the 10-day total egg count per cage. Any alternation in fecundity level relative to controls was considered a reporter of ovarian

dysfunction. Owing to the over-dispersion observed in the data, a quasi-Poisson regression was used to test for altered fecundity compared with controls on the basis of the main predictors of the presence of the candidate gene KD, presence of the 90 CGG repeats, and interaction term between those two genotypes.

RESULTS

Description of Study Cohort

Whole genome sequencing from 63 cases and 51 controls were analyzed. All self-identified as Caucasian. The mean onset of FXPOI was 29.7 years (SD, 4.9; range, 16–35) for cases, and the mean age at menopause was 51.6 years (SD, 1.8; range, 50–57) for controls (Fig. 1). The mean PM repeat size was not significantly different between cases (88.9 repeats; SD, 11.3; range, 56–117) and controls (88.7 repeats; SD, 21.1; range, 57–130) (P>.10), although the sd was significantly larger for controls (F-test; P<.001). As shown in Figure 1, cases more often had alleles in the mid-range of 80–100, reflecting the FXPOI high-risk repeat range (4, 6–8).

Analysis of Common Variants

GWAS of common variants.—We conducted a GWAS primarily as an overall QC measure, as we could only detect a common variant with a large effect size in our case/control dataset. The Q–Q plot of the logistic regression results indicated that there was no population stratification or other oddities of the data. No SNP exceeded Bonferroni-adjusted genome-wide significance as expected (Supplemental Fig. 1).

Age at menopause PRS analysis and its association with FXPOI.—We calculated a PRS to test the hypothesis that the polygenic component associated with natural age at menopause explains, in part, the variation in the risk of FXPOI (Methods). The training set used to derive the PRS for age at natural menopause was composed of 69,360 women of European ancestry (15). In that study, 54 SNPs across 44 regions were found to be genome-wide significant, with effect sizes ranging from 0.07 to 0.88 years/allele.

Using PRSice software (46), we calculated a PRS for our PM cases and controls using SNPs with P-value threshold sets for association with age at menopause, adjusting for the first five principal components and PM repeat size and repeat size squared. In this analysis, the maximum variance in the risk of FXPOI explained by the PRS was 7.9%. (Nagelkerke's R^2; P=.01; Fig. 2). This maximum was on the basis of the PRS calculated from associated SNPs that had a P-value <.0021 in the training GWAS dataset analyses (n = 1,099 of the 2,407,374 total SNPs). We estimated the odds ratio for case/control status for each quartile of PRS and found that the odds ratio for the highest quartile of PRS scores was statistically significant (odds ratio, 7.89; 95% confidence limit, 2.12–29.35; Supplemental Table 2).

Gene-Based Analyses of Rare Variants On the Basis of WGS Data

For the rare variant analysis, we examined variants at an MAF < 0.05 and used the kernel-based approach SKAT-O that optimizes between burden testing and SKAT models. We adjusted for PM repeat size and repeat size squared. For the untargeted gene approach,

we interrogated 6,752,810 variants in 25,404 genes. There were no genes that exceeded Bonferroni-adjusted statistical significance (adjusted for the number of genes tested). Thirty-four genes passed a threshold of nominal significance at $P<.001$ (Supplemental Table 3).

Two additional analyses were conducted on subsets of variants. First, SKAT-O analyses were performed filtering on variants located in exon–untranslated regions with an MAF < 0.05; this included 281,828 variants in 18,975 genes. Second, we filtered on rarer variants at an MAF < 0.01; these analyses were on the basis of 4,784,690 variants and 25,346 genes. Sixteen and 31 genes, respectively, passed the nominal statistical significance threshold of $P<.001$ (Supplemental Table 3).

Taking out the overlap from these three analyses, the resulting 75 genes with a $P<.001$ were then ranked on the basis of the literature with respect to roles in ovarian function or fertility and having a fly ortholog and TRiP line stock available. Out of these, 13 genes that met these criteria were chosen for further screening using the *Drosophila* PM model (Supplemental Table 4).

In addition, we conducted a candidate gene approach, interrogating the six RNA-binding genes that were previously identified in studies of the PM sequestration model. None of the gene-set analyses exceeded Bonferroni-adjusted statistical significance of P-value <.008 (Supplemental Table 6). Because these were strong candidate genes involved in etiology of FXTAS and because of the literature showing their involvement in ovarian function (Supplemental Table 5), we decided to include them in the *Drosophila* PM model screen.

Drosophila Fecundity as a Whole-Organism Functional Screen

Screening prioritized genes ranked from the WGS case/control study.—We first examined whether fecundity was altered in the 90-CGG-repeat model compared with controls (Supplemental Fig. 2). The controls that were examined included wild type (OregonR) alone as well as the cross progeny with the *nanos>Gal4* alone and with the 90 CGG repeats and the cross progeny of the two Bloomington TRiP background lines (Bloomington Stocks #36303 and #36304) with the *nanos>Gal4* alone and with the 90 CGG repeats (Supplemental Fig. 2). There were no differences in the observed level of fecundity between the control flies and the respective 90-CGG-repeat flies on the same background ($P = .22$) (Supplemental Fig. 2).

The 18 TRiP lines available for KD of the 13 WGS case/control prioritized genes along with their gene function in *Drosophila* are provided in Supplemental Table 1; four genes did not have lines that produced viable progeny when crossed with the germline-expressing *nanos>Gal4* line and were excluded from further studies. For the remaining nine genes (represented by 12 lines), four genotypes per prioritized gene were tested: background control with *nanos>Gal4* alone, background control with 90 CGG repeats, KD of the prioritized gene alone, and the double mutant (90 CGG and KD). An initial genetic screen with at least three replicate cages for each genotype, each containing five female flies, was performed, and the total number of eggs laid was measured over 10 days (Methods). Out of nine genes that were screened, we ranked the top three that showed the greatest difference between the KD alone and double mutant (Supplemental Fig. 3)—*SUMO1*, *KRR1*, and

PDHA2. In each case, increased levels of fecundity were observed for the double mutation flies.

To confirm the apparent differences observed in the initial screen, follow-up experiments were conducted to increase the sample size (results from at least 10 cages) for *SUMO1*, *KRR1*, and *PDHA2*. For the *SUMO1*, we confirmed a significant increase in fecundity for the double mutant compared with each of the other genotypes (Fig. 3). Using a quasi-Poisson regression model, we found no evidence for an effect of 90 CGG repeats alone or the *SUMO1* KD alone compared with the respective control; however, the interaction term related to the effect of both mutant genotypes together was statistically significant ($P<.05$) (Table 1). This same pattern was observed for *KRR1*, where the interaction term associated with the effect of the double mutant was statistically significant ($P<.03$) (Table 1). For *PDHA2,* a different pattern was observed. In this case, the effect of the KD itself significantly increased fecundity compared with the control genotype ($P<.0001$). There was no evidence for an interaction between the *PDHA2* KD and 90 CGG ($P>.10$) (Table 1).

Screening RNA-binding protein candidate genes drawn from the PM sequestration model.—We tested six RNA-binding genes that are reported to play a role in the PM sequestration mechanism as candidates (Supplemental Table 4) (28–30, 34). Effects on fecundity levels appeared to be more pronounced in a subset of these candidate genes. For example, for *CUGBP1* KD and one of two *DROSHA* KD models, total and near-total loss of fecundity, respectively, resulted. This was observed for both KD alone and the corresponding double mutant (Supplemental Fig. 4). Because this was a preliminary screen, we combined fecundity data from all RNA-binding protein genes and tested for differences between genotypes using a quasi-Poisson regression model. Fecundity in the group of KDs alone was significantly different from that in controls ($P = 2.38e^{-6}$), and the interaction term for KD/90CGG was not significant.

DISCUSSION

In this study, we took the first step to identify genetic variants that play a role in the variable expression of ovarian insufficiency among the women who carry a fragile X PM. Previous work suggested that modifying genetic risk factors do influence age at onset of FXPOI. Hunter et al. (14) showed a statistically significant contribution of an additive genetic component to explain the risk of FXPOI, and Spath et al. (7) showed an association of the average age of menopause among first-degree relatives of women with a PM and the risk of FXPOI. Both studies adjusted for the known large effect of PM repeat size on the risk of FXPOI. These findings, combined with studies showing associations of genetic variants for natural age at menopause (15–17) and for idiopathic POI in the general population (47, 48), motivated us to take a novel strategy that combined WGS and *Drosophila* genetics to identify highly ranked candidate genes that are primed for further study in mammalian systems. We based our studies on women who carried a PM and experienced FXPOI/age at menopause at the extreme tails of the onset distribution: 35 years (cases) and   50 years (controls) of age.

On the basis of studies that show a significant genetic component related to age at natural menopause, we first examined a PRS derived from common variants associated with lower age at natural menopause identified through a large GWAS of Day et al. (15). In that study, 21% of the variance in age at menopause was explained overall using 30,000 SNPs with $P<.05$. We found that the PRS explained approximately 7.9% of the variance in the risk for case/control status related to early FXPOI, adjusting for PM repeat size and repeat size squared. This result is consistent with our previous findings of an additive genetic component involved in the onset of FXPOI (7, 14). Further studies on the basis of a larger sample of PM carriers using age at FXPOI/menopause as the outcome are warranted to determine the predictive value of this PRS. Irrespective, our finding suggests that, even on the background of a large, single gene effect, the combined effect of common genetic variants is significant as a modifier of severity of FXPOI.

To examine rare variants as modifiers of the age of onset of FXPOI, we took an untargeted approach and compared WGS variants using several different filtering criteria. Thirteen genes were highly ranked using gene-set analyses (SKAT-O) and the literature (i.e., those with a role in ovarian function). On the basis of a *Drosophila* genetic screen using altered fecundity as an indicator of possible ovarian dysfunction, the germline KDs of *SUMO1* and *KRR1* were identified as interacting with the PM. In addition, the germline KD of *PDHA2* alone was shown to have an impact on fecundity, irrespective of the 90 CGG repeats.

Of these three highly ranked genes, *SUMO1* may be the most interesting. It plays a role in regulating granulosa cell apoptosis via sumoylation. Phenotype studies of the fragile X PM mouse models have all shown traits associated with reduced ovarian function. Overall, it appears that the original follicular pool is not disturbed but there is an increased rate of atresia/apoptosis (reviewed in Sherman et al. (49)). Buijsen et al. (33) characterized the Dutch exCGG-KI mouse model and found that the number of atretic large antral follicles was increased by almost ninefold in the older PM females (40-week-old assessment) and that recent ovulations had reduced the number of fresh corpora lutea. Conca Dioguardi et al. (50) found a similar phenotype in the 130-CGG-repeat knock-in FX PM mouse model. They further characterized the mitochondrial state of the granulosa cells and oocytes and found that the mice had decreased mitochondrial content, structurally abnormal mitochondria, and reduced expression of critical mitochondrial genes. Because *SUMO1* is knocked down in our fecundity experiment, it is possible that apoptosis is dysregulated in the fly ovaries, resulting in increased egg laying. This could be consistent with what is known about the *Drosophila* ortholog to *SUMO1*, smt3; smt3 is expressed in the germline and plays a role in ovarian follicle cells (51, 52). However, this is only speculation as we applied this fly model system strictly as a nonspecific reporter of ovarian function, not as an indicator of mechanism.

KRR1 (Ca^{2+}/calmodulin-dependent protein kinase) belongs to the family of RNA-binding proteins containing KH domains. It encodes a ribosome assembly factor and is associated with 90S particles and involved in 35S pre-rRNA processing (53, 54). It has been identified as a susceptibility locus for polycystic ovary syndrome (PCOS) through GWAS (reviewed by Jones et al. [55]), although its function related to PCOS is not understood. Expression studies in whole ovaries collected from bovine fetuses show that *KRR1* is expressed

throughout ovarian development (56). Its mRNA levels are found to be high in early stages of ovarian development and declined significantly at the later stages. A study by Pau et al. (57) examined the expression pattern of *KRR1* using a subcutaneous adipose tissue biopsy in 38 women with PCOS. The expression pattern suggested that *KRR1* confers the risk of PCOS through a metabolic or development mechanism. Expression levels at the *KRR1* locus were associated with increased expression of *GLIPR1* and *PHLDA1*. *PHLDA1* is a nuclear protein that plays a role in the antiapoptotic effects of insulin-like growth factor-1 (58). *GLIPR1* has proapoptotic activity in prostate cancer cells and is expressed in the testes and may have a role in sperm-oocyte interactions (59). In our data, we saw no effect of the *KRR1* KD on the levels of fecundity; only when combined with the 90-CGG genotype were fecundity levels increased compared with controls and the KD. Given that *KRR1* encodes an RNA-binding protein, perhaps, it recognizes the secondary structure resulting from the long 90-CGG-repeat tract, altering its ability to carry out its normal function in ovarian development. This would require more investigation in a mammalian system.

In addition to taking an untargeted approach, we tested RNA-binding proteins that had previously been associated with the sequestration mechanism associated with FXTAS (28–30, 34). Although we found no evidence that the variation in these genes play a modifying role the penetrance of FXPOI, the fecundity screen showed that several of these genes altered ovarian function. Results showed lower fecundity in the germline KD and the double mutant candidate KD and 90-CGG-repeat flies compared with controls, especially for *CUGBP1* (Supplemental Fig. 4). All of these RNA-binding proteins have canonical roles in the ovary (Supplemental Table 4); such roles appeared to be altered in the KD lines tested. The total loss of fecundity resulted for the *CUGBP1* KD as well as the corresponding KD/90CGG double mutant; however, it is unclear whether there is a genetic interaction between the KD and 90 CGG given the total loss of fecundity for the KD. The ovarian dysfunction observed for these RNA-binding proteins may not be directly involved in the PM sequestration method. Further research would be necessary to fully understand the associated mechanisms.

Our study has several limitations. First, age at onset for FXPOI or age at menopause is a phenotype that transitions over time and is sometimes difficult to accurately define. For this study, most cases were interviewed by a gynecologist (H.S.H.) to help better define onset of FXPOI (36). For controls, most were on the basis of self-report. However, because we took the extremes and did not depend on the specific age for this study, we believe that the phenotype is accurately classified. Next, we recognize that the sample size was small, both because of the rarity of the disorder and the limited resources available for WGS. To maximize power, we took the approach of drawing from the extreme tails of the onset of FXPOI/age at menopause distribution. A significant limitation of the genetic results, both the PRS and susceptibility gene identification, is that our study only included women who self-identified as Caucasian. Thus, the translation of these findings to other populations is compromised. Limitations for prioritizing genes from the WGS study include the following: relying on gene annotation to determine to which genes or pathways the variants belong, as well as information on whether these genes are expressed in ovarian tissues; basing our ranking on current literature to define a role in ovarian function, reducing the potential to identify novel genes; and limiting ranked genes to those with *Drosophila* ortholog and

available RNAi lines. Nonetheless, our results provide the first set of possible modifying genes for further study.

Two avenues for future studies naturally follow our findings. First, it will be significant to test the specific variants identified in the prioritized genes in the *Drosophila* model or in mammalian model systems to understand their role in disrupting gene function. Again, using model systems, it will be significant to understand each gene's role in ovarian function and how the altered gene product interacts with the PM. The other avenue of research relates to determining the predictive value of the identified genetic variation. Conducting studies in larger cohorts of women of different ethnicities who carry a PM along with their relatives will begin to determine the level of penetrance related to the specific genetic variant or to the PRS. Investigating whether the identified genetic factors are associated with the full spectrum of the disorder (e.g., age at FXPOI/menopause), not just the extremes, is additionally a significant next step. Lastly, examining these prioritized genes in cohorts of women diagnosed with idiopathic POI is warranted. Such studies may begin to identify subgroups of individuals with these particular disrupted pathways.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

REFERENCES

1. Palacios S, Henderson VW, Siseles N, Tan D, Villaseca P. Age of menopause and impact of climacteric symptoms by geographical region. Climacteric 2010;13:419–28. [PubMed: 20690868]

2. Welt CK, Smith PC, Taylor AE. Evidence of early ovarian aging in fragile X premutation carriers. J Clin Endocrinol Metab 2004;89:4569–74. [PubMed: 15356064]

3. Nelson LM. Clinical practice. Primary ovarian insufficiency. N Engl J Med 2009;360:606–14. [PubMed: 19196677]

4. Allen EG, Sullivan AK, Marcus M, Small C, Dominguez C, Epstein MP, et al. Examination of reproductive aging milestones among women who carry the FMR1 premutation. Hum Reprod 2007;22:2142–52. [PubMed: 17588953]

5. Sherman SL. Premature ovarian failure in the fragile X syndrome. Am J Med Genet 2000;97:189–94. [PubMed: 11449487]

6. Sullivan AK, Marcus M, Epstein MP, Allen EG, Anido AE, Paquin JJ, et al. Association of FMR1 repeat size with ovarian dysfunction. Hum Reprod 2005; 20:402–12. [PubMed: 15608041]

7. Spath MA, Feuth TB, Smits AP, Yntema HG, Braat DD, Thomas CM, et al. Predictors and risk model development for menopausal age in fragile X premutation carriers. Genet Med 2011;13:643–50. [PubMed: 21597380]

8. Ennis S, Ward D, Murray A. Nonlinear association between CGG repeat number and age of menopause in FMR1 premutation carriers. Eur J Hum Genet 2006;14:253–5. [PubMed: 16251893]

9. Allen EG, Glicksman A, Tortora N, Charen K, He W, Amin A, et al. FXPOI: pattern of AGG interruptions does not show an association with age at amenorrhea among women with a premutation. Front Genet 2018;9:292. [PubMed: 30123240]

10. Tejada MI, Garcia-Alegria E, Bilbao A, Martinez-Bouzas C, Beristain E, Poch M, et al. Analysis of the molecular parameters that could predict the risk of manifesting premature ovarian failure in female premutation carriers of fragile X syndrome. Menopause 2008;15:945–9. [PubMed: 18427356]

11. Bione S, Benedetti S, Goegan M, Menditto I, Marozzi A, Ferrari M, et al. Skewed X-chromosome inactivation is not associated with premature ovarian failure in a large cohort of Italian patients. Am J Med Genet A 2006;140:1349–51. [PubMed: 16691602]

12. Rodriguez-Revenga L, Madrigal I, Badenas C, Xuncla M, Jimenez L, Mila M. Premature ovarian failure and fragile X female premutation carriers: no evidence for a skewed X-chromosome inactivation pattern. Menopause 2009; 16:944–9. [PubMed: 19373114]

13. Spath MA, Nillesen WN, Smits AP, Feuth TB, Braat DD, van Kessel AG, et al. X chromosome inactivation does not define the development of premature ovarian failure in fragile X premutation carriers. Am J Med Genet A 2010; 152A:387–93. [PubMed: 20101683]

14. Hunter JE, Epstein MP, Tinker SW, Charen KH, Sherman SL. Fragile X-associated primary ovarian insufficiency: evidence for additional genetic contributions to severity. Genet Epidemiol 2008;32:553–9. [PubMed: 18357616]

15. Day FR, Ruth KS, Thompson DJ, Lunetta KL, Pervjakova N, Chasman DI, et al. Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. Nat Genet 2015;47:1294–303. [PubMed: 26414677]

16. Stolk L, Perry JR, Chasman DI, He C, Mangino M, Sulem P, et al. Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways. Nat Genet 2012;44:260–8. [PubMed: 22267201]

17. Perry JR, Hsu YH, Chasman DI, Johnson AD, Elks C, Albrecht E, et al. DNA mismatch repair gene MSH6 implicated in determining age at natural menopause. Hum Mol Genet 2014;23:2490–7. [PubMed: 24357391]

18. Primerano B, Tassone F, Hagerman RJ, Hagerman P, Amaldi F, Bagni C. Reduced FMR1 mRNA translation efficiency in fragile X patients with premutations. RNA 2002;8:1482–8. [PubMed: 12515381]

19. Tassone F, Hagerman RJ, Taylor AK, Gane LW, Godfrey TE, Hagerman PJ. Elevated levels of FMR1 mRNA in carrier males: a new mechanism of involvement in the fragile-X syndrome. Am J Hum Genet 2000;66:6–15. [PubMed: 10631132]

20. Tassone F, Hagerman RJ, Taylor AK, Mills JB, Harris SW, Gane LW, et al. Clinical involvement and protein expression in individuals with the FMR1 premutation. Am J Med Genet 2000;91:144–52. [PubMed: 10748416]

21. Tassone F, Hagerman PJ. Expression of the FMR1 gene. Cytogenet Genome Res 2003;100:124–8. [PubMed: 14526172]

22. Kenneson A, Zhang F, Hagedorn CH, Warren ST. Reduced FMRP and increased FMR1 transcription is proportionally associated with CGG repeat number in intermediate-length and premutation carriers. Hum Mol Genet 2001;10:1449–54. [PubMed: 11448936]

23. Peprah E, He W, Allen E, Oliver T, Boyne A, Sherman SL. Examination of FMR1 transcript and protein levels among 74 premutation carriers. J Hum Genet 2010;55:66–8. [PubMed: 19927162]

24. Pieretti M, Zhang FP, Fu YH, Warren ST, Oostra BA, Caskey CT, et al. Absence of expression of the FMR-1 gene in fragile X syndrome. Cell 1991;66:817–22. [PubMed: 1878973]

25. Hagerman RJ, Hagerman P. Fragile X-associated tremor/ataxia syndrome -features, mechanisms and management. Nat Rev Neurol 2016;12: 403–12. [PubMed: 27340021]

26. Handa V, Saha T, Usdin K. The fragile X syndrome repeats form RNA hairpins that do not activate the interferon-inducible protein kinase, PKR, but are cut by Dicer. Nucleic Acids Res 2003;31:6243–8. [PubMed: 14576312]

27. Zumwalt M, Ludwig A, Hagerman PJ, Dieckmann T. Secondary structure and dynamics of the r(CGG) repeat in the mRNA of the fragile X mental retardation 1 (FMR1) gene. RNA Biol 2007;4:93–100. [PubMed: 17962727]

28. Sellier C, Freyermuth F, Tabet R, Tran T, He F, Ruffenach F, et al. Sequestration of DROSHA and DGCR8 by expanded CGG RNA repeats alters micro-RNA processing in fragile X-associated tremor/ataxia syndrome. Cell Rep 2013;3:869–80. [PubMed: 23478018]

29. Sellier C, Rau F, Liu Y, Tassone F, Hukema RK, Gattoni R, et al. Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. EMBO J 2010;29:1248–61. [PubMed: 20186122]

30. Sofola OA, Jin P, Qin Y, Duan R, Liu H, de Haro M, et al. RNA-binding proteins hnRNP A2/B1 and CUGBP1 suppress fragile X CGG premutation repeat-induced neurodegeneration in a Drosophila model of FXTAS. Neuron 2007;55:565–71. [PubMed: 17698010]

31. Jin P, Duan R, Qurashi A, Qin Y, Tian D, Rosser TC, et al. Pur alpha binds to rCGG repeats and modulates repeat-mediated neurodegeneration in a Drosophila model of fragile X tremor/ataxia syndrome. Neuron 2007;55: 556–64. [PubMed: 17698009]

32. Todd PK, Oh SY, Krans A, He F, Sellier C, Frazer M, et al. CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. Neuron 2013;78:440–55. [PubMed: 23602499]

33. Buijsen RA, Visser JA, Kramer P, Severijnen EA, Gearing M, Charlet-Berguerand N, et al. Presence of inclusions positive for polyglycine containing protein, FMRpolyG, indicates that repeat-associated non-AUG translation plays a role in fragile X-associated primary ovarian insufficiency. Hum Reprod 2016;31:158–68. [PubMed: 26537920]

34. Aumiller V, Graebsch A, Kremmer E, Niessing D, Forstemann K. Drosophila Pur-a binds to trinucleotide-repeat containing cellular RNAs and translocates to the early oocyte. RNA Biol 2012;9:633–43. [PubMed: 22614836]

35. Jin P, Zarnescu DC, Zhang F, Pearson CE, Lucchesi JC, Moses K, et al. RNA-mediated neurodegeneration caused by the fragile X premutation rCGG repeats in Drosophila. Neuron 2003;39:739–47. [PubMed: 12948442]

36. Hipp HS, Charen KH, Spencer JB, Allen EG, Sherman SL. Reproductive and gynecologic care of women with fragile X primary ovarian insufficiency (FXPOI). Menopause 2016;23:993–9. [PubMed: 27552334]

37. Johnston HR, Chopra P, Wingo TS, Patel V. International Consortium on Brain and Behavior in 22q11.2 Deletion Syndrome, Epstein MP, et al. PEMapper and PECaller provide a simplified approach to whole-genome sequencing. Proc Natl Acad Sci U S A 2017;114:E1923–32. [PubMed: 28223510]

38. Kotlar AV, Trevino CE, Zwick ME, Cutler DJ, Wingo TS. Bystro: rapid online variant annotation and natural-language filtering at whole-genome scale. Genome Biol 2018;19:14. [PubMed: 29409527]

39. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Giga Science 2015;4:7. [PubMed: 25722852]

40. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020;581:434–43. [PubMed: 32461654]

41. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am J Hum Genet 2012;91:224–37. [PubMed: 22863193]

42. Wu MC, Kraft P, Epstein MP, Taylor DM, Chanock SJ, Hunter DJ, et al. Powerful SNP-set analysis for case-control genome-wide association studies. Am J Hum Genet 2010;86:929–42. [PubMed: 20560208]

43. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. Gigascience 2019;8:giz082. [PubMed: 31307061]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

44. Daenzer JM, Sanders RD, Hang D, Fridovich-Keil JL. UDP-galactose 4'-epim-erase activities toward UDP-Gal and UDP-GalNAc play different roles in the development of Drosophila melanogaster. PLoS Genet 2012;8: e1002721. [PubMed: 22654673]

45. Armstrong AR. Drosophila melanogaster as a model for nutrient regulation of ovarian function. Reproduction 2019;159:R69–82.

46. Euesden J, Lewis CM, O'Reilly PF. PRSice: polygenic risk score software. Bioinformatics 2015;31:1466–8. [PubMed: 25550326]

47. Rossetti R, Ferrari I, Bonomi M, Persani L. Genetics of primary ovarian insufficiency. Clin Genet 2017;91:183–98. [PubMed: 27861765]

48. Jin M, Yu Y, Huang H. An update on primary ovarian insufficiency. Sci China Life Sci 2012;55:677–86. [PubMed: 22932883]

49. Sherman SL, Curnow EC, Easley CA, Jin P, Hukema RK, Tejada MI, et al. Use of model systems to understand the etiology of fragile X-associated primary ovarian insufficiency (FXPOI). J Neurodev Disord 2014;6:26. [PubMed: 25147583]

50. Conca Dioguardi C, Uslu B, Haynes M, Kurus M, Gul M, Miao DQ, et al. Granulosa cell and oocyte mitochondrial abnormalities in a mouse model of fragile X primary ovarian insufficiency. Mol Hum Reprod 2016;22:384–96. [PubMed: 26965313]

51. Talamillo A, Herboso L, Pirone L, Perez C, Gonzalez M, Sanchez J, et al. Scavenger receptors mediate the role of SUMO and Ftz-f1 in Drosophila steroido-genesis. PLoS Genet 2013;9:e1003473. [PubMed: 23637637]

52. Hashiyama K, Shigenobu S, Kobayashi S. Expression of genes involved in sumoylation in the Drosophila germline. Gene Expr Patterns 2009;9:50–3. [PubMed: 18755298]

53. Grandi P, Rybin V, Bassler J, Petfalski E, Strauss D, Marzioch M, et al. 90S preribosomes include the 35S pre-rRNA, the U3 snoRNP, and 40S subunit processing factors but predominantly lack 60S synthesis factors. Mol Cell 2002; 10:105–15. [PubMed: 12150911]

54. Sasaki T, Toh-E A, Kikuchi Y. Yeast Krr1p physically and functionally interacts with a novel essential Kri1p, and both proteins are required for 40S ribosome biogenesis in the nucleolus. Mol Cell Biol 2000;20:7971–9. [PubMed: 11027267]

55. Jones MR, Goodarzi MO. Genetic determinants of polycystic ovary syndrome: progress and future directions. Fertil Steril 2016;106:25–32. [PubMed: 27179787]

56. Hartanti MD, Rosario R, Hummitzsch K, Bastian NA, Hatzirodos N, Bonner WM, et al. Could perturbed fetal development of the ovary contribute to the development of polycystic ovary syndrome in later life? PLoS One 2020;15:e0229351. [PubMed: 32078641]

57. Pau CT, Mosbruger T, Saxena R, Welt CK. Phenotype and tissue expression as a function of genetic risk in polycystic ovary syndrome. PLoS One 2017; 12:e0168870. [PubMed: 28068351]

58. Toyoshima Y, Karas M, Yakar S, Dupont J, Lee Helman H, LeRoith D. TDAG51 mediates the effects of insulin-like growth factor I (IGF-I) on cell survival. J Biol Chem 2004;279:25898–904. [PubMed: 15037619]

59. Gibbs GM, Lo JC, Nixon B, Jamsai D, O'Connor AE, Rijal S, et al. Glioma pathogenesis-related 1-like 1 is testis enriched, dynamically modified, and redistributed during male germ cell maturation and has a potential role in sperm-oocyte binding. Endocrinology 2010;151:2331–42. [PubMed: 20219979]
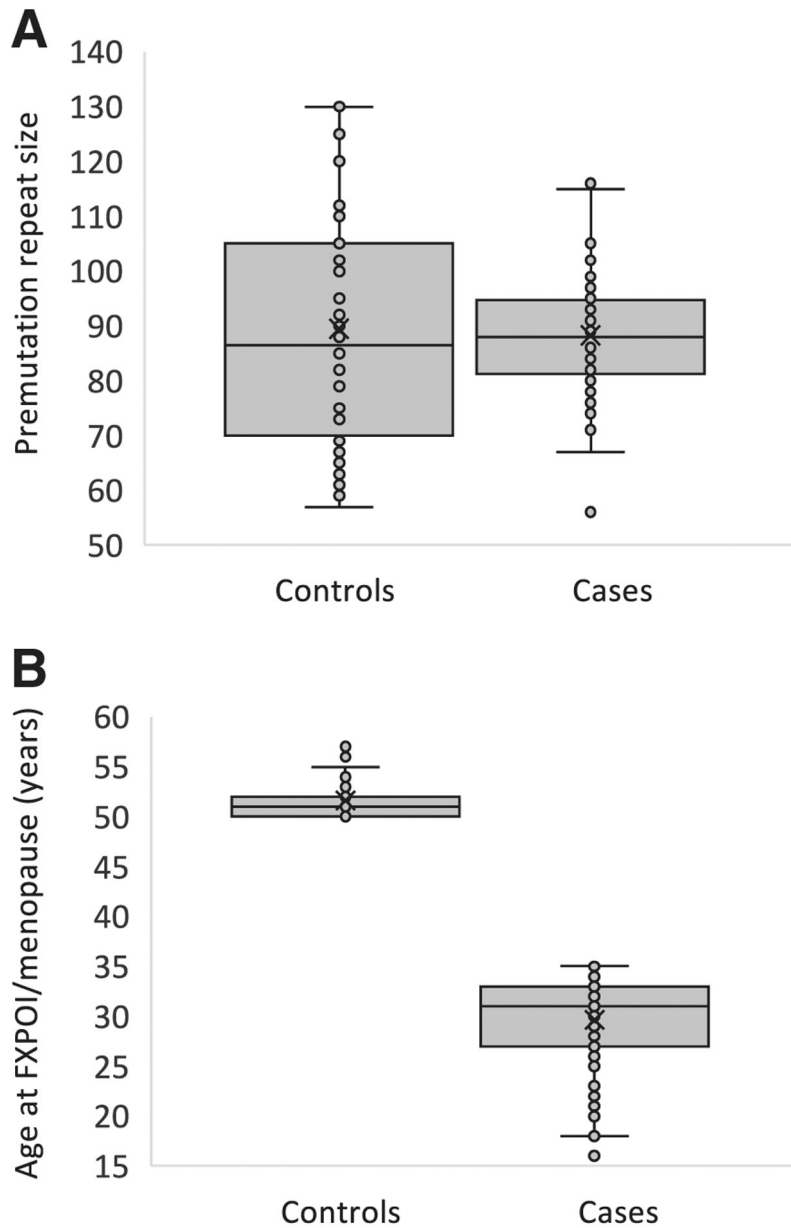
FIGURE 1.
Distribution of cohort before quality control (QC). **(A)** Distribution of premutation repeat size alleles among cases (women with a diagnosis of fragile X-associated primary ovarian insufficiency before the age of 35 years; n = 63) and controls (women with a premutation who experienced menopause after the age of 50 years; n = 51). **(B)** Distribution of age (years) at diagnosis of fragile X-associated primary ovarian insufficiency (cases) and age of menopause (controls).
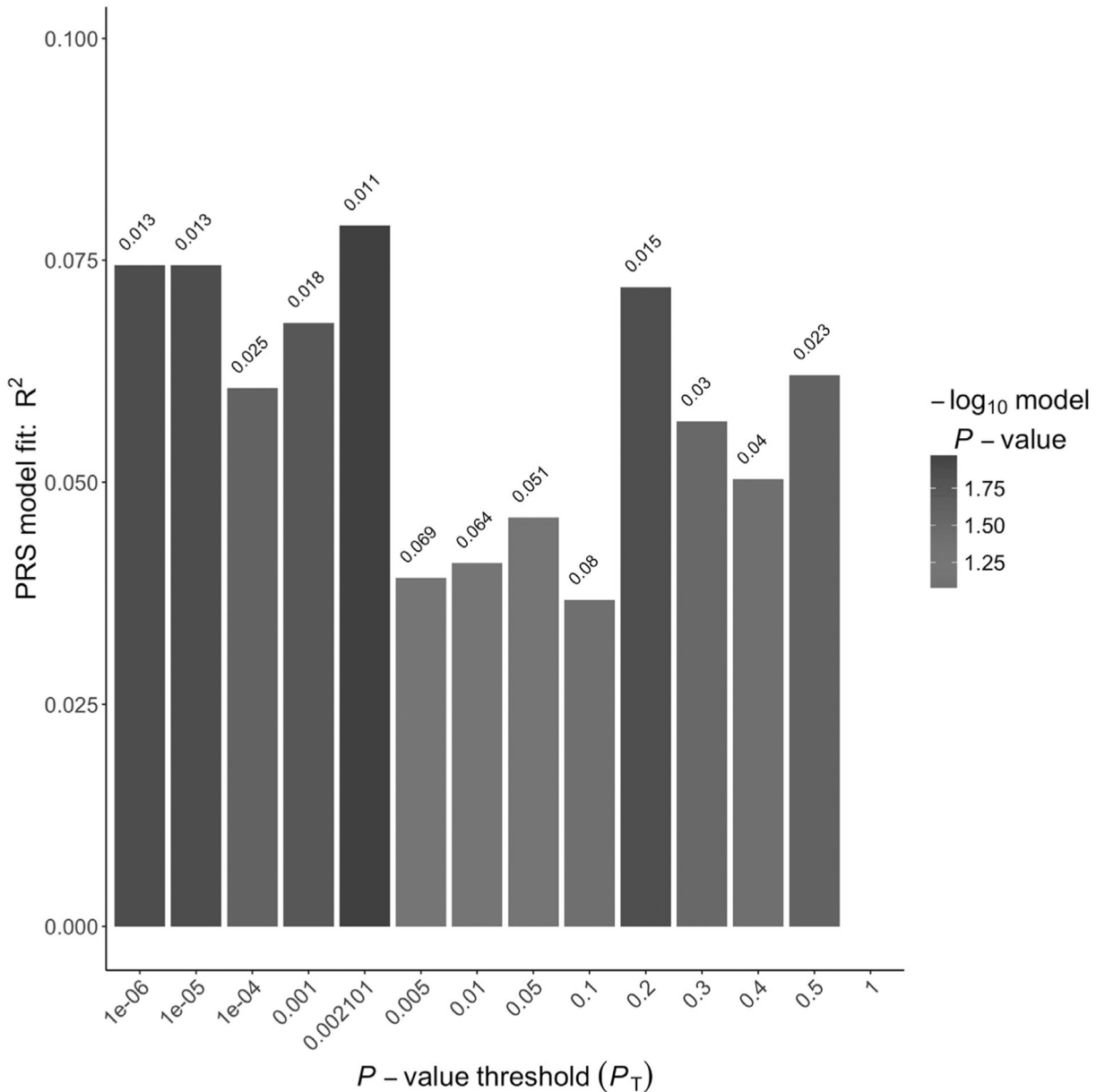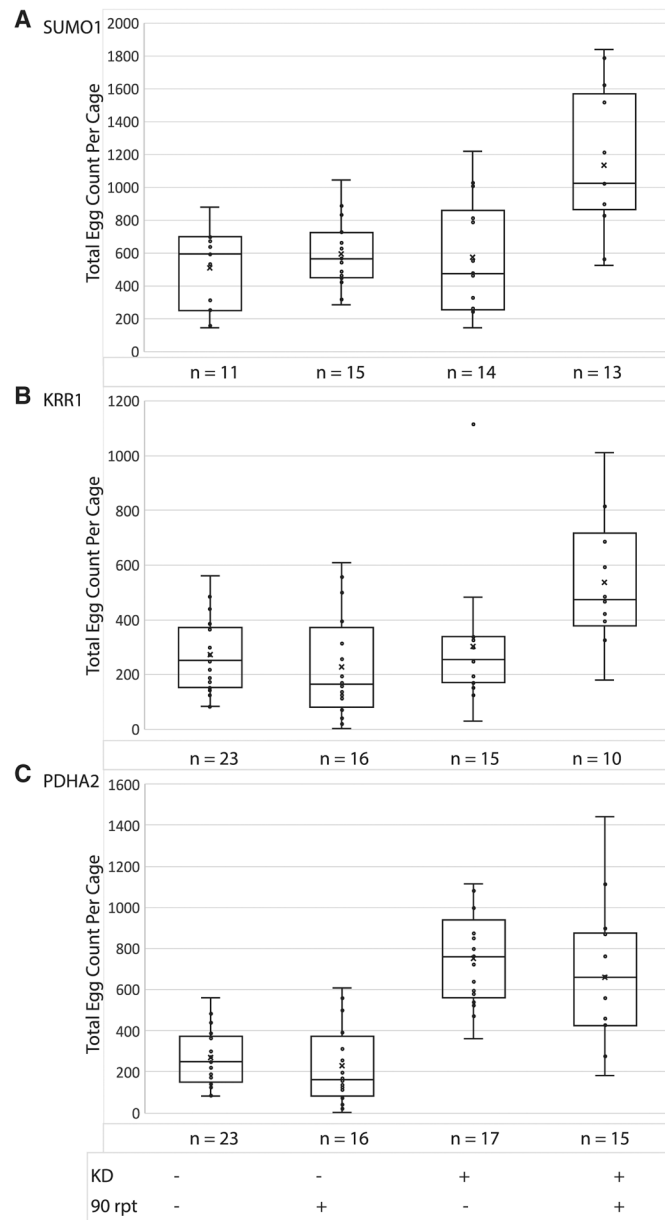
FIGURE 2.
Polygenic risk score analysis reveals a maximum Nagelkerke's R^2 of 7.9% at a threshold of *P*-values <.0021 in the discovery GWAS set (15). On the X-axis, from left to right, *P*-value thresholds generated from the discovery GWAS represent the most stringent to the most liberal and, thus, include the smallest number to the largest number of single nucleotide polymorphisms, respectively, used in the polygenic risk score calculation. The *P*-values above the bars represent the statistical significance of the polygenic risk score explaining the risk variance in fragile X-associated primary ovarian insufficiency case/control status on the basis of the related Nagelkerke's R^2 value.

FIGURE 3.
Follow-up fecundity testing on top three candidates drawn from the whole genome sequencing analyses. Controls included here are the cross progeny of the corresponding Bloomington TRiP background line (Bloomington Stocks # 36303 and # 36304) with the *nanos>Gal4* alone and with the 90 CGG repeat. The number of cage replicates for each genotype is indicated by "n."

TABLE 1

Results from the quasi-Poisson regression model to examine the fecundity levels by genotype for top three candidates drawn from the whole genome sequencing analyses. The independent variables in the model included the presence of the 90 CGG repeats (CGG), presence of a knockdown, and the interaction term between those two (CGG*KD)

SUMO1

Variable	Beta-coefficient	Std. error	t value	Pr(>|t|)
(Intercept)	6.23	0.16	39.16	<2e−16
CGG	0.15	0.20	0.76	0.45
KD	0.12	0.21	0.59	0.56
CGG*KD	0.53	0.26	2.03	0.05

KRR1

Variable	Beta-coefficient	Std. error	t value	Pr(>|t|)
(Intercept)	5.60	0.14	38.91	<2e−16
CGG	−0.17	0.24	−0.73	0.47
KD	0.12	0.22	0.54	0.59
CGG*KD	0.74	0.33	2.24	0.03

PDHA2

Variable	Beta-coefficient	Std. error	t value	Pr(>|t|)
(Intercept)	6.00	0.13	42.09	<2.00E−16
CGG	−0.17	0.22	−0.79	0.43
KD	1.012	0.16	6.23	3.53E−08
CGG*KD	0.04	0.26	0.16	0.88

Note: KD knockdown; Pr(>|t|) p value.