

ORIGINAL ARTICLE

Multivariate, region-based genetic analyses of facets of reproductive aging in White and Black women

Lawrence F. Bielak¹  | Patricia A. Peyser¹ | Jennifer A. Smith^{1,2} | Wei Zhao¹ | Edward A. Ruiz-Narvaez³ | Sharon L. R. Kardia¹ | Sioban D. Harlow¹

¹Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA

²Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, Michigan, USA

³Department of Nutritional Sciences, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA

Correspondence

Lawrence F. Bielak, Department of Epidemiology, School of Public Health, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109, USA.
Email: lfbielak@umich.edu

Funding information

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) and for the SWAN Genomic Analyses and SWAN Legacy (U01AG017719). 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. SDH gratefully acknowledge use of the services and facilities of the Population Studies Center at the University of Michigan, funded by NICHD Center Grant R24 HD041028

Abstract

Background: Age at final menstrual period (FMP) and the accompanying hormone trajectories across the menopause transition do not occur in isolation, but likely share molecular pathways. Understanding the genetics underlying the endocrinology of the menopause transition may be enhanced by jointly analyzing multiple interrelated traits.

Methods: In a sample of 347 White and 164 Black women from the Study of Women's Health Across the Nation (SWAN), we investigated pleiotropic effects of 54 candidate genetic regions of interest (ROI) on 5 menopausal traits (age at FMP and premenopausal and postmenopausal levels of follicle stimulation hormone and estradiol) using multivariate kernel regression (Multi-SKAT). A backward elimination procedure was used to identify which subset of traits were most strongly associated with a specific ROI.

Results: In White women, the 20 kb ROI around rs10734411 was significantly associated with the multivariate distribution of age at FMP, premenopausal estradiol, and postmenopausal estradiol (omnibus p -value = .00004). This association did not replicate in the smaller sample of Black women.

Conclusion: This study using a region-based, multiple-trait approach suggests a shared genetic basis among multiple facets of reproductive aging.

KEYWORDS

estradiol, follicle stimulating hormone, genetics, menopause

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Molecular Genetics & Genomic Medicine* published by Wiley Periodicals LLC.

1 | BACKGROUND

Ovarian aging is a complex, highly orchestrated phenomenon that reflects declining ovarian reserve (Burger et al., 2007). Several cohort studies of midlife women have described key characteristics of ovarian aging including age at menopause (as defined by age at final menstrual period (FMP)), and the timing and longitudinal trajectories of change in estradiol (E2) and follicle stimulating hormone (FSH) levels during the menopause transition (Burger et al., 2007; Dennerstein et al., 2007; Finkelstein et al., 2020; Freeman et al., 2005; Randolph Jr et al., 2011; Sowers et al., 2008a, 2008b). The transmenopause (Greendale et al., 2012) refers to the period from approximately 2 years before the FMP to approximately 2 years after the FMP and is characterized by a sharp increase in levels of FSH and a sharp decline in levels of E2, with the peak and nadir of hormone levels varying across women (Burger et al., 2007; Dennerstein et al., 2007; Freeman et al., 2005; Randolph Jr et al., 2011; Sowers et al., 2008a, 2008b).

Several candidate gene studies have been conducted on age at natural menopause, and serum E2 and FSH levels (Dunning et al., 2004; Grigороva et al., 2008, 2010; Prescott et al., 2012; Ruth et al., 2016; Schüring et al., 2013). Although large genome-wide association studies (GWAS) have been conducted for age at natural menopause, few GWAS have been conducted for FSH and E2 levels in midlife women since few studies of the menopausal transition have both measures of hormones and DNA available in a large sample of women.

A GWAS meta-analysis of age at natural menopause, conducted in nearly 70,000 White women, identified 54 independent single nucleotide polymorphisms (SNPs) in 44 genomic loci that were genome-wide significantly associated ($p < 5 \times 10^{-8}$) with age at natural menopause (Day et al., 2015). Recently, all 54 loci were replicated in a meta-analysis of age at natural menopause in ~200,000 women of European Ancestry (Ruth et al., 2021). The 54 identified loci identified by Day et al. (2015) harbored genes involved in DNA damage response as well as primary ovarian insufficiency and delayed puberty. The genome-wide significant SNPs explained approximately 6% of the variance in age at natural menopause, and the variance explained increased to 21% using independent SNPs with at least marginal significance ($p < .05$).

Importantly, age at FMP and the accompanying hormone trajectories across the menopause transition do not occur in isolation, but likely share molecular pathways. Understanding the genetics underlying the endocrinology of the menopause transition may be enhanced by jointly analyzing multiple interrelated traits while accounting for between trait correlations rather than by using a

univariate approach that independently tests one trait at a time. Consideration of pleiotropic effects, where the same genetic factor can influence multiple traits, can increase power and expand understanding of related traits. While power for detecting individual SNPs associated with an outcome is limited because the causal SNP is not likely to be the typed or imputed SNP, considering the joint effect of multiple SNPs in a region may be advantageous because several SNPs in the region may be in linkage disequilibrium (LD) with the causal SNP (Wu et al., 2010). Thus, a region-based approach utilizes information from different but correlated SNPs in a region and may have increased power to detect an association (Wu et al., 2010).

Therefore, we used Multi-SKAT (Dutta et al., 2018), a general framework that extends mixed effect models, to jointly test associations of multiple menopausal traits (i.e., age at FMP and premenopausal and postmenopausal levels of FSH and E2) with selected genetic regions in White and Black women from the Study of Women's Health Across the Nation (SWAN). The selected genetic regions were based on the 54 independent SNPs previously identified to be significantly associated ($p < 5 \times 10^{-8}$) with age at natural menopause (Day et al., 2015).

2 | METHODS

2.1 | The Study of Women's health across the nation (SWAN)

2.1.1 | Ethical compliance

The study protocol was approved by the Institutional Review Boards at each study site. Participants provided written, informed consent at each visit.

The SWAN cohort includes 3302 women (1550 White, 935 Black, 286 Hispanic, 250 Chinese, and 281 Japanese) enrolled in 1996/1997 at 7 clinical sites. Since the baseline visit, women have been followed approximately annually for 15 follow-up visits, with the study remaining in contact with 75% of surviving participants. Details of the recruitment and enrollment have been described elsewhere (Sowers et al., 2000). In brief, a cross-sectional screening survey identified women eligible for the cohort study. In addition to residence in the geographic area of the site, eligibility criteria included being age 42–52 years at baseline. Women had to have an intact uterus, not be pregnant or lactating, and have had at least one menstrual period and not used reproductive hormones in the past 3 months.

During follow-up visits 5 and 6, whole blood was collected at six of the participating sites to generate genetic materials including extracted DNA and immortalized B-lymphocyte cells which were stored at the SWAN

Repository. A separate informed consent was obtained for the DNA collection. At the participating sites, 1757 (88.3%) of the 1988 women who attended these clinic visits consented to provide genetic materials, representing 52% of the White, 44% of the Black, and 60% of the Chinese and Japanese women in the original cohort. Immortalized cell lines were developed successfully for 1588 (90.4%), 1536 of which (96.7%) were processed into distributable diluted, extracted DNA. Of those with available DNA, 1464 (95.3%; 740 White, 368 Black, 139 Chinese, 146 Japanese) were successfully genotyped at the genotyping center.

2.2 | Study design

For the current analyses, we required that women have complete data for all measures described below. We performed the analyses in two stages. Stage 1 discovery analyses were conducted in White women and Stage 2 replication analyses were conducted in Black women. Stage 1 analyses were carried out for all 54 genetic regions described below, whereas only genetic regions with suggestive levels of significance ($p < .05$) in the Stage 1 analyses were analyzed in Stage 2. We chose the White women for discovery analyses because of their larger sample size compared to Black women. Due to sample size concerns, Chinese and Japanese women were not included in the current study.

2.3 | Measures

Each visit included interviewer-administered and self-administered questionnaires that assessed information on menstrual characteristics and current hormone therapy use, sociodemographic characteristics including education, and lifestyle including smoking. Physical assessments included measurement of height and weight, and a fasting blood draw was obtained. At baseline, women's history of contraceptive use was ascertained. Women also maintained a monthly menstrual calendar where they recorded prospectively the days that they bled until they reached their FMP or had a hysterectomy.

2.3.1 | Hormone assays

Serum E2 assays were conducted in duplicate and FSH assays in singlicate using an ACS-180 automated analyzer (Bayer Diagnostics Corp., Norwood, MA). Serum E2 concentrations were measured with a modified, off-line ACS-180 (E2-6) immunoassay. Inter- and intra-assay coefficients of variation averaged 10.6% and 6.4%, respectively,

over the assay range, and the lower limit of detection was 1 pg/ml. Serum FSH concentrations were measured with a two-site chemiluminometric immunoassay. Inter- and intra-assay coefficients of variation were 12.0% and 6.0%, respectively, and the lower limit of detection was 1.1 IU/liter. If the blood draw was prior to menopause, we also recorded whether the measurement was taken during cycle days 2–5 or outside of this window.

2.3.2 | Genotype data

Genotyping methods and quality control procedures are described in Supplementary Table S1. Samples were genotyped on the Illumina Multi-Ethnic Global Array (MEGA A1; genome build GRCh37/hg19) at the Center for Inherited Disease Research (CIDR) at Johns Hopkins University. Imputation to the 1000G Phase 3 v5 reference panel was performed using the Michigan Imputation Server (Das et al., 2016). SWAN genetic data are available from dbGaP (accession number: phs001470). The King-robust (Manichaikul et al., 2010) method was used to select a maximum set of unrelated women by removing one from each pair of women who showed cryptic relatedness at 3rd degree or closer within White and Black women separately. Principal component analysis was used to further select an unrelated homogenous analysis sample by excluding outliers who deviated more than 2 standard deviations (SD) from the mean on eigenvector 1 or 2. Then, genetic principal components (PCs) were recalculated in the remaining White ($n = 740$) and Black ($n = 368$) women separately. Next, 54 genetic regions of interest (ROI) were defined by all variants that fell within ± 10 kb of each of the 54 independent genome-wide significant SNPs identified by Day et al. (2015). While the choice of window size is arbitrary, we selected a window of 10 kb on either side of each lead SNP to include potential regulatory regions.

2.3.3 | Traits of interest

Age at menopause was defined as age at the FMP, with the FMP defined retrospectively after 12 months of amenorrhea. Surgical menopause was defined by report of either hysterectomy or bilateral oophorectomy. The FMP could not be determined in women who began hormone therapy and had no subsequent untreated menstrual bleeds. Women who had surgical menopause or whose FMP could not be determined because of hormone use or missing data were excluded from this analysis. For E2 and FSH, we assessed women's levels at the start and end of the transmenopause, using the measurement that was closest to the date 2 years before and 2 years after the FMP,

respectively (Randolph Jr et al., 2011), in order to capture values before the accelerated rise of FSH and the sharp decline in E2 marking the late menopausal transition and values once the changes plateaued after menopause. Due to skewed distributions, all hormone measurements were log-transformed prior to analyses.

2.4 | Statistical analysis

For the current analyses, we required that women have complete data for all measures. Of the 740 White women and 368 Black women, 12 White and 18 Black women were missing baseline risk factor data; 281 White and 114 Black women were missing age at FMP; and 100 White and 72 Black women were missing hormone data. Therefore, the analytical samples for Stage 1 and Stage 2 included 347 White women and 164 Black women, respectively.

Medians and interquartile ranges were calculated for each trait as well as for continuous baseline demographic and risk factor variables. Frequency and percentages were calculated for discrete baseline demographic and risk factor variables. Differences between White and Black women were evaluated with two-sample nonparametric median tests for continuous variables and chi-squared tests for categorical variables. The Pearson correlation coefficient among the traits of interest was estimated.

2.4.1 | SKAT-O

We tested for region-based associations between each ROI and each of the five traits of interest using the optimal unified sequence kernel association test (SKAT-O) (Lee et al., 2012; Wu et al., 2011). SKAT-O performs both a genetic burden test as well as the SNP-Set (Sequence) Kernel Association Test (SKAT) test and adaptively selects the best linear combination of both tests to maximize test power. The burden test is a method that evaluates whether a composite score of the number of minor alleles for the variants in an ROI is associated with the outcome. Burden tests are optimal when all the rare variants in the ROI have identical effect sizes and directions. SKAT is a score-based variance component test that evaluates the joint effect of multiple variants in an ROI on an outcome of interest. SKAT assumes that the effect size of each individual SNP in the region follows an arbitrary distribution with mean zero. The test statistic assesses whether the variance of this distribution deviates from zero, testing the hypothesis that at least one variant in the ROI is associated with the outcome. The contribution of individual variants to the test statistic can be weighted by minor allele frequency (MAF). We used SKAT-O with a weighted

kernel [Beta(1,25)] that up-weights rare variants while applying some weight to variants with MAF 1%–5% and negligible weight to common variants.

We performed analyses separately for each trait, adjusting for baseline body mass index (BMI; weight in kilograms divided by height in meters squared), smoking status (current/former versus never), education (high school degree or less, some college, college degree, or more), oral contraceptive use (yes/no), as well as study site and the first five genetic PCs. The minimum SKAT-O *p*-value across all 5 traits was defined as the “minPhen” (Conneely & Boehnke, 2007) for each of the 54 ROIs. To adjust for multiple testing, a Bonferroni correction was used to determine statistical significance of the minPhen for 54 ROIs (minPhen <0.00093; i.e., 0.05/54 = 0.00093).

2.4.2 | Multi-SKAT

We tested for pleiotropic effects of each ROI on the five traits of interest using multivariate kernel regression (Multi-SKAT) (Dutta et al., 2018). Multi-SKAT models the association of variants in an ROI on traits through a kernel matrix and performs a variance component test of association. Multi-SKAT is flexible in modeling the relationships between sets of traits and an ROI of variants using both a trait kernel and a genomic kernel. Three trait kernels were analyzed: 1) homogeneous kernel in which the effect sizes on different traits are assumed to be homogeneous; 2) heterogeneous kernel in which the effect sizes on different traits are assumed to be uncorrelated among themselves; 3) trait covariance kernel which assumes that the correlation among effect sizes is proportional to that among the residual traits after adjusting for nongenetic covariates. The combined results of the different trait kernels were evaluated with a minimum *p* value-based omnibus test (minPcom) (Dutta et al., 2018). Models were adjusted using the same variables as in the SKAT-O analyses described above. To adjust for multiple testing, a Bonferroni correction was used to determine statistical significance for 54 ROIs (minPcom <0.00093; i.e., 0.05/54 = 0.00093). Since the current analyses are exploratory, we further investigated ROIs considered to be suggestive with a less stringent minPcom <0.05.

2.4.3 | Backward trait elimination

After identifying ROIs at least suggestively associated in the multivariate analyses, we performed a backward elimination procedure to identify which subset of traits were most strongly associated with the specific ROI (Ionita-Laza et al., 2014). This procedure iteratively removes

traits, one at a time, based on the minPcom of the reduced set of traits. Removing a strongly associated trait will likely result in an increase of the minPcom while removing a trait that is contributing noise will likely lead to a decrease in the minPcom. For Step 1: Start with the full set of five traits, $V_c = \{v_1, \dots, v_5\}$, and compute the minPcom. Step 2: Remove each of the traits one at a time from V_c and compute the corresponding minPcom for the remaining traits. Step 3: If the minimum minPcom of a reduced set is less than the minPcom for V_c , then remove the trait that led to the smallest minPcom. Continuing iteratively, repeat Steps 2 and 3; if the current p -value cannot be improved, then return the current set of traits.

2.4.4 | Backward variant elimination

Backward variant elimination was used to help us determine which variants were contributing signal versus noise to the overall association of an ROI with the determined subset of multivariate traits. For each of the final Multi-SKAT models identified in the backward trait elimination procedure described above, we reran the final model leaving out one variant at a time for each variant in the ROI. If the minPcom decreased after removing a specific variant, then it could be assumed that the variant was adding noise to the model that included all remaining variants in the ROI. Conversely, if the minPcom increased after removing a specific variant, then it could be assumed that the variant was adding signal to the model that included all remaining variants in the ROI. We followed a similar iterative procedure as described above, removing variants with the smallest minPcom, one at a time from the ROI. Rather than stopping when the minPcom could not be improved, we continued to remove variants until only two variants remained in the dataset. The backward variant elimination results were presented graphically by plotting the $-\log_{10}(\text{minPcom})$ versus the iteration in which the variants were removed from the ROI. Thus, each point in the plot reflects the minPcom at the iteration including the subset of variants remaining in the ROI.

The minPcom from the backward variant elimination is not easily interpretable due to the large and variable number of variants in each ROI. Therefore, we performed a simulation study to estimate the distribution of the ROI-specific minPcom under the null hypothesis to determine the Type 1 error rate for the minPcom in the original data. We simulated 100 data sets by permuting the traits/covariates together and conducted Multi-SKAT analyses for each set of 100 permuted traits/covariates using the same final model of traits determined in the original backward trait analyses. We then used backward variant elimination on each simulated dataset and estimated 95% confidence

intervals for the empirical minPcoms. The Type 1 error rate was defined as the proportion of empirical minPcoms that were less than the observed minPcom of the best subset of variants in the original analysis for the respective ROI. Finally, the Type 1 error rate at $\alpha = 0.05$ was Bonferroni corrected for the number of ROIs evaluated with backward variant elimination.

For the best subset of biallelic variants in an ROI, we examined pairwise LD statistics for the respective population using LDmatrix (Machiela & Chanock, 2015). Next, these variants were thinned by LD using an R^2 threshold of 0.1 and MAF threshold of 0.01, to identify a set of independent variants in the ROI. Finally, we queried the *Atlas of GWAS Summary Statistics* (Watanabe et al., 2019) to identify associations between these independent variants and any traits at $p < .001$.

2.4.5 | Stage 2 replication analyses

Multi-SKAT analyses were conducted for the ROIs at least suggestively associated ($p < .05$) in the Stage 1 Multi-SKAT analyses. For these analyses, the same subset of traits most strongly associated with the specific ROI in Stage 1 backward trait elimination was investigated in Stage 2 analyses. Due to differences in linkage disequilibrium structure and allele frequencies between White and Black women, backward variant elimination was not conducted in the Stage 2 replication analyses.

3 | RESULTS

3.1 | Participant characteristics

Characteristics of women in the current investigation are presented in Table 1. In the Stage 1 discovery sample of 347 White women, the median (IQR) age at FMP was 52.0 (3.8) years. Over one-half (60%) of women had a college or post-college education and over three-quarters (76.4%) had used oral contraceptives; 15% were current smokers. Median (IQR) BMI was 26.0 (9.5) kg/m^2 . The median (IQR) premenopausal and postmenopausal E2 was 50.1 pg/ml (88.4) and 16.1 pg/ml (12.2), respectively. The median (IQR) premenopausal and postmenopausal FSH was 25.4 IU/L (28.4) and 105.1 IU/L (56.5), respectively.

In the Stage 2 replication sample of 164 Black women, the median (IQR) age at FMP was 52.0 (3.9) years. 11% of women had a college or post-college education and approximately three-quarters (74.4%) used oral contraceptives prior to baseline; 28% were current smokers. Median (IQR) BMI was 30.5 (10.9) kg/m^2 . The median (IQR) premenopausal and postmenopausal E2 was 46.4 pg/ml (71.9)

TABLE 1 Descriptive statistics of women in the Study of Women's Health Across the Nation (SWAN) with complete data for traits of interest, genetic regions of interest (ROI), and risk factors

	Stage 1 discovery White women (N = 347)	Stage 2 replication Black women (N = 164)	<i>p</i> *
	Median (IQR)	Median (IQR)	
Age at FMP (years)	52.0 (3.8)	52.0 (3.9)	.9485
E2 premenopausal (pg/ml)	50.1 (88.4)	46.4 (71.9)	.7276
E2 postmenopausal (pg/ml)	16.1 (12.2)	21.1 (14.0)	<.0001
FSH premenopausal (IU/L)	25.4 (28.4)	25.6 (30.0)	.9758
FSH postmenopausal (IU/L)	105.1 (56.5)	81.6 (60.2)	<.0001
BMI (kg/m ²)	26.0 (9.5)	30.5 (10.9)	<.0001
	N (%)	N (%)	
Education			<.0001
High school degree or less	50 (14.4)	45 (27.4)	
Some college	89 (25.6)	81 (49.3)	
College degree or more	208 (60.0)	38 (11.0)	
History of oral contraceptive use	265 (76.4)	122 (74.4)	.6262
Current smoker	52 (15.0)	46 (28.0)	.0005

Abbreviations: BMI, body mass index; E2, estradiol; FMP, final menstrual period; FSH, follicle stimulating hormone.

**P*: *p*-value from two-sample nonparametric median tests for continuous variables and chi-squared tests for categorical variables.

and 21.1 pg/ml (14.0), respectively. The median (IQR) premenopausal and postmenopausal FSH was 25.6 IU/L (30.0) and 81.6 IU/L (60.2), respectively.

Median postmenopausal E2 was significantly lower in White compared to Black women (16.1 pg/ml vs. 21.1 pg/ml, respectively; $p < .0001$), and median postmenopausal FSH was significantly higher in White compared to Black women (105.1 IU/L vs. 81.6 IU/L, respectively; $p < .0001$). Median BMI was significantly lower in White compared to Black women (26.0 vs. 30.5 kg/m², respectively; $p < .0001$). There were significant differences ($p < .0001$) in educational attainment between White and Black women. The percentage of women who were current smokers was significantly less among White versus Black women (15.0 vs 28.0%, respectively; $p < .0001$).

3.2 | Correlation among age at menopause and hormone levels

Pearson correlation coefficients among age at FMP and log-transformed hormone levels are presented in Table 2. In the Stage 1 discovery sample of White women, premenopausal measures of E2 and FSH were negatively correlated ($r = -0.594$ [$p < .001$]), as were postmenopausal measures of E2 and FSH ($r = -0.428$ [$p < .001$]). Postmenopausal E2 was negatively correlated with premenopausal FSH ($r = -0.168$ [$p < .01$]). Premenopausal

and postmenopausal measures of FSH were positively correlated ($r = 0.184$ [$p < .001$]).

In the Stage 2 replication sample of Black women, premenopausal measures of E2 and FSH were negatively correlated ($r = -0.556$ [$p < .001$]), as were postmenopausal measures of E2 and FSH ($r = -0.303$ [$p < .001$]). Premenopausal and postmenopausal measures of FSH were positively correlated ($r = 0.203$ [$p < .001$]).

3.3 | Stage 1 discovery in white women

3.3.1 | Distribution of variant types across 54 ROIs

The mean (SD) number of all variants per ROI across all ROIs was 215.65 (62.13) (Supplementary Table S2). The mean (SD) number of monomorphic variants per ROI was 46.06 (24.31). The mean (SD) number of rare (MAF < 0.01), low frequency (0.01 ≤ MAF < 0.05), and common (MAF ≥ 0.05) variants per ROI was 102.06 (36.86), 20.07 (13.27), and 47.46 (26.19), respectively.

3.3.2 | SKAT-O for each trait of interest and ROI

Supplementary Table S3 shows SKAT-O *p*-values and the minPhen for associations of each ROI with

TABLE 2 Correlation among the age at FMP and four log-transformed hormone levels in the Study of Women's Health Across the Nation (SWAN) sample (347 White women and 164 Black women)

		Age at FMP (years)	E2 premenopausal (pg/ml)	E2 postmenopausal (pg/ml)	FSH premenopausal (IU/L)	FSH postmenopausal (IU/L)
Age at FMP (years)	White	1.000				
	Black	1.000				
E2 premenopausal (pg/ml)	White	-0.095	1.000			
	Black	-0.039	1.000			
E2 postmenopausal (pg/ml)	White	0.024	0.099	1.000		
	Black	0.014	-0.068	1.000		
FSH premenopausal (IU/L)	White	0.057	-0.594***	-0.168**	1.000	
	Black	0.090	-0.556***	0.026	1.000	
FSH postmenopausal (IU/L)	White	0.006	0.063	-0.428***	0.184***	1.000
	Black	0.092	0.141	-0.303***	0.203**	1.000

Abbreviations: E2, estradiol; FMP, final menstrual period; FSH, follicle stimulating hormone.

** $.001 \leq p < .01$; *** $p < .001$.

TABLE 3 Stage 1 discovery p -values for significant and suggestive 20Kb ROIs associated with age at FMP and 4 log-transformed hormone levels in 347 White women from the Study of Women's Health Across the Nation (SWAN). These are the ROIs that will be further explored with backward trait elimination

ROI	CHR	POS	minPhen	SKAT.Omni	Burden.Omni	minPcom
rs707938	6	31,729,359	0.01522	0.05720	0.01569	0.02873
rs10734411	11	32,541,784	0.00003	0.00059	0.10599	0.00151
rs3741604	12	66,696,410	0.00015	0.02203	0.01747	0.03182

Notes: minPhen: minimum SKAT-O p -value from single trait analyses across all 5 traits (i.e., age at FMP and four log-transformed hormonal measures). SKAT.Omni: Multi-SKAT minimum p value based on omnibus test across all SKAT kernels. Burden.Omni: Multi-SKAT minimum p value based on omnibus test across all Burden-type kernels. minPcom: Multi-SKAT minimum p value based on omnibus test across all SKAT and Burden-type kernels.

Abbreviations: CHR, chromosome; FMP, final menstrual period; ROI, region of interest; POS, position of SNP (build GRCh37/hg19).

each trait of interest. Quantile-quantile (QQ) plots for the minPhen for each trait of interest are shown in Supplementary Figure S1. Since the ROIs were selected for previously identified variants associated with age at FMP, the plots appear to show inflation. The lambdas ranged from 1.22 to 2.32.

Two ROIs, around rs10734411 (minPhen = 0.00003) and around rs3741604 (minPhen = 0.00015), were each statistically significant with at least one of the five traits of interest. The ROI around rs10734411 was significantly associated with premenopausal E2 (SKAT-O p -value = .00003), and there was suggestive evidence of an association with age at FMP (SKAT-O p -value = .01220). The ROI around rs3741604 was significantly associated with postmenopausal FSH (SKAT-O p -value = .00015), and there was suggestive evidence of an association with premenopausal E2 (SKAT-O p -value = .03642) and postmenopausal E2 (SKAT-O p -value = .00248). The ROI around rs707938 was suggestively associated with postmenopausal E2 (SKAT-O p -value = .01522).

3.3.3 | Multi-SKAT

Supplementary Table S4 shows Multi-SKAT results for associations of each ROI with all five traits. The QQ plot for the minPcom is shown in Supplementary Figure S2. Table 3 shows the minPhen from the single trait analyses and Multi-SKAT results for three ROIs with minPcom <0.05. The minPcom from the Multi-SKAT analyses including all five traits of interest was always greater than minPhen which indicates the lowest SKAT-O p -value across the individual traits. The ROI with the overall strongest association in Multi-SKAT analyses was around rs10734411 (minPcom = 0.00151); the SKAT.Omni p -value for this ROI was 0.00059.

3.3.4 | Backward trait elimination

The three ROIs with Multi-SKAT minPcom <0.05 are presented in Table 4. The best multivariate reduced model

TABLE 4 Stage 1 discovery backward trait elimination results for top ROIs in Table 3 showing the traits and Multi-SKAT *p*-values for the best multivariate model in 347 White women from the Study of Women's Health Across the Nation (SWAN)

ROI	Traits in best multivariate model	SKAT.Omni	BURDEN.Omni	minPcom
rs707938	Age at FMP E2 premenopausal FSH premenopausal FSH postmenopausal	0.03290	0.00452	0.00924
rs10734411	Age at FMP E2 premenopausal E2 postmenopausal	0.00003	0.01528	0.00004
rs3741604	E2 premenopausal FSH postmenopausal	0.00200	0.00251	0.00370

Notes: SKAT.Omni: Multi-SKAT minimum *p* value based on omnibus test across all SKAT kernels. Burden.Omni: Multi-SKAT minimum *p* value based on omnibus test across all Burden-type kernels. minPcom: Multi-SKAT minimum *p* value based on omnibus test across all SKAT and Burden-type kernels. Abbreviations: E2, estradiol; FMP, final menstrual period; FSH, follicle stimulating hormone; ROI, region of interest.

for the ROI around rs10734411 was statistically significant (minPcom = 0.00004) and included the traits age at FMP, premenopausal E2, and postmenopausal E2. While the omnibus SKAT test for this model was statistically significant (SKAT.Omni = 0.00003), the omnibus burden test was only suggestively significant (Burden.Omni = 0.01528). After backward trait elimination analyses, the ROIs around rs707938 and rs3741604 were suggestively significantly associated (minPcom = 0.00924 and 0.00370, respectively), each with a different subset of traits as shown in Table 4.

3.3.5 | Backward variant elimination

The influence of single variants was investigated for the best set of traits identified in the backward trait elimination analyses for each of the three ROIs investigated above. Figure 1 shows the change in minPcom as variants were iteratively removed one-by-one for each of the three ROIs. Supplementary Figure S3 shows the distribution of the ROI-specific minPcoms under the null hypothesis as determined in the simulation study of 100 permuted datasets. After Bonferroni correction for three ROIs, statistical significance for the Type 1 error rate was defined as 0.01667 (i.e., $0.05/3 = 0.01667$).

The best subset of variants in the ROI around rs10734411 (minPcom = 2.44×10^{-15}) included 58 variants out of 221 polymorphic variants in the original data. In the simulation study, the mean (95% CI) minPcom was 0.0009 (0.0005, 0.0014) and the minimum minPcom was 1.33×10^{-15} . There was one empirical minPcom less than the observed minPcom resulting in a Type 1 error rate of 0.01.

The best subset of variants in the ROI around rs707938 (minPcom = 2.2×10^{-9}) included 42 variants out of 137 polymorphic variants in the original data. In

the simulation study, the mean (95% CI) minPcom was 0.0009 (0.0001, 0.0018) and the minimum minPcom was 1.40×10^{-12} . There were six empirical minPcoms less than the observed minPcom resulting in a Type 1 error rate of 0.06.

The best subset of variants in the ROI around rs3741604 (minPcom = 8.87×10^{-7}) included 94 variants out of 162 polymorphic variants in the original data. In the simulation study, the mean (95% CI) minPcom was 0.0066 (0.0032, 0.0101) and the minimum minPcom was 3.33×10^{-12} . There were two empirical minPcoms less than the observed minPcom resulting in a Type 1 error rate of 0.02.

3.4 | Stage 2 replication in black women

The three ROIs at least suggestively associated in the Stage 1 Multi-SKAT analyses (rs707938, rs10734411, and rs3741604) were examined in Stage 2 replication analyses in 164 Black women from SWAN.

3.4.1 | Distribution of variant types across three ROIs

The mean (SD) number of all variants per ROI across the three ROIs was 235.33 (47.01) (Supplementary Table S5). The mean (SD) number of monomorphic variants per ROI was 16.33 (23.97). The mean (SD) number of rare, low frequency, and common variants per ROI was 95.33 (21.96), 45.67 (15.53), and 78.00 (33.78), respectively.

3.4.2 | Multi-SKAT

Table 5 shows Multi-SKAT results for associations of each of the three ROIs with the best subset of traits as found

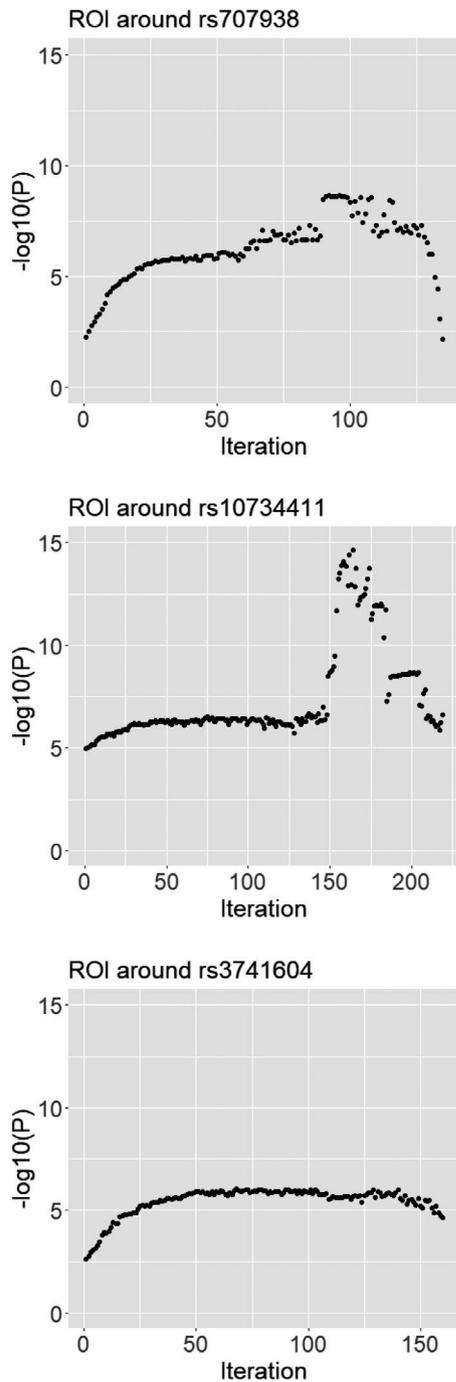


FIGURE 1 Backward variant elimination: Change in minPcom as variants were iteratively removed one-by-one for each of the three regions of interest (ROI) in Stage 1 discovery in 347 White women from the Study of Women's Health Across the Nation (SWAN). minPcom : Multi-SKAT minimum p value based on omnibus test across all SKAT and Burden-type kernels

in Stage 1 discovery backward trait elimination analyses. None of the 3 ROIs reached a suggestive level of evidence for an association with the indicated traits of interest in Stage 2 replication among 164 Black women.

4 | DISCUSSION

This study jointly analyzed associations between five interrelated menopausal traits and selected genetic ROIs related to age at natural menopause (Day et al., 2015) using a multivariate kernel regression approach. We used backward elimination to identify the best subset of traits associated with an ROI as well as the best subset of variants within the ROI. After backward trait elimination analyses in 347 White women from the SWAN Study, the ROI around rs10734411 was statistically significantly associated ($\text{minPcom} = 0.00004$) with age at FMP, premenopausal E2, and postmenopausal E2 in a multivariate model. Suggestive levels of evidence were found for ROIs around rs707938 and rs3741604 with $\text{minPcom} = 0.00924$ and 0.00370 , respectively. These findings suggest a pleiotropic relationship between these ROIs and sets of menopausal traits as evaluated in the current study.

4.1 | Pleiotropy

Investigations of genetic pleiotropy, the association of a single gene or locus with more than one trait, may improve our understanding of the underlying genetic architecture of complex traits (Schaid et al., 2016). Biologic pleiotropy occurs when a single gene has a direct biologic effect on multiple traits (Schaid et al., 2016; Solovieff et al., 2013). A single variant in a gene may be associated with different traits or different variants in a gene may be associated with different traits. Mediated pleiotropy may be observed when one trait is causally associated with a second trait and the gene that is associated with the first trait is indirectly associated with the second trait. A gene may also appear to be associated with multiple traits due to spurious associations or various biases such as trait misclassification or ascertainment bias.

As reviewed by Solovieff et al. (2013), one approach to investigating pleiotropy is to compare the univariate marginal associations of a locus with multiple traits. Another approach is to reduce the number of traits by estimating principal components and then conducting the analyses with the principal components. In the current analyses, we used a multivariate regression approach to regress multiple traits on an ROI. The multivariate approach is generally more powerful than a univariate approach (Galesloot et al., 2014).

4.2 | ROI around rs10734411

In the current study, the ROI around rs10734411 was significantly associated ($\text{minPcom} = 0.00004$) with the

TABLE 5 Stage 2 replication Multi-SKAT *p*-values in 164 Black women from the Study of Women's Health Across the Nation (SWAN) for top ROIs and traits remaining in the best model in Stage 1 discovery (as shown in Table 4)

ROI	Traits in best multivariate model	SKAT.Omni	BURDEN.Omni	minPcom
rs707938	Age at FMP E2 premenopausal FSH premenopausal FSH postmenopausal	0.2300	0.5318	0.3979
rs10734411	Age at FMP E2 premenopausal E2 postmenopausal	0.0879	0.2343	0.1565
rs3741604	E2 premenopausal FSH postmenopausal	0.8544	0.5685	0.8056

Notes: SKAT.Omni: Multi-SKAT minimum *p* value based on omnibus test across all SKAT kernels. Burden.Omni: Multi-SKAT minimum *p* value based on omnibus test across all Burden-type kernels. minPcom: Multi-SKAT minimum *p* value based on omnibus test across all SKAT and Burden-type kernels. Abbreviations: E2, estradiol; FMP, final menstrual period; FSH, follicle stimulating hormone; ROI, region of interest.

multivariate distribution of age at FMP and pre- and postmenopausal E2 levels in 347 White women. This multivariate finding did not replicate in Black women (minPcom = 0.1565). Importantly, the number of women in the sample of Black women ($N = 164$) was less than one-half of the number of women in the sample of White women ($n = 347$). Additionally, there are known differences in LD patterns and allele frequencies between White and Black populations (Reich et al., 2001).

In univariate SKAT-O analyses (Supplementary Table S3), the ROI around rs10734411 was significantly associated with postmenopausal E2 (minPhen = 0.00003) and marginally associated with age at FMP (minPhen = 0.01220) but not with premenopausal E2 (minPhen = 0.06163). Thus, it appears that the univariate association with postmenopausal E2 drives the observed multivariate association with age at FMP and pre- and postmenopausal E2.

The closest gene to rs10734411 is eukaryotic translation initiation factor 3 (*EIF3M*; 11p13; OMIM accession number: 609641; GenBank: NM_006360.6, NP_006351.2) (Day et al., 2015). This gene encodes a protein that is part of the eukaryotic translation initiation factor 3 complex (eIF-3) required for protein synthesis. This gene is overexpressed in the ovary in women.

The backward variant elimination conducted in this study is one approach to identifying a subset of variants in an ROI for further consideration with computational fine mapping and/or follow-up functional studies (Ionita-Laza et al., 2014). The best subset of variants in the ROI around rs10734411 included 58 variants (minPcom = 2.44×10^{-15}). The index variant, rs10734411 (MAF = 0.49), was not in the subset of variants found to be most significant after backward variant elimination. Due to limited sample sizes, we were not able to conduct meaningful investigations of the associations between individual variants within the ROI and the traits of interest.

We constructed a heatmap matrix of pairwise LD statistics (Supplementary Figure S4) for the 50 biallelic variants in the ROI using Utah Residents from North and West Europe (CEU) as the reference population (Machiela & Chanock, 2015). These 50 variants were then thinned by LD using an R^2 threshold of 0.1 and MAF threshold of 0.01, revealing 12 independent variants. Three of these variants (rs11031842, rs181823223, rs12285345) were in LD ($R^2 = 0.58, 0.14, \text{ and } 0.32$, respectively) with the index variant (rs10734411) that defines this ROI. Among the remaining nine variants, six (rs182674327, rs35371236, rs192792544, rs567330231, rs201699741, and rs76500960) were not associated with any traits at $p < .001$ in the *Atlas of GWAS Summary Statistics* (Watanabe et al., 2019). Three variants (rs11031851, rs558036469, and rs186872950), however, were associated at $p < .001$ with traits in several domains. These domains include respiratory, neurological, endocrine, and immunological and suggest traits for future investigation of pleiotropic effects.

Importantly, to place these univariate findings in context, our Multi-SKAT results are based on analyses of a set of variants in the ROI, not single individual variants, and the Multi-SKAT test up-weights rare variants and down-weights common variants. Furthermore, the Multi-SKAT test evaluates the association of a set of variants with the multivariate distribution of multiple traits, not with the distribution of a single univariate trait.

4.3 | ROIs around rs707938 and rs3741604

Suggestive levels of evidence were found for ROIs around rs707938 and rs3741604. The closest gene to rs707938 is MutS Homolog 5 (*MSH5*; 6p21.33, OMIM accession number: 603382; GenBank: NG_011611.1) (Day et al., 2015). Polymorphisms in this gene have been linked to premature

ovarian failure. The closest gene to rs3741604 is DNA Helicase B (*HELB*; 12q14.3, OMIM accession number: 614539; GenBank: NM_001370285.1, NP_001357214.1) (Day et al., 2015). This gene is involved in DNA replication, repair, recombination, and transcription.

4.4 | Previous studies

In a previous paper evaluating polygenic risk scores (PRS) for age at menopause in SWAN, we reported that an increased menopause PRS was significantly associated with later age at menopause, as well as with length of the reproductive lifespan and length of the menopausal transition in White women, but not Black women (Zhao et al., 2021). These findings also suggest genetic pleiotropy of traits related to reproductive aging.

A few single nucleotide polymorphisms (SNPs), most notably in the *CYP19A1* gene in European ancestry and multi-ancestry studies (Eriksson et al., 2018; Prescott et al., 2012), are reported to be significantly associated with E2 levels. A recent GWAS from the Twins UK study ($n = 2913$), which included about 11% perimenopausal and 52% postmenopausal women, reported an association between E2 levels and an intronic SNP, rs117585797, in *ANO2* (Ruth et al., 2016). *FSHB* is a key locus associated with both FSH level and age at menopause (Grigorova et al., 2010; Rull et al., 2018; Ruth et al., 2016; Schüring et al., 2013). In the Twins UK GWAS, the C allele of an *FSHB* intergenic SNP, rs11031005 (chromosome:11; GRCh37/hg19 position:30,226,356), was associated with lower FSH levels (Ruth et al., 2016). This same SNP is in perfect LD ($R^2 = 1.0$; $D' = 1.0$) with rs11031006 (chromosome:11; GRCh37/hg19 position:30,226,528) which was among the top genome-wide significant loci associated with delayed menopause in the Day et al. study (Day et al., 2015). In the current analyses, the ROI around rs11031006 was not significantly associated with age at FMP (SKAT-O p -value = .07188) or premenopausal FSH (SKAT-O p -value = .07882) (Supplementary Table S4).

A multi-ancestry meta-analysis of age at menopause was conducted in The Population Architecture using Genomics and Epidemiology (PAGE) Study (Fernández-Rhodes et al., 2018). Participants consisted of African American, Hispanic/Latina, Asian American, and American Indian/Alaska Native women. Genetic data included ~200,000 SNPs of primarily cardiometabolic loci genotyped on the *MetaboChip* (Illumina, Inc., San Diego, CA, USA). Genetic data for only 14 (approximately 25%) of the previously identified genetic loci associated with age at natural menopause were available. While two of these genetic loci were significantly associated with age at natural menopause in Hispanic/Latina women, none

were significantly associated with age at menopause in the 20,398 African American participants.

As FSH and E2 fluctuate across the menopausal transition and within a menstrual cycle, one annual measure may not precisely assess between-woman differences, which may have weakened the associations with hormone levels, particularly prior to menopause. Large-scale genetic studies with frequent hormonal assessments are needed to better understand the underlying genetic architecture of the menopausal transition. Further, the menopause GWAS used to identify the 54 ROIs included only European ancestry populations. Future large-scale GWAS that focus on non-European or multi-ancestry cohorts are needed to better understand the genetic risk profiles of reproductive aging in those populations. In the current analyses, we used a weighted kernel [Beta(1,25)] that weights variants in the ROI based on their MAFs. Kernels are flexible and future work could incorporate alternate weighting schemes such as weighting on functional information (Lee et al., 2012).

4.5 | Limitations and strengths

Black women were more likely than women of other race/ethnicities to have had a hysterectomy at the time of SWAN enrollment making them ineligible to participate, which may have led to differential patterns of selection bias. Black women were also somewhat less likely to participate in the SWAN Genetic Study, but participants and nonparticipants did not differ by age at menopause, or by education, smoking status, BMI, or contraceptive use. The small sample size limited study power and missing information on age at menopause and hormone levels, largely due to use of hormones, further limited power. SWAN is the largest cohort to date to document the natural history of the menopausal transition and one of the few such cohorts to have obtained genome-wide genotype data; thus, we are currently unable to replicate our findings in external cohorts. Strengths of this study include the multi-ancestry population, the prospectively collected data on reproductive aging traits that ensures more precise measures of menopause timing, and availability of serial hormone measurements enabling assessment of hormone levels before and after the FMP. Another strength is the use of a multiple traits simultaneously.

5 | CONCLUSIONS

In conclusion, this Multi-SKAT analysis suggests a shared genetic basis among multiple facets of reproductive aging, including age at menopause and specific

hormone levels before and after the FMP. Larger scale genetic studies with multiple hormone measurements across the menopausal transition are needed, as are studies that more explicitly consider the shared genetic basis of complex biological events such as reproductive aging.

ACKNOWLEDGMENTS

Clinical Centers: University of Michigan, Ann Arbor—Siobán Harlow, PI 2011–present; MaryFran Sowers, PI 1994–2011; Massachusetts General Hospital, Boston, MA—Joel Finkelstein, PI 1999–present; Robert Neer, PI 1994–1999; Rush University, Rush University Medical Center, Chicago, IL—Howard Kravitz, PI 2009–present; Lynda Powell, PI 1994–2009; University of California, Davis/Kaiser—Ellen Gold, PI; University of California, Los Angeles—Gail Greendale, PI; Albert Einstein College of Medicine, Bronx, NY—Carol Derby, PI 2011–present, Rachel Wildman, PI 2010–2011; Nanette Santoro, PI 2004–2010; University of Medicine and Dentistry—New Jersey Medical School, Newark—Gerson Weiss, PI 1994–2004; and the University of Pittsburgh, Pittsburgh, PA—Karen Matthews, PI.

NIH Program Office: National Institute on Aging, Bethesda, MD—Chhanda Dutta 2016–present; Winifred Rossi 2012–2016; Sherry Sherman 1994–2012; Marcia Ory 1994–2001; National Institute of Nursing Research, Bethesda, MD—Program Officers.

Central Laboratory: University of Michigan, Ann Arbor—Daniel McConnell (Central Ligand Assay Satellite Services).

SWAN Repository: University of Michigan, Ann Arbor—Siobán Harlow 2013–Present; Dan McConnell 2011–2013; MaryFran Sowers 2000–2011.

Coordinating Center: University of Pittsburgh, Pittsburgh, PA—Maria Mori Brooks, PI 2012–present; Kim Sutton-Tyrrell, PI 2001–2012; New England Research Institutes, Watertown, MA—Sonja McKinlay, PI 1995–2001.

Steering Committee: Susan Johnson, Current Chair; Chris Gallagher, Former Chair.

We thank the study staff at each site and all the women who participated in SWAN.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest.

DATA AVAILABILITY STATEMENT

SWAN genetic data are available from dbGaP (accession number: phs001470).

ORCID

Lawrence F. Bielak  <https://orcid.org/0000-0002-3443-8030>

REFERENCES

- Burger, H. G., Hale, G., Robertson, D., & Dennerstein, L. (2007). A review of hormonal changes during the menopausal transition: Focus on findings from the Melbourne Women's midlife health project. *Human Reproduction Update*, *13*, 559–565. <https://doi.org/10.1093/humupd/dmm020>
- Conneely, K. N., & Boehnke, M. (2007). So many correlated tests, so little time! Rapid adjustment of *p* values for multiple correlated tests. *The American Journal of Human Genetics*, *81*, 1158–1168. <https://doi.org/10.1086/522036>
- Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A. E., Kwong, A., Vrieze, S. I., Chew, E. Y., Levy, S., McGue, M., Schlessinger, D., Stambolian, D., Loh, P.-R., Iacono, W. G., Swaroop, A., Scott, L. J., Cucca, F., Kronenberg, F., Boehnke, M., ... Fuchsberger, C. (2016). Next-generation genotype imputation service and methods. *Nature Genetics*, *48*, 1284–1287. <https://doi.org/10.1038/ng.3656>
- Day, F. R., Ruth, K. S., Thompson, D. J., Lunetta, K. L., Pervjakova, N., Chasman, D. I., Stolk, L., Finucane, H. K., Sulem, P., Bulik-Sullivan, B., Esko, T., Johnson, A. D., Elks, C. E., Franceschini, N., He, C., Altmaier, E., Brody, J. A., Franke, L. L., Huffman, J. E., ... Murray, A. (2015). Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nature Genetics*, *47*, 1294–1303. <https://doi.org/10.1038/ng.3412>
- Dennerstein, L., Lehert, P., Burger, H. G., & Guthrie, J. R. (2007). New findings from non-linear longitudinal modelling of menopausal hormone changes. *Human Reproduction Update*, *13*, 551–557. <https://doi.org/10.1093/humupd/dmm022>
- Dunning, A. M., Dowsett, M., Healey, C. S., Tee, L., Luben, R. N., Folkerd, E., Novik, K. L., Kelemen, L., Ogata, S., Pharoah, P. D. P., Easton, D. F., Day, N. E., & Ponder, B. A. J. (2004). Polymorphisms associated with circulating sex hormone levels in postmenopausal women. *Journal of the National Cancer Institute*, *96*, 936–945. <https://doi.org/10.1093/jnci/djh167>
- Dutta, D., Scott, L., Boehnke, M., & Lee, S. (2018). Multi-SKAT: General framework to test for rare-variant association with multiple phenotypes. *Genetic Epidemiology*, *43*, 4–23. <https://doi.org/10.1002/gepi.22156>
- Eriksson, A. L., Perry, J., Coviello, A. D., Delgado, G. E., Ferrucci, L., Hoffman, A. R., Huhtaniemi, I. T., Ikram, M. A., Karlsson, M. K., Kleber, M. E., Laughlin, G. A., Liu, Y., Lorentzon, M., Lunetta, K. L., Mellström, D., Murabito, J. M., Murray, A., Nethander, M., Nielson, C. M., ... Ohlsson, C. (2018). Genetic determinants of circulating estrogen levels and evidence of a causal effect of estradiol on bone density in men. *The Journal of Clinical Endocrinology and Metabolism*, *103*, 991–1004. <https://doi.org/10.1210/jc.2017-02060>
- Fernández-Rhodes, L., Malinowski, J. R., Wang, Y., Tao, R., Pankratz, N., Jeff, J. M., Yoneyama, S., Carty, C. L., Setiawan, V. W., Le Marchand, L., Haiman, C., Corbett, S., Demerath, E., Heiss, G., Gross, M., Buzkova, P., Crawford, D. C., Hunt, S. C., Rao, D. C., ... North, K. (2018). The genetic underpinnings of variation in ages at menarche and natural menopause among women from the multi-ethnic population architecture using genomics and epidemiology (PAGE) Study: A trans-ethnic meta-analysis. *PLoS One*, *13*, e0200486. <https://doi.org/10.1371/journal.pone.0200486>

- Finkelstein, J. S., Lee, H., Karlamangla, A., Neer, R. M., Sluss, P. M., Burnett-Bowie, S. M., Darakananda, K., Donahoe, P. K., Harlow, S. D., Prizand, S. H., Joffe, H., Kumar, A., Martin, D. E., McConnell, D., Merrilat, S., Morrison, A., Pastore, L. M., Randolph, J. F., Greendale, G. A., & Santoro, N. (2020). Antimüllerian hormone and impending menopause in late reproductive age: The Study of Women's health across the nation. *Journal of Clinical Endocrinology and Metabolism*, *105*, e1862–e1871. <https://doi.org/10.1210/clinem/dgz283>
- Freeman, E., Sammel, M., Gracia, C., Kapoor, S., Lin, H., Liu, L., & Nelson, D. (2005). Follicular phase hormone levels and menstrual bleeding status in the approach to menopause. *Fertility and Sterility*, *83*, 383–392. <https://doi.org/10.1016/j.fertnstert.2004.06.066>
- Galesloot, T. E., van Steen, K., Kiemeny, L. A., Janss, L. L., & Vermeulen, S. H. (2014). A comparison of multivariate genome-wide association methods. *PLoS One*, *9*, e95923. <https://doi.org/10.1371/journal.pone.0095923>
- Greendale, G. A., Sowers, M., Han, W., Huang, M. H., Finkelstein, J. S., Crandall, C. J., Lee, J. S., & Karlamangla, A. S. (2012). Bone mineral density loss in relation to the final menstrual period in a multiethnic cohort: Results from the Study of Women's health across the nation (SWAN). *Journal of Bone and Mineral Research*, *27*, 111–118. <https://doi.org/10.1002/jbmr.534>
- Grigorova, M., Punab, M., Ausmees, K., & Laan, M. (2008). FSHB promoter polymorphism within evolutionary conserved element is associated with serum FSH level in men. *Human Reproduction*, *23*, 2160–2166. <https://doi.org/10.1093/humrep/den216>
- Grigorova, M., Punab, M., Poolamets, O., Kelgo, P., Ausmees, K., Korrovits, P., Vihljajev, V., & Laan, M. (2010). Increased prevalence of the -211 T allele of follicle stimulating hormone (FSH) β subunit promoter polymorphism and lower serum FSH in infertile men. *The Journal of Clinical Endocrinology & Metabolism*, *95*, 100–108. <https://doi.org/10.1210/jc.2009-1010>
- Ionita-Laza, I., Capanu, M., De Rubeis, S., McCallum, K., & Buxbaum, J. D. (2014). Identification of rare causal variants in sequence-based studies: Methods and applications to VPS13B, a gene involved in Cohen syndrome and autism. *PLoS Genetics*, *10*, e1004729. <https://doi.org/10.1371/journal.pgen.1004729>
- Lee, S., Emond, M. J., Bamshad, M. J., Barnes, K. C., Rieder, M. J., Nickerson, D. A., NHLBI GO Exome Sequencing Project—ESP Lung Project Team, Christiani, D. C., Wurfel, M. M., & Lin, X. (2012). Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *American Journal of Human Genetics*, *91*, 224–237. <https://doi.org/10.1016/j.ajhg.2012.06.007>
- Machiela, M. J., & Chanock, S. J. (2015). LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*, *31*, 3555–3557. <https://doi.org/10.1093/bioinformatics/btv402>
- Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W. M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics*, *26*, 2867–2873. <https://doi.org/10.1093/bioinformatics/btq559>
- Prescott, J., Thompson, D. J., Kraft, P., Chanock, S. J., Audley, T., Brown, J., Leyland, J., Folkard, E., Doody, D., Hankinson, S. E., Hunter, D. J., Jacobs, K. B., Dowsett, M., Cox, D. G., Easton, D. F., & de Vivo, I. (2012). Genome-wide association study of circulating estradiol, testosterone, and sex hormone-binding globulin in postmenopausal women. *PLoS One*, *7*, e37815. <https://doi.org/10.1371/journal.pone.0037815>
- Randolph, J. F., Jr., Zheng, H., Sowers, M. R., Crandall, C., Crawford, S., Gold, E. B., & Vuga, M. (2011). Change in follicle-stimulating hormone and estradiol across the menopausal transition: Effect of age at the final menstrual period. *The Journal of Clinical Endocrinology & Metabolism*, *96*, 746–754. <https://doi.org/10.1210/jc.2010-1746>
- Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J., Lavery, T., Kouyoumjian, R., Farhadian, S. F., Ward, R., & Lander, E. S. (2001). Linkage disequilibrium in the human genome. *Nature*, *411*, 199–204. <https://doi.org/10.1038/35075590>
- Rull, K., Grigorova, M., Ehrenberg, A., Vaas, P., Sekavin, A., Nõmmemees, D., Adler, M., Hanson, E., Juhanson, P., & Laan, M. (2018). FSHB–211 G>T is a major genetic modulator of reproductive physiology and health in childbearing age women. *Human Reproduction*, *33*, 954–966. <https://doi.org/10.1093/humrep/dey057>
- Ruth, K. S., Campbell, P. J., Chew, S., Lim, E. M., Hadlow, N., Stuckey, B. G., Brown, S. J., Feenstra, B., Joseph, J., Surdulescu, G. L., Zheng, H. F., Richards, J. B., Murray, A., Spector, T. D., Wilson, S. G., & Perry, J. R. (2016). Genome-wide association study with 1000 genomes imputation identifies signals for nine sex hormone-related phenotypes. *European Journal of Human Genetics*, *24*, 284–290. <https://doi.org/10.1038/ejhg.2015.102>
- Ruth, K. S., Day, F. R., Hussain, J., Martínez-Marchal, A., Aiken, C. E., Azad, A., Thompson, D. J., Knoblochova, L., Abe, H., Tarry-Adkins, J. L., Gonzalez, J. M., Fontanillas, P., Claringbould, A., Bakker, O. B., Sulem, P., Walters, R. G., Terao, C., Turon, S., Horikoshi, M., ... Perry, J. R. B. (2021). Genetic insights into biological mechanisms governing human ovarian ageing. *Nature*, *596*, 393–397. <https://doi.org/10.1038/s41586-021-03779-7>
- Schaid, D. J., Tong, X., Larrabee, B., Kennedy, R. B., Poland, G. A., & Sinnwell, J. P. (2016). Statistical methods for testing genetic pleiotropy. *Genetics*, *204*, 483–497. <https://doi.org/10.1534/genetics.116.189308>
- Schüring, A. N., Busch, A. S., Bogdanova, N., Gromoll, J., & Tüttelmann, F. (2013). Effects of the FSH- β -subunit promoter polymorphism -211G->T on the hypothalamic-pituitary-ovarian axis in normally cycling women indicate a gender-specific regulation of gonadotropin secretion. *Journal of Clinical Endocrinology and Metabolism*, *98*, E82–E86. <https://doi.org/10.1210/jc.2012-2780>
- Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M., & Smoller, J. W. (2013). Pleiotropy in complex traits: Challenges and strategies. *Nature Reviews. Genetics*, *14*, 483–495. <https://doi.org/10.1038/nrg3461>
- Sowers, M., Crawford, S., Sternfeld, B., Morganstein, D., Gold, E., Greendale, G., Evans, D. A., Neer, R., Matthews, K., Sherman, S., Lo, A., Weiss, G., & Kelsey, J. (2000). SWAN: A multi-center, multi-ethnic, community-based cohort study of women and the menopausal transition. In R. A. Lobo, J. Kelsey, & R. Marcus (Eds.), *Menopause: Biology and pathobiology* (pp. 175–188). Academic Press ISBN 0124537901, 9780124537903.
- Sowers, M. R., Zheng, H., McConnell, D., Nan, B., Harlow, S., & Randolph, J. F., Jr. (2008b). Follicle stimulating hormone and its rate of change in defining menopause transition stages. *The*

- Journal of Clinical Endocrinology & Metabolism*, 93, 3958–3964. <https://doi.org/10.1210/jc.2008-0482>
- Sowers, M. R., Zheng, H., McConnell, D., Nan, B., Harlow, S. D., & Randolph, J. F., Jr. (2008a). Estradiol rates of change in relation to the final menstrual period in a population-based cohort of women. *The Journal of Clinical Endocrinology & Metabolism*, 93, 3847–3852. <https://doi.org/10.1210/jc.2008-1056>
- Watanabe, K., Stringer, S., Frei, O., Umićević Mirkov, M., de Leeuw, C., Polderman, T. J. C., van der Sluis, S., Andreassen, O. A., Neale, B. M., & Posthuma, D. (2019). A global overview of pleiotropy and genetic architecture in complex traits. *Nature Genetics*, 51, 1339–1348. <https://doi.org/10.1038/s41588-019-0481-0>
- Wu, M. C., Kraft, P., Epstein, M. P., Taylor, D. M., Chanock, S. J., Hunter, D. J., & Lin, X. (2010). Powerful SNP-set analysis for case-control genome-wide association studies. *American Journal of Human Genetics*, 86, 929–942. <https://doi.org/10.1016/j.ajhg.2010.05.002>
- Wu, M. C., Lee, S., Cai, T., Li, Y., Boehnke, M., & Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *American Journal of Human Genetics*, 89, 82–93. <https://doi.org/10.1016/j.ajhg.2011.05.029>
- Zhao, W., Smith, J. A., Bielak, L. F., Ruiz-Narvaez, E. A., Yu, M., Hood, M. H., Peyser, P. A., Kardia, S. L. R., & Harlow, S. D. (2021). Associations between polygenic risk scores for age at menarche and menopause, reproductive timing, and serum hormone levels in multiple race/ethnic groups. *Menopause*, 28, 819–828. <https://doi.org/10.1097/GME.0000000000001775>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Bielak, L. F., Peyser, P. A., Smith, J. A., Zhao, W., Ruiz-Narvaez, E. A., Kardia, S. L. & Harlow, S. D. (2022). Multivariate, region-based genetic analyses of facets of reproductive aging in White and Black women. *Molecular Genetics & Genomic Medicine*, 10, e1896. <https://doi.org/10.1002/mgg3.1896>