Human Reproduction, Vol.28, No.6 pp. 1695–1706, 2013
Advanced Access publication on March 18, 2013 doi:10.1093/humrep/det071

human reproduction

ORIGINAL ARTICLE ***Reproductive epidemiology***

Replication of genetic loci for ages at menarche and menopause in the multi-ethnic Population Architecture using Genomics and Epidemiology (PAGE) study

C.L. Carty[1,*], K.L. Spencer[2], V.W. Setiawan[3], L. Fernandez-Rhodes[4], J. Malinowski[5], S. Buyske[6,7], A. Young[1], N.W. Jorgensen[8], I. Cheng[9], C.S. Carlson[1], K. Brown-Gentry[5], R. Goodloe[5], A. Park[10], N.I. Parikh[11], B. Henderson[3], L. Le Marchand[12], J. Wactawski-Wende[13], M. Fornage[14], T.C. Matise[6], L.A. Hindorff[15], A.M. Arnold[8], C.A. Haiman[3], N. Franceschini[4], U. Peters[1], and D.C. Crawford[5]

[1]Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA [2]Heidelberg University, Tiffin, OH, USA [3]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA [4]Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA [5]Center for Human Genetics Research, Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN, USA [6]Department of Genetics, Rutgers University, Piscataway, NJ, USA [7]Department of Statistics, Rutgers University, Piscataway, NJ, USA [8]Department of Biostatistics, University of Washington, Seattle, WA, USA [9]Cancer Prevention Institute of California, Fremont, CA, USA [10]Department of Obstetrics & Gynecology, Washington Hospital Center, Georgetown University School of Medicine, Washington, DC, USA [11]John A. Burns School of Medicine and Cardiovascular Division, University of Hawaii, The Queens Medical Center, Honolulu, HI, USA [12]University of Hawaii Cancer Center, Honolulu, HI, USA [13]Department of Social and Preventive Medicine, University at Buffalo, The State University of New York, Buffalo, NY, USA [14]Health Science Center, University of Texas, Houston, TX, USA [15]Office of Population Genomics, NHGRI, NIH, Bethesda, MD, USA

*Correspondence address. E-mail: ccarty@fhcrc.org

Submitted on October 3, 2012; resubmitted on December 7, 2012; accepted on January 29, 2013

STUDY QUESTION: Do genetic associations identified in genome-wide association studies (GWAS) of age at menarche (AM) and age at natural menopause (ANM) replicate in women of diverse race/ancestry from the Population Architecture using Genomics and Epidemiology (PAGE) Study?

SUMMARY ANSWER: We replicated GWAS reproductive trait single nucleotide polymorphisms (SNPs) in our European descent population and found that many SNPs were also associated with AM and ANM in populations of diverse ancestry.

WHAT IS KNOWN ALREADY: Menarche and menopause mark the reproductive lifespan in women and are important risk factors for chronic diseases including obesity, cardiovascular disease and cancer. Both events are believed to be influenced by environmental and genetic factors, and vary in populations differing by genetic ancestry and geography. Most genetic variants associated with these traits have been identified in GWAS of European-descent populations.

STUDY DESIGN, SIZE, DURATION: A total of 42 251 women of diverse ancestry from PAGE were included in cross-sectional analyses of AM and ANM.

MATERIALS, SETTING, METHODS: SNPs previously associated with ANM ($n = 5$ SNPs) and AM ($n = 3$ SNPs) in GWAS were genotyped in American Indians, African Americans, Asians, European Americans, Hispanics and Native Hawaiians. To test SNP associations with ANM or AM, we used linear regression models stratified by race/ethnicity and PAGE sub-study. Results were then combined in race-specific fixed effect meta-analyses for each outcome. For replication and generalization analyses, significance was defined at $P < 0.01$ for ANM analyses and $P < 0.017$ for AM analyses.

MAIN RESULTS AND THE ROLE OF CHANCE: We replicated findings for AM SNPs in the *LIN28B* locus and an intergenic region on 9q31 in European Americans. The *LIN28B* SNPs (rs314277 and rs314280) were also significantly associated with AM in Asians, but not in other race/ethnicity groups. Linkage disequilibrium (LD) patterns at this locus varied widely among the ancestral groups. With the exception of an intergenic SNP at 13q34, all ANM SNPs replicated in European Americans. Three were significantly associated with ANM in other race/ethnicity populations: rs2153157 (6p24.2/*SYCP2L*), rs365132 (5q35/*UIMC1*) and rs16991615 (20p12.3/*MCM8*). While rs1172822 (19q13/*BRSK1*) was not significant in the populations of non-European descent, effect sizes showed similar trends.

LIMITATIONS, REASONS FOR CAUTION: Lack of association for the GWAS SNPs in the non-European American groups may be due to differences in locus LD patterns between these groups and the European-descent populations included in the GWAS discovery studies; and in some cases, lower power may also contribute to non-significant findings.

WIDER IMPLICATIONS OF THE FINDINGS: The discovery of genetic variants associated with the reproductive traits provides an important opportunity to elucidate the biological mechanisms involved with normal variation and disorders of menarche and menopause. In this study we replicated most, but not all reported SNPs in European descent populations and examined the epidemiologic architecture of these early reported variants, describing their generalizability and effect size across differing ancestral populations. Such data will be increasingly important for prioritizing GWAS SNPs for follow-up in fine-mapping and resequencing studies, as well as in translational research.

STUDY FUNDING/COMPETING INTEREST(S): The Population Architecture Using Genomics and Epidemiology (PAGE) program is funded by the National Human Genome Research Institute (NHGRI), supported by U01HG004803 (CALiCo), U01HG004798 (EAGLE), U01HG004802 (MEC), U01HG004790 (WHI) and U01HG004801 (Coordinating Center), and their respective NHGRI ARRA supplements. The authors report no conflicts of interest.

Key words: menopause / menarche / genome-wide association study / race/ethnicity / single nucleotide polymorphism

Introduction

Menarche and menopause define a woman's reproductive lifespan and are important risk factors for chronic disease. Menarche marks the onset of the first menstrual period and its early occurrence has been associated with behavioral and psychosocial disorders during adolescence (Hayward *et al.*, 1997) and adverse outcomes during adulthood including type 2 diabetes, cardiovascular disease and breast, ovarian and endometrial cancers (Hsieh *et al.*, 1990; Lakshman *et al.*, 2008; Lakshman *et al.*, 2009; He *et al.*, 2010; Stockl *et al.*, 2011; Cramer, 2012). These associations may in part be mediated by obesity in children, which is a predictor of both early menarche and adulthood obesity (Freedman *et al.*, 2003). Ongoing decreases in age at menarche (AM) worldwide have motivated the search for factors triggering puberty, many of which are believed to be environmental (Parent *et al.*, 2003). However, genetic factors are also important (Gajdos *et al.*, 2010). AM heritability (h^2), which is the proportion of variation in AM explained by genetics, is relatively high with h^2 estimates ranging from 0.24 to 0.73 (Kaprio *et al.*, 1995; Anderson *et al.*, 2007; Gajdos *et al.*, 2010).

Age at natural menopause (ANM) is defined as the age at cessation of natural ovarian function (not occurring as a result of surgery, i.e. bilateral oophorectomy, radiation treatment or chemotherapeutic agents). Early menopause, typically described as menopause occurrence before the age of 45 years, is associated with increased risk of osteoporosis, coronary heart disease, stroke and all-cause mortality (Kritz-Silverstein and Barrett-Connor, 1993; Cooper and Sandler, 1998; Lisabeth *et al.*, 2009; Wellons *et al.*, 2012) whereas late menopause, defined as menopause occurring at the age of 54 years or older, is associated with an increased risk of breast and endometrial cancers (Hsieh *et al.*, 1990; Cramer, 2012). ANM heritability varies in the literature, ranging from 0.31 to 0.87 (Murabito *et al.*, 2005). Although environmental and lifestyle factors, such as smoking, have been reported to influence timing of menopause (Gold *et al.*, 2001), in general they explain only a small proportion of the variation in ANM (van Noord *et al.*, 1997; Voorhuis *et al.*, 2010). Family history is one of the strongest predictors of ANM, (Morris *et al.*, 2011) and thus motivates the search for genetic risk factors.

To date, genome-wide association studies (GWASs) have identified genetic variants associated with reproductive lifespan traits, but have been conducted almost exclusively in European-descent populations (He *et al.*, 2009; Perry *et al.*, 2009; Stolk *et al.*, 2009; Sulem *et al.*, 2009) or in limited samples of non-European women (Liu *et al.*, 2009; Chen *et al.*, 2012) despite the fact that AM (Freedman *et al.*, 2002; Wu *et al.*, 2002; Anderson *et al.*, 2003; Gajdos *et al.*, 2010) and ANM (Bromberger *et al.*, 1997; Gold *et al.*, 2001) vary widely by race/ethnicity. For example, natural menopause occurs at a median age of 51.4 years in women of European ancestry while, on average, it occurs earlier in African American and Hispanic women, and later in women of Japanese ancestry (Bromberger *et al.*, 1997; te Velde and Pearson, 2002; Henderson *et al.*, 2008; He and Murabito, 2012). Reasons for racial differences in reproductive traits remain uncertain, though they may be a result of genetic variation, nutritional status (Parent *et al.*, 2003) or secular trends in anthropometric traits, such as height and weight, which also vary by race (Freedman *et al.*, 2002). Additional research is needed to understand the genetic variation underlying reproductive traits in diverse populations. Therefore, we sought to replicate GWAS findings for AM and ANM in our large sample of European Americans and to investigate whether these associations are also consistent in women of diverse race/ethnicity, including American Indians, African Americans, Asians, Hispanics and Native Hawaiians from the Population Architecture using Genomics (PAGE) study (Matise *et al.*, 2011).

Methods

Study participants

The PAGE study consists of a coordinating center and four study sites: (i) Causal Variants Across the Life Course (CALiCo), a consortium of epidemiologic cohorts including the Atherosclerosis Risk in Communities Study (ARIC), Coronary Artery Risk Development in Young Adults (CARDIA), Cardiovascular Health Study (CHS), the Hispanic Community Health Study/Study of Latinos (SOL) and Strong Heart Study (SHS); (ii) Epidemiologic Architecture for Genes Linked to Environment (EAGLE) accessing three National Health and Nutrition Examination Surveys (NHANES); (iii) the Multiethnic Cohort (MEC) study and (iv) the Women's Health Initiative (WHI).

The current analysis includes European American (EA), African American (AA), Hispanic, American Indian (Am Ind) and Native Hawaiian women from ARIC, CARDIA, CHS, SHS, EAGLE (NHANES III only), MEC and WHI with available genotype and AM or ANM data and who consented to use of genetic data. The PAGE cohorts have been described previously (Matise *et al.*, 2011), but briefly: ARIC is a prospective study of atherosclerotic cardiovascular disease in 15 792 men and women aged 45–64 at the baseline examination in 1987–1989 (The ARIC Investigators, 1989). ARIC recruited predominantly EA and AA individuals from four communities: Forsyth County, NC; Jackson, MI; suburban areas of Minneapolis, MN and Washington County, MD. CARDIA is a multicenter longitudinal study of the development and determinants of cardiovascular disease over time in 5115 young adults initially aged 18–30 years at the baseline examination in 1985–1986 (Hughes *et al.*, 1987; Cutter *et al.*, 1991). AA and EA adults were recruited from four US cities (Birmingham, AL; Chicago, IL; Minneapolis, MN; and Oakland, CA). The CHS is a population-based longitudinal study of risk factors for cardiovascular disease and stroke in adults 65 years of age or older, recruited at four field centers (Forsyth County, NC; Sacramento County, CA; Washington County, MD; Pittsburgh, PA) (Fried *et al.*, 1991). Overall, 5201 predominantly EA individuals were recruited in 1989–1990 from random samples of Medicare eligibility lists, followed by an additional 687 AA recruited in 1992–1993 (total $n = 5888$). The EAGLE study accessed DNA samples ($n = 7159$) from NHANES III, phase 2, collected between 1991 and 1994 (Chang *et al.*, 2009). NHANES is a US population-based, cross-sectional survey of Americans (infants to the elderly) ascertained regardless of health status, conducted by the National Center for Health Statistics at the Centers for Disease Control and Prevention. DNA samples collected for NHANES participants aged 12 years or older are linked to surveys of demographic, health and lifestyle factors as well as to data from a physical examination and laboratory measures. MEC is a population-based prospective cohort study of over 215 000 men and women in Hawaii and California aged 45–75 years at baseline (1993–1996) and primarily of five ancestries (Kolonel *et al.*, 2000). Participants eligible for the current analysis were females who provided blood samples for nested case–control studies of breast, colorectal, ovarian, endometrial or other cancers for genetic association and biomarker studies in PAGE ($n = 7216$). The SHS is a population-based study of risk factors for cardiovascular disease among Am Ind (Lee *et al.*, 1990; North *et al.*, 2003). Participating communities include seven tribes in southwestern Oklahoma (Apache, Caddo, Comanche, Delaware, Fort Sill Apache, Kiowa and Wichita), three tribes in Arizona (Gila River and Salt River Pima/Maricopa and Akchin Pima/Papago) and three tribes from South/North Dakota (Oglala Sioux, Cheyenne River Sioux and Spirit Lake Communities). The WHI is a large national health study of postmenopausal women which includes a series of clinical trials designed to assess chronic disease prevention and an observational study designed to study risk factors for chronic diseases. A total of 161 808 women aged 50–79 years old were recruited from 40 US clinical centers between 1993 and 1998 (The Women's Health Initiative Study Group, 1998). A subset of women ($n \sim 24000$) from the WHI clinical trials and observational study were selected for genotyping in PAGE and are included in these analyses.

All studies collected self-identified race/ethnicity information by questionnaire. Additional details about the Asian and Hispanic race/ethnicity groups follow. In MEC, 100% of Asian participants were of Japanese descent. Of the WHI participants self-identifying as 'Asian or Pacific Islander', 93% provided more specific ancestry information; the majority reported Japanese ancestry, though mixed and other Asian ancestry (including Chinese, Filipino, Korean and Vietnamese) were also reported. While EAGLE included only self-identified Mexican Americans, Hispanics in WHI and MEC were predominantly Mexican American, but may also include other Latino groups. Institutional review boards at each participating study site approved all procedures, and written informed consent was obtained from all participants.

Outcomes

Similar to previous GWAS, our goal was to measure genetic associations within the normal population distributions of AM and ANM. AM was defined as the age when menstrual periods started, in years. AM was treated as a continuous variable in analyses, however, in MEC, values were recorded in categories defined by 2-year intervals, for example 15–16 years. In this case, we assigned AM as the mean value of these intervals, for example 15.5 years. Data collection in WHI was continuous except for the lower and upper categories for which women were assigned a value of 0.5 years lower or higher than the category value, respectively. For example, for the maximum category of 17 years or older, women were assigned a value of 17.5 years. In studies with continuous data for AM, women reporting extreme values (AM <8 years or >19 years) were excluded from the analyses.

ANM was defined as the age at cessation of regular menstrual periods. Women with hysterectomy before onset of natural menopause or who reported menopause onset as a result of surgery (bilateral oophorectomy), radiation treatment or chemotherapy and women reporting ANM <40 years or >60 years were excluded from these analyses. In addition, in WHI and CHS, women who reported use of hormone replacement before the onset of natural menopause were also excluded. In MEC, ANM was collected as a categorical variable. As described for AM, the mean value of the category interval was assigned for ANM in MEC.

Genotyping

Using a Catalog of Published Genome-Wide Association Studies (Hindorff *et al.*, 2009) (www.genome.gov/gwastudies), we selected SNPs associated with AM and ANM in prior GWAS (as of February 2010) for genotyping in the PAGE study. Thus, SNPs reported in recent GWAS of these traits (Elks *et al.*, 2010; Stolk *et al.*, 2012) are not included. Some reported SNPs were in high linkage disequilibrium (LD) in the studied populations, in which case, we chose only one SNP. In total, we investigated three menarche SNPs (rs314277, rs314280 and rs7861820) reflecting two loci and five menopause SNPs (rs1172822, rs16991615, rs2153157, rs365132 and rs7333181) reflecting five loci. Study-specific DNA extraction, genotyping methods and quality control are described in detail elsewhere (Matise *et al.*, 2011). As reported previously (Matise *et al.*, 2011), principal components of continental ancestry (African, European and Asian) were derived from a panel of ancestry informative genetic markers (Kosoy *et al.*, 2009) or from existing genome-wide SNP data in each PAGE study separately. Genetic ancestry information was not available for the EAGLE samples.

Statistical analysis

SNPs were coded assuming an additive mode of inheritance. Race/ethnicity-specific linear regression models were used to test for SNP associations with the continuous outcomes of AM or ANM. Both AM and ANM models were adjusted for principal components of ancestry to control for potential confounding by population stratification, and also for other cohort-specific covariates, such as enrollment site, as appropriate. In addition, ANM models were adjusted for birth year. In sensitivity analyses, we investigated adjustment for birth year and BMI in AM models to consider potential SNP associations (controlling for BMI) in which adult BMI serves as a weak proxy of childhood BMI. Tests of association were performed separately in each of the participating cohorts and then combined in race-specific fixed effect meta-analyses for each outcome. Of note, data for Native Hawaiians and for Am Ind (AM outcome) were only available in one study, MEC or WHI respectively, and thus did not undergo meta-analysis. All meta-analyses were performed using inverse-variance weighted models in METAL (Willer *et al.*, 2010), and between-study effect size consistency was assessed using the *Q*-test (χ^2 *P*-value) and the I^2 metric (Higgins and Thompson, 2002), where low I^2 suggests that there is little between-study variability that cannot be explained by chance. Higgins *et al.* (2003) suggest the classification of low and high I^2 at 25 and 75%, respectively. For SNPs with $I^2 \geq 60\%$ or χ^2 *P*-value <0.05, we present study-specific results. For replication and generalization analyses, significance was defined at $P < 0.017$ and $P < 0.010$ to account for multiple tests ($n = 3$ and 5) for each of the AM and ANM outcomes, respectively.

MEC and WHI samples included large numbers of cancer cases and controls. Because female-related cancers, such as breast cancer, may be related to our outcomes, in sensitivity analyses we tested for interaction by cancer case/control status in the MEC and WHI samples. Additionally in WHI, we compared the cancer case/control sample results with the large portion of the sample not selected with respect to cancer status. No significant differences in SNP associations by cancer case/control status (data not shown) were found in either study and thus cancer samples were retained in the analyses.

Using Quanto (Gauderman and Morrison, 2006), we estimated the statistical power for each SNP in each racial/ethnic population (Supplementary data, Table SI) to detect the previously reported effect size based on the European descent discovery population.

Results

Age at menarche

The mean AM by race/ethnicity ranged from 12.5 to 13.2 years (Table I). We replicated the AM association for the intergenic 9q31

Table I Baseline characteristics of women in the age at menarche analyses.

PAGE cohort	ARIC	CARDIA	EAGLE	MEC	WHI	Total
EA (*n*)	4368	898	1546	2121	16 091	25 024
AM (in years)	12.9 ± 1.6	12.7 ± 1.4	12.8 ± 1.7	13.0 ± 1.6	12.6 ± 1.5	
Enrollment (age in years)	53.9 ± 5.7	25.5 ± 3.4	50.2 ± 22.1	59.7 ± 8.4	64.3 ± 7.0	
BMI (in kg/m^2)	26.5 ± 5.4	23.1 ± 4.4	26.3 ± 6.1	25.7 ± 5.4	28.0 ± 5.8	
AA (*n*)	2333	556	1179	2123	2350	8541
AM (in years)	12.9 ± 1.7	12.5 ± 1.6	12.6 ± 1.9	13.2 ± 1.7	12.6 ± 1.7	
Enrollment (age in years)	53.3 ± 5.7	24.5 ± 3.9	35.6 ± 17.7	60.5 ± 8.8	61.4 ± 7.0	
BMI (in kg/m^2)	30.8 ± 6.5	25.9 ± 6.5	28.3 ± 7.4	29.0 ± 5.9	31.6 ± 6.8	
Asians (*n*)	n/a	n/a	n/a	2650	1297	3947
AM (in years)				13.1 ± 1.7	12.7 ± 1.6	
Enrollment (age in years)				59.7 ± 8.2	62.1 ± 7.3	
BMI (in kg/m^2)				24.0 ± 4.0	25.5 ± 5.0	
Hispanics (*n*)	n/a	n/a	1026	1693	1053	3772
AM (in years)			12.6 ± 1.7	13.1 ± 1.7	12.7 ± 1.6	
Enrollment (age in years)			36.5 ± 18.3	59.3 ± 7.4	60.4 ± 6.8	
BMI (in kg/m^2)			27.5 ± 6.2	28.4 ± 5.7	29.2 ± 6.2	
Am Indians (*n*)	n/a	n/a	n/a	n/a	233	233
AM (in years)					12.8 ± 1.8	
Enrollment (age in years)					62.0 ± 7.2	
BMI (in kg/m^2)					30.3 ± 6.1	
N Hawaiians (*n*)	n/a	n/a	n/a	734	n/a	734
AM (in years)				12.8 ± 1.7		
Enrollment age (in years)				56.2 ± 7.5		
BMI (in kg/m^2)				29.3 ± 6.9		

PAGE, population architecture using genomics and epidemiology study, AM, age at menarche; EA, European Americans; AA, African Americans; Am Indians, American Indians; N Hawaiians, Native Hawaiians; ARIC: atherosclerosis risk in communities study; CARDIA, coronary artery risk development in young adults; EAGLE, epidemiologic architecture for genes linked to environment; MEC, multiethnic cohort study; WHI, women's health initiative.
Characteristics are presented as the mean ± SD unless otherwise indicated.

SNP, rs7861820, in our European descent sample (Table II; $P = 1.5E-6$). Betas were similar (same direction) in Asians, Hispanics and Native Hawaiians, but were not significant. We also tested two SNPs located in the 6q21 region, rs314277 and rs314280, for their associations with AM. Rs314280, located upstream of *LIN28B,* was significantly associated with AM in EA ($P = 2E-08$) and Asians ($P = 0.001$). In AA, the effect size was comparable, but not significant ($P = 0.16$). Effect sizes for this SNP were larger, but also not significant in Native Hawaiians ($P = 0.02$) and Am Ind ($P = 0.15$); in Hispanics, the association was also null ($P = 0.98$). Findings for rs314277, located in intron 2 of the *LIN28B* gene, also varied by race/ethnicity. The SNP was significantly associated with AM in EA ($P = 0.001$) and Native Hawaiians ($P = 0.016$). In EA, Hispanics and Asians, effect sizes were of similar magnitude to those reported by He *et al.* (2009) in that each additional A allele was associated with an ~1–2 month(s) increase in AM. These *LIN28B* SNPs are not strongly correlated in the PAGE populations, with r^2 ranging from 0.09 to 0.26.

In conditional analyses including both *LIN28B* SNPs, we found that adjustment of intronic rs314277 for rs314280 substantially attenuates the association in EA and Asians, both reducing the effect estimate and also rendering rs314277 insignificant in the model (Table III). However, adjustment of rs314280 with rs314277 only slightly attenuates the association of rs314280 with AM in these groups. Similar, but weaker findings are seen in the smaller sample of Native Hawaiians. In Hispanics, associations are stronger for rs314277 than rs214280 in both models. Allele frequencies for both SNPs varied widely by ancestral group in PAGE. For example, the T allele frequency for rs314280 was 0.75 in AA and 0.30 in Asians. We also observed population variation in the LD (r^2) between these SNPs and in the region (not shown).

In sensitivity analyses, adjustment of AM models for year of birth and BMI at study enrollment had only small effects on results (Supplementary data, Table SII).

Age at natural menopause

Mean ANM ranged from 48.0 to 51.1 years in the PAGE populations (Table IV). All five ANM loci, except for rs7333181 (13q34), replicated in EA and of those replicating, three SNPs were significantly associated with ANM in the other race/ethnicity groups: rs2153157 (6p24.2/*SYCP2L*), rs365132 (5q35/*UIMC1*) and rs16991615 (20p12.3/*MCM8*) (Table V). We observed significant heterogeneity ($P < 0.05$) for the synonymous *UIMC1* SNP rs365132 in EA, which was associated with ~4.8 months later ANM (on average) in MEC and WHI, but was not significant in CHS or ARIC. Between-study heterogeneity ($I^2 = 74$) was also seen for the smaller Am Ind sample, though the study-specific results were not significant. While rs1172822 (19q13/*BRSK1*) was not significant in the non-EA groups, effect sizes in all groups (except AA) were comparable to the EA estimates. In AA, we observed between-study heterogeneity, though results for all studies were insignificant. The coded allele frequency (CAF) for this SNP was notably different in the Asian population (average CAF = 0.05) compared with 0.25–0.36 in other populations. The most robust associations were found for rs16991615 (20p12.3/*MCM8*) in which we replicated the association in EA ($P = 8E-15$), as well as observed significant associations in Hispanics ($P = 0.0004$) and Am Ind ($P = 0.001$). Notably, in the Am Ind population, each additional copy of the A allele was associated with a 2.6 ± 0.8 year increase in menopause age.

Table II Age at menarche SNP association results in women of diverse ancestry.

SNP, locus, coded/other allele	GWAS reference study[a]		PAGE study results for AM					
	CAF	Beta ± SE (in years)	Race/ ethnicity	*n*	Ave CAF	[b]Beta ± SE (years)	*P*-value	I^2
rs314277 *LIN28B* (intron2) 6q21 A/C	0.14	0.16 ± 0.02	EA	24 932	0.15	0.07 ± 0.02	**0.0001**	20
			AA	7994	0.38	0.002 ± 0.03	0.9589	52
			Asians	3929	0.05	0.18 ± 0.08	0.0304	0
			Hispanics	3743	0.11	0.09 ± 0.06	0.1482	0
			Am Indians	233	0.17	−0.07 ± 0.24	0.7757	n/a
			N Hawaiians	729	0.06	0.47 ± 0.20	0.0160	n/a
rs314280 *LIN28B* (upstream) 6q21 T/C	0.45	0.09 ± 0.02	EA	24 819	0.45	0.08 ± 0.01	**2.0E–08**	0
			AA	7999	0.75	0.05 ± 0.03	0.1551	30
			Asians	3934	0.30	0.13 ± 0.04	**0.0012**	60
			Hispanics	3751	0.37	0.001 ± 0.04	0.9843	0
			Am Indians	231	0.42	0.25 ± 0.17	0.1446	n/a
			N Hawaiians	728	0.39	0.22 ± 0.10	0.0218	n/a
rs7861820 intergenic 9q31.2 C/T	0.48	−0.09 ± 0.02	EA	23 371	0.49	−0.07 ± 0.01	**1.5E–06**	0
			AA	6830	0.89	0.02 ± 0.05	0.6461	0
			Asians	3920	0.78	−0.04 ± 0.05	0.4370	42
			Hispanics	2734	0.49	−0.09 ± 0.05	0.0643	0
			Am Indians	232	0.50	0.12 ± 0.17	0.4854	n/a
			N Hawaiians	726	0.55	−0.09 ± 0.09	0.3123	n/a

Ref, reference; CAF, coded allele frequency; Ave, average.
P-values significant after multiple testing correction for three tests are given in bold.
[a]Reference genome-wide association study (GWAS) of AM described for all SNPs is He *et al.* (2009).
[b]Beta reflects change in AM in years for each additional copy of the coded allele.

Table III ***LIN28B*** **SNPs conditional model results in the PAGE Study.**[a]

Race/ethnicity group	*n*	LD between SNPs (r^2)[b]	SNP	Single SNP model meta-analysis results		Conditional model[c] meta-analysis results	
				Beta ± SE	*P*-value	Beta ± SE	*P*-value
EA	24 720	0.21	rs314277	0.073 ± 0.019	1.20e−4	0.030 ± 0.021	0.1587
			rs314280	0.077 ± 0.014	1.46e−8	0.067 ± 0.015	1.10e−5
AA	7974	0.21	rs314277	−0.001 ± 0.030	0.9871	−0.021 ± 0.031	0.4968
			rs314280	0.044 ± 0.032	0.1610	0.055 ± 0.035	0.1177
Asians	3942	0.13	rs314277	0.185 ± 0.083	0.0256	0.095 ± 0.089	0.2851
			rs314280	0.133 ± 0.041	0.0012	0.114 ± 0.044	0.0106
Hispanics	3722	0.23	rs314277	0.110 ± 0.062	0.0781	0.137 ± 0.070	0.0501
			rs314280	0.003 ± 0.040	0.9416	−0.038 ± 0.045	0.4067
Am Indians	231	0.26	rs314277	−0.049 ± 0.240	0.8372	−0.363 ± 0.283	0.1993
			rs314280	0.253 ± 0.174	0.1446	0.328 ± 0.202	0.1034
N Hawaiians	735	0.09	rs314277	0.470 ± 0.197	0.0168	0.379 ± 0.205	0.0650
			rs314280	0.206 ± 0.095	0.0301	0.154 ± 0.099	0.1214

LD, linkage disequilibrium.
[a]Only women with data for both SNPs are included in conditional analyses and so numbers may differ from Table I.
[b]LD estimates based on the study with the largest sample size within a given race/ethnicity group.
[c]Conditional model includes both rs314277 and rs314280 using coded alleles as described previously.

Discussion

Insights into the genetic basis of the reproductive lifespan may provide a better understanding of mechanisms related to puberty and menopause and their variation in diverse populations. In this study, we evaluated associations of known genetic variants with normal variation in AM and ANM in diverse race/ethnicity populations. We successfully replicated several GWAS findings for menarche (*LIN28B* and 9q31.2) and menopause (*BRSK1*, *MCM8*, *SYCP2L* and *UIMC1*) in European descent participants and found that several, but not all, loci also generalized to diverse race/ethnicity groups. In general, SNPs that replicated in our EA population had effect sizes similar to the original reports; SNPs generalizing to other race/ethnicity groups had comparable effect sizes although were less precise, due to smaller sample sizes.

Recent GWAS have identified the 6q21/*LIN28B* locus association with AM in European descent populations (He *et al.*, 2009; Ong *et al.*, 2009; Perry *et al.*, 2009; Sulem *et al.*, 2009; Elks *et al.*, 2010). We tested two *LIN28B* SNPs: rs314277 located in intron 2 and rs314280 in the regulatory region rich with transcription factor-binding sites (Ernst *et al.*, 2011) upstream of *LIN28B*. The lin-28 homolog B (*LIN28B*) gene is an important member of the Lin28/let-7 axis, which plays a key role in growth, proliferation and metabolism at the cellular level, as well as in development, insulin sensitivity and oncogenesis in whole organisms (Thornton and Gregory, 2012). In humans, *LIN28B* has also been associated with body height (Gudbjartsson *et al.*, 2008; Lettre *et al.*, 2008), which is consistent with the observation that rapid height and weight increases precede puberty (Parent *et al.*, 2003). Interestingly, Ong *et al.* (2009) identified a *LIN28B* SNP rs314276 (not tested in PAGE) associated with earlier menarche, faster pre-pubertal growth spurt in girls and shorter adult height in women; however, in their analysis of European women, rs314277 was not significantly associated with AM. The *LIN28B* SNPs replicated (at $P < 0.017$) in our EA, Asian and Native Hawaiian populations. Our effect sizes for these SNPs (betas = 0.07–0.13) were comparable to previous reports (He *et al.*, 2009; Sulem *et al.*, 2009). However, findings for these SNPs in AA, Hispanics and Am Ind were null, in spite of reasonable power to detect the reported association in AA (>90%) and Hispanics (>71%). Rs314280 is in moderately high LD ($r^2 = 0.63–0.96$) in HapMap CEU and Asian (CHB + JPT) populations with other intronic and 5' region *LIN28B* SNPs, including rs314276 which was also associated with AM (Ong *et al.*, 2009) as described above, rs167359 and rs314274, but is in lower LD with these SNPs in HapMap3 Mexican Americans ($r^2 = 0.44–0.48$) and AA ($r^2 = 0.18–0.30$). Our conditional analyses of the *LIN28B* SNPs suggest that the AM association for intronic SNP rs314277 does not appear to be independent from rs314280. Also, adjustment of rs314280 for rs314277 attenuates its association with AM, though it is still significant, suggesting that the regulatory region SNP rs314280 is more likely to be in high LD with the causal variant(s). However, this observation does not appear to be consistent for PAGE Hispanics. Overall, neither of the *LIN28B* SNPs appear to have strong and consistent effects across all populations; fine mapping of this region may help to identify common causal variant(s).

The lack of association for the *LIN28B* SNPs that we observed in AA (in spite of high power) is consistent with data suggesting that LD extends over significantly shorter distances in African populations than in non-African populations (Reich *et al.*, 2001). Additionally many GWAS SNPs, such as the ones we studied, tend to be proxies of untyped functional variants, due to typical chip design (Bhangale *et al.*, 2008). Thus, the lack of generalization to AA may be due to differential LD with the actual causal SNP or, potentially, differential genetic background effects in AA as described in Carlson *et al.* (in revision 2012).

In European descent women, we replicated all ANM loci reported in two recent GWAS (He *et al.*, 2009; Stolk *et al.*, 2009) with the exception of one intergenic SNP at 13q34. This SNP was also not replicated in a more recent GWAS published in 2012 (Stolk *et al.*, 2012),

Table IV Baseline characteristics of women in the age at natural menopause analyses.

PAGE Cohort	ARIC	CHS	MEC	SHS	WHI	Total
EA (*n*)	1628	959	1037	n/a	7563	11 187
ANM (in years)	48.4 ± 3.7	49.3 ± 4.3	50.7 ± 3.1		50.5 ± 4.0	
Enrollment age (in years)	57.7 ± 4.3	72.4 ± 5.5	61.4 ± 7.4		65.0 ± 6.8	
BMI (in kg/m^2)	26.5 ± 5.2	26.0 ± 4.8	25.3 ± 5.2		28.0 ± 5.9	
Current smoker (*n* (%))	385 (23.7)	110 (11.5)	126 (12.2)		659 (8.7)	
AA (*n*)	626	196	735	n/a	876	2433
ANM (in years)	48.2 ± 3.9	48.9 ± 4.7	50.6 ± 3.2		50.2 ± 4.4	
Enrollment age (in years)	57.3 ± 4.5	73.5 ± 5.9	62.8 ± 7.3		61.9 ± 6.8	
BMI (in kg/m^2)	31.0 ± 6.7	29.2 ± 6.1	28.8 ± 5.9		31.2 ± 6.8	
Current smoker [*n* (%)]	146 (23.3)	22 (11.2)	121 (16.5)		115 (13.1)	
Asians (*n*)	n/a	n/a	1317	n/a	707	2024
ANM (in years)			51.1 ± 3.0		50.3 ± 4.0	
Enrollment age (in years)			62.1 ± 6.7		62.9 ± 7.2	
BMI (in kg/m^2)			23.7 ± 3.8		25.2 ± 5.0	
Current smoker [(*n* (%)]			88 (6.7)		38 (5.4)	
Hispanics (*n*)	n/a	n/a	737	n/a	420	1157
ANM (in years)			50.5 ± 3.0		49.7 ± 3.9	
Enrollment age (in years)			60.3 ± 6.5		60.5 ± 6.9	
BMI (in kg/m^2)			28.2 ± 5.5		29.2 ± 6.2	
Current smoker [*n* (%)]			79 (10.7)		33 (7.9)	
Am Indians (*n*)	n/a	n/a	n/a	861	76	937
ANM (in years)				48.0 ± 4.2	49.8 ± 4.2	
Enrollment age (in years)				57.6 ± 7.6	61.6 ± 7.2	
BMI (in kg/m^2)				31.6 ± 6.9	29.6 ± 6.5	
Current smoker [*n* (%)]				225 (27.0)	7 (9.2)	
N Hawaiians (*n*)	n/a	n/a	300	n/a	n/a	300
ANM (in years)			50.8 ± 3.1			
Enrollment age (in years)			58.7 ± 6.7			
BMI (in kg/m^2)			28.2 ± 6.4			
Current smoker [(*n* (%)]			55 (18.3)			

ANM, age at natural menopause; CHS, cardiovascular health study; SHS, strong heart study. Characteristics are presented as the mean ± SD unless otherwise indicated.

which included ~2500 PAGE samples of European descent from ARIC and CHS. In a sensitivity analysis, exclusion of these 2500 samples from our total 11, 150 ANM samples had little effect on ANM results, with the exception of the rs365132, which became more significant ($P = 8.0E-12$) and no longer showed evidence of heterogeneity in the meta-analysis (Supplementary data, Table SIII). In spite of low allele frequencies, we observed robust associations in diverse ancestral groups for a non-synonymous SNP, rs16991615, in the mini-chromosome maintenance 8 (*MCM8*) gene. The protein encoded by *MCM8* is believed to be important in the cell division cycle; however, rs16991615 is not predicted by the sorting tolerant from intolerant algorithm (SIFT) (Kumar *et al.*, 2009) or PolyPhen2 (Adzhubei *et al.*, 2010) to be damaging to the protein. This locus also was recently replicated in a study of Hispanic women ($n = 3468$), which included about 370 women from PAGE (Chen *et al.*, 2012). Three of the five ANM SNPs generalized to diverse ancestral groups, and findings for the intronic SNP rs1172822 in *BRSK1* showed similar trends in all ancestral groups, but were not statistically significant in the non-European descent populations. Samples sizes were smaller in the ANM than AM analyses, mainly because of exclusions for non-natural menopause.

Some limitations of our study deserve mention. We relied on self-report and recall of ages at menarche and menopause, requiring women to report events that in some cases occurred many years ago. However, recall of AM or age at menopause (natural or induced) has been shown to be fairly reproducible in mainly European descent populations with high overall validity (± 1 year) compared with recorded age at the time of occurrence (Bean *et al.*, 1979; Colditz *et al.*, 1987; den Tonkelaar, 1997). Importantly, we would not expect any potential recall bias to be differential with respect to the studied genotypes. Our study is unique in describing genetic associations in large numbers of women differing by race/ethnicity. Yet,

Table V Age at natural menopause SNP association results in women of diverse ancestry.

SNP, locus, coded/other allele	GWAS reference study			PAGE study results for ANM					
	Ref	CAF	aBeta ± SE (in years)	Race/ethnicity	*n*	Ave CAF	[a]Beta ± SE (in years)	*P*-value	I^2
rs365132 *UIMC1* (syn) 5q35.2 T/G	[b]	0.49	0.39 ± 0.052	[d]EA	11 147				59
				ARIC	1625	0.50	0.01 ± 0.13	0.9602	
				CHS	954	0.51	0.09 ± 0.19	0.6261	
				MEC	1007	0.51	0.38 ± 0.14	**0.0069**	
				WHI	7561	0.50	0.40 ± 0.09	**8.3E−5**	
				AA	2237	0.79	0.46 ± 0.14	**0.0011**	0
				Asians	2008	0.48	0.22 ± 0.10	0.0281	0
				Hispanics	1145	0.42	0.25 ± 0.14	0.0904	37
				[d]Am Indians	929				74
				SHS	853	0.27	0.31 ± 0.23	0.1867	
				WHI	76	0.46	−1.51 ± 0.87	0.0920	
				N Hawaiians	299	0.57	−0.03 ± 0.25	0.9070	n/a
rs2153157 *SYCP2L* (intron) 6p24.2 T/C	[b]	0.49	0.29 ± 0.052	EA	11 161	0.49	0.17 ± 0.05	**0.0007**	0
				AA	2102	0.72	0.02 ± 0.13	0.9006	0
				Asians	2018	0.69	0.25 ± 0.11	0.0262	0
				Hispanics	1152	0.42	0.31 ± 0.14	0.0241	0
				Am Indians	933	0.24	0.64 ± 0.24	**0.0072**	0
				N Hawaiians	297	0.50	0.04 ± 0.26	0.8744	n/a
rs7333181 intergenic 13q34 A/G	[c]	0.12	0.52 ± 0.093	EA	10 110	0.13	0.06 ± 0.08	0.4677	0
				AA	1566	0.10	0.41 ± 0.27	0.1277	0
				Asians	707	0.04	0.17 ± 0.56	0.7680	0
				Hispanics	419	0.09	−0.06 ± 0.49	0.9042	0
				Am Indians	936	0.05	0.08 ± 0.57	0.8839	0
				N Hawaiians	n/a				
rs1172822 *BRSK1* (intron) 19q13.42 T/C	[b,c]	0.39	−0.39 ± 0.060	EA	11 146	0.36	−0.36 ± 0.05	**2.3E−11**	0
				[d]AA	2230				60
				ARIC	539	0.27	0.32 ± 0.26	0.2190	
				CHS	144	0.29	−0.31 ± 0.63	0.6238	
				MEC	673	0.27	0.01 ± 0.20	0.9574	
				WHI	874	0.28	−0.69 ± 0.33	0.0990	
				Asians	2021	0.05	−0.29 ± 0.23	0.2064	0
				Hispanics	1152	0.33	−0.27 ± 0.14	0.0560	0
				Am Indians	924	0.33	−0.27 ± 0.21	0.1951	n/a
				N Hawaiians	300	0.25	−0.41 ± 0.29	0.1556	n/a
rs16991615 *MCM8* (non-syn) 20p12.3 A/G	[b]	0.06	1.07 ± 0.110	EA	10 955	0.07	0.82 ± 0.11	**7.8E−15**	21
				AA	2129	0.02	0.57 ± 0.47	0.2276	33
				Asians	1978	0.001	1.25 ± 1.72	0.4690	0
				Hispanics	1143	0.06	1.05 ± 0.30	**0.0004**	0
				Am Indians	934	0.02	2.63 ± 0.80	**0.0010**	0
				N Hawaiians	294	0.03	−0.41 ± 0.78	0.6006	n/a

Ref, reference. *P*-values significant after multiple testing correction for five tests are given in bold.
[a]Beta reflects change in the ANM in years for each additional copy of the coded allele.
[b]Reference GWAS of ANM described is He *et al.* (2009).
[c]Reference GWAS of ANM described is Stolk *et al.* (2009).
[d]Due to high between-study heterogeneity ($I^2 > 60\%$ and/or χ^2 *P*-value <0.05), study-specific results are reported separately.

we acknowledge that some heterogeneity within our specified race/ethnicity groups may exist. We also observed moderate-to-high between-study heterogeneity for a few SNPs; in particular, the ANM analyses of rs365132 showed significant between-study heterogeneity in EA and Am Ind, indicating that the meta-analysis effect size is not a good summary of the study-specific estimates. In EA, the SNP was significantly associated with ANM in MEC and WHI, but not in ARIC or CHS. Reasons for this heterogeneity are not clear; the SNP remained significant regardless of cancer status in WHI and MEC (in sensitivity analyses). In Am Ind, coded allele frequencies notably differed between WHI and SHS, as did effect sizes. However, the WHI sample size is small and although the effect size in SHS was comparable to the other groups, the SNP did not replicate in either study and thus the interpretation (null result) remains the same.

A final limitation relates to study power to replicate previously reported findings in the non-European populations. While power for EA was excellent (>99% for all SNPs), we had low power to replicate findings in Am Ind or Native Hawaiians (power for all SNPs was

<60%) (Supplementary data, Table SI). Power for AA ranged from very high for the AM *LIN28B* SNPs (>90%) to modest for other SNPs, ranging from 56 to 83%. Hispanics had modest power for AM SNPs (60–82%) and more variable power for ANM SNPs (19–86%). Similarly, power for Asians varied highly by SNP (8–87%) due to varying allele frequencies and sample sizes. Overall, low power for some of the SNP-race/ethnicity group analyses may have limited our ability to replicate previously published GWAS findings, which also may be inflated due to ascertainment bias, termed the winner's curse (Zollner and Pritchard, 2007).

In spite of these limitations, our study is unique in describing genetic associations with important markers of the reproductive lifespan in diverse US racial/ethnic groups. To our knowledge, such data on Am Ind and Native Hawaiians have not been presented previously and also are very limited in US Hispanics and AA (Chen *et al.*, 2012; Dvornyk and Waqar-ul-Haq, 2012; He and Murabito, 2012; Spencer *et al.*, 2013). Generalization does not appear to depend on whether SNPs are reported to be functional, though one SNP, rs16991615, which did generalize to diverse groups, is a non-synonymous variant in *MCM8*. Indeed, robust replication and generalization were seen for several intronic SNPs as well as an SNP in the regulatory region of *LIN28B*. Generalization also does not appear to be consistent by race—we observed varying generalization by SNP, outcome and race/ethnicity group and for several SNPs, observed large differences in allele frequencies between the ancestral groups. In spite of reasonable power for the AM and ANM analyses, we did not replicate most associations in the large PAGE AA population, perhaps due to different patterns in LD between African and non-African populations, as described above.

The discovery of genetic variants associated with the reproductive lifespan provides an important opportunity to elucidate the biological mechanisms involved with both normal variation and disorders of menarche and menopause. As a next step in this process, we have examined the epidemiologic architecture of these early reported variants, describing their generalizability and effect size across differing ancestral populations in the USA. Our findings suggest that many but not all reproductive trait loci transfer across women of diverse ancestry. In particular, many SNP associations do not appear to extend to AA women. Inclusion of diverse ethnic groups, such as AA, in genetic discovery studies and fine-mapping studies of GWAS-identified loci is needed to identify population-specific variants and facilitate identification of causal variants. Furthermore, given that race/ethnicity may reflect a combination of different social and environmental factors in addition to genetic factors, investigation of how environmental factors interact with genetic variants to alter timing of AM and ANM in diverse populations may be warranted.

Supplementary data

Supplementary data are available at http://humrep.oxfordjournals.org/.

Acknowledgements

The PAGE consortium thanks the staff and participants of all PAGE studies for their important contributions.

Authors' roles

A.M.A., C.L.C., K.L.S. contributed to study design; I.C., L.L.M., M.F., N.F., U.P., V.M.S. contributed to data collection; B.H., C.A.H., C.S.C., D.C.C., J.W.W., K.L.S. contributed to study execution; A.Y., C.L.C., K.B.G., M.F., N.W.J., R.G., S.B., V.M.S. contributed to analysis; C.L.C. contributed to manuscript drafting; A.Y., C.L.C., L.A.H., L.F.R., N.F., V.W.S. contributed to data interpretation and A.M.A., A.P., C.A.H., D.C.C., J.M., J.W.W., L.A.H., L.F.R., N.F., N.P., T.C.M., U.P. contributed to critical review of the manuscript.

Funding

(a) The Population Architecture Using Genomics and Epidemiology (PAGE) program is funded by the National Human Genome Research Institute (NHGRI), supported by U01HG004803 (CALiCo), U01HG004798 (EAGLE), U01HG004802 (MEC), U01HG004790 (WHI) and U01HG004801 (Coordinating Center), and their respective NHGRI ARRA supplements. The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. The complete list of PAGE members can be found at http://www.pagestudy.org. (b) The data and materials included in this report result from a collaboration between the following studies: The 'Epidemiologic Architecture for Genes Linked to Environment (EAGLE)' is funded through the NHGRI PAGE program (U01HG004798-01 and its NHGRI ARRA supplement). The study participants derive from the National Health and Nutrition Examination Surveys (NHANES), and these studies are supported by the CDC. The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the CDC. The Multiethnic Cohort study (MEC) characterization of epidemiological architecture is funded through the NHGRI PAGE program (U01HG004802 and its NHGRI ARRA supplement). The MEC study is funded through the NCI (R37CA54281, R01 CA63, P01CA33619, U01CA136792 and U01CA98758). Funding support for the 'Epidemiology of putative genetic variants: The Women's Health Initiative' study is provided through the NHGRI PAGE program (U01HG004790 and its NHGRI ARRA supplement). The WHI program is funded by the NHLBI; NIH; and US HHS through contracts N01WH22110, 24152, 32100-2, 32105-6, 32108-9, 32111-13, 32115, 32118-32119, 32122, 42107-26, 42129-32 and 44221. The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: http://www.whiscience.org/publications/WHI_investigators_shortlist.pdf. Funding support for the Genetic Epidemiology of Causal Variants across the Life Course (CALiCo) program was provided through the NHGRI PAGE program (U01HG004803 and its NHGRI ARRA supplement). The following studies contributed to this manuscript and are funded by the following agencies: the Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by NHLBI contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C and HHSN268201100012C); R01HL087641; R01HL59367; R01HL086694; NHGRI contract U01HG004402; and NIH contract HHSN268200625226C. Infrastructure was partly supported by Grant Number

UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. The authors thank the staff and participants of the ARIC study for their important contributions. The Coronary Artery Risk Development in Young Adults (CARDIA) study is supported by the following: NIH, NHLBI contracts: N01-HC-95095; N01-HC-48047; N01-HC-48048; N01-HC-48049; N01-HC-48050; N01-HC-45134; N01-HC-05187; and N01-HC-45205. The Cardiovascular Health Study (CHS) is supported by contracts N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01-HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133, N01-HC-85239, HHSN268201200036C, grants U01HL080295 and R01 HL087652, HL105756 from NHLBI, with additional contribution from the NINDS See also http://www.chs-nhlbi.org/pi.htm. DNA handling and genotyping was supported in part by National Center of Advancing Translational Technologies CTSI grant UL1TR000124 and NIDDK grant DK063491 to the Southern California Diabetes Endocrinology Research Center and Cedars-Sinai Board of Governors' Chair in Medical Genetics (JIR). The Strong Heart Study (SHS) is supported by NHLBI grants U01 HL65520, U01 HL41642, U01 HL41652, U01 HL41654 and U01 HL65521. The opinions expressed in this paper are those of the author(s) and do not necessarily reflect the views of the Indian Health Service. Assistance with phenotype harmonization, SNP selection and annotation, data cleaning, data management, integration and dissemination and general study coordination was provided by the PAGE Coordinating Center (U01HG004801-01 and its NHGRI ARRA supplement). NIMH also contributes to the support for the Coordinating Center. L.F.R. received funding from the Ruth L. Kirschstein National Research Service Award, NIH, NHLBI # 5T32HL007055.

Conflict of interest

None declared.

References

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;**7**:248–249.

Anderson SE, Dallal GE, Must A. Relative weight and race influence average age at menarche: results from two nationally representative surveys of US girls studied 25 years apart. *Pediatrics* 2003;**111**:844–850.

Anderson CA, Duffy DL, Martin NG, Visscher PM. Estimation of variance components for age at menarche in twin families. *Behav Genet* 2007;**37**:668–677.

Bean JA, Leeper JD, Wallace RB, Sherman BM, Jagger H. Variations in the reporting of menstrual histories. *Am J Epidemiol* 1979;**109**:181–185.

Bhangale TR, Rieder MJ, Nickerson DA. Estimating coverage and power for genetic association studies using near-complete variation data. *Nat Genet* 2008;**40**:841–843.

Bromberger JT, Matthews KA, Kuller LH, Wing RR, Meilahn EN, Plantinga P. Prospective study of the determinants of age at menopause. *Am J Epidemiol* 1997;**145**:124–133.

Carlson CS, Matise TC, North KE, Haiman CA, Fesinmeyer MD, Buyske S, Schumacher FR, Peters U, Franceschini N, Ritchie MD *et al.* Generalization and attenuation of association results from European GWAS in Populations of non-European Ancestry: the PAGE Study. PLoS Biol *(in revision)* 2013.

Chang MH, Lindegren ML, Butler MA, Chanock SJ, Dowling NF, Gallagher M, Moonesinghe R, Moore CA, Ned RM, Reichler MR *et al.* Prevalence in the United States of selected candidate gene variants: Third National Health and Nutrition Examination Survey, 1991–1994. *Am J Epidemiol* 2009;**169**:54–66.

Chen CT, Fernandez-Rhodes L, Brzyski RG, Carlson CS, Chen Z, Heiss G, North KE, Woods NF, Rajkovic A, Kooperberg C *et al.* Replication of loci influencing ages at menarche and menopause in Hispanic women: the Women's Health Initiative SHARe Study. *Hum Mol Genet* 2012;**21**:1419–1432.

Colditz GA, Stampfer MJ, Willett WC, Stason WB, Rosner B, Hennekens CH, Speizer FE. Reproducibility and validity of self-reported menopausal status in a prospective cohort study. *Am J Epidemiol* 1987;**126**:319–325.

Cooper GS, Sandler DP. Age at natural menopause and mortality. *Ann Epidemiol* 1998;**8**:229–235.

Cramer DW. The epidemiology of endometrial and ovarian cancer. *Hematol Oncol Clin North Am* 2012;**26**:1–12.

Cutter GR, Burke GL, Dyer AR, Friedman GD, Hilner JE, Hughes GH, Hulley SB, Jacobs DR Jr, Liu K, Manolio TA *et al.* Cardiovascular risk factors in young adults. The CARDIA baseline monograph. *Control Clin Trials* 1991;**12**:1S–77S.

den Tonkelaar I. Validity and reproducibility of self-reported age at menopause in women participating in the DOM-project. *Maturitas* 1997;**27**:117–123.

Dvornyk V, Waqar-ul-Haq. Genetics of age at menarche: a systematic review. *Hum Reprod Update* 2012;**18**:198–210.

Elks CE, Perry JR, Sulem P, Chasman DI, Franceschini N, He C, Lunetta KL, Visser JA, Byrne EM, Cousminer DL *et al.* Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. *Nat Genet* 2010;**42**:1077–1085.

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 2011;**473**:43–49.

Freedman DS, Khan LK, Serdula MK, Dietz WH, Srinivasan SR, Berenson GS. Relation of age at menarche to race, time period, and anthropometric dimensions: the Bogalusa Heart Study. *Pediatrics* 2002;**110**:e43.

Freedman DS, Khan LK, Serdula MK, Dietz WH, Srinivasan SR, Berenson GS. The relation of menarcheal age to obesity in childhood and adulthood: the Bogalusa heart study. *BMC Pediatr* 2003;**3**:3.

Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, Kuller LH, Manolio TA, Mittelmark MB, Newman A *et al.* The Cardiovascular Health Study: design and rationale. *Ann Epidemiol* 1991;**1**:263–276.

Gajdos ZK, Henderson KD, Hirschhorn JN, Palmert MR. Genetic determinants of pubertal timing in the general population. *Mol Cell Endocrinol* 2010;**324**:21–29.

Gauderman W, Morrison J. *QUANTO 1.2.4. A Computer Program for Power and Sample Size Calculations for Genetic-epidemiology Studies* (http://hydra.usc.edu/gxe), 2006.

Gold EB, Bromberger J, Crawford S, Samuels S, Greendale GA, Harlow SD, Skurnick J. Factors associated with age at natural menopause in a multiethnic sample of midlife women. *Am J Epidemiol* 2001;**153**:865–874.

Gudbjartsson DF, Walters GB, Thorleifsson G, Stefansson H, Halldorsson BV, Zusmanovich P, Sulem P, Thorlacius S, Gylfason A, Steinberg S *et al.* Many sequence variants affecting diversity of adult human height. *Nat Genet* 2008;**40**:609–615.

Hayward C, Killen JD, Wilson DM, Hammer LD, Litt IF, Kraemer HC, Haydel F, Varady A, Taylor CB. Psychiatric risk associated with early puberty in adolescent girls. *J Am Acad Child Adolesc Psychiatry* 1997;**36**:255–262.

He C, Murabito JM. Genome-wide association studies of age at menarche and age at natural menopause. *Mol Cell Endocrin* 2012. [Epub ahead of print].

He C, Kraft P, Chen C, Buring JE, Pare G, Hankinson SE, Chanock SJ, Ridker PM, Hunter DJ, Chasman DI. Genome-wide association studies identify loci associated with age at menarche and age at natural menopause. *Nat Genet* 2009;**41**:724–728.

He C, Zhang C, Hunter DJ, Hankinson SE, Buck Louis GM, Hediger ML, Hu FB. Age at menarche and risk of type 2 diabetes: results from 2 large prospective cohort studies. *Am J Epidemiol* 2010;**171**:334–344.

Henderson KD, Bernstein L, Henderson B, Kolonel L, Pike MC. Predictors of the timing of natural menopause in the Multiethnic Cohort Study. *Am J Epidemiol* 2008;**167**:1287–1294.

Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;**21**:1539–1558.

Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;**327**:557–560.

Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA* 2009;**106**:9362–9367.

Hsieh CC, Trichopoulos D, Katsouyanni K, Yuasa S. Age at menarche, age at menopause, height and obesity as risk factors for breast cancer: associations and interactions in an international case–control study. *Int J Cancer* 1990;**46**:796–800.

Hughes GH, Cutter G, Donahue R, Friedman GD, Hulley S, Hunkeler E, Jacobs DR Jr, Liu K, Orden S, Pirie P *et al*. Recruitment in the coronary artery disease risk development in young adults (Cardia) study. *Control Clin Trials* 1987;**8**:68S–73S.

Kaprio J, Rimpela A, Winter T, Viken RJ, Rimpela M, Rose RJ. Common genetic influences on BMI and age at menarche. *Hum Biol* 1995; **67**:739–753.

Kolonel LN, Henderson BE, Hankin JH, Nomura AM, Wilkens LR, Pike MC, Stram DO, Monroe KR, Earle ME, Nagamine FS. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. *Am J Epidemiol* 2000;**151**:346–357.

Kosoy R, Nassir R, Tian C, White PA, Butler LM, Silva G, Kittles R, Alarcon-Riquelme ME, Gregersen PK, Belmont JW *et al*. Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Hum Mutat* 2009;**30**:69–78.

Kritz-Silverstein D, Barrett-Connor E. Early menopause, number of reproductive years, and bone mineral density in postmenopausal women. *Am J Public Health* 1993;**83**:983–988.

Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009;**4**:1073–1081.

Lakshman R, Forouhi N, Luben R, Bingham S, Khaw K, Wareham N, Ong KK. Association between age at menarche and risk of diabetes in adults: results from the EPIC-Norfolk cohort study. *Diabetologia* 2008; **51**:781–786.

Lakshman R, Forouhi NG, Sharp SJ, Luben R, Bingham SA, Khaw KT, Wareham NJ, Ong KK. Early age at menarche associated with cardiovascular disease and mortality. *J Clin Endocrinol Metab* 2009; **94**:4953–4960.

Lee ET, Welty TK, Fabsitz R, Cowan LD, Le NA, Oopik AJ, Cucchiara AJ, Savage PJ, Howard BV. The Strong Heart Study. A study of cardiovascular disease in American Indians: design and methods. *Am J Epidemiol* 1990;**132**:1141–1155.

Lettre G, Jackson AU, Gieger C, Schumacher FR, Berndt SI, Sanna S, Eyheramendy S, Voight BF, Butler JL, Guiducci C *et al*. Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat Genet* 2008;**40**:584–591.

Lisabeth LD, Beiser AS, Brown DL, Murabito JM, Kelly-Hayes M, Wolf PA. Age at natural menopause and risk of ischemic stroke: the Framingham heart study. *Stroke* 2009;**40**:1044–1049.

Liu YZ, Guo YF, Wang L, Tan LJ, Liu XG, Pei YF, Yan H, Xiong DH, Deng FY, Yu N *et al*. Genome-wide association analyses identify SPOCK as a key novel gene underlying age at menarche. *PLoS Genet* 2009;**5**:e1000420.

Matise TC, Ambite JL, Buyske S, Carlson CS, Cole SA, Crawford DC, Haiman CA, Heiss G, Kooperberg C, Marchand LL *et al*. The next PAGE in understanding complex traits: design for the analysis of Population Architecture Using Genetics and Epidemiology (PAGE) study. *Am J Epidemiol* 2011;**174**:849–859.

Morris DH, Jones ME, Schoemaker MJ, Ashworth A, Swerdlow AJ. Familial concordance for age at natural menopause: results from the Breakthrough Generations Study. *Menopause* 2011;**18**:956–961.

Murabito JM, Yang Q, Fox C, Wilson PW, Cupples LA. Heritability of age at natural menopause in the Framingham Heart Study. *J Clin Endocrinol Metab* 2005;**90**:3427–3430.

North KE, Howard BV, Welty TK, Best LG, Lee ET, Yeh JL, Fabsitz RR, Roman MJ, MacCluer JW. Genetic and environmental contributions to cardiovascular disease risk in American Indians: the strong heart family study. *Am J Epidemiol* 2003;**157**:303–314.

Ong KK, Elks CE, Li S, Zhao JH, Luan J, Andersen LB, Bingham SA, Brage S, Smith GD, Ekelund U *et al*. Genetic variation in LIN28B is associated with the timing of puberty. *Nat Genet* 2009;**41**:729–733.

Parent AS, Teilmann G, Juul A, Skakkebaek NE, Toppari J, Bourguignon JP. The timing of normal puberty and the age limits of sexual precocity: variations around the world, secular trends, and changes after migration. *Endocrine Rev* 2003;**24**:668–693.

Perry JR, Stolk L, Franceschini N, Lunetta KL, Zhai G, McArdle PF, Smith AV, Aspelund T, Bandinelli S, Boerwinkle E *et al*. Meta-analysis of genome-wide association data identifies two loci influencing age at menarche. *Nat Genet* 2009;**41**:648–650.

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R *et al*. Linkage disequilibrium in the human genome. *Nature* 2001;**411**:199–204.

Spencer KL, Malinowski JR, Carty CL, Franceschini N, Fernández-Rhodes L, Young A, Cheng I, Ritchie MD, Haiman CA, Wilkens LR *et al*. Genetic variation and the reproductive lifespan: African American Women from the population architecture using genomics and epidemiology (PAGE) study. *PLoS ONE* 2013;**8**:e55258.

Stockl D, Meisinger C, Peters A, Thorand B, Huth C, Heier M, Rathmann W, Kowall B, Stockl H, Doring A. Age at menarche and its association with the metabolic syndrome and its components: results from the KORA F4 study. *PLoS ONE* 2011;**6**:e26076.

Stolk L, Zhai G, van Meurs JB, Verbiest MM, Visser JA, Estrada K, Rivadeneira F, Williams FM, Cherkas L, Deloukas P *et al*. Loci at chromosomes 13, 19 and 20 influence age at natural menopause. *Nat Genet* 2009;**41**:645–647.

Stolk L, Perry JR, Chasman DI, He C, Mangino M, Sulem P, Barbalic M, Broer L, Byrne EM, Ernst F *et al*. Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways. *Nat Genet* 2012;**44**:260–268.

Sulem P, Gudbjartsson DF, Rafnar T, Holm H, Olafsdottir EJ, Olafsdottir GH, Jonsson T, Alexandersen P, Feenstra B, Boyd HA *et al*. Genome-wide association study identifies sequence variants on 6q21 associated with age at menarche. *Nat Genet* 2009;**41**:734–738.

te Velde ER, Pearson PL. The variability of female reproductive ageing. *Hum Reprod Update* 2002;**8**:141–154.

The ARIC Investigators. The Atherosclerosis risk in communities (ARIC) study: design and objectives. *Am J Epidemiol* 1989;**129**:687–702.

The Women's Health Initiative Study Group. Design of the Women's health initiative clinical trial and observational study. *Control Clin Trials* 1998;**19**:61–109.

Thornton JE, Gregory RI. How does Lin28 let-7 control development and disease? *Trends Cell Biol* 2012;**22**:474–482.

van Noord PA, Dubas JS, Dorland M, Boersma H, te Velde E. Age at natural menopause in a population-based screening cohort: the role of menarche, fecundity, and lifestyle factors. *Fertil Steril* 1997;**68**:95–102.

Voorhuis M, Onland-Moret NC, van der Schouw YT, Fauser BC, Broekmans FJ. Human studies on genetics of the age at natural menopause: a systematic review. *Hum Reprod Update* 2010; **16**:364–377.

Wellons M, Ouyang P, Schreiner PJ, Herrington DM, Vaidya D. Early menopause predicts future coronary heart disease and stroke: the Multi-ethnic Study of Atherosclerosis. *Menopause* 2012;**19**:1081–1087.

Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010;**26**:2190–2191.

Wu T, Mendola P, Buck GM. Ethnic differences in the presence of secondary sex characteristics and menarche among US girls: the Third National Health and Nutrition Examination Survey, 1988–1994. *Pediatrics* 2002;**110**:752–757.

Zollner S, Pritchard JK. Overcoming the winner's curse: estimating penetrance parameters from case-control data. *Am J Hum Genet* 2007;**80**:605–615.