Human Reproduction, Vol.36, No.7, pp. 1999–2010, 2021
Advance Access Publication on May 22, 2021 doi:10.1093/humrep/deab086

human reproduction

ORIGINAL ARTICLE ***Reproductive genetics***

Large trans-ethnic meta-analysis identifies *AKR1C4* as a novel gene associated with age at menarche

C. Sarnowski [1,2,32,*], D.L. Cousminer[3,4,5], N. Franceschini[6], L.M. Raffield[7], G. Jia[8], L. Fernández-Rhodes[9], S.F.A. Grant[3,4,10,11], H. Hakonarson [3,4,10,12], L.A. Lange[13], J. Long[8], T. Sofer [14,15], R. Tao[16,17], R.B. Wallace[18], Q. Wong[19], G. Zirpoli [20,21], E. Boerwinkle[22], J.P. Bradfield[3,4,23], A. Correa[24,25,26], C.L. Kooperberg[27], K.E. North[6,28], J.R. Palmer[20,21], B.S. Zemel[10,29], W. Zheng [8], J.M. Murabito[30,31], and K.L. Lunetta[1]

[1]Boston University School of Public Health, Boston, MA, USA [2]Department of Epidemiology, Human Genetics and Environmental Sciences, The University of Texas Health Science Center at Houston School of Public Health, Houston, TX, USA [3]Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA [4]Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA [5]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA [6]Department of Epidemiology, University of North Carolina at Chapel Hill Gillings School of Global Public Health, Chapel Hill, NC, USA [7]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA [8]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN, USA [9]Department of Biobehavioral Health, College of Health and Human Development, Pennsylvania State University, University Park, PA, USA [10]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA [11]Division of Endocrinology, The Children's Hospital of Philadelphia, Philadelphia, PA, USA [12]Division of Pulmonary Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA, USA [13]Department of Medicine, University of Colorado Denver, Anschutz Medical Campus, Aurora, CO, USA [14]Departments of Medicine and of Biostatistics, Harvard University, Boston, MA, USA [15]Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital, Boston, MA, USA [16]Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, USA [17]Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA [18]University of Iowa College of Public Health, Iowa City, IA, USA [19]Department of Biostatistics, University of Washington, Seattle, WA, USA [20]Slone Epidemiology Center at Boston University, Boston, MA, USA [21]Section of Hematology/Oncology, Department of Medicine, Boston University School of Medicine, Boston, MA, USA [22]Human Genetic Center and Department of Epidemiology, The University of Texas School of Public Health, Houston, TX, USA [23]Quantinuum Research, LLC, Wayne, PA, USA [24]Department of Medicine, University of Mississippi Medical Center, Jackson, MS, USA [25]Department of Pediatrics, University of Mississippi Medical Center, Jackson, MS, USA [26]Department of Population Health Science, University of Mississippi Medical Center, Jackson, MS, USA [27]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA [28]Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill, NC, USA [29]Division of Gastroenterology, Hepatology and Nutrition, The Children's Hospital of Philadelphia, Philadelphia, PA, USA [30]National Heart Lung and Blood Institute and Boston University's Framingham Heart Study, Framingham, MA, USA [31]Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA [32]Present address: Department of Epidemiology, Human Genetics and Environmental Sciences, University of Texas Health Science Center School of Public Health, 1200 Pressler St, Houston, TX 77030, USA

*Correspondence address. Department of Epidemiology, Human Genetics and Environmental Sciences, The University of Texas Health Science Center at Houston School of Public Health, 1200 Pressler St, TX 77030, Houston, USA. Tel: +1-713-500-9004; E-mail: Chloe.Sarnowski@uth.tmc.edu https://orcid.org/0000-0002-6090-7099

Submitted on November 30, 2020; resubmitted on March 12, 2021; editorial decision on March 22, 2021

STUDY QUESTION: Does the expansion of genome-wide association studies (GWAS) to a broader range of ancestries improve the ability to identify and generalise variants associated with age at menarche (AAM) in European populations to a wider range of world populations?

SUMMARY ANSWER: By including women with diverse and predominantly non-European ancestry in a large-scale meta-analysis of AAM with half of the women being of African ancestry, we identified a new locus associated with AAM in African-ancestry participants, and generalised loci from GWAS of European ancestry individuals.

WHAT IS KNOWN ALREADY: AAM is a highly polygenic puberty trait associated with various diseases later in life. Both AAM and diseases associated with puberty timing vary by race or ethnicity. The majority of GWAS of AAM have been performed in European ancestry women.

STUDY DESIGN, SIZE, DURATION: We analysed a total of 38 546 women who did not have predominantly European ancestry backgrounds: 25 149 women from seven studies from the ReproGen Consortium and 13 397 women from the UK Biobank. In addition, we used an independent sample of 5148 African-ancestry women from the Southern Community Cohort Study (SCCS) for replication.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Each AAM GWAS was performed by study and ancestry or ethnic group using linear regression models adjusted for birth year and study-specific covariates. ReproGen and UK Biobank results were meta-analysed using an inverse variance-weighted average method. A trans-ethnic meta-analysis was also carried out to assess heterogeneity due to different ancestry.

MAIN RESULTS AND THE ROLE OF CHANCE: We observed consistent direction and effect sizes between our meta-analysis and the largest GWAS conducted in European or Asian ancestry women. We validated four AAM loci (1p31, 6q16, 6q22 and 9q31) with common genetic variants at $P < 5 \times 10^{-7}$. We detected one new association (10p15) at $P < 5 \times 10^{-8}$ with a low-frequency genetic variant lying in *AKR1C4*, which was replicated in an independent sample. This gene belongs to a family of enzymes that regulate the metabolism of steroid hormones and have been implicated in the pathophysiology of uterine diseases. The genetic variant in the new locus is more frequent in African-ancestry participants, and has a very low frequency in Asian or European-ancestry individuals.

LARGE SCALE DATA: N/A

LIMITATIONS, REASONS FOR CAUTION: Extreme AAM (<9 years or >18 years) were excluded from analysis. Women may not fully recall their AAM as most of the studies were conducted many years later. Further studies in women with diverse and predominantly non-European ancestry are needed to confirm and extend these findings, but the availability of such replication samples is limited.

WIDER IMPLICATIONS OF THE FINDINGS: Expanding association studies to a broader range of ancestries or ethnicities may improve the identification of new genetic variants associated with complex diseases or traits and the generalisation of variants from European-ancestry studies to a wider range of world populations.

STUDY FUNDING/COMPETING INTEREST(S): Funding was provided by CHARGE Consortium grant R01HL105756-07: Gene Discovery For CVD and Aging Phenotypes and by the NIH grant U24AG051129 awarded by the National Institute on Aging (NIA). The authors have no conflict of interest to declare.

Key words: trans-ethnic meta-analysis / age at menarche / GWAS / *AKR1C4* / ReproGen Consortium / UK Biobank

Introduction

The timing of menarche (age at menarche, AAM) is a highly polygenic trait (twin study heritability >50%) (Anderson *et al.*, 2007; Morris *et al.*, 2011) associated for women with various diseases later in life (Day *et al.*, 2015). Early menarche (defined as puberty timing before 9 years old) is associated with a wide range of later health outcomes, including increased risk of breast cancer (Kotsopoulos *et al.*, 2005; Collaborative Group on Hormonal Factors in Breast Cancer, 2012), obesity (Freedman *et al.*, 2003) and type 2 diabetes (Lakshman *et al.*, 2008; Elks *et al.*, 2013). Late menarche (defined as a puberty timing age after 18 years old) is also associated with later life outcomes, including lower fertility (Komura *et al.*, 1992; Weghofer *et al.*, 2013) and osteoporosis. In a large prospective study of UK women, both early and late menarche were associated with increased vascular disease risk (Canoy *et al.*, 2015).

AAM and diseases associated with puberty timing both vary by race or ethnicity. For example, African-American (AA) girls have an earlier AAM compared to European-American girls (Wu *et al.*, 2002; Anderson *et al.*, 2003; Chumlea *et al.*, 2003). African-American women tend to have twice the prevalence of chronic diseases (known to be related to early AAM, such as obesity and type 2 diabetes), as well as higher prevalence of hypertension, compared to non-Hispanic White women (Virani *et al.*, 2020).

The largest genome-wide association study (GWAS) of AAM, performed in up to ~370 000 women of European ancestry (EA) from the ReproGen Consortium, 23andMe, the UK Biobank (UKBB) and deCODE, identified 389 independent signals (Day *et al.*, 2017). In this study, the total-chip captured heritability, calculated in the UKBB only, was 32%.

Few GWAS or genetic studies have been performed for AAM in multi-ethnic populations or in women of diverse ancestries or ethnicities including Hispanic/Latina women and women of African ancestry (Carty *et al.*, 2013; Demerath *et al.*, 2013; Spencer *et al.*, 2013; Fernandez-Rhodes *et al.*, 2018; Horikoshi *et al.*, 2018). A GWAS of AAM performed in 18 089 African-American (AA) women provided the first evidence of cross-ancestry generalisation of menarche loci identified in European GWAS (60%) (Demerath *et al.*, 2013). This study used SNP arrays designed to capture genetic variants that are common in European populations. Thus, a substantial fraction of genetic variants that are more common in African populations is likely to have been missed or imprecisely tagged using imputation with the Hapmap reference panel. A later study that included 42 251 women of diverse ancestries replicated the 6q16 and 9q31 AAM regions reported in European descent populations and compared their effect sizes across differing ancestral populations (Carty *et al.*, 2013). Another trans-ethnic study of AAM using 45 364 women of diverse ancestries and the Metabochip array (Voight *et al.*, 2012) supported

the transferability of loci discovered in European women to women of other ancestries and identified new trans-ethnic associations at novel and at established loci (Fernandez-Rhodes *et al.*, 2018). A recent study from the Biobank Japan (BBJ) conducted in up to 67 029 women of Japanese ancestry reported 10 loci associated with puberty timing and showed large differences in allele frequencies and effect estimates between variants reported in Japanese and European GWAS (Horikoshi *et al.*, 2018). The genotyping array-based heritability estimated in this study for AAM was 13%. The lack of ancestral diversity of these genetic studies (with very few South Asian women included) and/or the use of low genome-wide coverage arrays have limited their ability to explore lower frequency genetic variation. The two mentioned AAM GWAS or meta-analysis of AAM GWAS (Demerath et al. and Horikoshi et al.) were conducted in only one population group (African-American and East Asian, respectively). No large meta-analysis of GWAS has been performed to date for AAM in multi-ethnic populations or in women of diverse ancestries or ethnicities. Large-scale trans-ethnic approaches for AAM conducted in women with diverse ancestries may help to better understand ethnic differences in pubertal timing and its associated chronic diseases.

In this study, we performed a large-scale trans-ethnic GWAS meta-analysis of AAM in women with diverse and predominantly non-European ancestry or ethnicity (African, Hispanic/Latina, East Asian and South Asian) from the ReproGen Consortium and the UKBB to identify novel associations that may be ancestry-specific, and to generalise GWAS regions reported by European studies to a wider range of world populations.

Materials and methods

Populations and participants

We included in our trans-ethnic meta-analysis women who did not have predominantly European ancestry backgrounds from seven studies from the ReproGen Consortium (ARIC: The Atherosclerosis Risk in Communities Study (Anonymous, 1989); BMDCS: The Bone Mineral Density in Childhood Study (Zemel *et al.*, 2011); BWHS: Black Women's Health Study (Palmer *et al.*, 2013); CHOP: Children's Hospital of Philadelphia; JHS: The Jackson Heart Study (JHS) (Carpenter *et al.*, 2004; Fuqua *et al.*, 2005; Keku *et al.*, 2005; Taylor *et al.*, 2005; Wilson *et al.*, 2005); HCHS/SOL: Hispanic Community Health Study/The Study of Latinos (Laurie *et al.*, 2010; Sorlie *et al.*, 2010); and WHI: The Women's Health Initiative (Anonymous, 1998)) and from the UKBB (O'Connell *et al.*, 2014; Galinsky *et al.*, 2016; Bycroft *et al.*, 2018; Zhou *et al.*, 2018). A description of each study is included in Supplementary Table SI. Details about the association analyses performed in the UKBB are included in the Supplementary Text. AAM was self-reported, recorded in whole years and analysed as a quantitative trait. AAM younger than 9 or older than 18 were excluded from analysis.

Genotyping and imputation

Studies used the most dense imputation reference panel available to them at the time of analyses, either 1000 Genomes (1 kG) or the Haplotype Reference Consortium (HRC). Three ReproGen studies provided results for chromosome X (BWHS, JHS and HCHS/SOL). A description of the reference panel used by each study is provided in Supplementary Table SII.

Genome-wide association study

Each study evaluated the association of single nucleotide genetic variants with AAM under an additive model. Covariates included birth year and additional study specific covariates to control for population structure (principal components). Studies with multiple distinct ancestries analysed each ancestry separately. The minimum sample size for each ancestry/phenotype combination for inclusion in this study was fixed to 100.

Quality control (QC) of GWAS results

QC for each study was performed using EasyQC (Winkler *et al.*, 2014). Study-specific GWAS results were filtered based on an imputation quality score greater or equal to 0.30 and an effective number of minor alleles (minor allele count $\times$ imputation quality score) greater or equal to 20.

Trans-ethnic meta-analysis

GWAS results across all studies from the ReproGen diverse ancestry sample were meta-analysed using METAL (Willer *et al.*, 2010) using an inverse variance-weighted average method. Results for which at least two studies contributed were then meta-analysed with the UKBB diverse ancestry sample GWAS using METAL. Genomic control correction was applied to all studies included in the meta-analyses. Variants with a minor allele frequency (MAF) less than 0.005 in the final meta-analysis were excluded. To investigate heterogeneity due to different ancestry at associated loci, we additionally performed a trans-ethnic meta-analysis using MR-MEGA and ancestry-stratified GWAS results (Magi *et al.*, 2017).

Replication

For the novel loci identified, we sought replication in an independent sample of 5148 African-ancestry women from the Southern Community Cohort Study (SCCS) (Signorello *et al.*, 2005; Kowalski *et al.*, 2019). A description of this study is provided in Supplementary Table SI and a description of the association analysis is included in the Supplementary Text.

Results

Populations and participants

We included in our trans-ethnic meta-analysis a total of 38 546 women (Table I): 25 149 women from seven studies from the ReproGen diverse ancestry sample and 13 397 women from the UKBB diverse ancestry samples, representing four major ancestry or ethnic groups (African: N = 19 028, Hispanic/Latina: N = 10 661, East Asian: N = 2095 and South Asian: N = 3205). A description of each study is included in Supplementary Table SI. Principal Component (PC) plots generated using output from MR-MEGA are presented in Supplementary Fig. S1.

Table I Description of the 38 546 women included in the meta-analysis of genome-wide association studies of age at menarche from the diverse ancestry sample of the ReproGen Consortium and the UK Biobank (UKBB).

		Studies[a]	N	Age at Menarche, mean (SD) or median [25–75%]
UKBB (N = 13 197)	ALL		13 397	13.09 (1.70)
Diverse ancestry sample	African (N = 4540)		4540	13.15 (1.74)
	South Asian (N = 3205)		3205	13.22 (1.69)
	East Asian (N = 2095)		2095	12.96 (1.61)
ReproGen (N = 25 149)	African American or African-ancestry (N = 14 488)	ARIC	1780	12.90 (1.70)
Diverse ancestry sample		BMDCS	146	12.13 [10.11-16.47]
		BWHS	2448	13.30 (1.60)
		CHOP	620	11.86 [9-18]
		WHI	8239	12.60 (1.60)
		JHS	1255	12.74 (1.73)
	Hispanic/Latina (N = 10 661)	WHI	3494	12.50 (1.60)
		HCHS/SOL	7167	12.60 (1.80)
ReproGen + UKBB Diverse ancestry sample			38 546	

[a]ARIC, The Atherosclerosis Risk in Communities Study; BMDCS, The Bone Mineral Density in Childhood Study; BWHS, Black Women's Health Study; CHOP, Children's Hospital of Philadelphia; JHS, The Jackson Heart Study (JHS); HCHS/SOL, Hispanic Community Health Study/The Study of Latinos; WHI, The Women's Health Initiative; UKBB, The UK Biobank.

Main results

Quantile-Quantile and Manhattan plots are presented in Supplementary Figs. S2 and S3. We detected one new association with a low-frequency (MAF = 0.008) genetic variant (10p15) and validated three known AAM loci for associated common genetic variants (6q16, 6q22, and 9q31) at the genome-wide threshold ($P < 5 \times 10^{-8}$). We did not observe heterogeneity (due to different ancestry or residual) between ReproGen and UKBB effect sizes for our main findings and observed consistent effect sizes and direction of effects between the two diverse ancestry samples for the most significantly associated signal in each locus (Table II and Supplementary Table SIII). We also extracted the results for the lead genetic variants at the four main genetic loci in the Pan-UKBB publicly-available ancestry-specific AAM GWAS results and did not observe significant heterogeneity in effect size (Supplementary Table SIV).

Except at the 9q31 locus, the sentinel variant differed between our trans-ethnic meta-analysis and the Day *et al.* (2017) EA GWAS (N∼252k) (Day *et al.*, 2017). The sentinel signal at 6q16 was rs2095812-G (MAF = 0.29, beta = 0.09, $P = 9.1 \times 10^{-11}$), at 6q22 it was rs9401883-G (MAF = 0.44, beta = −0.08, $P = 2.4 \times 10^{-8}$), and at 9q31 it was rs7852169-G (MAF = 0.21, beta = 0.09, $P = 1.3 \times 10^{-8}$). The 6q16 lead variant (rs2095812) was in high ($r^2 \geq 0.98$) linkage disequilibrium (LD) with the known GWAS variant (rs395962) in European, African or Asian ancestry individuals. The 6q22 lead variant (rs9401883) was in high LD ($r^2 > 0.75$) with the known GWAS variant (rs4897178) in European and Asian individuals but not in African individuals ($r^2 \leq 0.40$). LD estimated in the 1000 G Phase 3 v5 reference panel or in the UKBB is available in Supplementary Tables SV and SVI by ancestry. Regional association plots in 6q16, 6q22 and 9q31 are presented in Supplementary Figs. S4, S5, and S6.

The new and low MAF signal at 10p15, rs182498797-A (MAF = 0.008, beta = 0.53, $P = 1.7 \times 10^{-8}$), was only observed in African-ancestry studies. The variant is monomorphic in other populations from the 1000 Genomes Project data. The frequency of the minor A allele was similar in ReproGen African-American or African-ancestry women and in UKBB African-ancestry women (MAF∼1%), whereas the allele was extremely rare in UKBB European ancestry women (MAF = 4×10^{-6}). The variant was monomorphic in UKBB East and South Asian women. We found a consistent effect in the same direction for rs182498797-A (MAF = 0.02) in an independent sample of African-ancestry women from SCCS (one-sided test $P = 0.03$; meta-analysis result: beta = 0.51, $P = 3.2 \times 10^{-9}$). Regional association and forest plots at 10p15 are provided in Supplementary Fig. S7 and Fig. 1. The rs182498797 genetic variant lies in *AKR1C4*, a gene that belongs to the aldo-keto reductase family 1 (*AKR1*). We found associations for genetic variants lying in the cluster of *AKR1* genes at 10p15 (*AKR1C1-4*) and in *AKR1D1* in our AAM meta-analysis and in the published AAM GWAS (the Day *et al.*, 2017 EA (N∼252k) and in the Horikoshi *et al.*, 2018 Biobank Japan). The most associated genetic variants at these loci are presented in Supplementary Table SVII. We were not able to look-up rs182498797-A (or genetic variants in LD) in GTEx as this variant is extremely rare in non-African ancestry individuals, and availability of eQTL data in African-ancestry populations is limited. In addition, we saw evidence at the sub-genome-wide suggestive threshold ($P < 5 \times 10^{-7}$) for one additional known GWAS AAM locus driven by common genetic variants (1p31), and we detected two new loci, driven by a common variant (8p11) or a low frequency variant (Xq21) (Supplementary Table SVIII). The 1p31 lead

Table II Main association results ($P < 5 \times 10^{-8}$) from the fixed-effects meta-analysis of genome-wide association studies of age at menarche in 38 546 women from the diverse ancestry samples of the ReproGen Consortium and the UK Biobank (UKBB).

							ReproGen+UKBB diverse ancestry meta-analysis (N = 38 546)				ReproGen diverse ancestry meta-analysis (N = 25 149)				UKBB diverse ancestry GWAS (N = 13 197)			
Locus	Known/ Novel	rsid	Chr	Build 37 Position	ALT/REF[a]	Nearest Gene	EAF	Beta[b]	SE	*P*-value	EAF	Beta	SE	*P*-value	EAF	Beta	SE	*P*-value
6q16	Known	rs2095812	6	105 383 978	C/G	*LIN28B*	0.71	−0.09	0.01	9.1E−11	0.71	−0.10	0.02	1.2E−08	0.70	−0.08	0.02	9.0E−04
6q22	Known	rs9401883	6	126 797 111	A/G	*CENPW*	0.56	0.08	0.01	2.4E−08	0.59	0.07	0.02	3.0E−05	0.50	0.10	0.03	5.9E−05
9q31	Known	rs7852169	9	114 318 394	C/G	*PTGR1*	0.79	−0.09	0.02	1.3E−08	0.77	−0.08	0.02	1.2E−05	0.81	−0.11	0.03	1.0E−04
10p15	Novel	rs182498797	10	5 257 282	A/G	*AKR1C4*	0.008	0.53	0.09	1.7E−08	0.01	0.52	0.11	1.8E−06	0.004	0.56	0.18	1.5E−03

[a]ALT/REF: alternate/reference alleles; EAF: effect (ALT) allele frequency.
[b]The effect estimates (betas) are per year of age at menarche.

variant (rs7526762) was in high LD ($r^2 \geq 0.98$) with the known GWAS variant (rs11210476) in European, African or Asian ancestry individuals (Supplementary Tables SV and SVI). A regional association plot at 1p31 is provided in Supplementary Fig. S8. Regional association and forest plots at Xq21 and 8p11 are provided in Supplementary Figs. S9, S10, S11, S12, and S13. We observed consistent direction and magnitude of effects between ReproGen and UKBB for the most associated signal at each locus (Supplementary Table SVIII).

The novel signal at Xq21 (rs112344779-T) was detected in UKBB only. The variant was present in both BWHS and JHS ReproGen studies but did not pass our filtering criteria (effective number of minor alleles $\geq$ 20) in JHS to be included in the ReproGen meta-analysis. However, the direction of effect was consistent between UKBB and ReproGen (Supplementary Fig. S11). The minor allele T was more frequent in African-ancestry women (MAF $\geq$ 1%) than in East Asian, South Asian and European women (MAF $\leq 6 \times 10^{-5}$). We identified in the UKBB a second distinct signal in this locus (rs139960405-T, $P_{cond} = 1.8 \times 10^{-6}$) that was more frequent in African-ancestry women (MAF $\geq$ 1%) than in East Asian, South Asian and European women (MAF $\leq$ 0.002). This association did not replicate in SCCS. However, the variant was only available in a subset of participants with whole genome sequence data (N = 671) and the estimates could be unstable.

Evaluation of previously published AAM GWAS signals

We superimposed the known GWAS variants reported in previous AAM GWAS (the Day *et al.*, 2017 EA (N~252k) and the Horikoshi *et al.*, 2018 Biobank Japan) on our Manhattan plot (Supplementary Figs. S14 and S15). We also looked-up our main signals in the Day *et al.*, 2017 EA (N~252k) and in the Horikoshi *et al.*, 2018 Biobank Japan AAM GWAS (Supplementary Tables SIX and SX) and observed consistent direction and size of effects. Among the 24 172 variants (at ~200 loci) that were genome-wide significant in either our meta-analysis or in the Day *et al.*, 2017 EA AAM GWAS, we found that 82% of the variants had consistent direction of effects. Among the 362 variants that were genome-wide significant in either our meta-analysis or in the Horikoshi *et al.* (2018) Biobank Japan AAM GWAS, we found that 90% of the signals had consistent direction of effects. Comparison of effect sizes between our meta-analysis and the published AAM GWAS for the genetic variants passing the genome-wide threshold in at least one analysis is provided in Fig. 2.

The majority of discordant variants were common (MAF > 0.01) for both the Day *et al.*, 2017 EA (N~252k) and the Horikoshi *et al.*, 2018 Biobank Japan AAM GWAS, but the proportion of low-frequency or rare variants among the discordant variants was significantly higher than among concordant variants for the Day *et al.*, 2017 EA (N~252k) AAM GWAS (1% vs. 0.4%; $P = 4.2 \times 10^{-9}$).

The effect size estimate of the novel signal at 8p11 (rs10096592) was similar in the Day *et al.*, 2017 EA (N~252k) GWAS, although the variant was much less frequent and not significantly associated with AAM in the Day study (Supplementary Table SX). The minor allele T was more frequent in African-ancestry women (MAF = 17%), than in East Asian, South Asian, Hispanic and European women (MAF $\leq$ 2%). This association did not replicate in SCCS. However, the frequency of the minor allele T was much higher in SCCS (38%) compared to ReproGen or UKBB (18%).

Discussion

In this paper, we performed a large-scale trans-ethnic meta-analysis of AAM in 38 546 women from the diverse ancestry sample of the ReproGen Consortium and the UKBB. We were able to identify one novel association (10p15) with a genetic variant that was more common in African-ancestry participants and had very low frequencies in European or Asian ancestry individuals and to validate and generalise the association of common genetic variants in four known AAM GWAS loci (1p31, 6q16, 6q22 and 9q31) originally discovered in Europeans.

We checked the ranking of these four AAM GWAS loci among the 389 lead variants reported by the largest European AAM GWAS (Day *et al.*, 2017). Except at the 6q16 locus that was ranked first, the three other regions were not ranked as the top associated loci in the European AAM GWAS. This suggests that different regions may explain more of the AAM variance in the diverse ancestry sample of

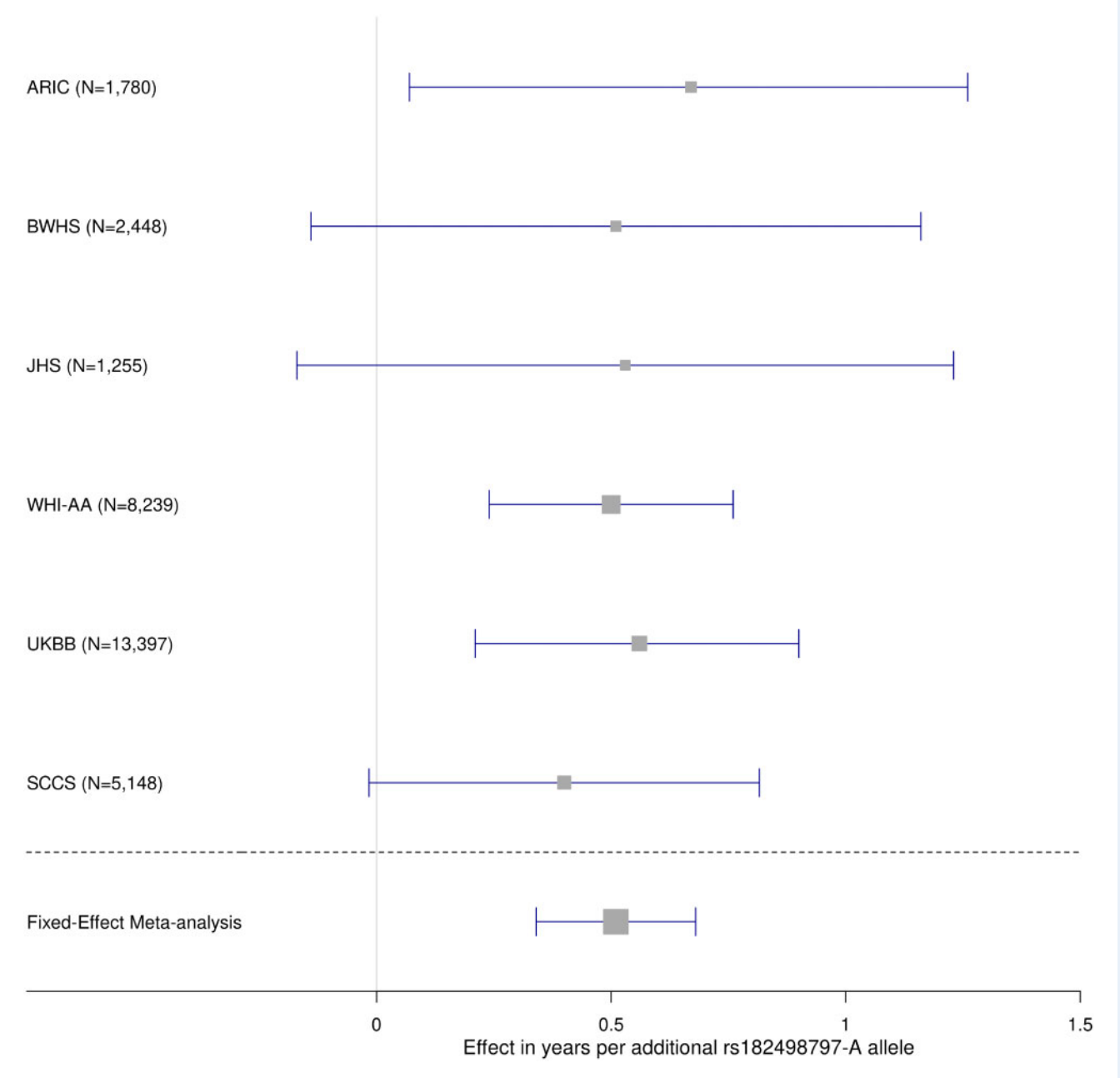

Figure 1. Forest-plot displaying the study-specific results for the new variant passing the genome-wide threshold (rs182498797-A) at 10p15, which was more common in African-ancestry participants. ARIC, The Atherosclerosis Risk in Communities Study; BWHS, Black Women's Health Study; JHS, The Jackson Heart Study; WHI, The Women's Health Initiative; UKBB, The UK Biobank; SCCS, Southern Community Cohort Study (one-sided test, $P = 0.03$.).

ReproGen and UKBB than in European populations, and/or could be the result of different LD patterns in Europeans compared with other population groups such as at the 6q22 locus.

The novel genome-wide significant signal identified at 10p15, rs182498797-A (MAF = 0.008, $P = 1.7 \times 10^{-8}$), lies in the intron 7 of the aldo-keto reductase Family 1, member C4 (*AKR1C4*) gene. This gene encodes a member of the aldo/keto reductase (AKR) superfamily. The AKR1C subfamily includes four human enzymes (AKR1C1-AKR1C4), which share high percentages of amino-acid identities (84–98%). AKR1C enzymes are expressed in different tissues, while AKR1C4 is mainly liver specific. Human AKRs (AKR1C1–AKR1C4 and AKR1D1) play essential roles in the metabolism of all steroid hormones (Rizner and Penning, 2014). Indeed, these enzymes regulate the activity/metabolism of androgens, oestrogens and progesterone, and the occupancy and transactivation of the corresponding receptors (Penning *et al.*, 2000). Genetic variants

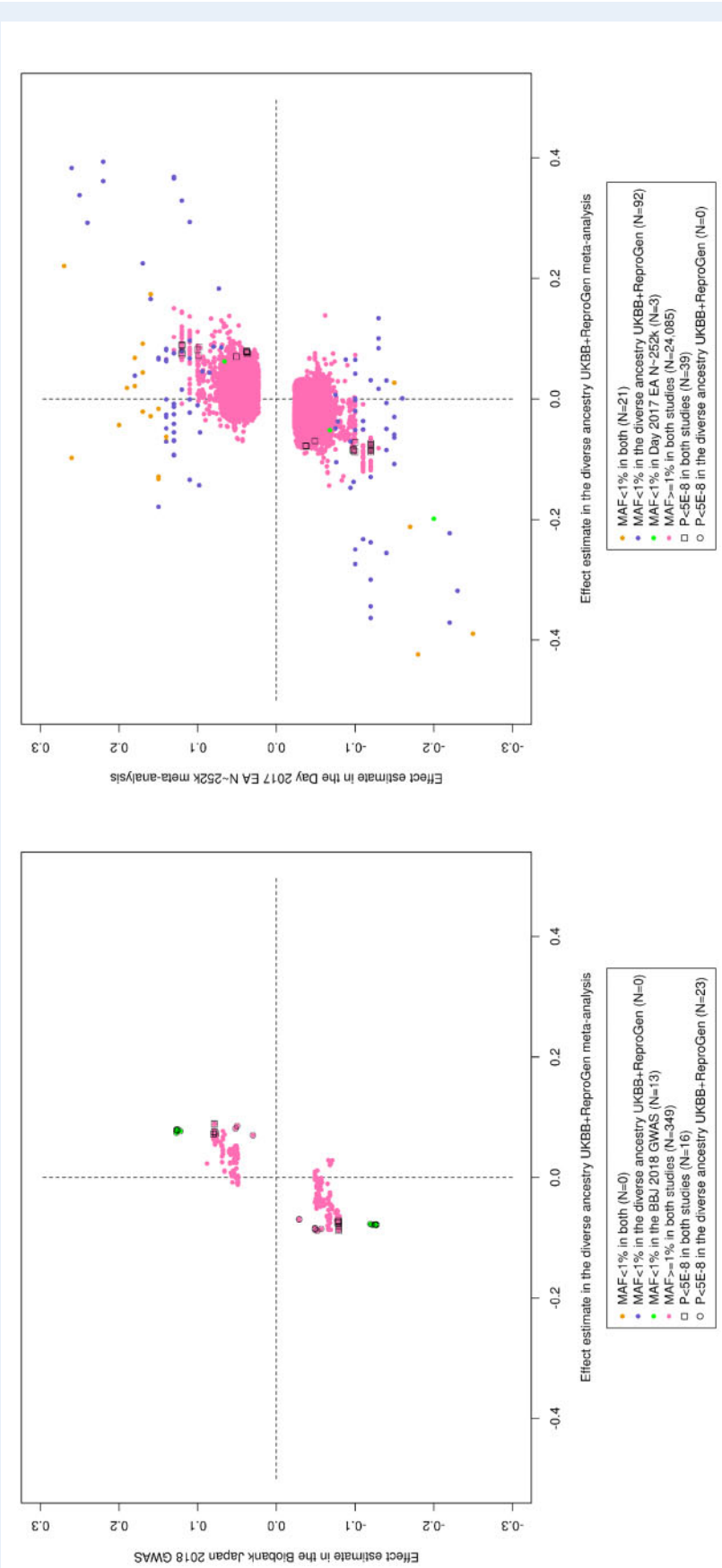

Figure 2. Comparison of effect sizes for genetic variants passing the genome-wide threshold ($P = 5 \times 10^{-8}$). The meta-analysis of age at menarche (AAM) in 38 546 women from the diverse ancestry samples of the ReproGen Consortium and the UK Biobank was compared with (**A**) the Horikoshi *et al.* (2018) Biobank Japan AAM GWAS for those present in both studies (plot on the left), and (**B**) the Day *et al.* (2017) EA (N~252k) AAM GWAS for those present in both studies (plot on the right). Genetic variants were coloured according to the minor allele frequency (MAF) in each study. On the plot on the left, the black outline circle indicates that the genetic variant was genome-wide significant in the diverse ancestry UKBB+ReproGen AAM meta-analysis. The black outline square indicates that the genetic variant was genome-wide significant in both the diverse ancestry UKBB+ReproGen AAM meta-analysis and the Horikoshi *et al.* (2018) Biobank Japan AAM GWAS. All the other variants were genome-wide significant in the Horikoshi *et al.* (2018) Biobank Japan AAM GWAS. On the plot on the right, the black outline square indicates that the genetic variant was genome-wide significant in both the diverse ancestry UKBB+ReproGen AAM meta-analysis and the Day *et al.* (2017) EA (N~252k) AAM GWAS. All the other variants were genome-wide significant in the Day *et al.* (2017) EA (N~252k) AAM GWAS.

in *AKR1C4* have been reported to be associated with blood metabolite ratios (Shin *et al.*, 2014), triglycerides levels (Willer *et al.*, 2013; Spracklen *et al.*, 2017; Hoffmann *et al.*, 2018; Ripatti *et al.*, 2020) and haemoglobin levels (Oskarsson *et al.*, 2020) in European-ancestry or trans-ethnic GWAS.

The products of AKR activity have been implicated in prostate disease (Stanbrough *et al.*, 2006), breast cancer (Lewis *et al.*, 2004; Lord *et al.*, 2005), obesity (Blouin *et al.*, 2005), polycystic ovary disease (Qin *et al.*, 2006) and delayed onset of puberty in humans (Rittner *et al.*, 1997). Altered expression of individual *AKR1C* genes is related to the development of prostate, breast, endometrial and cervical cancers as well as endometriosis. AKR1C enzymes (AKR1C1-AKR1C3) are involved in processes (disturbed prostaglandins, oestrogen and retinoid metabolism and actions) that are implicated in the pathophysiology of uterine diseases (endometrial and cervical cancers, uterine myoma and endometriosis) (Rizner, 2012). One recent GWAS identified *AKR1C3* as a novel epithelial ovarian cancer locus in women of African ancestry (Manichaikul *et al.*, 2020). A recent and large UKBB GWAS of testosterone levels and related sex hormone traits reported several genetic variants in the region of *AKR1C4* (rs79717793 and rs7475279) associated with testosterone and sex hormone-binding globulin measurement (Ruth *et al.*, 2020).

Quantitative trait loci (QTL) for important pig reproductive traits (age of puberty, nipple number and ovulation rate and plasma follicle-stimulating hormone) have been identified on pig chromosome 10q near the telomere, which is homologous to human chromosome 10p15 and contains an AKR gene cluster (Nonneman and Rohrer, 2003). One study of pigs found that two genetic variants in *AKR1C4* were significantly associated with nipple number and another was possibly associated with age at puberty (Nonneman *et al.*, 2006).

We identified suggestive associations that were more common in participants of African ancestry at 8p11 and Xq21. The Xq21 lead variants that were identified in the UKBB lie at 840 kb and 14 kb respectively from and tag the *DIAPH2* gene (Supplementary Figs. S9 and S10), which plays a role in the development and normal function of the ovaries. Defects in *DIAPH2* have been linked to premature ovarian Failure 2 (Genesio *et al.*, 2015). *DIAPH2* has been reported to have a critical role in pubertal and reproductive deficiencies in humans (Jedidi *et al.*, 2019). The lead variant at 8p11 lies at 18 kb from and tags the *ADAM2* gene (Supplementary Fig. 12). The encoded protein is a subunit of an integral sperm membrane glycoprotein called fertilin, which plays an important role in sperm-oocyte interactions (Sabetian *et al.*, 2014; Sun *et al.*, 2019). This gene has also been associated with fertility phenotypes in mice (Cho *et al.*, 1998; Nishimura *et al.*, 2001).

The strengths of our study are the large sample size and the ancestral diversity of participants, with approximately 50% of participants of African ancestry, the population with broadest diversity. Limitations include the fact that we analysed only single nucleotide variants and restricted our analyses to variants with a MAF greater or equal to 0.5% and to participants with an AAM ranging between 9 and 18 years. Additionally, women may not fully recall their AAM as most of the studies were conducted many years later. Further studies with whole genome sequence data or using newer multi-ancestry imputation panels such as TOPMed (Taliun *et al.*, 2021) will have the potential in the future to identify rare variation influencing AAM. Finally, by pooling all population groups in the UKBB to maximise the sample size, we may have missed some ancestry-specific signals that are more common in all population groups. Further studies in women with diverse and predominantly non-European ancestry are needed to confirm and extend our findings. However, the availability of such replication samples with AAM information is limited.

In conclusion, our large-scale trans-ethnic meta-analysis of AAM in women of diverse and predominantly non-European ancestry identified a new locus associated with AAM in African-ancestry participants. *AKR1C4* is a strong candidate gene for puberty timing and related disorders. We also generalised the associations in four known GWAS loci. These findings demonstrate that expanding GWAS studies to a broader range of ancestries may improve the ability to identify new genetic variants associated with complex diseases or traits and to generalise variants from European-ancestry studies to a wider range of world populations.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

The summary statistics from the association analyses that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgements

The data and materials included in this report result from a collaboration between the following studies.

ARIC

The Atherosclerosis Risk in Communities study was funded in whole or in part with Federal funds from the National Heart, Lung and Blood Institute, National Institutes of Health, Department of Health and Human Services (contract numbers HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700004I and HHSN268201700005I), R01HL087641, R01HL059367 and R01HL086694 from the National Human Genome Research Institute contract U01HG004402, and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.

BMDCS

Funding was from the National Institutes of Health (Grant no. R01 HD588860), Eunice Kennedy Shriver National Institute of Child Health and Human Development (Grant nos. N01-HD-1-3228, -3329, -3330, -3331, -3332, -3333), the Clinical and Translational Science Awards Program (Grant no. 8 UL1 TR0000770), and NIH/NICHD grant 1K99HD099330-01 (to D.C.) We appreciate the dedication of the study participants and their families, and the support of Dr. Karen

Winer, Scientific Director of the Bone Mineral Density in Childhood Study.

We acknowledge the contribution of the Bone Mineral Density in Childhood Group (BMDCS) and particularly the following study group members: Heidi J. Kalkwarf, John A. Shepherd, Joan M. Lappe, Sharon E. Oberfield and Vicente Gilsanz.

BWHS

The Black Women's Health Study was funded by grants from the National Cancer Institute, National Institutes of Health U01CA164974, R01CA098663 and P01CA151135. The views expressed in their manuscript are those of the authors and do not necessarily represent the views of the National Cancer Institute, the National Institutes of Health or the U.S. Department of Health and Human Services. The funders had no role in the design and conduct of the study, in the collection, analysis and interpretation of the data, and in the preparation, review or approval of the manuscript. We thank the BWHS staff and study participants for their important contributions.

CHOP

CHOP was financially supported by an Institute Development Award from the Children's Hospital of Philadelphia, a Research Development Award from the Cotswold Foundation, NIH grant R01 HD056465 and the Daniel B. Burke Endowed Chair for Diabetes Research (SFAG).

CHOP authors thank the network of primary care clinicians and the patients and families for their contribution to this project and to clinical research facilitated by the Pediatric Research Consortium (PeRC) at The Children's Hospital of Philadelphia. R. Chiavacci, E. Dabaghyan, A. (Hope) Thomas, K. Harden, A. Hill, C. Johnson-Honesty, C. Drummond, S. Harrison, F. Salley, C. Gibbons, K. Lilliston, C. Kim, E. Frackelton, G. Otieno, K. Thomas, C. Hou, K. Thomas and M.L. Garris provided expert assistance with genotyping and/or data collection and management. The authors would also like to thank S. Kristinsson, L.A. Hermannsson and A. Krisbjörnsson of Raförninn for extensive software design and contribution. Dr Cousminer currently works at GSK.

JHS

The Jackson Heart Study (JHS) was supported and conducted in collaboration with Jackson State University (HHSN268201800013I), Tougaloo College (HHSN268201800014I), the Mississippi State Department of Health (HHSN268201800015I) and the University of Mississippi Medical Center (HHSN268201800010I, HHSN268201800011I and HHSN268201800012I), and contracts from the National Heart, Lung, and Blood Institute (NHLBI) and the National Institute on Minority Health and Health Disparities (NIMHD). The authors also wish to thank the staff and participants of the JHS. The views expressed in their manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute, the National Institutes of Health or the U.S. Department of Health and Human Services. The funders had no role in the design and conduct of the study, in the collection, analysis and interpretation of the data, and in the preparation, review or approval of the manuscript.

HCHS/SOL

The Hispanic Community Health Study/Study of Latinos was carried out as a collaborative study supported by contracts from the National Heart, Lung and Blood Institute (NHLBI) to the University of North Carolina (N01-HC65233), University of Miami (N01-HC65234), Albert Einstein College of Medicine (N01-HC65235), Northwestern University (N01-HC65236) and San Diego State University (N01-HC65237). The following contributed to the HCHS/SOL through a transfer of funds to the NHLBI: National Center on Minority Health and Health Disparities, the National Institute of Deafness and Other Communications Disorders, the National Institute of Dental and Craniofacial Research (NIDCR), the National Institute of Diabetes and Digestive and Kidney Diseases, the National Institute of Neurological Disorders and Stroke and the Office of Dietary Supplements. The Genetic Analysis Center at the University of Washington was supported by NHLBI and NIDCR contracts (HHSN268201300005C AM03 and MOD03).

WHI

The WHI program was funded by the National Heart, Lung and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201600018C, HHSN268201600001C, HHSN268201600002C, HHSN268201600003C and HHSN268201600004C. 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: https://www-whi-org.s3.us-west-2.amazonaws.com/wp-content/uploads/WHI-Investigator-Long-List.pdf.

UK Biobank

This research was conducted using the UK Biobank Resource under Application Number 42614.

SCCS

This research was funded by National Institutes of Health grants (R01 CA202981 and U01 CA202979).

Authors' roles

Participation in study design: C.S., N.F., K.L.L., J.M.M. Analysis: C.S., D.L.C., N.F., L.M.R., G.J., J.P.B. and K.L.L. Manuscript drafting: C.S., K.L.L. and J.M.M. Critical review and manuscript comments: all authors.

Funding

Funding was provided by CHARGE Consortium grant R01HL105756-07: Gene Discovery For CVD and Aging Phenotypes. Work was also partly funded by the NIH grant U24AG051129 awarded by the National Institute on Aging (NIA).

Conflict of interest

The authors have no conflict of interest to declare.

References

Anonymous. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol* 1989;**4**:687–702.

Anonymous. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. *Control Clin Trials* 1998;**1**:61–109.

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.

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.

Blouin K, Blanchette S, Richard C, Dupont P, Luu-The V, Tchernof A. Expression and activity of steroid aldoketoreductases 1C in omental adipose tissue are positive correlates of adiposity in women. *Am J Physiol Endocrinol Metab* 2005;**288**:E398–E404.

Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J. *et al*. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018;**562**:203–209.

Canoy D, Beral V, Balkwill A, Wright FL, Kroll ME, Reeves GK, Green J, Cairns BJ. Million Women Study Collaborators. Age at menarche and risks of coronary heart and other vascular diseases in a large UK cohort. *Circulation* 2015;**131**:237–244.

Carpenter MA, Crow R, Steffes M, Rock W, Heilbraun J, Evans G, Skelton T, Jensen R, Sarpong D. Laboratory, reading center, and coordinating center data management methods in the Jackson Heart Study. *Am J Med Sci* 2004;**328**:131–144.

Carty CL, Spencer KL, Setiawan VW, Fernandez-Rhodes L, Malinowski J, Buyske S, Young A, Jorgensen NW, Cheng I, Carlson CS. *et al*. Replication of genetic loci for ages at menarche and menopause in the multi-ethnic Population Architecture using Genomics and Epidemiology (PAGE) study. *Hum Reprod* 2013;**28**:1695–1706.

Cho C, Bunch DO, Faure JE, Goulding EH, Eddy EM, Primakoff P, Myles DG. Fertilization defects in sperm from mice lacking fertilin beta. *Science* 1998;**281**:1857–1859.

Chumlea WC, Schubert CM, Roche AF, Kulin HE, Lee PA, Himes JH, Sun SS. Age at menarche and racial comparisons in US girls. *Pediatrics* 2003;**111**:110–113.

Collaborative Group on Hormonal Factors in Breast Cancer. Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. *Lancet Oncol* 2012;**11**:1141–1151.

Day FR, Elks CE, Murray A, Ong KK, Perry JR. Puberty timing associated with diabetes, cardiovascular disease and also diverse health outcomes in men and women: the UK Biobank study. *Sci Rep* 2015;**5**:11208.

Day FR, Thompson DJ, Helgason H, Chasman DI, Finucane H, Sulem P, Ruth KS, Whalen S, Sarkar AK, Albrecht E. *et al*. Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat Genet* 2017;**6**:834–841.

Demerath EW, Liu CT, Franceschini N, Chen G, Palmer JR, Smith EN, Chen CT, Ambrosone CB, Arnold AM, Bandera EV. *et al*. Genome-wide association study of age at menarche in African-American women. *Hum Mol Genet* 2013;**16**:3329–3346.

EasyQC: www.genepi-regensburg.de/easyqc/ (30 November 2020, date last accessed).

Elks CE, Ong KK, Scott RA, van der Schouw YT, Brand JS, Wark PA, Amiano P, Balkau B, Barricarte A, Boeing H, The InterAct Consortium *et al*. Age at menarche and type 2 diabetes risk: the EPIC-InterAct study. *Diabetes Care* 2013;**36**:3526–3534.

Fernandez-Rhodes L, Malinowski JR, Wang Y, Tao R, Pankratz N, Jeff JM, Yoneyama S, Carty CL, Setiawan VW, Le Marchand L. *et al*. 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* 2018;**13**:e0200486.

Freedman DS, Khan LK, Serdula MK, Dietz WH, Srinivasan SR, Berenson GS. Bogalusa heart study. The relation of menarcheal age to obesity in childhood and adulthood: the Bogalusa heart study. *BMC Pediatr* 2003;**3**:3–2431–3-3. Epub 2003 Apr 30.

Fuqua SR, Wyatt SB, Andrew ME, Sarpong DF, Henderson FR, Cunningham MF, Taylor HA. Jr. Recruiting African-American research participation in the Jackson Heart Study: methods, response rates, and sample description. *Ethn Dis* 2005;**4** Suppl 6:S6-18-29.

Galinsky KJ, Bhatia G, Loh PR, Georgiev S, Mukherjee S, Patterson NJ, Price AL. Fast principal-component analysis reveals convergent evolution of ADH1B in Europe and East Asia. *Am J Hum Genet* 2016;**98**:456–472.

Genesio R, Mormile A, Licenziati MR, De Brasi D, Leone G, Balzano S, Izzo A, Bonfiglio F, Conti A, Fioretti G. *et al*. Short stature and primary ovarian insufficiency possibly due to chromosomal position effect in a balanced X;1 translocation. *Mol Cytogenet* 2015;**8**:50–015. 0154-3. eCollection 2015.

Hoffmann TJ, Theusch E, Haldar T, Ranatunga DK, Jorgenson E, Medina MW, Kvale MN, Kwok PY, Schaefer C, Krauss RM. *et al*. A large electronic-health-record-based genome-wide study of serum lipids. *Nat Genet* 2018;**50**:401–413.

Horikoshi M, Day FR, Akiyama M, Hirata M, Kamatani Y, Matsuda K, Ishigaki K, Kanai M, Wright H, Toro CA. *et al*. Elucidating the genetic architecture of reproductive ageing in the Japanese population. *Nat Commun* 2018;**9**:1977–018-04398-z.

Jedidi I, Ouchari M, Yin Q. Sex chromosomes-linked single-gene disorders involved in human infertility. *Eur J Med Genet* 2019;**62**:103560.

Keku ERW, Taylor HA Jr, Garrison R, Wyatt SB, Richard M, Jenkins B, Reeves L, Sarpong D. Cardiovascular disease event classification in the Jackson Heart Study: methods and procedures. *Ethn Dis* 2005;**4** Suppl 6:S6-62-70.

Kowalski MH, Qian H, Hou Z, Rosen JD, Tapia AL, Shan Y, Jain D, Argos M, Arnett DK, Avery C. *et al*. Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. *PLoS Genet* 2019;**12**:e1008500.

Komura H, Miyake A, Chen CF, Tanizawa O, Yoshikawa H. Relationship of age at menarche and subsequent fertility. *Eur J Obstet Gynecol Reprod Biol* 1992;**44**:201–203.

Kotsopoulos J, Lubinski J, Lynch HT, Neuhausen SL, Ghadirian P, Isaacs C, Weber B, Kim-Sing C, Foulkes WD, Gershoni-Baruch R. *et al*. Age at menarche and the risk of breast cancer in BRCA1 and BRCA2 mutation carriers. *Cancer Causes Control* 2005;**16**:667–674.

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.

Laurie CC, Doheny KF, Mirel DB, Pugh EW, Bierut LJ, Bhangale T, Boehm F, Caporaso NE, Cornelis MC, Edenberg HJ, for the GENEVA Investigators *et al*. Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet Epidemiol* 2010;**34**:591–602.

Lewis MJ, Wiebe JP, Heathcote JG. Expression of progesterone metabolizing enzyme genes (AKR1C1, AKR1C2, AKR1C3, SRD5A1, SRD5A2) is altered in human breast carcinoma. *BMC Cancer* 2004; **4**: 27.

Lord SJ, Mack WJ, Van Den Berg D, Pike MC, Ingles SA, Haiman CA, Wang W, Parisky YR, Hodis HN, Ursin G. Polymorphisms in genes involved in estrogen and progesterone metabolism and mammographic density changes in women randomized to postmenopausal hormone therapy: results from a pilot study. *Breast Cancer Res* 2005;**7**:R336–R344.

Magi R, Horikoshi M, Sofer T, Mahajan A, Kitajima H, Franceschini N, McCarthy MI, Cogent-Kidney Consortium TC, Morris AP, COGENT-Kidney Consortium, T2D-GENES Consortium. Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum Mol Genet* 2017;**26**:3639–3650.

Manichaikul A, Peres LC, Wang XQ, Barnard ME, Chyn D, Sheng X, Du Z, Tyrer J, Dennis J, Schwartz AG, the African American Breast Cancer Consortium (AABC). *et al*. Identification of novel epithelial ovarian cancer loci in women of African ancestry. *Int J Cancer* 2020;**146**:2987–2998.

METAL: http://csg.sph.umich.edu/abecasis/metal/ (30 November 2020, date last accessed).

Morris DH, Jones ME, Schoemaker MJ, Ashworth A, Swerdlow AJ. Familial concordance for age at menarche: analyses from the Breakthrough Generations Study. *Paediatr Perinat Epidemiol* 2011; **25**:306–311.

MR-MEGA: https://genomics.ut.ee/en/tools/mr-mega/ (13 February 2021, date last accessed).

Nishimura H, Cho C, Branciforte DR, Myles DG, Primakoff P. Analysis of loss of adhesive function in sperm lacking cyritestin or fertilin beta. *Dev Biol* 2001;**233**:204–213.

Nonneman DJ, Rohrer GA. Comparative mapping of a region on chromosome 10 containing QTL for reproduction in swine. *Anim Genet* 2003;**34**:42–46.

Nonneman DJ, Wise TH, Ford JJ, Kuehn LA, Rohrer GA. Characterization of the aldo-keto reductase 1C gene cluster on pig chromosome 10: possible associations with reproductive traits. *BMC Vet Res* 2006;**2**:28.

O'Connell J, Gurdasani D, Delaneau O, Pirastu N, Ulivi S, Cocca M, Traglia M, Huang J, Huffman JE, Rudan I. *et al*. A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet* 2014;**10**:e1004234.

O'Connell J, Sharp K, Shrine N, Wain L, Hall I, Tobin M, Zagury JF, Delaneau O, Marchini J. Haplotype estimation for biobank scale datasets. *Nat Genet* 2016;**48**:817–820.

Oskarsson GR, Oddsson A, Magnusson MK, Kristjansson RP, Halldorsson GH, Ferkingstad E, Zink F, Helgadottir A, Ivarsdottir EV, Arnadottir GA. *et al*. Predicted loss and gain of function mutations in ACO1 are associated with erythropoiesis. *Commun Biol* 2020;**3**:189.

Palmer JR, Ruiz-Narvaez EA, Rotimi CN, Cupples LA, Cozier YC, Adams-Campbell LL, Rosenberg L. Genetic susceptibility loci for subtypes of breast cancer in an African American population. *Cancer Epidemiol Biomarkers Prev* 2013;**22**:127–134.

Pan-UK Biobank: https://pan.ukbb.broadinstitute.org/ (13 February 2021, date last accessed).

Penning TM, Burczynski ME, Jez JM, Hung CF, Lin HK, Ma H, Moore M, Palackal N, Ratnam K. Human 3alpha-hydroxysteroid dehydrogenase isoforms (AKR1C1-AKR1C4) of the aldo-keto reductase superfamily: functional plasticity and tissue distribution reveals roles in the inactivation and formation of male and female sex hormones. *Biochem J* 2000;**351**:67–77.

Qin K, Ehrmann DA, Cox N, Refetoff S, Rosenfield RL. Identification of a functional polymorphism of the human type 5 17beta-hydroxysteroid dehydrogenase gene associated with polycystic ovary syndrome. *J Clin Endocrinol Metab* 2006;**91**:270–276.

ReproGen Consortium: https://www.reprogen.org/ (30 November 2020, date last accessed).

Ripatti P, Ramo JT, Mars NJ, Fu Y, Lin J, Soderlund S, Benner C, Surakka I, Kiiskinen T, Havulinna AS. *et al*. Polygenic hyperlipidemias and coronary artery disease risk. *Circ Genom Precis Med* 2020; **2**:e002725.

Rittner HL, Lee PD, Blum WF, Doerr HG, Steiss J, Kreuder J, Rascher W, Kiess W. Developmental patterns of serum 3 alpha-androstanediol glucuronide. *J Endocrinol Invest* 1997;**20**:138–143.

Rizner TL. Enzymes of the AKR1B and AKR1C subfamilies and uterine diseases. *Front Pharmacol* 2012;**3**:34.

Rizner TL, Penning TM. Role of aldo-keto reductase family 1 (AKR1) enzymes in human steroid metabolism. *Steroids* 2014;**79**:49–63.

Ruth KS, Day FR, Tyrrell J, Thompson DJ, Wood AR, Mahajan A, Beaumont RN, Wittemans L, Martin S, Busch AS, The Endometrial Cancer Association Consortium *et al*. Using human genetics to understand the disease impacts of testosterone in men and women. *Nat Med* 2020;**26**:252–258.

Sabetian S, Shamsir MS, Naser MA. Abu Naser M. Functional features and protein network of human sperm-egg interaction. *Syst Biol Reprod Med* 2014;**60**:329–337.

Scalable and Accurate Implementation of Generalized mixed model (SAIGE): https://github.com/weizhouUMICH/SAIGE/ (30 November 2020, date last accessed).

Shin S-Y, Fauman EB, Petersen A-K, Krumsiek J, Santos R, Huang J, Arnold M, Erte I, Forgetta V, Yang T-P, The Multiple Tissue Human Expression Resource (MuTHER) Consortium *et al*. An atlas of genetic influences on human blood metabolites. *Nat Genet* 2014;**46**:543–550.

Signorello LB, Hargreaves MK, Steinwandel MD, Zheng W, Cai Q, Schlundt DG, Buchowski MS, Arnold CW, McLaughlin JK, Blot WJ. Southern community cohort study: establishing a cohort to investigate health disparities. *J Natl Med Assoc* 2005;**7**:972–979.

Sorlie PD, Aviles-Santa LM, Wassertheil-Smoller S, Kaplan RC, Daviglus ML, Giachello AL, Schneiderman N, Raij L, Talavera G, Allison M. *et al*. Design and implementation of the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol* 2010; **20**:629–641.

Spencer KL, Malinowski J, Carty CL, Franceschini N, Fernandez-Rhodes L, Young A, Cheng I, Ritchie MD, Haiman CA, Wilkens L. *et al*. Genetic variation and reproductive timing: African American women from the Population Architecture using Genomics and Epidemiology (PAGE) Study. *PLoS One* 2013;**2**:e55258.

Spracklen CN, Chen P, Kim YJ, Wang X, Cai H, Li S, Long J, Wu Y, Wang YX, Takeuchi F. *et al*. Association analyses of East Asian individuals and trans-ancestry analyses with European individuals reveal new loci associated with cholesterol and triglyceride levels. *Hum Mol Genet* 2017;**9**:1770–1784.

Stanbrough M, Bubley GJ, Ross K, Golub TR, Rubin MA, Penning TM, Febbo PG, Balk SP. Increased expression of genes converting adrenal androgens to testosterone in androgen-independent prostate cancer. *Cancer Res* 2006;**66**:2815–2825.

Sun TC, Wang JH, Wang XX, Liu XM, Zhang CL, Hao CF, Ma WZ, Deng SL, Liu YX. Effects of sperm proteins on fertilization in the female reproductive tract. *Front Biosci (Landmark Ed)* 2019;**24**: 735–749.

Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, Taliun SAG, Corvelo A, Gogarten SM, Kang HM. *et al*. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* 2021;**7845**:290–299.

Taylor HA Jr, Wilson JG, Jones DW, Sarpong DF, Srinivasan A, Garrison RJ, Nelson C, Wyatt SB. Toward resolution of cardiovascular health disparities in African Americans: design and methods of the Jackson Heart Study. *Ethn Dis* 2005;**4** Suppl 6:S6-4-17.

UK Biobank: https://www.ukbiobank.ac.uk/ (30 November 2020, date last accessed).

Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, Chamberlain AM, Chang AR, Cheng S, Delling FN. *et al*. Heart Disease and Stroke Statistics – 2020 Update: a report from the American Heart Association. *Circulation* 2020;**9**:e139–e596.

Voight BF, Kang HM, Ding J, Palmer CD, Sidore C, Chines PS, Burtt NP, Fuchsberger C, Li Y, Erdmann J. *et al*. The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet* 2012;**8**:e1002793.

Weghofer A, Kim A, Barad DH, Gleicher N. Age at menarche: a predictor of diminished ovarian function? *Fertil Steril* 2013;**100**: 1039–1043.

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

Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, Mora S. *et al*. Discovery and refinement of loci associated with lipid levels. *Nat Genet* 2013;**11**: 1274–1283.

Wilson JG, Rotimi CN, Ekunwe L, Royal CD, Crump ME, Wyatt SB, Steffes MW, Adeyemo A, Zhou J, Taylor HA Jr. *et al*. Study design for genetic analysis in the Jackson Heart Study. *Ethn Dis* 2005;**4** Suppl 6:S6-30-37.

Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, Magi R, Ferreira T, Fall T, Graff M, Justice AE, Genetic Investigation of Anthropometric Traits (GIANT) Consortium *et al*. Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc* 2014;**9**:1192–1212.

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.

Zemel BS, Kalkwarf HJ, Gilsanz V, Lappe JM, Oberfield S, Shepherd JA, Frederick MM, Huang X, Lu M, Mahboubi S. *et al*. Revised reference curves for bone mineral content and areal bone mineral density according to age and sex for black and non-black children: results of the bone mineral density in childhood study. *J Clin Endocrinol Metab* 2011;**96**:3160–3169.

Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, LeFaive J, VandeHaar P, Gagliano SA, Gifford A. *et al*. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet* 2018; **50**:1335–1341.