

ASSOCIATION STUDIES ARTICLE

Genome-wide association study of anti-Müllerian hormone levels in pre-menopausal women of late reproductive age and relationship with genetic determinants of reproductive lifespan

Katherine S. Ruth^{1,†}, Ana Luiza G. Soares^{2,3}, Maria-Carolina Borges^{2,3}, A. Heather Eliassen⁴, Susan E. Hankinson^{4,5}, Michael E. Jones⁶, Peter Kraft⁷, Hazel B. Nichols⁸, Dale P. Sandler⁹, Minouk J. Schoemaker⁶, Jack A. Taylor¹⁰, Anne Zeleniuch-Jacquotte¹¹, Deborah A. Lawlor^{2,3,12}, Anthony J. Swerdlow^{6,13} and Anna Murray^{1,*}

¹Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, UK, ²MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, UK, ³Population Health Science, Bristol Medical School, University of Bristol, Bristol, UK, ⁴Department of Epidemiology, Harvard T.H. Chan School of Public Health, and Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA, ⁵Department of Biostatistics and Epidemiology, School of Public Health and Health Sciences, University of Massachusetts, Amherst, MA, USA, ⁶Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK, ⁷Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA, ⁸Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA, ⁹Epidemiology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA, ¹⁰Epigenetic and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA, ¹¹Department of Population Health and Perlmutter Cancer Center, New York University School of Medicine, New York, NY, USA, ¹²National Institute for Health Research Bristol Biomedical Research Centre, University Hospitals Bristol NHS Foundation Trust and University of Bristol, Bristol, UK and ¹³Division of Breast Cancer Research, The Institute of Cancer Research, London, UK

*To whom correspondence should be addressed at: Genetics of Complex Traits, University of Exeter Medical School, RILD Level 3, Royal Devon & Exeter Hospital, Barrack Road, Exeter, EX2 5DW, UK. Tel: +44 (0)1392408257; Email: A.Murray@exeter.ac.uk

Abstract

Anti-Müllerian hormone (AMH) is required for sexual differentiation in the fetus, and in adult females AMH is produced by growing ovarian follicles. Consequently, AMH levels are correlated with ovarian reserve, declining towards menopause when the oocyte pool is exhausted. A previous genome-wide association study identified three genetic variants in and around the

[†]Katherine S. Ruth, <http://orcid.org/0000-0003-4966-9170>

Received: October 29, 2018. Revised: January 8, 2019. Accepted: January 10, 2019

© The Author(s) 2019. Published by Oxford University Press. All rights reserved.

For Permissions, please email: journals.permissions@oup.com

AMH gene that explained 25% of variation in AMH levels in adolescent males but did not identify any genetic associations reaching genome-wide significance in adolescent females. To explore the role of genetic variation in determining AMH levels in women of late reproductive age, we carried out a genome-wide meta-analysis in 3344 pre-menopausal women from five cohorts (median age 44–48 years at blood draw). A single genetic variant, rs16991615, previously associated with age at menopause, reached genome-wide significance at $P = 3.48 \times 10^{-10}$, with a per allele difference in age-adjusted inverse normal AMH of 0.26 standard deviations (SD) (95% confidence interval (CI) [0.18,0.34]). We investigated whether genetic determinants of female reproductive lifespan were more generally associated with pre-menopausal AMH levels. Genetically-predicted age at menarche had no robust association but genetically-predicted age at menopause was associated with lower AMH levels by 0.18 SD (95% CI [0.14,0.21]) in age-adjusted inverse normal AMH per one-year earlier age at menopause. Our findings provide genetic support for the well-established use of AMH as a marker of ovarian reserve.

Introduction

Anti-Müllerian hormone (AMH) is a member of the transforming growth factor-beta superfamily that regulates the growth and development of ovarian follicles in females and is required for sexual differentiation in the fetus, causing regression of the Müllerian ducts during testis development (1). In males, AMH is required for testes development and function and levels increase rapidly shortly after birth, peaking at 6 months of age, and then decline to low levels in during puberty (2). In women, AMH is produced by the granulosa cells of growing follicles and levels are correlated with the number of growing follicles and are used as a clinical measure of ovarian reserve (3). AMH levels increase in women from birth until their 20s, before declining gradually with age until levels are undetectable after menopause when ovarian reserve is exhausted (1,3–6). Since AMH levels are stable throughout the menstrual cycle, they can be used as a measure of fertility in women of late reproductive age and to predict response to fertility treatment (7).

AMH levels vary widely between women and genetic variation is thought to be important, though few genetic studies have been conducted. Rare AMH mutations have been found with functional effects on AMH signalling (8,9), while polymorphisms in AMH or the gene coding for its receptor, AMHR2, have been associated with response to ovarian stimulation, infertility, follicle recruitment, primary ovarian insufficiency and polycystic ovary syndrome in candidate gene studies (10).

A previous genome-wide association study (GWAS) in 1360 adolescent males and 1455 adolescent females from a single cohort identified three genetic variants in and around the AMH gene that were independently associated with higher levels of AMH in adolescent males ($P = 2 \times 10^{-49}$ to $P = 3 \times 10^{-8}$ for each variant when jointly included in the regression model) (11). None of these variants showed strong evidence of statistical association in adolescent females ($P = 8 \times 10^{-4}$ to $P = 0.9$ for each variant when jointly included in the regression model), with considerably weaker effect estimates than in males. For all three variants there was strong statistical evidence of a sex difference ($P_{\text{HET}} = 3 \times 10^{-4}$ to $P_{\text{HET}} = 6 \times 10^{-12}$), with the three cumulatively explaining 24.5% of the variation in AMH levels in males compared with 0.8% in females. No cohorts were available for replication of this initial study and it is unknown whether the weak or absent association in adolescent females persists into older ages, as would be expected since differences in ovarian decline result in variation in AMH levels between women.

We undertook a GWAS meta-analysis of 3344 women from five cohorts—the Generations Study, Sister Study, Nurses' Health Study, Nurses' Health Study II and Avon Longitudinal Study of Parents and Children (ALSPAC) – to investigate genetic determinants of AMH levels in pre-menopausal women of late reproductive age (median age at blood draw 44–48 years). We aimed to

identify novel genetic variants associated with AMH levels and to explore the effects of published genetic variants associated with AMH levels in previous GWAS.

Results

AMH is associated with a single significant signal in a known menopause locus

In our genome-wide meta-analysis (Table 1), a single genetic variant in the MCM8 gene at 20p12.3 reached genome-wide significance at $P < 5 \times 10^{-8}$ (rs16991615, $P = 3.48 \times 10^{-10}$) (Fig. 1, Supplementary Material, Fig. S1). Within each of the five genotyped cohorts, we inverse-normally transformed AMH (to ensure normality of the residuals in the association analysis) and tested the association of over 11 million autosomal genetic variants imputed to HRC r1.1 2016 (12) adjusted for age and genetic relatedness (13). We performed inverse variance weighted (IVW) meta-analyses of the genome-wide results from the five cohorts, filtering our results to include only variants present in three or more of the five cohorts analysed, resulting in a total of 8.4 million variants in the final results dataset (summary statistics to be made available through the EMBL-EBI GWAS catalogue at <https://www.ebi.ac.uk/gwas/summary-statistics>). A total of 242 variants had $P < 1 \times 10^{-5}$, resulting in 24 signals following distance-based clumping of variants within 500 kb, with the top 10 signals presented in Table 2. The minor A allele of rs16991615 increased age-adjusted inverse normal AMH by 0.26 standard deviations (SD) per allele (95% confidence interval (CI) [0.18,0.34], $P = 3.48 \times 10^{-10}$) (Table 2).

Variants previously shown to be strongly associated with AMH levels in adolescent males had weak effects in pre-menopausal adult women in our study

For three genetic variants in and around the AMH gene that were previously found to be independently associated with higher levels of AMH in adolescent males (11), we estimated the effects in pre-menopausal women when the variants were jointly included in the regression model (joint model), by carrying out approximate conditional analyses using the software GCTA (14). To allow comparison between our results and the original study's estimates, we generated effect estimates for age-adjusted inverse normal AMH in the adolescent males and females from the original study sample (ALSPAC offspring), since results from the original study were unadjusted and presented in natural log-transformed AMH.

Table 1. AMH levels and age of women in each of the five cohorts included in the genome-wide analysis

Study	n	Median (interquartile range)	
		AMH (pmol/L)	Age (years) at blood draw
Generations Study	379	3.9 (0.8,11.7)	44 (40,48)
Sister Study	438	1.2 (0.1,6.0)	48 (45,51)
Nurses' Health Studies (Illumina) ¹	225	5.5 (1.5,12.4)	45 (42,48)
Nurses' Health Studies (OncoArray) ¹	417	6.7 (2.5,14.9)	44 (41,46)
ALSPAC	1885	2.0 (0.4, 5.2)	46 (44,49)
Total	3344		

¹Data from the Nurses' Health Study and Nurses' Health Study II were combined and genotyped on two different genome arrays, which were included as separate sub-samples in this analysis.

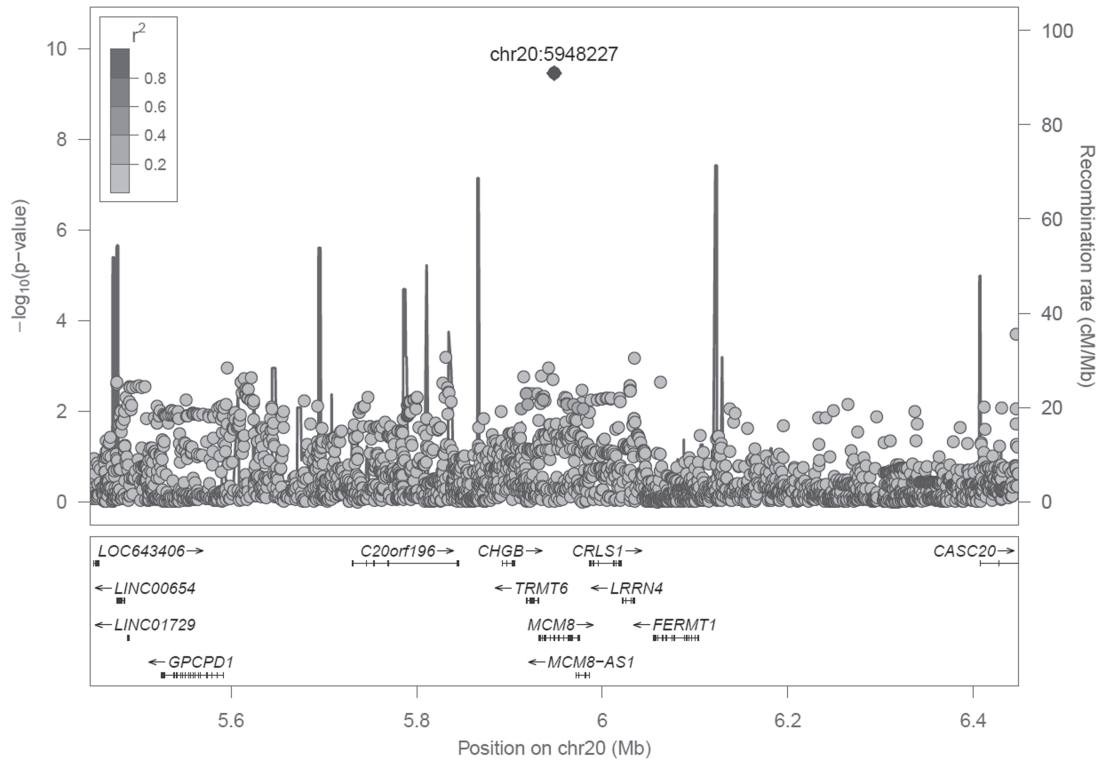


Figure 1. Association statistics with age-adjusted inverse normal AMH for variants within 500 kb of rs16991615 (chr20:5948227) showing linkage disequilibrium with the top variant. Note: Linkage disequilibrium shown is in unrelated Europeans in UK Biobank.

Table 2. Top 10 signals from the genome-wide analysis of age-adjusted inverse normal AMH in pre-menopausal women

SNPID	Chr	Pos	EA/OA/EAF	Beta (SE)	P-value	Dir.	Het. P	Imp. qual
rs16991615	20	5948227	A/G/0.068	0.26 (0.04)	3.5×10^{-10}	+++++	0.14	0.94
rs62236881	22	29450193	A/G/0.008	0.85 (0.16)	1.1×10^{-7}	++++?	0.77	0.84
rs186783371	5	88062223	T/A/0.011	0.64 (0.12)	1.4×10^{-7}	++++?	0.64	0.61
rs35829351	11	6120686	G/A/0.449	0.11 (0.02)	2.7×10^{-7}	++++-	0.02	0.89
rs10732995	1	175111574	T/C/0.966	0.29 (0.06)	3.9×10^{-7}	+++++	0.86	0.93
rs76673357	20	17831206	G/T/0.056	0.22 (0.04)	5.0×10^{-7}	+++++	0.47	0.89
rs7168070	15	93910220	T/C/0.329	0.11 (0.02)	5.5×10^{-7}	+++++	0.98	0.89
rs62237617	22	28761148	T/C/0.011	0.74 (0.15)	6.5×10^{-7}	++++?	0.89	0.82
rs141456816	17	10362307	C/T/0.013	0.50 (0.10)	6.7×10^{-7}	++++-	0.44	0.66
rs6743761	2	10393228	A/G/0.482	0.09 (0.02)	2.7×10^{-6}	++++-	0.11	0.99

Beta = difference in mean age-adjusted inverse normal AMH (SD) per allele. Chr = chromosome; Dir. = positive (+) or negative (-) direction of effect in ALSPAC, Sister Study, Generations Study, Nurses' Health Studies (OncoArray), Nurses' Health Studies (Illumina) respectively, with "?" indicating that the variant was absent; EA = effect allele; EAF = weighted average effect allele frequency across the studies; Het. P = P-value from Cochran's Q-test of heterogeneity of effects across the studies; Imp. qual = mean imputation quality across the studies; Pos = position in hg19/GRCh37; OA = other allele; SE = standard error.

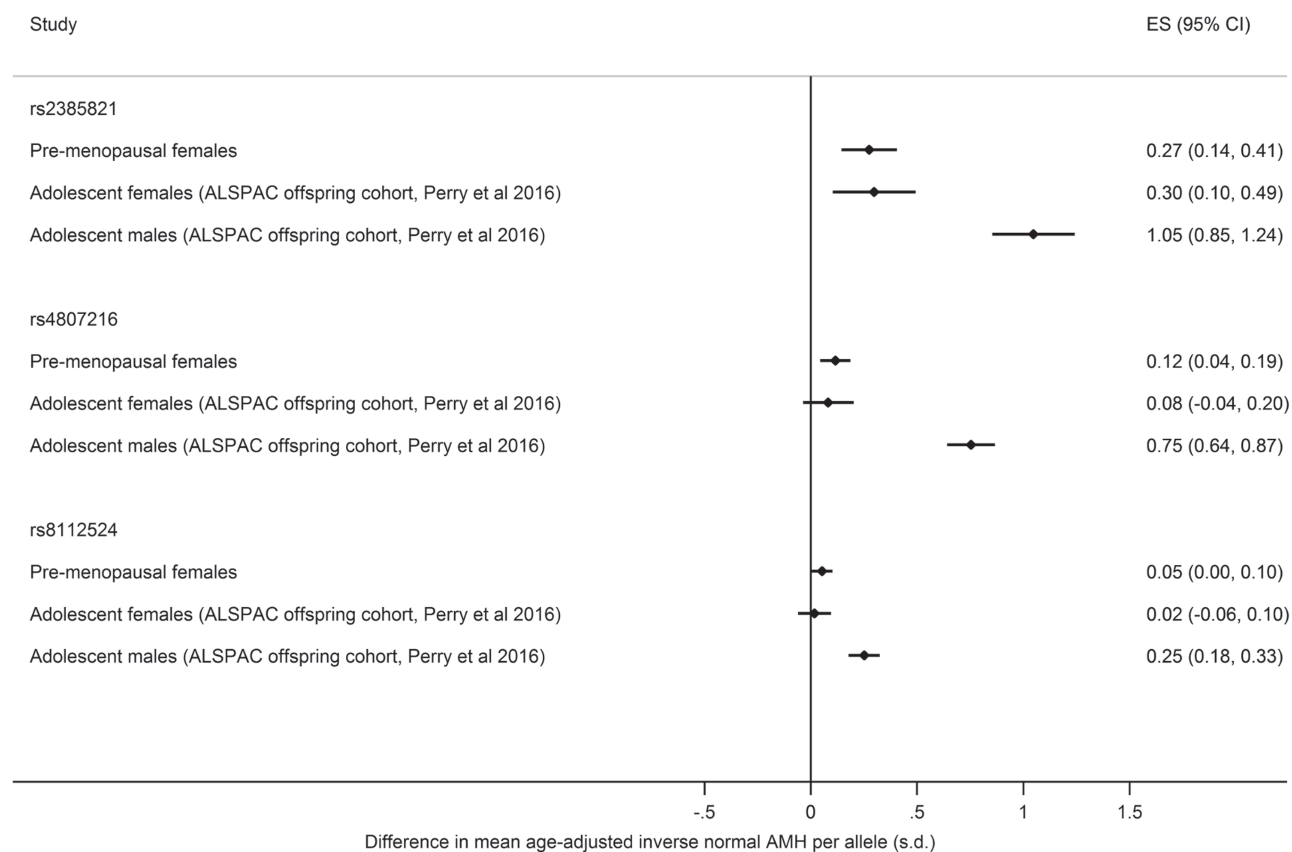


Figure 2. Comparison of ESs of three genetic variants previously associated with higher levels of AMH (11) when jointly included in the regression model: effect in adolescent males and females from the ALSPAC offspring cohort (previous study, Perry et al 2016 (11)) and pre-menopausal women (current study). Note: Effect is difference in mean AMH per allele in standard deviations of age-adjusted inverse normal AMH. ES = effect size.

The effect estimates from the joint model for the three published genetic variants were directionally concordant across adolescent males and females (ALSPAC offspring cohort) and the pre-menopausal women in the current study (from five cohorts including ALSPAC mothers), but had about one-fifth of the effect on the level of AMH compared with the effect in adolescent males ($P_{HET} < 0.001$) (Fig. 2). The weak or null effect sizes (ESs) for rs4807216, rs8112524 and rs2385821 were similar in adolescent and pre-menopausal females ($P_{HET} > 0.05$ for all). Genetic variant rs2385821 had the strongest effects in females of the three variants from the previous publication (11), but did not reach genome-wide significance in pre-menopausal females (for joint model, per allele difference in age-adjusted inverse normal AMH of 0.27 SD (95% CI [0.13,0.41]), $P = 4.0 \times 10^{-5}$) (Supplementary Material, Table S1).

Genetic variants for early menopause are associated with reduced levels of AMH

Since the only genetic variant to reach genome-wide significance in our study (rs16991615) has previously been reported as associated with menopause timing (15–17), we investigated the association of AMH levels with all 56 genetic variants associated with menopause timing (17). The effect estimates of these genetic variants on AMH level were positively correlated with the published effects on age at menopause ($r = 0.83$) and there were consistent directions of effect for 50/56 variants ($P = 1 \times 10^{-9}$ for binomial sign test; $\chi^2_{56} = 194.39$,

$P = 4 \times 10^{-17}$ for global chi-squared test of association) (Fig. 3, Supplementary Material, Table S2). There were no obvious outliers among the 56 menopause timing variants and generally variants with large effects on age at menopause also had large effects on AMH levels.

Two-sample Mendelian randomization analysis by IVW and Egger estimation supported a causal relationship between genetically-predicted age at menopause and pre-menopausal AMH level (Fig. 3, Supplementary Material, Table S3). For a one-year increase in genetically-predicted age at menopause, age-adjusted inverse normal AMH was increased by 0.18 SD (95% CI [0.14,0.21]) with no horizontal pleiotropy detected by Egger analysis (in Egger analysis, 0.20 SD (95% CI [0.13,0.27]) age-adjusted inverse normal AMH per one-year increase in genetically-predicted age at menopause, P -intercept = 0.49). This relationship remained similar even when rs16991615 was excluded from the analysis (Supplementary Material, Table S3).

Genetic variants for age at menarche are not associated with AMH levels

We investigated the effect of genetic variants associated with age at menarche on AMH levels (18), since menarche marks the start of cyclic selection of ovarian follicles from the growing follicle pool. We identified 327 of 389 published independent signals (18) in our meta-analysis results. For the 327 variants, there was little correlation between the published

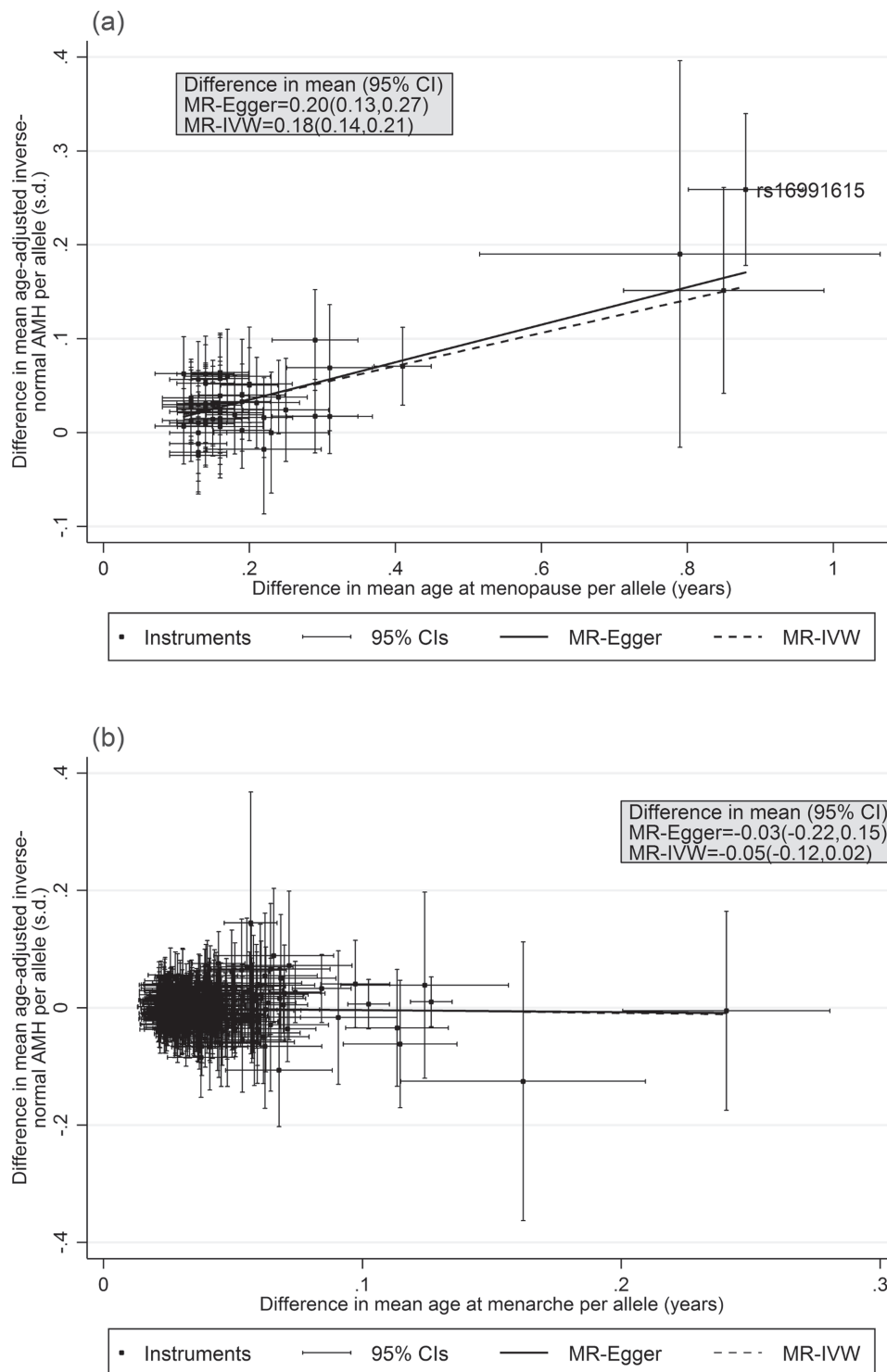


Figure 3. IVW and Egger two-sample Mendelian randomization analyses of the effect of genetically-predicted (A) age at menopause and (B) age at menarche on age-adjusted inverse normal AMH levels in pre-menopausal women. Note: Difference in mean age-adjusted inverse-normal AMH per one-year increase in genetically-predicted age at menarche or menopause.

effect on age at menarche and the effect on AMH levels ($r = -0.05$) and the directions of effects were not consistent, with 158/327 (48%) in the same direction ($P = 0.58$ for binomial sign test; $\chi^2_{327} = 328.62$, $P = 0.46$ for global chi-squared test of association) (Fig. 3, Supplementary Material, Table S4). Two-sample Mendelian randomization analysis by IVW and Egger

estimation found no causal relationship between age at menarche and AMH level (difference in mean age-adjusted inverse normal AMH per one-year increase in genetically-predicted age at menopause was for IVW, -0.05 SD (95% CI $[-0.12, 0.02]$) and for Egger, -0.03 SD (95% CI $[-0.22, 0.15]$, P -intercept = 0.87) (Fig. 3, Supplementary Table S3).

Genetic variant for follicle-stimulating hormone levels is not associated with AMH levels

Since levels of follicle-stimulating hormone (FSH) and luteinizing hormone rise around menopause, we tested the association of a genetic variant at the *FSHB* locus that affects levels of these hormones (19) with AMH levels in pre-menopausal women. The *FSHB* promoter polymorphism (rs10835638; $-211G > T$) was not associated with AMH levels (per allele difference in age-adjusted inverse normal AMH of 0.01 SD [95% CI $[-0.05, 0.07]$], $P = 0.79$).

Sensitivity analysis

Results for genetic variants with $P < 5 \times 10^{-5}$ in the main analysis were well correlated whether we adjusted for age or not ($r = 0.99$) (Supplementary Material, Fig. S2), used our favoured inverse normal transformation or a natural log transformation (as in reference (11)) ($r = 1.00$) (Supplementary Material, Fig. S3), and when we excluded women whose AMH level was imputed as it was below the lower limit of detection ($r = 0.99$) (Supplementary Material, Fig. S4).

Discussion

Our results indicate that variation in AMH levels in pre-menopausal women is contributed to by the underlying biology of ovarian reserve, as shown by the correlation between genetically-predicted age at menopause and AMH levels, supporting the use of AMH as a means of measuring ovarian reserve. The only signal passing genome-wide significance in our analyses was rs16991615 in *MCM8*, a published menopause timing variant (15–17), with the same allele associated with earlier menopause and lower AMH levels. Genome-wide analyses of menopause timing, a proxy measure for ovarian reserve, have identified 56 genetic variants and highlighted the importance of DNA damage response pathways during follicle formation *in utero* and for follicle maintenance during a woman's lifetime (17). Additionally, 389 genetic variants have been identified for menarche timing, the age at which cyclic selection of ovarian follicles from the pool of growing, AMH-producing follicles starts (18). Therefore, it is plausible that genetic determinants of menarche and menopause timing could affect ovarian reserve and influence AMH levels in pre-menopausal women, many years prior to menopause. For genetic variants associated with age at menopause, the published effect estimates were positively correlated with the effects on AMH levels and there was evidence of a causal relationship between genetically-predicted earlier menopause and lower pre-menopausal AMH levels, which remained even when rs16991615 was excluded. We did not find any association between genetically-predicted age at menarche and pre-menopausal AMH levels. We interpret these results as suggesting that AMH levels in pre-menopausal women are determined by declining ovarian reserve as a result of reproductive ageing but not menarche timing, and that women with lower AMH are nearer to the end of their reproductive lifespan.

Variant rs16991615 has previously been found to be associated with differences in menopause timing by 0.9 years per allele and is a missense variant in exon 9 of *MCM8* (E341K), which is required for homologous recombination (20). Other mutations in *MCM8* causing reduced double strand break repair have been found in women with premature ovarian failure (21) and follicle development is arrested at an early stage in *MCM8* knockout mice (22). Pathway analysis has shown that the menopause

timing variants identified from genome-wide analyses are enriched for genes involved in DNA damage response, including double-strand break repair during meiosis, suggesting that the genetic determinants of age at menopause act during ovarian follicle formation or maintenance, potentially affecting ovarian reserve from before birth until menopause (18). Therefore, it seems likely that rs16991615 affects AMH levels through differences in ovarian reserve.

Although the three published GWAS signals in and around the *AMH* gene (11) did not reach genome-wide significance when jointly included in the regression analysis, they did show directional consistency and were nominally associated in the pre-menopausal women in our study. The previous GWAS of adolescent females included fewer samples ($n = 1455$) than our analysis ($n = 3344$), hence we were better powered to detect the effects of these variants in females. The three published variants for AMH had smaller effects in pre-menopausal women compared with adolescent males but were consistent with the associations seen previously in adolescent females. The strongest signal in the GWAS of AMH levels in adolescents, rs4807216, was not associated with age at menopause in the most recent genome-wide meta-analysis (per allele difference in age at menopause of 0.05 years [95% CI $[-0.05, 0.15]$], $P = 0.37$) (17). Differences in genetic regulation of AMH levels in males and females are plausible given AMH's different function in men and women. In males, AMH is required for regression of the Müllerian ducts during testis development in the fetus, and is involved in testicular development and function (2). In females, AMH is produced by granulosa cells of primary, pre-antral and small antral follicles, inhibiting both the further recruitment of primordial follicles from the follicle pool and also FSH-dependent selection of follicles for growth during the menstrual cycle (3,7,23,24). AMH expression starts *in utero* at 36 week's gestation, peaking at around 25 years before declining to undetectable levels at menopause (3,5,7). In our analysis, a polymorphism in the promoter of *FSHB* ($-211G > T$) that affects FSH levels (25–27) had no effect on AMH levels in pre-menopausal women, supporting the absence of direct negative feedback of FSH on AMH.

AMH levels vary widely between women (6), reflecting factors such as variation in ovarian reserve, age and ethnicity (13). We controlled for age and ethnicity by adjusting for age and restricting our analyses to genetically European individuals. Adjustment for age will remove a source of variation in AMH level, the effect of the decrease in the primordial follicle pool with age, highlighting the effect of genetic variants responsible for variation in the initial size of the primordial follicle pool or that either accelerate or protect against loss of ovarian reserve with age.

We would have been unable to detect low frequency variants or those with a smaller ES since we were only powered (>80%) to detect a variant with a MAF of 5% and an effect of 0.36 SD or greater in the sample size analysed. We were unable to evaluate the effect of time from study participation to menopause on our results, to investigate whether the association of the menopause variants was modified by proximity to menopause, since we did not have sufficient follow-up data. Future analysis should consider stratifying by participants' time to menopause. However, such analyses would require large numbers of women who had pre-menopausal AMH measurements (AMH levels are generally undetectable post-menopause), recorded age at menopause and varying times of follow-up since age at menopause. We are not aware of any such study currently but with continued follow-up of the women included in this study such analyses may be possible.

This study confirms genetically that AMH levels are a marker for ovarian decline and reproductive ageing in pre-menopausal women. In addition to its use as a marker of fertility, there is evidence that AMH is a biomarker of breast cancer risk in pre-menopausal women. In a recent study, odds of pre- and postmenopausal breast cancer were 60% higher in women in the highest quartile of AMH level compared with the lowest, even after adjusting for potential confounders such as age (28). Our study suggests that these effects could be mediated through preserved ovarian reserve, or a correlate of ovarian reserve, as a result of delayed reproductive ageing, supported by findings from a large scale genomic analysis that showed a causal effect of later menopause on increased risk of breast cancer by 6% per year (17) and strong epidemiological evidence that later age at menopause increases risk of breast cancer (29). Our study provides genetic evidence that underlying biological factors responsible for reproductive ageing contribute to AMH levels in pre-menopausal women and are likely to be the main driver for the observed associations of AMH with health outcomes.

Materials and Methods

Studies included

The central analysis team at University of Exeter Medical School coordinated data collection from the five studies. We included 3344 women who had pre-menopausal AMH levels measured, who were participants in the Generations Study (30), the Sister Study (31), the Nurses' Health Study, the Nurses' Health Study II (32) and ALSPAC (33–35) (Table 1) (Supplementary Information). For the Generations Study, the Sister Study and the Nurses' Health Studies, genotype and phenotype data were provided to the central analysis team for quality control, cleaning and analysis. For the ALSPAC study, quality control and genotype-phenotype analyses were undertaken in house and summary descriptive, GWAS and sensitivity analyses statistics were provided to the central analysis team for meta-analysis.

Genetic data

In the Generations Study, Sisters and the Nurses' Health Studies, samples were genotyped on the OncoArray array (Table 1). For the Nurses' Health Studies, a further 225 samples were genotyped on an Illumina array. For each cohort and array type, data were cleaned using a standard quality control process in PLINK v1.9 (www.cog-genomics.org/plink/1.9/) (36). SNPs were removed if they were poorly genotyped (missing in >5% samples) or were not in Hardy-Weinberg equilibrium ($P < 1 \times 10^{-6}$). Samples were removed if they were poorly genotyped (missing >5% SNPs), were a sex mismatch or were outliers in terms of heterozygosity. Within each cohort, samples that were related to each other as 3rd degree relatives or closer were identified and the sample with the greater proportion of missing SNPs was removed. Principal component analysis was carried out in FlashPCA in order to identify and remove genetically non-European samples from the analysis. SNPs with MAF > 1% aligned to the correct strand in HRC v1.1 were used for imputation. Genotypes for chromosomes (chr) 1–22 were phased using SHAPEIT and imputed to HRC r1.1 2016 using the University of Michigan Imputation Server (<https://imputationserver.sph.umich.edu/>) (12,37,38).

ALSPAC mothers were genotyped using the Illumina Human660W-Quad array at Centre National de Génotypage and genotypes were called with Illumina GenomeStudio. Quality control was performed in PLINK v1.07 (39) by removing poorly

genotyped SNPs (missing in >5% samples), not in Hardy-Weinberg equilibrium ($P < 1 \times 10^{-6}$), or that had MAF < 1%. Samples were removed if they were poorly genotyped (missing >5% SNPs), had indeterminate X chr heterozygosity or extreme autosomal heterozygosity. Samples showing evidence of population stratification were identified by multidimensional scaling of genome-wide identity by state (IBS) pairwise distances using the four HapMap populations as a reference, and then excluded (IBS > 0.125). Haplotypes were estimated using SHAPEIT (v2.r644) (40) and phased haplotypes were then imputed to HRC panel (12) using IMPUTE V3.

AMH phenotype

For the Generations Study, Sister Study and the Nurses' Health Studies, AMH levels were measured in blood samples taken from pre-menopausal women before breast cancer incidence by the individual studies as part of a collaborative, prospective study of AMH and breast cancer risk (28). Serum AMH levels were measured using an ultrasensitive ELISA (Ansh Labs, Webster, TX) (Sister Study) or a picoAMH enzyme-linked immunoabsorbent assay (Ansh Labs, Webster, TX) (Generations Study and the Nurses' Health Studies, and samples below limit of detection of ELISA in Sister Study) (28). In ALSPAC, blood samples were taken following a standardized protocol in women who attended a series of clinic assessments starting about 18 years after the index pregnancy and fasted (overnight or a minimum of 6 h for those assessed in the afternoon) serum AMH levels were measured using the Beckman Coulter AMH Gen II ELISA assay (34,35).

For samples with AMH below the lower limit of detection, levels were imputed: for the Generations Study, the value was the midpoint between zero and the lower limit of detection (0.00821 pmol/L); for the Sister Study, missing values were imputed as 0.0015 ng/mL to be consistent with the previous analysis (28); for ALSPAC, measured AMH values <0.01 ng/mL were imputed to be 0.01 ng/mL. A small number of women ($n = 24$) in the Nurses' Health Studies with AMH below the lower limit of detection (2.038 pg/mL) were excluded from the analyses. For all studies, measured values of AMH were converted to pmol/L using $1 \text{ pg/mL} = 0.00714 \text{ pmol/L}$ and $1 \text{ ng/mL} = 7.14 \text{ pmol/L}$. AMH was transformed by inverse normal transformation, in which the rank of the AMH value was converted to the z-score for the corresponding quantile of a standard normal distribution. This was done in order to approximate a normal distribution to ensure normality of the residuals in the association analysis and to rank the measured AMH values from each cohort to a consistent scale, avoiding the issue of consistently higher/lower AMH measurements due to the assay used. AMH levels for each study are summarized in Table 1.

Genome-wide analysis

Genome-wide linear regression analysis was carried separately for each of the five cohorts for autosomal genetic variants with imputation quality >0.4 assuming an additive model. Age at time the blood sample was taken was included as a covariate since age is known to be strongly associated with AMH levels and was negatively correlated with AMH level in exploratory analysis (median age in each study is summarized in Table 1). For the Generations Study, Sisters and Nurses' Health Study, analysis was carried out using GEMMA 0.94.1 (41), which calculates a genetic relationship matrix to account for cryptic relatedness

and population stratification between the samples. The genetic relationship matrix was created from a pruned list of uncorrelated SNPs created in PLINK 1.9 (www.cog-genomics.org/plink/1.9/) (36) using `-indep-pairwise`, excluding regions of long-range linkage disequilibrium, based on variants with $MAF > 0.01$, excluding variants with $r^2 > 0.5$ (window size of 1000, calculated in steps of 50). The analysis included approximately 11.8 million genetic variants for the Generations Study, 12.9 million for the Sister Study, 12.3 million for Nurses' Health Study OncoArray and 11.1 million for Nurses' Health Study Illumina. For ALSPAC, the analysis was carried out in SNPTESTv2.5 (42) adjusting for the top 10 principal components of ancestry which resulted in approximately 14.7 million SNPs.

Standard error (SE) weighted meta-analysis of the individual GWAS results was carried out in METAL (version 2011-03-25) (43) with genomic control applied to account for inflation due to any remaining population stratification. Genetic variants included in the meta-analysis had imputation >0.4 and minor allele count >5 (calculated from allele frequencies), resulting in a total of 11.2 million autosomal SNPs that were analysed. Approximately 8.4 million variants were present in three or more of the five datasets analysed and were included in our final results. We identified independent signals as being suggestive of genome-wide association if they had $P < 1 \times 10^{-5}$ and were more than 500 kb from another signal; from these, we identified signals reaching genome-wide significance at $P < 5 \times 10^{-8}$.

Manhattan and quantile–quantile plots for the genome-wide association results were created using the package *qqman* (44) in R (The R Foundation for Statistical Computing). LocusZoom v1.4 (45) was used to plot the association statistics with age-adjusted inverse normal AMH for variants within 500 kb of the top variant, showing linkage disequilibrium. Linkage disequilibrium was calculated in PLINK v1.9 (www.cog-genomics.org/plink/1.9/) (36) from best guess genotypes for 1000 Genomes Phase 3/HRC imputed variants in ~340 000 unrelated Europeans from the UK Biobank study (46).

Generation of age-adjusted inverse normal effect estimates in ALSPAC offspring cohort

For three published genetic variants that were associated with AMH levels in adolescent males in the ALSPAC offspring cohort (11), we generated effect estimates for age-adjusted inverse normal transformed AMH in the original study sample, since the original published estimates were unadjusted and presented in natural log-transformed AMH. Analyses were carried in SNPTEST v2.5 (42) adjusting for the top 10 principal components of ancestry and age and excluding the most extreme 1% of measured AMH values, resulting in 1312 males and 1421 females from the ALSPAC offspring cohort for analysis. Other methods were as described previously (11).

Estimation of joint effects of variants in and around the AMH gene

We used GCTA (version 1.25.0), using the command `-cojo-joint` (14), to carry out an approximate conditional analysis to estimate the joint effects of three genetic variants in and around the AMH gene (11). Linkage disequilibrium between the variants was estimated using a random sample of 8569 white British individuals from the UK Biobank May 2015 interim release of imputed genetic data (47).

Comparison of effects for published traits with AMH results

We compared the published effect estimates for 56 genetic variants associated with age at menopause (17) and 389 genetic variants associated with age at menarche (18) with their effects on AMH level in our analysis, by carrying out a Binomial sign test of directional consistency and calculating Pearson correlation coefficients. To explore whether age at menopause or age at menarche causes differences in AMH levels, we used the genetic variants associated with menopause and menarche timing as instruments for age at menopause and age at menarche in two-sample Mendelian Randomization analyses. We used the Stata package *mrrobust* (48) to carry out IVW and Egger (which takes account of horizontal pleiotropy (49)) analyses. Analyses were carried out in Stata MP 13.0 and Stata SE 14.2 (StataCorp, TX, USA).

Sensitivity analysis

The genome-wide analysis was repeated without adjustment for age, using a natural log transformation [to be consistent with the previously published GWAS (11)], and also excluding women whose AMH level was imputed as it was below the lower limit of detection. We compared ESs in the main analysis with estimates from these alternate analyses for genetic variants with $P < 5 \times 10^{-5}$ in the main analysis.

Supplementary Material

Supplementary Material is available at HMG online.

Acknowledgements

We thank Breast Cancer Now and the Institute of Cancer Research for support and funding of the Generations Study, and the study participants, study staff, and the doctors, nurses and other health care providers and health information sources who have contributed to the study. The Institute of Cancer Research acknowledges National Health Service funding to the NIHR Biomedical Research Centre. We would like to thank the participants and staff of the Nurses' Health Study and the Nurses' Health Study II. This research has been conducted using the UK Biobank Resource. We would like to thank Dr R.N. Beaumont and Dr S.E. Jones (Genetics of Complex Traits, University of Exeter Medical School) for assistance with calculating linkage disequilibrium in UK Biobank. The authors would like to acknowledge the use of the University of Exeter High-Performance Computing (HPC) facility in carrying out this work. We are extremely grateful to all the families who took part in the ALSPAC study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The work presented here is that of the authors. The views expressed in this publication are those of the author(s) and not necessarily those of the UK Medical Research Council, Wellcome Trust, British Heart Foundation, the National Institute for Health Research or the UK National Health Service or Department of Health.

Conflict of Interest statement. D.A.L. has received support from Roche Diagnostics and Medtronic for research unrelated to that presented here. A.M. and K.S.R. received support from Randox

Laboratories Ltd for research unrelated to that presented here. No other conflicts of interest are declared.

Funding

Breast Cancer Now [to Generations Study]; British Heart Foundation [SP/07/008/24066 to ALSPAC/D.A.L.]; Gillings Family Foundation [to K.S.R.]; National Institute of Health Research [NF-SI-0611-10196 to D.A.L.]; Roche Diagnostics [to ALSPAC]; The Institute of Cancer Research [to Generations Study]; UK Medical Research Council [102215/2/13/2, G1001357 to ALSPAC/DAL; MC_UU_00011/6 to M.C.B., A.L.G.S. and D.A.L.; MR/P014054/1 to M.C.B.]; UK National Health Service [to NIHR Biomedical Research Centre Institute of Cancer Research]; University of Bristol [to ALSPAC and the MRC Integrative Epidemiology Unit]; US National Cancer Institute [CA186107, CA176726, CA49449, CA67262, CA178949 to the Nurses' Health Study and the Nurses' Health Study II; CA178949 to A.Z.-J.]; Wellcome Trust [102215/2/13/2, WT088806 and WT092830/Z/10 to ALSPAC/D.A.L.]; NIHR Biomedical Research Centre at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol [to D.A.L.].

References

- La Marca, A., Broekmans, F.J., Volpe, A., Fauser, B.C. and MacKlon, N.S. (2009) Anti-Müllerian hormone (AMH): what do we still need to know? *Hum. Reprod.*, **24**, 2264–2275.
- Matuszczak, E., Hermanowicz, A., Komarowska, M. and Debek, W. (2013) Serum AMH in physiology and pathology of male gonads. *Int. J. Endocrinol.*, **2013**, 128907.
- Visser, J.A., Schipper, I., Laven, J.S.E. and Themmen, A.P.N. (2012) Anti-Müllerian hormone: an ovarian reserve marker in primary ovarian insufficiency. *Nat. Rev. Endocrinol.*, **8**, 331–341.
- Hagen, C.P., Aksglaede, L., Sørensen, K., Mouritsen, A., Andersson, A.M., Petersen, J.H., Main, K.M. and Juul, A. (2012) Individual serum levels of anti-Müllerian hormone in healthy girls persist through childhood and adolescence: a longitudinal cohort study. *Hum. Reprod.*, **27**, 861–866.
- Kelsey, T.W., Anderson, R.A., Wright, P., Nelson, S.M. and Wallace, W.H.B. (2012) Data-driven assessment of the human ovarian reserve. *Mol. Hum. Reprod.*, **18**, 79–87.
- de Kat, A.C., van der Schouw, Y.T., Eijkemans, M.J.C., Herber-Gast, G.C., Visser, J.A., Verschuren, W.M.M. and Broekmans, F.J.M. (2016) Back to the basics of ovarian aging: a population-based study on longitudinal anti-Müllerian hormone decline. *BMC Med.*, **14**, 151.
- Meczekalski, B., Czyzyk, A., Kunicki, M., Podfigurna-Stopa, A., Plociennik, L., Jakiel, G., Maciejewska-Jeske, M. and Lukaszuk, K. (2016) Fertility in women of late reproductive age: the role of serum anti-Müllerian hormone (AMH) levels in its assessment. *J. Endocrinol. Invest.*, **39**, 1259–1265.
- Alvaro Mercadal, B., Imbert, R., Demeestere, I., Gervy, C., De Leener, A., Englert, Y., Costagliola, S. and Delbaere, A. (2015) AMH mutations with reduced in vitro bioactivity are related to premature ovarian insufficiency. *Hum. Reprod.*, **30**, 1196–1202.
- Gorsic, L.K., Kosova, G., Werstein, B., Sisk, R., Legro, R.S., Hayes, M.G., Teixeira, J.M., Dunaif, A. and Urbanek, M. (2017) Pathogenic anti-Müllerian hormone variants in polycystic ovary syndrome. *J. Clin. Endocrinol. Metab.*, **102**, 2862–2872.
- Pabalan, N., Montagna, E., Singian, E., Tabangay, L., Jarjanazi, H., Barbosa, C.P. and Bianco, B. (2016) Associations of polymorphisms in anti-Müllerian hormone (AMH Ile49Ser) and its type II receptor (AMHR2 -482 A>G) on reproductive outcomes and polycystic ovary syndrome: a systematic review and meta-analysis. *Cell. Physiol. Biochem.*, **39**, 2249–2261.
- Perry, J.R.B., McMahon, G., Day, F.R., Ring, S.M., Nelson, S.M. and Lawlor, D.A. (2016) Genome-wide association study identifies common and low-frequency variants at the AMH gene locus that strongly predict serum AMH levels in males. *Hum. Mol. Genet.*, **25**, 382–388.
- McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K. et al. (2016) A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.*, **48**, 1279–1283.
- La Marca, A., Grisendi, V. and Griesinger, G. (2013) How much does AMH really vary in normal women? *Int. J. Endocrinol.*, **2013**.
- Yang, J., Lee, S.H., Goddard, M.E. and Visscher, P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.*, **88**, 76–82.
- Stolk, L., Zhai, G., van Meurs, J. B. J., Verbiest, M. M. P. J., Visser, J. A., Estrada, K., Rivadeneira, F., Williams, F. M., Cherkas, L., Deloukas, P. et al. (2009) Loci at chromosomes 13, 19 and 20 influence age at natural menopause. *Nat. Genet.*, **41**, 645–647.
- Stolk, L., Perry, J.R.B., Chasman, D.I., He, C., Mangino, M., Sulem, P., Barbalic, M., Broer, L., Byrne, E.M., Ernst, F. et al. (2012) Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways. *Nat. Genet.*, **44**, 260–268.
- Day, F.R., Ruth, K.S., Thompson, D.J., Lunetta, K.L., Pervjakova, N., Chasman, D.I., Stolk, L., Finucane, H.K., Sulem, P., Bulik-Sullivan, B. et al. (2015) Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nat. Genet.*, **47**, 1294–1303.
- Day, F.R., Thompson, D.J., Helgason, H., Chasman, D.I., Finucane, H., Sulem, P., Ruth, K.S., Whalen, S., Sarkar, A.K., Albrecht, E. et al. (2017) Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat. Genet.*, **49**, 834–841.
- Ruth, K.S., Campbell, P.J., Chew, S., Lim, E.M., Hadlow, N., Stuckey, B.G.A., Brown, S.J., Feenstra, B., Joseph, J., Surdulescu, G.L. et al. (2016) Genome-wide association study with 1000 genomes imputation identifies signals for nine sex hormone-related phenotypes. *Eur. J. Hum. Genet.*, **24**, 284–290.
- Nishimura, K., Ishiai, M., Horikawa, K., Fukagawa, T., Takata, M., Takisawa, H. and Kanemaki, M.T. (2012) Mcm8 and Mcm9 form a complex that functions in homologous recombination repair induced by DNA interstrand crosslinks. *Mol. Cell*, **47**, 511–522.
- AlAsiri, S., Basit, S., Wood-Trageser, M.A., Yatsenko, S.A., Jeffries, E.P., Surti, U., Ketterer, D.M., Afzal, S., Ramzan, K., Faiyaz-Ul Haque, M. et al. (2015) Exome sequencing reveals MCM8 mutation underlies ovarian failure and chromosomal instability. *J. Clin. Invest.*, **125**, 258–262.
- Lutzmann, M., Grey, C., Traver, S., Ganier, O., Maya-Mendoza, A., Ranisavljevic, N., Bernex, F., Nishiyama, A., Montel, N., Gavois, E. et al. (2012) MCM8- and MCM9-deficient mice reveal gametogenesis defects and genome instability due to impaired homologous recombination. *Mol. Cell*, **47**, 523–534.
- Visser, J.A. and Themmen, A.P.N. (2014) Role of anti-Müllerian hormone and bone morphogenetic proteins in

- the regulation of FSH sensitivity. *Mol. Cell Endocrinol.*, **382**, 460–465.
24. Kevenaar, M.E., Themmen, A.N., Laven, J.E., Sonntag, B., Fong, S.L., Uitterlinden, A.G., de Jong, F.H., Pols, H.P., Simoni, M. and Visser, J.A. (2007) Anti-Müllerian hormone and anti-Müllerian hormone type II receptor polymorphisms are associated with follicular phase estradiol levels in normo-ovulatory women. *Hum. Reprod.*, **22**, 1547–54.
 25. Grigorova, M., Punab, M., Ausmees, K. and Laan, M. (2008) FSHB promoter polymorphism within evolutionary conserved element is associated with serum FSH level in men. *Hum. Reprod.*, **23**, 2160–2166.
 26. Schuring, A.N., Busch, A.S., Bogdanova, N., Gromoll, J. and Tuttelmann, F. (2013) Effects of the FSH-beta-subunit promoter polymorphism -211G->T on the hypothalamic-pituitary-ovarian axis in normally cycling women indicate a gender-specific regulation of gonadotropin secretion. *J. Clin. Endocrinol. Metab.*, **98**, E82–E86.
 27. La Marca, A., Papaleo, E., Alviggi, C., Ruvolo, G., De Placido, G., Candiani, M., Cittadini, E., De Michele, F., Moriondo, V., Catellani, V. et al. (2013) The combination of genetic variants of the FSHB and FSHR genes affects serum FSH in women of reproductive age. *Hum. Reprod.*, **28**, 1369–1374.
 28. Ge, W., Clendenen, T.V., Afanasyeva, Y., Koenig, K.L., Agnoli, C., Brinton, L.A., Dorgan, J.F., Eliassen, A.H., Falk, R.T., Hallmans, G. et al. (2018) Circulating anti Müllerian hormone and breast cancer risk: a study in ten prospective cohorts. *Int. J. Cancer*, **10.1002/ijc.31249**.
 29. Collaborative Group on Hormonal Factors in Breast Cancer (2012) Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. *Lancet Oncol.*, **13**, 1141–1151.
 30. Swerdlow, A.J., Jones, M.E., Schoemaker, M.J., Hemming, J., Thomas, D., Williamson, J. and Ashworth, A. (2011) The breakthrough generations study: design of a long-term UK cohort study to investigate breast cancer aetiology. *Br. J. Cancer*, **105**, 911–917.
 31. Nichols, H.B., Baird, D.D., Stanczyk, F.Z., Steiner, A.Z., Troester, M.A., Whitworth, K.W. and Sandler, D.P. (2015) Anti-Müllerian hormone concentrations in premenopausal women and breast cancer risk. *Cancer Prev. Res.*, **8**, 528–534.
 32. Eliassen, A.H., Zeleniuch-Jacquotte, A., Rosner, B. and Hankinson, S.E. (2016) Plasma anti-Müllerian hormone concentrations and risk of breast cancer among premenopausal women in the Nurses' Health Studies. *Cancer Epidemiol. Biomarkers Prev.*, **25**, 854–860.
 33. Boyd, A., Golding, J., Macleod, J., Lawlor, D.A., Fraser, A., Henderson, J., Molloy, L., Ness, A., Ring, S. and Smith, G.D. (2013) Cohort profile: the 'Children of the 90s'—the index offspring of the avon longitudinal study of parents and children. *Int. J. Epidemiol.*, **42**, 111–127.
 34. Fraser, A., Macdonald-Wallis, C., Tilling, K., Boyd, A., Golding, J., Davey Smith, G., Henderson, J., Macleod, J., Molloy, L., Ness, A. et al. (2013) Cohort profile: the avon longitudinal study of parents and children: ALSPAC mothers cohort. *Int. J. Epidemiol.*, **42**, 97–110.
 35. Wang, Q., Ferreira, D.L.S., Nelson, S.M., Sattar, N., Ala-Korpela, M. and Lawlor, D.A. (2018) Metabolic characterization of menopause: cross-sectional and longitudinal evidence. *BMC Med.*, **16**, 1–12.
 36. Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M. and Lee, J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, **4**. <https://doi.org/10.1186/s13742-015-0047-8>
 37. Loh, P.-R., Danecek, P., Palamara, P.F., Fuchsberger, C., Y., A.R., H., K.F., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G.R. et al. (2016) Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.*, **48**, 1443–1448.
 38. Das, S., Forer, L., Schonherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E., Levy, S., McGue, M. et al. (2016) Next-generation genotype imputation service and methods. *Nat. Genet.*, **48**, 1284–1287.
 39. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J. et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
 40. O'Connell, J., Gurdasani, D., Delaneau, O., Pirastu, N., Ulivi, S. et al. (2014) A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet.*, **10**, e1004234. <https://doi.org/10.1371/journal.pgen.1004234>.
 41. Zhou, X. and Stephens, M. (2012) Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.*, **44**, 821–824.
 42. Marchini, J. and Howie, B. (2010) Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.*, **11**, 499–511.
 43. Willer, C.J., Li, Y. and Abecasis, G.R. (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**, 2190–2191.
 44. Turner, S.D. (2018) qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J. Open Source Softw.*, **3**, 731.
 45. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Glied, T.P., Boehnke, M., Abecasis, G.R., Willer, C.J. and Frisman, D. (2011) LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*, **27**, 2336–2337.
 46. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliot, L.R., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., Connell, J.O. et al. (2017) Genome-wide genetic data on ~500,000 UK Biobank participants, bioRxiv <http://dx.doi.org/10.1101/166298>.
 47. UK Biobank (2015) Genotype imputation and genetic association studies of UK Biobank.
 48. Spiller, W., Davies, N.M. and Palmer, T.M. (2017) Software application profile: mrrobust—a tool for performing two-sample summary mendelian randomization analyses. bioRxiv [dx.doi.org/10.1101/14212](https://doi.org/10.1101/14212).
 49. Bowden, J., Smith, G.D. and Burgess, S. (2015) Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.*, **44**, 512–525.