A genome-wide association study of anti-Müllerian hormone (AMH) levels in Samoan women

Erdogan-Yildirim Z[1,2*], Carlson JC[2,3], Krishnan M[2,†], Zhang JZ[3], Lambert-Messerlian G[4], Naseri T[5,6,7], Viali S[6,8,9], Hawley NL[9], McGarvey ST[7], Weeks DE[2,3], Minster RL[2]

Author Affiliations:

[1] Center for Craniofacial and Dental Genetics, Department of Oral and Craniofacial Sciences, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA 15261, USA

[2] Department of Human Genetics, School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA

[3] Department of Biostatistics and Health Data Science, School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA

[4] Department of Obstetrics and Gynecology, and Pathology and Laboratory Medicine, Brown University, Providence, RI 02912, USA

[5] Naseri & Associates Public Health Consultancy Firm and Family Health Clinic, Apia, Samoa

[6] Oceania University of Medicine, Apia, Samoa

[7] International Health Institute and Department of Epidemiology, School of Public Health, Brown University, Providence, RI 02903, USA

[8] School of Medicine, National University of Samoa, Apia, Samoa

[9] Department of Chronic Disease Epidemiology, School of Public Health, Yale University, New Haven, CT 06520, USA

[*] First and corresponding author.

[†] Current affiliations: Department of Biobehavioral Health, College of Health and Human Development, Pennsylvania State University, State College, PA, USA, and Department of Epidemiology, University of North Carolina at Chapel Hill, NC, USA.

MESH terms:
Anti-Müllerian Hormone, Genome-Wide Association Study, Reproduction, Reproductive Health, Samoa, Pacific Island People

Abstract

Study question: Can a genome-wide association study (GWAS) and transcriptome-wide association study (TWAS) help identify genetic variation or genes associated with circulating anti-Müllerian hormone (AMH) levels in Samoan women?

Summary answer: We identified eleven genome-wide suggestive loci (strongest association signal in *ARID3A* 19-946163-G-C [$p = 2.32 \times 10^{-7}$]) and seven transcriptome-wide significant genes (*GINS2, SENP3, USP7, TUSC3, MAFA, METTL4, NDFIP1* [all with a $p < 2.50 \times 10^{-6}$]) associated with circulating AMH levels in Samoan women.

What is known already: Three prior GWASs of AMH levels identified eight loci in premenopausal women of European ancestry *(AMH, MCM8, TEX41*, *CHECK2, CDCA7*, *EIF4EBP1, BMP4* and an uncharacterized non-coding RNA gene *CTB-99A3.1*), among which the *MCM8* locus was shared among all three studies.

Study design, size, duration: We included a sample of 1,185 women from two independently recruited samples: a family study ($n = 212$; [age: 18 to 40 years]) recruited in 2002–03 from Samoa and American Samoa; and the Soifua Manuia Study ($n = 973$; age: 25 to 51 years), a cross-sectional population-based study recruited in 2010 from Samoa.

Participants/materials, setting, methods: Serum AMH levels were measured using enzyme linked immunosorbent assays (ELISA). We performed GWASs in the two participant samples using a Cox mixed-effects model to account for AMH levels below detectable limits and adjusted for centered age, centered age², polity, and kinship via kinship matrix. The summary statistics were then meta-analyzed using a fixed-effect model. We annotated the variants with $p < 1 \times 10^{-5}$ and calculated posterior probability of causality for prioritization. We further annotated variants using FUMA and performed colocalization and transcriptome-wide association analysis. We also assessed whether any previously reported loci were replicated in our GWAS.

Main results and the role of chance: We identified eleven novel genome-wide suggestive loci ($p < 1 \times 10^{-5}$) associated with AMH levels and replicated *EIF4EBP1,* a previously reported AMH locus, in the GWAS. The lead variant in *ARID3A*, 19-946163-G-C is in high linkage disequilibrium ($r^2 = 0.79$) with the known age-at-menopause variant 19-950694-G-A. Nearby *KISS1R* is a biologically plausibility causal gene in the region; kisspeptin regulates ovarian follicle development and has been linked to AMH levels. Further investigation of the *ARID3A* locus is warranted.

Limitations, reasons for caution: The main limitations of our study are the small sample size for a GWAS and the use of the transcription model trained on mostly European samples from the Genotype Tissue Expression (GTEx) project, which may have led to reduced power to detect genotype-expression associations. Our findings need to be validated in larger Polynesian cohorts.

Wider implications of the findings: In addition to replicating one of the eight previously discovered AMH loci, we identified new suggestive associations. It is known that the inclusion of founder populations aids in the discovery of novel loci. These findings could enhance our understanding of AMH and AMH-related reproductive phenotypes (ovarian reserve, age at menopause, premature ovarian failure, and polycystic ovary syndrome) and help build a screening approach for women at risk for these phenotypes using genetically predicted AMH levels.

Study funding/competing interest(s): This work was funded by NIH grants R01-HL093093 (PI: S.T.M.), R01-HL133040 (PI: R.L.M.), and T90-DE030853 (PI: C.S. Sfeir). Molecular data for the Trans-Omics in Precision Medicine (TOPMed) Program was supported by the National Heart, Lung and Blood Institute (NHLBI). The content is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health.

Introduction

Anti-Müllerian hormone (AMH) has an important role in ovarian biology and female reproductive health.[1] AMH is produced in women after birth and is exclusively secreted by the granulosa cells of developing ovarian follicles until the antral stage is reached [2–4]. It functions to promote the development of a dominant follicle while suppressing non-dominant follicles. Not surprisingly, serum AMH concentrations highly correlate with the number of ovarian follicles [5–8], making it one of the best-known biomarkers of ovarian reserve (defined as the capacity of the ovary to provide egg cells that are capable of fertilization resulting in a healthy and successful pregnancy) and a predictor of the oocyte outcomes of *in vitro* fertilization [9,10]. Serum AMH concentrations reflect the trajectory of female reproductive life span, peaking around age 25 and then gradually decreasing until becoming undetectable before menopause [9,11]. Hence, AMH also has utility in predicting time-to-menopause [5–8,12–14].

In recent years, researchers have focused on the clinical application of AMH as a surrogate marker to evaluate polycystic ovaries and to diagnose polycystic ovary syndrome (PCOS), a common disorder affecting fertility and metabolic health of women [11,13,15]. Studies have consistently shown that AMH levels are higher in all PCOS subtypes compared to normo-ovulatory women and women with polycystic ovarian morphology alone [16] and that AMH levels correlate with PCOS subtype and severity [17,18]. Hence, AMH could provide a non-invasive alternative to transvaginal ultrasound for antral follicle counts, particularly when the latter is not available or not feasible due to cost and or acceptability (i.e. due to cultural and psycho-social reasons, especially for adolescents) [19].

Our research group has been focused on the health of Samoans, a founder population, for more than thirty years. We have conducted several epidemiological studies to describe the influences of adiposity on Samoan women's reproductive health, specifically describing menstrual irregularity, hyperandrogenemia, and estimating prevalence of PCOS [20–22]. This is especially important given the high and rising levels of adiposity among Samoan women that are

characteristic of Pacific Islanders more broadly [23–27]. Recently, two distinct PCOS subtypes (metabolic vs. reproductive) with distinct genetic architecture have been described in individuals of European ancestry [28], and it was found that PCOS susceptibility loci differ between lean and overweight/obese cases [29]. We do not have diagnoses of PCOS in Samoan women, and so we cannot directly examine the genetic determinants of PCOS in this study. However, it could be fruitful to examine related phenotypes to begin to understand genetic determinants of reproductive health in this population.

There have been several studies of the genetic variation underlying AMH levels [30–32]. In addition to polymorphisms within the *AMH* gene itself [32–34], GWASs have mapped seven genes in women of European ancestry, of which four are implicated in cell cycle regulation (*MCM8* [31,32,34], *TEX41* [32,34], *CHECK2* [34], and *CDCA7* [32]). The other three genes are *EIF4EBP1* [34], *BMP4* [34], and an uncharacterized non-coding RNA gene (*CTB-99A3.1* [32]). The established loci explain about 13% to 15% of the single nucleotide variant (SNV)-based heritability [32,34]. A major limitation of the existing studies, however, is that the information on the genetic underpinnings of AMH have been derived solely from women of European ancestry [35,36]. More research is needed in diverse populations not only to enhance our understanding of the underlying biology but also to ensure access to adequate health care and effective treatment for these communities. Importantly, the inclusion of founder populations in genetic research is important since the reduced allelic heterogeneity in these groups can be advantageous to discover novel loci via genome-wide association studies [23–27].

In this study, we aimed to identify genetic determinants of circulating AMH levels via genome-wide and transcriptome-wide analyses in Samoan women.

Materials and methods

Study subjects

Two independent study samples comprised of a total of 1,185 Samoan women were selected for a GWAS to assess the genetic variation associated with circulating serum AMH levels. The first

sample of 212 women aged ≥ 18 years and < 40 years was drawn from a 2002–2003 family study of genetic linkage analysis of cardiometabolic traits (for sample flowchart see Supplementary Figure S1) [22,37,38]. The age range was limited to reproductive-aged women in the parent study [22] to avoid potential effects of perimenopause. Participants were recruited from villages across ‘Upolu and Savai‘i, the two largest islands of Samoa, and Tutuila, the largest island of American Samoa.

The second sample of 973 Samoan women aged ≥ 25 to ≤ 50 years was drawn from a 2010 cross-sectional population-based study (Soifua Manuia [in Samoan: “Good Health”] Study) of obesity and cardiometabolic health [21,24,39] (for sample see Supplementary Figure S1). Participants were recruited from thirty-three villages across ‘Upolu and Savai‘i. All participants completed a questionnaire surveying their health history and lifestyle factors related to cardiometabolic and reproductive health including socio-economic status, dietary intake, and physical activity [22,37,38].

In both studies, women who had a history of hysterectomy and/or ovariectomy or who were pregnant or lactating at the time of recruitment were excluded. Hormonal contraceptive use was not well captured in our cohorts for use as an exclusion criterion. The baseline characteristics between individuals with measured and unmeasured AMH levels are compared in Supplementary Table S2.

Anthropometric and biochemical measurements

The collection and measurement of anthropometric, cardiometabolic and lifestyle-related data have been described in detail before [22,24,39]. Whole blood samples were collected for genotyping and serum biomarker measurements after an overnight fast [22,24,39].

Serum AMH levels were measured using manual enzyme linked immunosorbent assays (ELISA) from Ansh Labs (Webster, TX). In the 2002–03 family study, AMH levels were determined for 198 participants using the picoAMH ELISA assay; AMH levels for 14 women were determined

with the Ultrasensitive (us)AMH/MIS ELISA assay. In the 2010 Soifua Manuia study, the picoAMH assay was used for women age ≥ 40 years old (n = 516) and the usAMH/MIS assay was used for women < 40 years old (n = 457) [21].The inter- and intraassay coefficients of variation were < 15%. The detection limit was 6 pg/mL and 0.08 ng/mL for the picoAMH assay and the usAMH/MIS assay, respectively. To harmonize measurements from the two different assays, the values from the picoAMH assay were rescaled to align them with values from the usAMH/MIS assay using this equation (Ansh Lab, insert AL124-i released on 2019-09-27, regression R^2 = 0.99, p < 0.0001):

usAMH/MIS assay (ng/mL) = (picoAMH assay (pg/mL) + 50.66) / 0.92 / 1000

Genotyping and imputation

Genotyping in 2002–03 family study was performed using the Global Screening Array-24 v.3.0 BeadChip (Illumina, CA, USA) with 644,880 SNVs including custom content pertinent to Samoans. For the 2010 Soifua Manuia study, 659,492 SNVs were genotyped genome-wide using Affymetrix 6.0 array. Quality control procedures were implemented for genotypes from both arrays following the guidelines outlined by Laurie et al. (2010). Detailed description of genotyping and quality control have been previously published [24,39].

Using a reference panel derived from 1,285 Samoan individuals with whole-genome sequencing [40], we performed imputation using minimac4 in both samples and removed variants with R^2 < 0.3, yielding an additional 16,744,117 (2002–03 family study) and 15,633,124 (2010 Soifua Manuia study) SNVs [41].

Ethical approval

Research protocols, informed consents and secondary analyses of both studies were approved by the Health Research Committee of the Samoan Ministry of Health and the Institutional Review Boards of Brown University as well as University of Cincinnati and University of Pittsburgh for the 2010 Soifua Manuia study and additionally the American Samoa Department of Health IRB for

the 2002–03 family study [22,24,39]. All participants were informed about their rights verbally in Samoan by trained research staff before obtaining their written consent [22,24,39].

Genome-wide association study

To account for AMH levels below the detection limit (2002–03 family study: n = 1; 2010 Soifua Manuia study: n = 169), we tested for association between genotype dosages and AMH levels using a Cox mixed-effects model as implemented in the R package {coxmeg} [42,43]. Since Cox regression is designed for right-truncated data, we used the reciprocal of the measured AMH levels [44]. We adjusted for fixed effects of centered age and centered age^2 (as well as polity for the 2002–03 family study) and for random effects of genetic relatedness using empirical kinship coefficients as estimated by PC-Relate [45]. Due to the genetic homogeneity of the sample [24], we did not adjust for principal components of ancestry.

The association results from both Samoan samples were meta-analyzed using a p value–based fixed-effect approach via METAL [46]. Before combining the two association studies, results from the 2002–03 family study and the 2010 Soifua Manuia study were filtered for minor allele frequency (MAF) ≥ 0.05 and MAF ≥ 0.01, respectively, to keep the minor allele count between the two samples close in range. SNVs with a p value ≤ 1×10^{-5} in the two pre–meta-analysis GWASs were tested for goodness of fit to Hardy–Weinberg equilibrium, and SNVs with $p < 0.0001$ were excluded from subsequent analyses. Conditional analyses to detect secondary signals were conducted by including the lead SNVs in each suggestively associated region as covariates in the original model.

Manhattan and QQ plots were created using the R package {fastman} [47]. Tests with p values ≤ 5×10^{-8} were considered genome-wide significant, and ≤ 1×10^{-5} were considered suggestive. Genomic positions are in Genome Reference Consortium Human Build 38 (hg38).

To refine the regions of associations and determine the most probable causal variant, each locus with a p value ≤ 1×10^{-5} was visualized with LocusZoom [48] and assigned a posterior

probability via Bayesian fine-mapping with PAINTOR v2.1. [49] using functional annotation from the Ensembl Variant Effect Predictor (VEP, [50]) and RegulomeDB [51]. This fine-mapping was carried out in a region ±500 kb around each lead SNV, accounting for Samoan-specific linkage disequilibrium (LD) structure.

To identify gene sets and pathways that are enriched for variants relevant to AMH variation, GWAS meta-analysis summary statistics were processed by the Functional Mapping and Annotation (FUMA) v1.5.2 genetic associations pipeline [52] after converting the genomic positions from UCSC hg38 to genomic build UCSC hg19 via liftOver [53]. In addition to functional annotation, we used FUMA to carry out gene-based analysis via Multi-marker Analysis of GenoMic Annotation (MAGMA) [54] and preliminary expression quantitative trait loci (eQTL) analysis. Independent loci were sets of SNVs with a GWAS p value $\leq 1 \times 10^{-5}$ and not in high LD ($r^2 < 0.6$) with other SNVs with p value $\leq 1 \times 10^{-5}$ on the same chromosome. Candidate SNVs for examination with FUMA were SNVs within ±500 kb of the lead SNV that had a GWAS p value < 0.05 and were in LD ($r^2 \geq 0.6$) with the lead SNVs. LD structure for FUMA analysis was calculated using all populations of the 1000G phase 3. Variants located outside the gene boundaries but relevant to the nearby gene are assigned to it by MAGMA. The annotation window size for proxy SNVs was set at 40 kb upstream and 10 kb downstream around each gene. The significance threshold for the gene-based test was set at $p = 2.50 \times 10^{-6}$ following Bonferroni correction for 20,000 tests. eQTL analysis interrogated whole blood, endocrine tissues (adrenal gland, hypothalamus, ovary, pancreas, pituitary, thyroid) and metabolic tissues (subcutaneous adipose tissue, liver) using data from GTEx v8 [55]. p values in the eQTL analyses were Bonferroni corrected based on the number of genes assessed.

Known AMH loci

We examined our GWAS summary statistics for evidence of association of AMH levels with the reported lead SNVs in the eight known AMH loci (*TEX41* [2-144887307-A-G] [32,34]; *CDCA7* [2-

173394597-C-T] [32]; *CTB-99A3.1* [5-146560687-G-A] [32]; *EIF4EBP1* [8-38015258-C-T] [34]; *BMP4* [14-53956049-G-T] [34]; *AMH* [19-2251818-T-C] [34]; *MCM8* [20-5967581-G-A] [31,32,34]; and *CHECK2* [22-28707610-T-C] [34]). A locus was considered replicated if the lead SNV was present in the meta-analysis, had a p value < 0.05, and had the same effect direction.

When the lead SNVs from these prior GWASs was absent from the meta-analysis, because the allele frequency was too low for inclusion in the meta-analysis, we also report the lead SNV within ±50 kb of the prior GWAS's lead SNV if it had $p < 0.05$. For each of these nearby SNVs we also calculated a per-region Bonferroni-corrected significance threshold. The threshold was corrected for the number of independent SNVs in the region as calculated by simpleM [56].

Transcriptome-wide association study

To assess the association of estimated gene expression levels based on genotypes with the phenotypes, we carried out a transcriptome-wide association study (TWAS) on meta-analysis summary statistics with the MetaXcan [57–59] suite of tools. We used the MASHR algorithm to predict gene expression, as it has demonstrated better performance than the Elastic Net algorithm [60]. We performed TWAS in whole blood, endocrine tissues (adrenal gland, hypothalamus, ovary, pancreas, pituitary, thyroid) and metabolic tissues (liver, subcutaneous adipose tissue) using S-PrediXcan and combined the results for each gene from all tested single-tissue models into a single aggregate statistic using S-MultiXcan [58]. The p value for statistical significance ($p = 2.61 \times 10^{-6}$) is adjusted for the number of genes tested (n = 19,158).

Colocalization analysis

Colocalization analyses were performed using fastENLOC [61–64]. For this, we converted the z scores from the meta-analysis to posterior inclusion probabilities (PIP) for causality via TORUS and colocalized 5,589,187 variants with eQTL data from GTEx within the nine issues listed above [65]. Precomputed GTEx multi-tissue annotations are available at https://github.com/xqwen/fastenloc [61–64].

Results

The socio-demographic characteristics and AMH levels for the two samples are presented in Table 1, and AMH levels by age are depicted in Supplementary Figure S2.

Table 1 Sample demographics and characteristics for quantitative trait

		2002–03 Family Study						2010 Soifua Manuia Study					
Variable		*n*	mean	sd	min	median	max	*n*	mean	sd	min	median	max
Age		212	28.3	6.8	18.0	28.1	39.8	973	39.3	7.6	25.0	40.7	50.9
BMI (kg/m^2)		212	34.0	8.5	20.4	32.9	69.0	971	34.8	6.8	18.0	34.4	59.9
AMH (ng/mL)†		212	3.90	6.01	0.06	2.81	77.5	973	1.64	2.65	0.06	0.59	25.8
AMH (ng/mL)††		211	3.91	6.02	0.08	2.82	77.5	804	1.97	2.81	0.06	0.97	25.8
Polity	American Samoa	60%						0%					
	Samoa	40%						100%					

†values below the assay limit of detection are winsorized to the detection limit values
††values below the assay limit of detection excluded

The GWAS of AMH levels had no genome-wide statistically significant associations ($p \leq 5 \times 10^{-8}$) but eleven loci with $p \leq 1 \times 10^{-5}$ were observed (Figure 1, Supplementary Figure S3, Table 2). The lead variants of each locus also had the highest posterior probability after fine-mapping for causal variants, except for the lead variant in *ARID3A* (Table 2). The quality score of all imputed lead variants were > 0.88. We confirmed via conditional analysis that no secondary signals were present in suggestive loci. We confirmed via conditional analysis that no secondary signals were present in suggestive loci. For each of the suggestive loci, the regional plots are shown in Supplementary Figure S4.

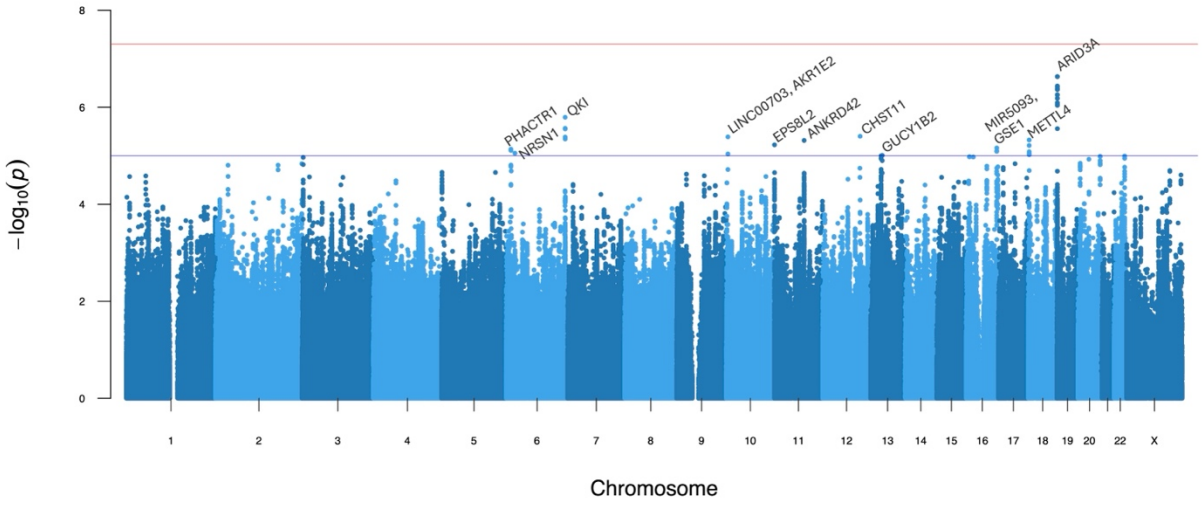

Figure 1. Manhattan plot for anti-Müllerian hormone levels
The solid red and blue lines denote the genome-wide significant and suggestive *p* value thresholds at $p < 5 \times 10^{-8}$ and $p < 1 \times 10^{-5}$, respectively. The peak SNV in each independent locus that surpassed the suggestive threshold is labeled with the nearby genes.

Table 2 Meta-analysis results for AMH levels presenting independent SNVs with *p* value ≤ 1×10^{-5}

Locus Information						2002–03 Family Study		2010 Soifua Manuia Study		Meta-Analysis	
Lead variant	Nearest Gene	Type	RDB	Samoan EAF	EUR EAF	β (SE)	*p*	β (SE)	*p*	*p*	PP
19-946163-G-C	*ARID3A*	intronic	5	0.453	0.123	−0.13 (0.13)	3.10×10^{-1}	−0.39 (0.08)	1.67×10^{-7}	2.32×10^{-7}	0.11
6-163620593-G-A	*QKI*	intergenic	5	0.132	*	−0.14 (0.18)	4.27×10^{-1}	−0.57 (0.12)	8.52×10^{-7}	1.61×10^{-6}	0.67
12-104584595-T-A	*CHST11*	intronic	6	0.845	0.524	−0.32 (0.18)	6.78×10^{-2}	−0.44 (0.10)	2.26×10^{-5}	3.98×10^{-6}	0.63
10-4474887-C-A	*AKR1E2*	intergenic	5	0.069	*	0.48 (0.24)	4.49×10^{-2}	0.63 (0.15)	3.36×10^{-5}	4.09×10^{-6}	0.40
18-2490805-C-T	*METTL4*	intergenic	7	0.787	0.346	−0.20 (0.15)	1.83×10^{-1}	−0.42 (0.09)	9.46×10^{-6}	4.73×10^{-6}	0.34
11-83219203-T-C	*ANKRD42*	intronic	4	0.285	0.347	−0.38 (0.14)	6.55×10^{-3}	−0.31 (0.08)	1.59×10^{-4}	4.83×10^{-6}	0.06
11-722202-G-C	*EPS8L2*	intronic	3a	0.412	0.194	0.33 (0.14)	1.57×10^{-2}	0.29 (0.08)	1.09×10^{-4}	5.96×10^{-6}	0.33
16-85420473-G-A	*GSE1*	intergenic	5	0.877	0.318	−0.27 (0.20)	1.87×10^{-1}	−0.48 (0.11)	1.39×10^{-5}	6.91×10^{-6}	0.59
6-12525440-G-A	*PHACTR1*	intergenic	7	0.097	0.119	0.09 (0.21)	6.78×10^{-1}	−0.67 (0.13)	2.71×10^{-7}	7.32×10^{-6}	0.17
6-23537402-C-G	*NRSN1*	intergenic	4	0.052	0.000	0.31 (0.26)	2.43×10^{-1}	0.82 (0.19)	1.31×10^{-5}	8.87×10^{-6}	0.77
13-51023905-C-T	*GUCY1B2*	intronic	7	0.117	0.279	0.79 (0.20)	5.73×10^{-5}	0.35 (0.12)	2.70×10^{-3}	9.84×10^{-6}	0.36

Each **variant** is presented with its gnomAD ID. We reported the effect allele (**EA**), its frequency (**EAF**), and the effect estimates (**β**) with standard error (**SE**). The **EUR EAF** reports the effect allele frequency in Europeans observed in 1000 Genome project as reported by gnomAD v4.[66] Asterisk (*) indicates that the variant is unobserved in 1000G. R^2 is the imputation quality in non-genotyped markers. **RDB** is the RegulomeDB functional class, with class scores ≤ 2 having greater probability of a transcriptional regulatory role. The posterior probability (**PP**) was calculated using PAINTOR.

We replicated one of the eight known AMH loci: *EIF4EBP1* (Supplementary Table S3, Supplementary Figure S7 and Supplementary Figure S8). Of the seven known AMH loci with lead variants not replicated here, the lead variants in three (*MCM8*, *CHECK2* and *CTB99A3.1*) were ultra-rare in Samoans (MAF < 0.0001). Within ±50 kb of two unreplicated loci (*AMH* and *TEX41*), we detected two SNVs with a significant p-value (2-144839456-A-T [$p = 0.0018$] and 19-2250470-G-A [$p = 0.0006$], respectively), however, they were not in LD with the known AMH loci.

The strongest GWAS association is in intron 3 of *ARID3A* at 19p13.3. The lead SNV, 19-946163-G-C ($p = 2.32 \times 10^{-7}$), and nearby SNVs are presented in Figure 2. The alternate allele of the lead SNV was associated with lower AMH levels (Supplementary Figure S5 and Supplementary Figure S6). This locus also harbors the known age-at-menopause variant 19-950694-G-A, observed in women of European ancestry, which is in high LD ($r^2 = 0.79$) with the lead AMH variant in this locus in this study [67,68]. Fine-mapping identified 19-982128-A-G, upstream of *WDR18*, as the most probable causal variant with a posterior probability (PP) of 0.34. The lead

AMH GWAS variant, 19-946163-G-C, (PP = 0.11) and age-at-menopause variant 19-950694-G-A (PP = 0.02) were the top two eQTLs affecting *ARID3A* expression in thyroid tissue ($p = 5.75 \times 10^{-7}$ and $p = 3.69 \times 10^{-7}$, respectively, both with an FDR = 2.60×10^{-10}).

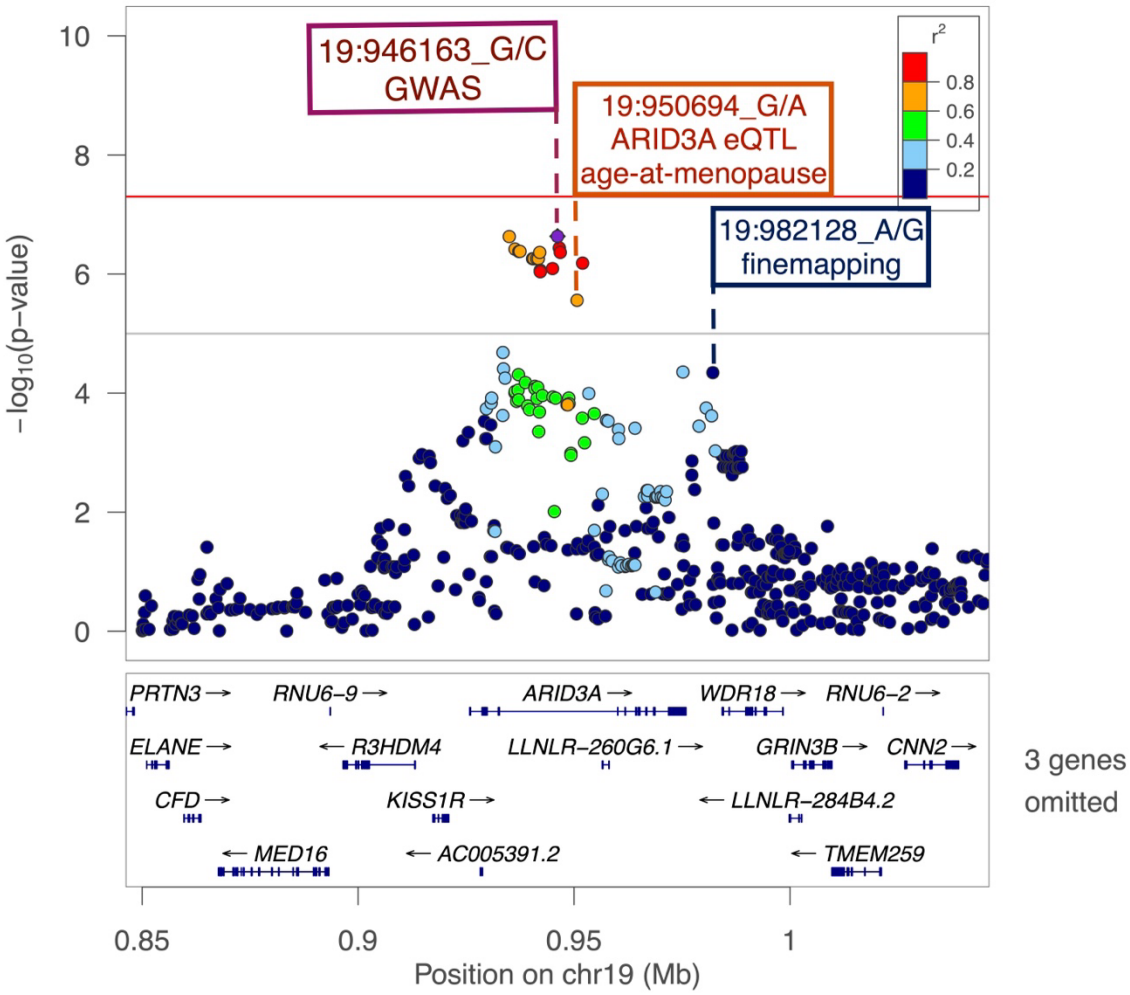

Figure 2. Regional plot for *ARID3A* locus

19-946163-G-C, the lead SNV, is a purple diamond. The color of all other SNVs reflects LD with the lead SNV as calculated in the Samoan samples. The solid red and grey line indicate the significance and suggestive thresholds at $p < 5 \times 10^{-8}$ and $p < 1 \times 10^{-5}$, respectively.

Gene-based analysis identified five significant associations with AMH levels: *ARID3A* ($p = 5 \times 10^{-10}$) and nearby *R3HDM4* ($p = 1.47 \times 10^{-9}$) as well as *P2RX6* ($p = 1.15 \times 10^{-6}$), *AC002472.1* ($p = 8.30 \times 10^{-7}$), and *PTPRB* ($p = 2.47\times 10^{-6}$) (Supplementary Figure S9).

There were seven transcriptome-wide significant genes observed in the TWAS of whole blood, endocrine tissues (adrenal gland, hypothalamus, ovary, pancreas, pituitary, thyroid), and metabolic tissues (liver, subcutaneous adipose tissue) in association with AMH levels. The strongest association of the seven was *GINS2*. *METTL4* was not only transcriptome-wide significance but also suggestively associated with AMH levels. The TWAS results are presented in Figure 3 and Supplementary Table S4.

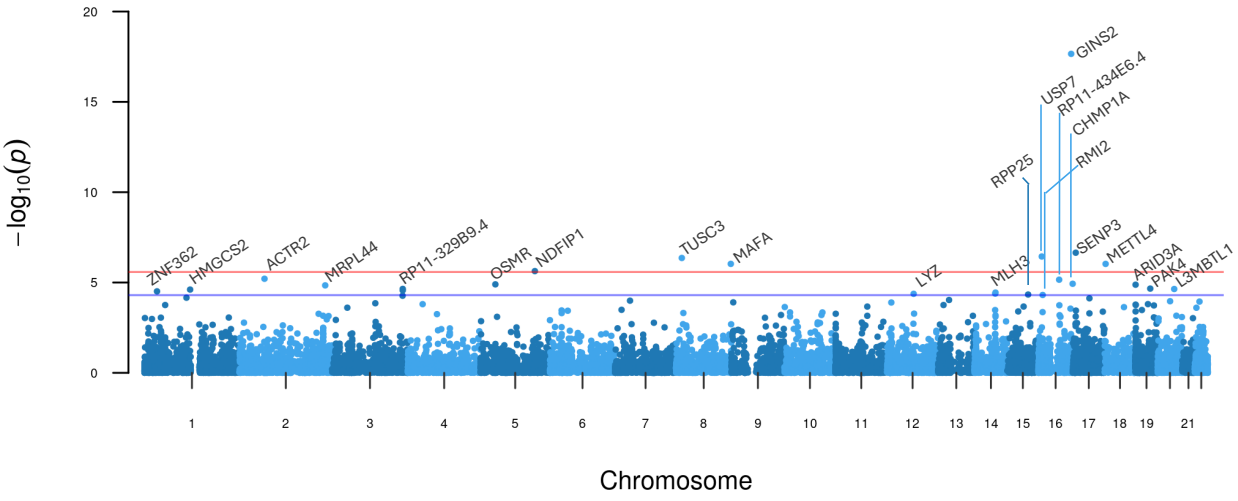

Figure 3 Manhattan plot of TWAS results for AMH

The solid red and blue lines denote the transcriptome-wide significant and suggestive *p* value thresholds at $p < 2.6 \times 10^{-6}$ and $p < 1 \times 10^{-5}$, respectively. The genes that surpassed the suggestive threshold are annotated.

We conducted a colocalization analysis using summary statistics from the GWAS and TWAS. In SNV-level and gene-level colocalization analysis, *ARID3A* had a locus-level colocalization probability (LCP) of 0.34 in the liver. All other colocalization probabilities for other genes and tissues were below the threshold of 0.30.

Discussion

Here we report the first study examining association between genetic variants and AMH levels in women from a Polynesian population. We identified eleven novel suggestive loci via GWAS (*ARID3A, QKI, CHST11, AKR1E2, METTL4, ANKRD42, EPS8L2, GSE1, EDN1, NRSN1,* and *GUCY1B2*) and seven significant genes via TWAS (*GINS2, SENP3, USP7, TUSC3, MAFA,*

METTL4, and *NDFIP1*). Additionally, we replicated the known AMH locus *EIF4EBP1* previously detected in women of European ancestry [34]. Among these findings, three genomic risk loci—*ARID3A, GSE1* and nearby *GINS2,* and *METTL4*—had associations in both GWAS and TWAS and are biologically plausible.

Our key GWAS finding was association of AMH with variants in the gene encoding AT-rich interaction domain 3A (*ARID3A*), a member of a family of proteins regulating chromatin binding. It was a also significant locus in gene-based analysis and was suggestively significant in the TWAS. The lead variant in *ARID3A* is located 4.5 kb upstream from a known age-at-menopause variant [67,68] and the two are in high LD. The biological link between AMH levels (a marker of ovarian follicle reserve) and age-at menopause is well recognized, as AMH testing is used to predict time-to-menopause in late reproductive age women in clinical practice [69,70]. Notably, both the lead *ARID3A* GWAS variant and the known age-at-menopause variant are positioned within GeneHancer [71] regulatory element GH19J000942, a target of which is the gene encoding KISS1 (kisspeptin) receptor (*KISS1R*). While there is not strong statistical evidence of gene-based or TWAS associations for *KISS1R* itself in this study, that does not rule it out as having a causal role in this locus. Kisspeptin/KISS1R signaling is pivotal in folliculogenesis [72]. Increased ovarian kisspeptin levels hinder the transition of primary follicle into antral stage by inhibiting the *FSHR* expression and increasing the AMH levels [72]. Observations in mice with conditional ablation of *Kiss1r* in oocytes suggest that the resulting deregulation may lead to premature ovarian failure [73]. Although the lead variant is in an intron of *ARID3A*, should this variant be causal or in LD with the causal variant, it may be acting through an effect on *KISS1R*.

In the TWAS, *GINS2* was the most significant gene and is within the *GSE1* locus on 16q24.1, which was suggestively significant in the GWAS. *GINS2* is highly conserved among eukaryotes and encodes one of the essential subunits that form the tetrameric Go-Ichi-Nii-San (GINS) complex, which has a key role in DNA replication [74,75]. *GINS2* is downregulated in ovaries of old-aged rhesus monkeys compared to young- and middle-aged ones and in atretic

bovine ovarian follicles compared to the healthy ones [76,77]. *GSE1*, an epigenetic regulator and a known oncogene, may also have a biological connection to AMH levels, as it has highest expression in the pituitary and the ovary [55,78]. The expression of *Gse1* in the primordial follicles is downregulated in estrogen receptor β (ERβ) knockout mice [79]. Loss of ERβ activates follicle growth and leads to early depletion of the ovary reserve, suggesting a regulatory role for *Gse1* in folliculogenesis [79]. AMH stimulates gonadotropin-releasing hormone (GnRH) expression in its role in the hypothalamic-pituitary-gonadal hormonal axis, and therefore, it is notable that *Gse1* is highly enriched in GnRH neurons upon gonadectomy [80,81].

The *METTL4* locus on 18p11.32 includes *METTL4* and *NDC80* and was identified by both GWAS and TWAS analyses (Supplementary Figure S4J). *METTL4* encodes methyltransferase 4, which is responsible for the adenine methylation involved in regulating RNA-splicing [82]. Epigenetic regulation via N^6-methyladenosine modifications plays an active role in response to environmental stressors (hypoxia, starvation, toxicants, etc.) and has been reported to be a relevant mechanism in the development of PCOS and premature ovarian insufficiency [83–87]. While *METTL3*, a paralog of *METTL4*, has been implicated in follicle development and fertility by regulating the stability of oocyte meiotic maturation-related transcripts in mice, not much is known about the role of *METTL4* in female reproduction [82,88,89]. Nearby *NDC80* encodes a component of the nuclear division cycle 80 kinetochore complex [90,91]. In mouse oocytes, this complex partakes in initiating oocyte maturation by enabling the spindle assembly required for G2/M transition by stabilizing cyclin B2 levels [91–95]. Consequently, the dormant primordial follicles arrested since birth at G2/M resume meiosis [91–94]. Association of variants near *METTL4* may indicate that genes that control the G2/M transition have a biological connection to modulation AMH levels.

We replicated one of the eight known AMH loci—intergenic variant 8-38015258-C-T (rs10093345) near *EIF4EBP1*. This SNV was also significantly associated with age-at-menopause in the UK Biobank [96]. Notably, lead variants in four known AMH loci—*AMH*, *CHECK2*,

CTB-99A3.1, and *MCM8*—are low-frequency or rare variants in Europeans; these same variants in Samoans are either rare with low Samoan-specific imputation quality ($R^2 < 0.10$) or are ultra-rare (MAF < 0.0007). Such population-specific susceptibility loci highlight differences in AMH genetic architecture between Samoan individuals and individuals of European ancestry and highlight the importance of diversifying study populations. Recently, Moolhuijsen et al. investigated whether *AMH* promoter variation affects serum AMH levels in PCOS patients of Northern European ancestry and observed an association between rs10406324 (19-2249113-G-A) and lower AMH levels that was independent of follicle count and other PCOS markers [33]. Similar findings were also observed in normo-ovulatory women [32]. This suggests that genetic factors can contribute to variation in AMH levels independent from follicle count and/or PCOS status. In this study, rs10406324 could not be analyzed due to low MAF (< 0.005) and poor imputation quality ($R^2 <$ 0.10).

Strengths and limitations

While our sample was small compared to other published GWASs, we were uniquely positioned to identify variants that may be rare in other populations but common in Samoans due to population founder effects. Additionally, the genetic homogeneity of the Soifua Manuia [24] sample could result in better power to detect variants associated with AMH levels.

Measuring AMH levels in older women is a challenge, as AMH levels fall below detectable ranges during perimenopause. We addressed this challenge by employing two AMH assays from Ansh Lab: picoAMH and ultra-sensitive AMH/MIS ELISA kits for older women and younger women, respectively. These assays utilize the same antibodies and calibrators, with picoAMH extending coverage for the lower ranges of the standard curve [97]. To reduce the heterogeneity, we used the conversion factor from Ansh Lab to harmonize the AMH levels between the two age groups. Additionally, we robustly accounted for the AMH values below detectable limits in our GWASs leveraging the Cox's proportional hazards regression model via coxmeg.

For GWAS findings, although we cannot entirely rule out the possibility of false positive (FP) results and acknowledge that a plausible story for those can be made easily [98], we aimed to mitigate this by prioritizing the findings with evidence from both GWAS and TWAS analyses. Furthermore, while we identified several significant TWAS associations, absence of Samoan-specific eQTLs and limited representation of diverse populations in GTEx could have led to reduced power and detection of fewer eQTL associations.

Conclusion

Overall, eleven loci were suggestively associated with AMH levels in a meta-analysis of Samoan women, several of which have plausible links to ovarian function/folliculogenesis (*KISS1R, GINS2,* and *NDC80*). This study provides valuable insights into the genetic variation of AMH in Samoan women and replicates the previously detected association of *EIF4BP1* with AMH levels. The putative novel findings in this study will need to be validated in additional larger studies. The identification of variations affecting AMH levels, such as seen in our findings, may also improve understanding of the biological underpinnings of AMH-related reproductive traits such as ovarian function, age at menopause, premature ovarian failure, and PCOS. Eventually, these findings may contribute to the development of screening tools measuring the genetic susceptibility for AMH-related traits.

Authors' role

Z.E.-Y. conceptualized the study, performed the analyses, created visual representations of the results, contributed to the interpretation of the results and took the lead in writing the manuscript. M.K. and J.Z.Z. imputed the genotypes. G.L.-M. conducted the biomarker analyses. G.L.-M., T.N., S.V., N.L.H., S.T.M., D.E.W. and N.L.H. provided resources and curated data. J.C.C. set up the analysis pipeline and supported methodological approach. D.E.W. and R.L.M. supervised the project, provided statistical and subject matter expertise, and helped develop the study design.

J.C.C., G.L.-M., T.N., S.V., N.L.H., S.T.M., D.E.W. and R.L.M. contributed to interpretation of the results and revision of the article. All authors read and approved the final manuscript.

Data Availability Statement

The full AMH GWAS summary statistics will made available after publication through the GWAS catalog (https://www.ebi.ac.uk/gwas/).

Acknowledgements

The authors thank all study participants for their participation and contribution to this research. We acknowledge the assistance of the Samoa Ministry of Health and the Samoa Bureau of Statistics for their guidance and support in the conduct of this study. We thank the local village officials for their help and the participants for their generosity. We further acknowledge helpful comments of the TOPMed Reproductive Health Working Group.

Funding

This work was funded by NIH Grants R01-HL093093 (PI: S.T.M.), R01-HL133040 (PI: R.L.M.), and T90-DE030853 (PI: C.S. Sfeir). Molecular data for the Trans-Omics in Precision Medicine (TOPMed) Program was supported by the National Heart, Lung and Blood Institute (NHLBI). Genome sequencing for the Soifua Manuia study, labeled as “NHLBI TOPMed: Genome-wide Association Study of Adiposity in Samoans” (phs000972) in the dbGaP, was performed at the Northwest Genomics Center (HHSN268201100037C) and the New York Genome Center (HHSN268201500016C). Core support including centralized genomic read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01-HL117626-02S1; contract HHSN268201800002I). Core support including phenotype harmonization, data management, sample-identity QC, and general program coordination were provided by the TOPMed Data Coordinating Center (R01-HL120393; U01-

HL120393; contract HHSN268201800001I). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest

None declared.

References

[1] Jost, P.A. (1948) Le controle hormonal de la différenciation du sexe. *Biological Reviews*, **23**, 201–236.

[2] Hagen, C.P., Aksglaede, L., Sørensen, K., Main, K.M., Boas, M., Cleemann, L., Holm, K., Gravholt, C.H., Andersson, A.M., Pedersen, A.T., *et al.* (2010) Serum levels of anti-Müllerian hormone as a marker of ovarian function in 926 healthy females from birth to adulthood and in 172 Turner syndrome patients. *J Clin Endocrinol Metab*, **95**, 5003–5010.

[3] Kuiri-Hänninen, T., Kallio, S., Seuri, R., Tyrväinen, E., Liakka, A., Tapanainen, J., Sankilampi, U. and Dunkel, L. (2011) Postnatal developmental changes in the pituitary-ovarian axis in preterm and term infant girls. *J Clin Endocrinol Metab*, **96**, 3432–3439.

[4] Weenen, C., Laven, J.S.E., von Bergh, A.R.M., Cranfield, M., Groome, N.P., Visser, J.A., Kramer, P., Fauser, B.C.J.M. and Themmen, A.P.N. (2004) Anti-Müllerian hormone expression pattern in the human ovary: potential implications for initial and cyclic follicle recruitment. *Mol Hum Reprod*, **10**, 77–83.

[5] Durlinger, A.L.L., Gruijters, M.J.G., Kramer, P., Karels, B., Kumar, T.R., Matzuk, M.M., Rose, U.M., De Jong, F.H., Uilenbroek, J.T.J., Grootegoed, J.A., *et al.* (2001) Anti-Müllerian hormone attenuates the effects of FSH on follicle development in the mouse ovary. *Endocrinology*, **142**, 4891–4899.

[6] Andersen, C.Y. and Byskov, A.G. (2006) Estradiol and regulation of anti-Müllerian hormone, inhibin-A, and inhibin-B secretion: analysis of small antral and preovulatory human follicles' fluid. *J Clin Endocrinol Metab*, **91**, 4064–4069.

[7] Christiansen, S.C., Eilertsen, T.B., Vanky, E. and Carlsen, S.M. (2016) Does AMH reflect follicle number similarly in women with and without PCOS? *PLoS One*, **11**, e0146739.

[8] Van Rooij, I.A.J., Broekmans, F.J.M., Te Velde, E.R., Fauser, B.C.J.M., Bancsi, L.F.J.M.M., De Jong, F.H. and Themmen, A.P.N. (2002) Serum anti-Müllerian hormone levels: a novel measure of ovarian reserve. *Human Reproduction*, **17**, 3065–3071.

[9] Kotanidis, L., Asimakoppoulos, B. and Nikolettos, N. (2013) Association between AMH, oocyte number and availability of embryos for cryopreservation in IVF. *In Vivo (Brooklyn)*, **27**, 877–880.

[10] Moolhuijsen, L.M.E. and Visser, J.A. (2020) Anti-Müllerian hormone and ovarian reserve: update on assessing ovarian function. *J Clin Endocrinol Metab*, **105**, 3361.

[11] Tal, R. and Seifer, D.B. (2017) Ovarian reserve testing: a user's guide. *Am J Obstet Gynecol*, **217**, 129–140.

[12] Dewailly, D., Robin, G., Peigne, M., Decanter, C., Pigny, P. and Catteau-Jonard, S. (2016) Interactions between androgens, FSH, anti-Müllerian hormone and estradiol during folliculogenesis in the human normal and polycystic ovary. *Hum Reprod Update*, **22**, 709–724.

[13] Fraissinet, A., Robin, G., Pigny, P., Lefebvre, T., Catteau-Jonard, S. and Dewailly, D. (2017) Use of the serum anti-Müllerian hormone assay as a surrogate for polycystic ovarian morphology: impact on diagnosis and phenotypic classification of polycystic ovary syndrome. *Human Reproduction*, **32**, 1716–1722.

[14] Richardson, S.J., Senikas, V. and Nelson, J.F. (1987) Follicular depletion during the menopausal transition: evidence for accelerated loss and ultimate exhaustion. *J Clin Endocrinol Metab*, **65**, 1231–1237.

[15] Lie Fong, S., Laven, J.S.E., Duhamel, A. and Dewailly, D. (2017) Polycystic ovarian morphology and the diagnosis of polycystic ovary syndrome: redefining threshold levels for follicle count and serum anti-Müllerian hormone using cluster analysis. *Human Reproduction*, **32**, 1723–1731.

[16] Abbara, A., Eng, P.C., Phylactou, M., Clarke, S.A., Hunjan, T., Roberts, R., Vimalesvaran, S., Christopoulos, G., Islam, R., Purugganan, K., *et al.* (2019) Anti-Müllerian hormone (AMH) in the diagnosis of menstrual disturbance due to polycystic ovarian syndrome. *Front Endocrinol (Lausanne)*, **10**, 656.

[17] Capuzzo, M. and La Marca, A. (2021) Use of AMH in the differential diagnosis of anovulatory disorders including PCOS. *Front Endocrinol (Lausanne)*, **11**, 616766.

[18] Sahmay, S., Aydin, Y., Oncul, M. and Senturk, L.M. (2014) Diagnosis of polycystic ovary syndrome: AMH in combination with clinical symptoms. *J Assist Reprod Genet*, **31**, 213–220.

[19] Singh, A.K. and Singh, R. (2015) Can anti-Müllerian hormone replace ultrasonographic evaluation in polycystic ovary syndrome? A review of current progress. *Indian J Endocrinol Metab*, **19**, 731–743.

[20] Maredia, H., Lambert-Messerlian, G.M., Palomaki, G.E., Viali, S., Hawley, N.L. and Mcgarvey, S.T. (2016) Cut-off levels for hyperandrogenemia among Samoan women: an improved methodology for deriving normative data in an obese population. *Clin Biochem*, **49**, 782–786.

[21] Maredia, H., Hawley, N.L., Lambert-Messerlian, G., Fidow, U., Reupena, M.S., Naseri, T., McGarvey, S.T. and McGarvey, S.T. (2018) Reproductive health, obesity, and cardiometabolic risk factors among Samoan women. *Am J Hum Biol*, **30**, e23106.

[22] Lambert-Messerlian, G., Roberts, M.B., Urlacher, S.S., Ah-Ching, J., Viali, S., Urbanek, M. and McGarvey, S.T. (2011) First assessment of menstrual cycle function and reproductive endocrine status in Samoan women. *Human Reproduction*, **26**, 2518–2524.

[23] Lowe, J.K., Maller, J.B., Pe'er, I., Neale, B.M., Salit, J., Kenny, E.E., Shea, J.L., Burkhardt, R., Smith, J.G., Ji, W., *et al.* (2009) Genome-wide association studies in an isolated founder population from the Pacific Island of Kosrae. *PLoS Genet*, **5**, e1000365.

[24] Minster, R.L., Hawley, N.L., Su, C.T., Sun, G., Kershaw, E.E., Cheng, H., Buhule, O.D., Lin, J., Reupena, M.S., Viali, S., *et al.* (2016) A thrifty variant in CREBRF strongly influences body mass index in Samoans. *Nat Genet*, **48**, 1049–1054.

[25] Hatzikotoulas, K., Gilly, A. and Zeggini, E. (2014) Using population isolates in genetic association studies. *Brief Funct Genomics*, **13**, 371–377.

[26] Uffelmann, E., Huang, Q.Q., Munung, N.S., de Vries, J., Okada, Y., Martin, A.R., Martin, H.C., Lappalainen, T. and Posthuma, D. (2021) Genome-wide association studies. *Nature Reviews Methods Primers*, **1**, 59.

[27] Harris, D.N., Kessler, M.D., Shetty, A.C., Weeks, D.E., Minster, R.L., Browning, S., Cochrane, E.E., Deka, R., Hawley, N.L., Reupena, M.S., *et al.* (2020) Evolutionary history of modern Samoans. *Proc Natl Acad Sci USA*, **117**, 9458–9465.

[28] Dapas, M., Lin, F.T.J., Nadkarni, G.N., Sisk, R., Legro, R.S., Urbanek, M., Geoffrey Hayes, M. and Dunaif, A. (2020) Distinct subtypes of polycystic ovary syndrome with novel genetic associations: an unsupervised, phenotypic clustering analysis. *PLoS Med*, **17**, e1003132.

[29] Burns, K., Mullin, B.H., Moolhuijsen, L.M.E., Laisk, T., Tyrmi, J.S., Cui, J., Actkins, K. V., Louwers, Y. V., Davis, L.K., Dudbridge, F., *et al.* (2024) Body mass index stratified meta-analysis of genome-wide association studies of polycystic ovary syndrome in women of European ancestry. *BMC Genomics*, **25**, 208.

[30] 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.

[31] Ruth, K.S., Soares, A.L.G., Borges, M.C., Eliassen, A.H., Hankinson, S.E., Jones, M.E., Kraft, P., Nichols, H.B., Sandler, D.P., Schoemaker, M.J., *et al.* (2019) 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. *Hum Mol Genet*, **28**, 1392–1401.

[32] Verdiesen, R.M.G., van der Schouw, Y.T., van Gils, C.H., Verschuren, W.M.M., Broekmans, F.J.M., Borges, M.C., Gonçalves Soares, A.L., Lawlor, D.A., Eliassen, A.H., Kraft, P., *et al.* (2022) Genome-wide association study meta-analysis identifies three novel loci for circulating anti-Müllerian hormone levels in women. *Human Reproduction*, **37**, 1069–1082.

[33] Moolhuijsen, L.M.E., Louwers, Y. V., McLuskey, A., Broer, L., Uitterlinden, A.G., Verdiesen, R.M.G., Sisk, R.K., Dunaif, A., Laven, J.S.E. and Visser, J.A. (2022) Association between an AMH promoter polymorphism and serum AMH levels in PCOS patients. *Human Reproduction*, **37**, 1544–1556.

[34] Pujol-Gualdo, N., Karjalainen, M.K., Võsa, U., Arffman, R.K., Mägi, R., Ronkainen, J., Laisk, T. and Piltonen, T.T. (2024) Circulating anti-Müllerian hormone levels in pre-menopausal women: novel genetic insights from a genome-wide association meta-analysis. *Human Reproduction*, **39**, 1564–1572.

[35] Seifer, D.B., Golub, E.T., Lambert-Messerlian, G., Benning, L., Anastos, K., Watts, D.H., Cohen, M.H., Karim, R., Young, M.A., Minkoff, H., *et al.* (2009) Variations in serum Müllerian inhibiting substance between white, black, and Hispanic women. *Fertil Steril*, **92**, 1674–1678.

[36] Gromski, P.S., Patil, R.S., Chougule, S.M., Bhomkar, D.A., Jirge, P.R. and Nelson, S.M. (2022) Ethnic discordance in serum anti-Müllerian hormone in European and Indian healthy women and Indian infertile women. *Reprod Biomed Online*, **45**, 979–986.

[37] Chen, S., Francioli, L.C., Goodrich, J.K., Collins, R.L., Kanai, M., Wang, Q., Alföldi, J., Watts, N.A., Vittal, C., Gauthier, L.D., *et al.* (2023) A genomic mutational constraint map using variation in 76,156 human genomes. *Nature 2023 625:7993*, **625**, 92–100.

[38] Kichaev, G., Bhatia, G., Loh, P.R., Gazal, S., Burch, K., Freund, M.K., Schoech, A., Pasaniuc, B. and Price, A.L. (2019) Leveraging polygenic functional enrichment to improve GWAS power. *Am J Hum Genet*, **104**, 65–75.

[39] Hawley, N.L., Minster, R.L., Weeks, D.E., Viali, S., Reupena, M.S., Sun, G., Cheng, H., Deka, R. and McGarvey, S.T. (2014) Prevalence of adiposity and associated cardiometabolic risk factors in the Samoan genome-wide association study. *American Journal of Human Biology*, **26**, 491–501.

[40] Carlson, J.C., Krishnan, M., Liu, S., Anderson, K.J., Zhang, J.Z., Yapp, T.-A.J., Chiyka, E.A., Dikec, D.A., Cheng, H., Naseri, T., *et al.* Improving imputation quality in Samoans through the integration of population-specific sequences into existing reference panels. *medRxiv*.

[41] Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., *et al.* (2016) Next-generation genotype imputation service and methods. *Nat Genet*, **48**, 1284–1287.

[42] He, L. and Kulminski, A.M. (2020) Fast algorithms for conducting large-scale gwas of age-at-onset traits using cox mixed-effects models. *Genetics*, **215**, 41–58.

[43] Cox, D.R. (1972) Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, **34**, 187–202.

[44] Dinse, G.E., Jusko, T.A., Ho, L.A., Annam, K., Graubard, B.I., Hertz-Picciotto, I., Miller, F.W., Gillespie, B.W. and Weinberg, C.R. (2014) Accommodating Measurements Below a Limit of Detection: A Novel Application of Cox Regression. *Am J Epidemiol*, **179**, 1018–1024.

[45] Gogarten, S.M., Sofer, T., Chen, H., Yu, C., Brody, J.A., Thornton, T.A., Rice, K.M. and Conomos, M.P. (2019) Genetic association testing using the GENESIS R/Bioconductor package. *Bioinformatics*, **35**, 5346–5348.

[46] Willer, C.J., Li, Y. and Abecasis, G.R. (2010) METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**, 2190–2191.

[47] Paria, S.S., Rahman, S.R. and Adhikari, K. (2022) fastman: a fast algorithm for visualizing GWAS results using Manhattan and Q-Q plots. *bioRxiv*.

[48] Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R. and Willer, C.J. (2010) LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics*, Oxford University Press, **26**, 2336–2337.

[49] Kichaev, G., Yang, W.-Y., Lindstrom, S., Hormozdiari, F., Eskin, E., Price, A., Kraft, P. and Pasaniuc, B. (2014) Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet*, **10**, e1004722.

[50] McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P. and Cunningham, F. (2016) The Ensembl Variant Effect Predictor. *Genome Biol*, **17**, 122.

[51] Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S., *et al.* (2012) Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*, **22**, 1790–1797.

[52] Watanabe, K., Taskesen, E., Van Bochoven, A. and Posthuma, D. (2017) Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*, **8**, 1826.

[53] Maintainer BP. (2022) liftOver: Changing genomic coordinate systems with rtracklayer::liftOver.

[54] de Leeuw, C.A., Mooij, J.M., Heskes, T. and Posthuma, D. (2015) MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol*, **11**, e1004219.

[55] The GTEx consortium (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science (1979)*, **369**, 1318–1330.

[56] Gao, X., Starmer, J. and Martin, E.R. (2008) A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genet Epidemiol*, **32**, 361–369.

[57] Barbeira, A.N., Dickinson, S.P., Bonazzola, R., Zheng, J., Wheeler, H.E., Torres, J.M., Torstenson, E.S., Shah, K.P., Garcia, T., Edwards, T.L., *et al.* (2018) Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun*, **9**, 1825.

[58] Barbeira, A.N., Pividori, M.D., Zheng, J., Wheeler, H.E., Nicolae, D.L. and Im, H.K. (2019) Integrating predicted transcriptome from multiple tissues improves association detection. *PLoS Genet*, **15**, e1007889.

[59] Carithers, L.J., Ardlie, K., Barcus, M., Branton, P.A., Britton, A., Buia, S.A., Compton, C.C., Deluca, D.S., Peter-Demchok, J., Gelfand, E.T., *et al.* (2015) A novel approach to high-quality postmortem tissue procurement: the GTEx project. *Biopreserv Biobank*, **13**, 311–317.

[60] Barbeira, A.N., Melia, O.J., Liang, Y., Bonazzola, R., Wang, G., Wheeler, H.E., Aguet, F., Ardlie, K.G., Wen, X. and Im, H.K. (2020) Fine-mapping and QTL tissue-sharing information improves the reliability of causal gene identification. *Genet Epidemiol*, **44**, 854–867.

[61] Hukku, A., Sampson, M.G., Luca, F., Pique-Regi, R. and Wen, X. (2022) Analyzing and reconciling colocalization and transcriptome-wide association studies from the perspective of inferential reproducibility. *Am J Hum Genet*, **109**, 825.

[62] Hukku, A., Pividori, M., Luca, F., Pique-Regi, R., Im, H.K. and Wen, X. (2021) Probabilistic colocalization of genetic variants from complex and molecular traits: promise and limitations. *Am J Hum Genet*, **108**, 25.

[63] Pividori, M., Rajagopal, P.S., Barbeira, A., Liang, Y., Melia, O., Bastarache, L., Park, Y.S., Consortium, Gte., Wen, X. and Im, H.K. (2020) PhenomeXcan: Mapping the genome to the phenome through the transcriptome. *Sci Adv*, **6**, eaba2083.

[64] Wen, X., Pique-Regi, R. and Luca, F. (2017) Integrating molecular QTL data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization. *PLoS Genet*, **13**, e1006646.

[65] Wen, X. (2016) Molecular QTL discovery incorporating genomic annotations using Bayesian false discovery rate control. *Ann Appl Stat*, **10**, 1619–1638.

[66] Chen, S., Francioli, L.C., Goodrich, J.K., Collins, R.L., Kanai, M., Wang, Q., Alföldi, J., Watts, N.A., Vittal, C., Gauthier, L.D., *et al.* (2023) A genomic mutational constraint map using variation in 76,156 human genomes. *Nature*, **625**, 92–100.

[67] Kichaev, G., Bhatia, G., Loh, P.R., Gazal, S., Burch, K., Freund, M.K., Schoech, A., Pasaniuc, B. and Price, A.L. (2019) Leveraging polygenic functional enrichment to improve GWAS power. *Am J Hum Genet*, **104**, 65–75.

[68] 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.

[69] de Kat, A.C., Broekmans, F.J.M. and Lambalk, C.B. (2021) Role of AMH in prediction of menopause. *Front Endocrinol (Lausanne)*, **12**, 1078.

[70] 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.

[71] Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Stein, T.I., Rosen, N., Kohn, A., Twik, M., Safran, M., *et al.* (2017) GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)*.

[72] Cao, Y., Li, Z., Jiang, W., Ling, Y. and Kuang, H. (2019) Reproductive functions of kisspeptin/KISS1R systems in the periphery. *Reproductive Biology and Endocrinology*, **17**, 65.

[73] Ruohonen, S.T., Gaytan, F., Usseglio Gaudi, A., Velasco, I., Kukoricza, K., Perdices-Lopez, C., Franssen, D., Guler, I., Mehmood, A., Elo, L.L., *et al.* (2022) Selective loss of kisspeptin signaling in oocytes causes progressive premature ovulatory failure. *Hum Reprod*, **37**, 806–821.

[74] Takayama, Y., Kamimura, Y., Okawa, M., Muramatsu, S., Sugino, A. and Araki, H. (2003) GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast. *Genes Dev*, **17**, 1153.

[75] Li, Z. and Xu, X. (2019) Post-translational modifications of the mini-chromosome maintenance proteins in DNA replication. *Genes (Basel)*, **10**, 331.

[76] Wei, H., Liu, X., Yuan, J., Li, L., Zhang, D., Guo, X., Liu, L. and Zhang, S. (2015) Age-specific gene expression profiles of rhesus monkey ovaries detected by microarray analysis. *Biomed Res Int*, **2015**, 625192.

[77] Hatzirodos, N., Irving-Rodgers, H.F., Hummitzsch, K. and Rodgers, R.J. (2014) Transcriptome profiling of the theca interna from bovine ovarian follicles during atresia. *PLoS One*, **9**, 99706.

[78] Meng, Z., Yang, Y., Li, S., Huang, L., Yao, Z., Chen, Y., Wang, J., Shen, Y., Liang, P., Zhang, H., *et al.* (2024) GSE1 promotes the proliferation and migration of lung adenocarcinoma cells by downregulating KLF6 expression. *Cell Biol Int*, **48**, 1490–1506.

[79] Lee, E.B., Chakravarthi, V.P., Mohamadi, R., Dahiya, V., Vo, K., Ratri, A., Fields, P.E., Marsh, C.A. and Rumi, M.A.K. (2024) Loss of ERβ disrupts gene regulation in primordial and primary follicles. *Int J Mol Sci*, **25**, 3202.

[80] Burger, L.L., Vanacker, C., Phumsatitpong, C., Wagenmaker, E.R., Wang, L., Olson, D.P. and Moenter, S.M. (2018) Identification of genes enriched in GnRH neurons by translating ribosome affinity purification and RNAseq in mice. *Endocrinology*, **159**, 1922–1940.

[81] Oride, A., Kanasaki, H., Tumurbaatar, T., Tumurgan, Z., Okada, H. and Kyo, S. (2021) Effect of anti-Müllerian hormone in hypothalamic Kiss-1- and GnRH-producing cell models. *Gynecol Endocrinol*, **37**, 841–847.

[82] Chen, H., Gu, L., Orellana, E.A., Wang, Y., Guo, J., Liu, Q., Wang, L., Shen, Z., Wu, H., Gregory, R.I., *et al.* (2020) METTL4 is an snRNA m6Am methyltransferase that regulates RNA splicing. *Cell Res*, **30**, 544–547.

[83] Sheng, Y., Pan, B., Wei, F., Wang, Y. and Gao, S. (2021) Case study of the response of N6-methyladenine DNA modification to environmental stressors in the unicellular eukaryote tetrahymena thermophila. *mSphere*, **6**, e0120820.

[84] Priya, K., Setty, M., Babu, U.V. and Pai, K.S.R. (2021) Implications of environmental toxicants on ovarian follicles: how it can adversely affect the female fertility? *Environmental Science and Pollution Research 2021 28:48*, **28**, 67925–67939.

[85] Xu, C.L., Tan, Q.Y., Yang, H., Li, C.Y., Wu, Z. and Ma, Y.F. (2022) Melatonin enhances spermatogonia activity through promoting KIAA1429-mediated m6A deposition to activate the PI3K/AKT signaling. *Reprod Biol*, **22**, 100681.

[86] Szukiewicz, D., Trojanowski, S., Kociszewska, A. and Szewczyk, G. (2022) Modulation of the inflammatory response in polycystic ovary syndrome (PCOS) - Searching for epigenetic factors. *Int J Mol Sci*, **23**, 14663.

[87] Ding, C., Zou, Q., Ding, J., Ling, M., Wang, W., Li, H. and Huang, B. (2018) Increased N6-methyladenosine causes infertility is associated with FTO expression. *J Cell Physiol*, **233**, 7055–7066.

[88] Mu, H., Zhang, T., Yang, Y., Zhang, D., Gao, J., Li, J., Yue, L., Gao, D., Shi, B., Han, Y., *et al.* (2021) METTL3-mediated mRNA N6-methyladenosine is required for oocyte and follicle development in mice. *Cell Death Dis*, **12**, 989.

[89] Kweon, S.M., Chen, Y., Moon, E., Kvederaviciutė, K., Klimasauskas, S. and Feldman, D.E. (2019) An adversarial DNA N6-methyladenine-sensor network preserves polycomb silencing. *Mol Cell*, **74**, 1138-1147.e6.

[90] Wei, R.R., Sorger, P.K. and Harrison, S.C. (2005) Molecular organization of the Ndc80 complex, an essential kinetochore component. *Proc Natl Acad Sci USA*, **102**, 5363–5367.

[91] Sanders, J.R. and Jones, K.T. (2018) Regulation of the meiotic divisions of mammalian oocytes and eggs. *Biochem Soc Trans*, **46**, 797–806.

[92] Sun, S.C., Zhang, D.X., Lee, S.E., Xu, Y.N. and Kim, N.H. (2011) Ndc80 regulates meiotic spindle organization, chromosome alignment, and cell cycle progression in mouse oocytes. *Microscopy and Microanalysis*, **17**, 431–439.

[93] Gui, L. and Homer, H. (2013) Hec1-dependent cyclin B2 stabilization Regulates the G2-M transition and early prometaphase in mouse oocytes. *Dev Cell*, **25**, 43–54.

[94] Takenoshita, Y., Hara, M. and Fukagawa, T. (2022) Recruitment of two Ndc80 complexes via the CENP-T pathway is sufficient for kinetochore functions. *Nat Commun*, **13**, 851.

[95] Lemonnier, T., Dupré, A. and Jessus, C. (2020) The G2-to-M transition from a phosphatase perspective: a new vision of the meiotic division. *Cell Div*, **15**, 9.

[96] Ojavee, S.E., Darrous, L., Patxot, M., Läll, K., Fischer, K., Mägi, R., Kutalik, Z. and Robinson, M.R. (2023) Genetic insights into the age-specific biological mechanisms governing human ovarian aging. *Am J Hum Genet*, **110**, 1549–1563.

[97] Su, H.I., Sammel, M.D., Homer, M. V., Bui, K., Haunschild, C. and Stanczyk, F.Z. (2014) Comparability of AMH levels among commercially available immunoassays. *Fertil Steril*, **101**, 1766.

[98] Biedrzycki, R.J., Sier, A.E., Liu, D., Dreikorn, E.N. and Weeks, D.E. (2019) Spinning convincing stories for both true and false association signals. *Genet Epidemiol*, **43**, 356–364.