

## Genome-wide association study meta-analysis identifies three novel loci for circulating anti-Müllerian hormone levels in women

### Authors:

Renée MG Verdiesen,<sup>1\*</sup> Yvonne T van der Schouw,<sup>1</sup> Carla H van Gils,<sup>1</sup> WM Monique Verschuren,<sup>1,2</sup> Frank JM Broekmans,<sup>3</sup> Maria C Borges,<sup>4,5</sup> Ana LG Soares,<sup>4,5</sup> Deborah A Lawlor,<sup>4-6</sup> A Heather Eliassen,<sup>7,8</sup> Peter Kraft,<sup>8,9</sup> Dale P Sandler,<sup>10</sup> Sioban D Harlow,<sup>11</sup> Jennifer A Smith,<sup>11</sup> Nanette Santoro,<sup>12</sup> Minouk J Schoemaker,<sup>13</sup> Anthony J Swerdlow,<sup>13, 14</sup> Anna Murray,<sup>15</sup> Katherine S Ruth,<sup>15</sup> N Charlotte Onland-Moret<sup>1\*\*</sup>

### Author affiliations:

<sup>1</sup> Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, PO Box 85500, 3508 GA, The Netherlands

<sup>2</sup> Centre for Nutrition, Prevention and Health Services, National Institute for Public Health and the Environment (RIVM), 3720 BA, Bilthoven, The Netherlands

<sup>3</sup> Department of Reproductive Medicine and Gynecology, University Medical Center Utrecht, Utrecht University, Utrecht, 3584 CX, The Netherlands

<sup>4</sup> MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, BS8 2BN, UK

<sup>5</sup> Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, BS8 2BN, UK

<sup>6</sup> Bristol NIHR Biomedical Research Centre, Bristol, UK

<sup>7</sup> Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston MA, 02115, USA

<sup>8</sup> Department of Epidemiology, Harvard TH Chan School of Public Health, Boston MA, 02115, USA

<sup>9</sup> Department of Biostatistics, Harvard TH Chan School of Public Health, Boston MA, 02115, USA

<sup>10</sup> Epidemiology Branch, National Institute of Environmental Health Sciences, Durham, NC, 27709, USA

<sup>11</sup> Department of Epidemiology, University of Michigan, Michigan, 48109-2029, USA

<sup>12</sup> Department of Obstetrics and Gynecology, University of Colorado School of Medicine, Colorado, 80045, USA

<sup>13</sup> Division of Genetics and Epidemiology, The Institute of Cancer Research, London, SM2 5NG, UK

<sup>14</sup> Division of Breast Cancer Research, The Institute of Cancer Research, London, SW7 3RP, UK

<sup>15</sup> University of Exeter Medical School, University of Exeter, Exeter, EX2 5DW, UK

**Correspondence:** \*r.m.g.verdiesen@umcutrecht.nl; \*\*n.c.onland@umcutrecht.nl

## Abstract

Anti-Müllerian hormone (AMH) is expressed by antral stage ovarian follicles in women. Consequently, circulating AMH levels are detectable until menopause. Variation in age-specific AMH levels has been associated with breast cancer and polycystic ovary syndrome (PCOS), amongst other diseases.

Identification of genetic variants underlying variation in AMH levels could provide clues about the physiological mechanisms that explain these AMH-disease associations. To date, only one variant in *MCM8* has been identified to be associated with circulating AMH levels in women. We aimed to identify additional variants for AMH through a GWAS meta-analysis including data from 7049 premenopausal women of European ancestry, which more than doubles the sample size of the largest previous GWAS. We identified four loci associated with AMH levels at  $p < 5 \times 10^{-8}$ : the previously reported *MCM8* locus and three novel signals in or near *AMH*, *TEX41*, and *CDCA7*. The strongest signal was a missense variant in the *AMH* gene (rs10417628). Most prioritized genes at the other three identified loci were involved in cell cycle regulation. Genetic correlation analyses indicated a strong positive correlation among SNPs for AMH levels and for age at menopause ( $r_g = 0.82$ ,  $FDR = 0.003$ ). Exploratory Mendelian randomization analyses did not support a causal effect of AMH on breast cancer or PCOS risk, but should be interpreted with caution as they may be underpowered and the validity of genetic instruments could not be extensively explored. In conclusion, we identified a variant in the *AMH* gene and three other loci that may affect circulating AMH levels in women.

## Introduction

Anti-Müllerian hormone (AMH) is generally known for its function in sexual differentiation, during which AMH signaling is essential for the regression of internal female reproductive organs in male embryos.<sup>1</sup> In women, AMH is expressed by granulosa cells of primary ovarian follicles, and AMH expression continues until the antral stage.<sup>2</sup> AMH becomes undetectable after menopause, when the ovarian reserve is depleted, and AMH can therefore be used as a marker for reproductive aging.<sup>3</sup>

Variation in age-specific circulating AMH levels has been associated with the occurrence of several non-communicable diseases, including breast cancer.<sup>4</sup> In addition, it has been suggested that AMH may be involved in the pathogenesis of polycystic ovary syndrome (PCOS).<sup>5</sup> Gaining more insight into genetic variation and biological mechanisms underlying inter-individual variation in AMH expression through genome-wide association studies (GWASs) could provide new clues regarding postnatal functions of AMH, and possibly, the etiologies of non-communicable diseases associated with AMH levels.

Previous GWASs on circulating AMH levels included either a mixture of male and female adolescents<sup>6</sup>, a very small study population<sup>7</sup>, or women of late reproductive age<sup>8</sup> in whom AMH levels are generally very low. Of the previous GWASs, only the largest (n = 3344) identified a single genetic variant for AMH levels in premenopausal women, at chromosome 20 (rs16991615)<sup>8</sup>, which is also associated with natural age at menopause.<sup>9; 10</sup> As sample sizes of previous GWASs were relatively small, a larger GWAS meta-analysis might lead to detection of more AMH variation loci. Moreover, as most variation in AMH levels in women is observed at ages 20 to 40 years<sup>11</sup>, including younger women will increase power to identify additional loci. Therefore, we aimed to identify additional genetic variants for AMH through a GWAS meta-analysis including 7049 premenopausal female participants. For that, we combined summary statistics from the AMH GWAS meta-analysis by Ruth et al.<sup>8</sup> with GWAS data from 3705 additional women of early and middle reproductive age from 3 different cohorts.

## Subjects and Methods

### Study population

We included data from 7049 premenopausal female participants (median age ranged from 15.3 to 48 years across cohorts; Table 1) of European ancestry. In addition to the data from the AMH GWAS meta-analysis by Ruth et al.<sup>8</sup> (n = 3344), we included data from the Doetinchem Cohort Study<sup>12; 13</sup> (n = 2084), the Study of Women's Health Across the Nation (SWAN)<sup>14</sup> (n = 425), and data from adolescent daughters of the Avon Longitudinal Study of Parents and Children (ALSPAC)<sup>15</sup> (n = 1196). The GWAS by Ruth et al. included data from the Generations Study<sup>16</sup>, Sister Study<sup>17</sup>, Nurses' Health Study<sup>18</sup>, Nurses' Health Study II<sup>19</sup>, and ALSPAC mothers<sup>20</sup>. For the current study, we requested summary statistics excluding data from ALSPAC mothers, as we wanted to analyze data from the ALSPAC mothers separately to investigate potential bias due to cryptic relatedness. More details about participating studies and the definitions used for the assessment of menopausal status are described in the Supplemental Methods and Supplemental Table 1. All studies received ethical approval from an institutional ethics committee.

### AMH measurements

Included studies measured AMH in either serum or plasma using different AMH ELISA assays. Also, the methodology for handling AMH measurements below the assay limit of detection (LOD) differed across studies. A detailed overview of these study-specific details has been included in Supplemental Table 1. Across studies, the percentage of measurements under the assay-specific LODs ranged from 0% to 24.2%.

## Genotyping and imputation

Extensive details on genotyping and imputation procedures for each participating study are presented in Supplemental Table 2. Briefly, samples of the Generations Study, Sister Study, and most samples of the Nurses' Health Studies, were genotyped using the OncoArray array.<sup>8</sup> The remaining 225 samples of the Nurses' Health Studies were genotyped using Illumina HumanHap550 and HumanHap610 arrays.<sup>8</sup> Samples of the Doetinchem Cohort Study were genotyped using the Illumina Infinium Global Screening Array-24 Kit (Illumina Inc., San Diego, California, United States of America). For genotyping of samples from ALSPAC mothers and daughters, the Illumina Human660W-Quad array and Illumina HumanHap550 quad genome-wide SNP genotyping platform were used, respectively. SWAN participants were genotyped using the Illumina Multi-Ethnic Global Array (MEGA A1). All participating studies performed sample and SNP QC prior to imputation, which was done using the Haplotype Reference Consortium (HRC) panel version r1.1 2016 (Supplemental Table 2).

## Association analyses

All studies converted AMH concentrations to pmol/L using  $1 \text{ pg/mL} = 0.00714 \text{ pmol/L}$ . As AMH levels are not normally distributed, AMH measurements were transformed using rank-based inverse normal transformation in all studies, as previously described.<sup>8</sup>

In all studies linear models were fitted, assuming additive SNP effects, adjusted for age at blood collection (years) (Supplemental Methods). Analyses were further adjusted for population stratification by inclusion of either 10 principal components (ALSPAC, SWAN) or a kinship matrix (Generations Study, Sister Study, Nurses' Health Studies, Doetinchem Cohort Study). In addition, we included summary statistics of the meta-analysis of the Generations Study, Sister Study, and Nurses' Health Studies, which was performed using METAL, as described elsewhere.<sup>8</sup> Separate association analyses were conducted

for the ALSPAC mothers and daughters, because of the large differences in both age and AMH distributions between these groups (Supplemental Methods).

Prior to meta-analysis, we performed file-level and meta-level QC on all summary statistics files to clean and check the data, and to identify potential study-specific problems. File-level and meta-level QC were conducted using the R package EasyQC (v9.2), following a previously published protocol<sup>21</sup> (Supplemental Methods). No study-specific issues were identified through these QC procedures (Figure S1-S5). In addition, we sought to confirm that inclusion of ALSPAC mothers and daughters as separate cohorts would not result in inflation of effect estimates due to cryptic relatedness (411 mother-daughter pairs were present in the ALSPAC data). We checked this through meta-analyzing only data of the two ALSPAC cohorts and checking both the corresponding QQ plot and calculating  $\lambda$ . Given the absence of genomic inflation ( $\lambda = 1.01$ , QQplot in Figure S6) we included summary statistics of both ALSPAC mothers and daughters in the meta-analysis.

We performed an inverse variance weighted meta-analysis using METAL (version 2011-03-25). Genomic control was applied for all included studies. SNPs with a minor allele frequency (MAF) < 1% and/or poor imputation quality (info score < 0.4 or  $r^2$  < 0.3, depending on which metric was provided) were excluded. As a result, 8,298,138 autosomal SNPs were included in this AMH GWAS meta-analysis. To assess if observed effect estimates were homogeneous across studies, we additionally performed a heterogeneity analysis in METAL.

To identify lead and secondary SNPs within genome-wide significant associated loci, we performed an approximate conditional and joint association analysis.<sup>22</sup> We used Genome-wide Complex Trait Analysis (GCTA)<sup>23</sup> (version 1.93.1f beta) to run a stepwise model selection procedure to select independently associated SNPs (cojo-slct) using the summary-level data. We estimated LD between SNPs using data of 4059 unrelated participants from the EPIC-NL cohort<sup>24</sup> as LD reference panel.

Because of the strong correlation between AMH and age, and the difference in both the AMH and age distributions in the ALSPAC daughters compared to the other included cohorts, we performed a sensitivity analysis in which we excluded the ALSPAC daughters. Furthermore, this sensitivity analysis served as an additional check that inclusion of both ALSPAC mothers and daughters did not cause identification of false-positive hits.

### **Gene-based genome-wide association analysis**

We performed a gene-based genome-wide association analysis using the MAGMA<sup>25</sup> implementation (v1.07) in the online Functional Mapping and Annotation of Genome-wide Association Studies (FUMA) platform (FUMA)<sup>26</sup> (parameter settings are listed in Table S3). For this analysis, SNPs located in gene bodies were aggregated to 18,896 protein coding genes (Ensembl build 92). MAGMA tests the joint association of all SNPs in each gene with inverse normal transformed AMH levels using a multiple linear regression approach, which takes LD between SNPs into account.<sup>25</sup> FUMA considered genes to be significantly associated with circulating AMH levels if  $p < 2.65 \times 10^{-6}$  (Bonferroni corrected p-value;  $0.05/18,896$ ).

### **Functional annotation using FUMA**

FUMA is an integrative web-based platform that uses 18 biological resources and can be used to functionally annotate lead variants from GWAS, and to prioritize the most likely causal SNPs and genes<sup>26</sup>. We used the SNP2GENE process integrated into FUMA (v1.3.6)<sup>26</sup> for the characterization of genomic loci and functional gene mapping (parameter settings are listed in Table S3). We included SNPs identified in our approximate conditional and joint analysis as predefined lead SNPs for the characterization of genomic risk loci. SNPs that were in LD with these lead SNPs ( $r^2 > 0.6$ ) within a 500kb

window based on the 1000G Phase 3 European reference panel population in FUMA, and a GWAS meta-analysis p-value < 0.05 were selected as candidate SNPs. Non-GWAS-tagged SNPs from the 1000G Phase 3 European reference that met these LD and distance criteria were also selected as candidate SNPs. Candidate SNPs were annotated based on Combined Annotation Dependent Depletion (CADD) scores<sup>27</sup>, Regulome DB scores<sup>28</sup>, and chromatin states<sup>29</sup> (Table S3). Positional mapping, eQTL mapping and chromatin interaction mapping were used to map SNPs to genes (Table S3). For chromatin states, eQTL mapping and chromatin interaction mapping we only selected tissues and cell types that are most likely to be involved in AMH expression and signaling (Table S3).

### **Pathway analysis using DEPICT**

We used the hypothesis-free pathway analysis tool DEPICT (v1)<sup>30</sup> to prioritize the most likely causal genes at associated loci, to highlight gene sets enriched in genes within associated loci, and to identify tissues/cell types that are implicated by the associated loci. For these analyses, we included all suggestive significant SNPs ( $p < 5 \times 10^{-6}$ ), which were clumped at LD  $r^2 < 0.1$  and a physical distance of 500kb using PLINK v.1.9 as part of the DEPICT pipeline.

### **LD Score Regression**

We estimated SNP heritability using the LD Hub web interface (v1.9.3)<sup>31</sup> for LD score regression<sup>32</sup>. In addition, we used LD Hub to estimate SNP-based genetic correlations between AMH and phenotypes that have been associated with AMH in observational studies. These genetic correlation analyses make use of GWAS summary statistics for all SNPs to estimate genetic covariance among SNPs for two traits.<sup>33</sup> Included phenotypes comprised reproductive traits, hormones (leptin), anthropometric traits, blood lipids, glycemic traits, metabolites, cardiometabolic traits, cancer, autoimmune diseases, bone mineral



density, aging and smoking behavior. Of the 597 UK Biobank traits in the LDHub database, we only included traits that corresponded to these phenotype categories, resulting in 345 comparisons. To correct for multiple testing, we calculated false discovery rates (FDR), using the `p.adjust` function in R (R package “stats”).<sup>34</sup> FDR adjusted p-values < 0.05 were considered to be significant.

## **Mendelian randomization**

In observational studies, AMH has been associated with breast cancer<sup>4</sup>, and PCOS<sup>35</sup>, amongst other diseases. As the exact function of AMH in the etiology of these diseases is unclear, and actual AMH levels are associated with predicted future age at menopause and current menopausal status, causality of these AMH-disease associations remains to be determined. Mendelian randomization (MR) is a method that may provide evidence for causality of observational associations.<sup>36</sup> Because our AMH GWAS meta-analysis only included women, and previous research suggests that genetic variants for inter-individual differences in AMH levels differ between males and females<sup>6</sup>, we performed MR analyses for the female-specific outcomes breast cancer and PCOS only. We performed two-sample MR analyses<sup>37</sup> using the R package `TwoSampleMR` (version 0.5.1)<sup>38</sup>. We included identified lead SNPs as genetic instruments for AMH. For the outcomes, we included summary statistics from the most recent largest GWASs for breast cancer ( $n = 228,951$ ; 122,977 cases)<sup>39</sup> and PCOS ( $n = 113,238$ ; 10,074 cases)<sup>40</sup>. Wald ratio estimates were calculated for individual SNPs and a random effects inverse variance weighted (IVW) meta-analysis approach was used to combine these estimates. To assess the strength of included genetic variants for AMH we calculated F-statistics corresponding to the IVW analyses, using the proportion of variance in AMH explained by the genetic variants, the sample size of the outcome GWASs, and the number of variants included<sup>41</sup>. We compared the overall MR estimate (i.e. IVW) to SNP-specific MR estimates (i.e. Wald ratio) since inconsistent estimates are indicative of horizontal

pleiotropy, which is a violation of the MR assumptions.<sup>38</sup> In addition, we tested for heterogeneity in causal effects amongst the genetic instruments using Cochrane's Q statistics and performed leave-one-out sensitivity analyses to assess the potential effect of outlying variants.

## Results

Descriptive statistics on age and AMH levels of the study participants included in this GWAS meta-analysis are presented per study in Table 1. Median AMH ranged from 1.1 pmol/L in SWAN to 26.1 pmol/L in ALSPAC daughters. Median age ranged from 15.3 years in ALSPAC daughters to 48 years in the Sister Study.

### Genome-wide association analysis

We identified four genome-wide significant lead SNPs ( $p < 5 \times 10^{-8}$ ) for inverse normally transformed AMH, in four loci (Table 2, Figure 1, Figure S7-8). Approximate conditional and joint analysis did not reveal secondary signals. In addition to the previously reported locus on chromosome 20 (rs16991615, nearest gene: *MCM8*), we identified 1 locus on chromosome 19 (nearest gene: *AMH*) and 2 loci on chromosome 2 (nearest genes: *TEX41* and *CDCA7*). The strongest signal was rs10417628 on chromosome 19, which is physically located in the *AMH* gene ( $\beta = -0.34$ ,  $se = 0.05$ ,  $p = 1.2 \times 10^{-11}$ ) (Figure S8A). Combined the four lead SNPs explained 1.47% of the variance in AMH levels.

In the sensitivity analysis excluding ALSPAC daughters, all four loci from the main analysis remained genome-wide significant, and an additional locus at chromosome 5 (rs116090962, nearest gene: *CTB-99A3.1*) was identified ( $\beta = 0.38$ ,  $se = 0.07$ ,  $p = 6.0 \times 10^{-9}$ ) (Table S4).

### Gene-based genome-wide association analysis

Gene-based genome-wide association analysis, which tested associations between 18,896 protein coding genes and inverse normal transformed AMH, highlighted the following four significant genes: *AMH*, *BMP4*, *EIF4EBP1*, and *GAB2* (Supplemental Figure 9).

## Functional Annotation using FUMA

Through the SNP2GENE process, FUMA identified 82 candidate SNPs that were in LD with the four identified lead SNPs (Supplemental Table 5). These candidate SNPs were used for the prioritization of genes.

In total, 12 genes were mapped to the locus of the previously identified SNP on chromosome 20 (rs16991615) (Supplemental Table 6), of which *MCM8* and *CRLS1* were prioritized based on eQTL mapping (Figure 2A). *CRLS1* was the only gene prioritized based on both eQTL mapping and chromatin interactions. However, as rs16991615 is a missense variant located in exon 9 of the *MCM8* gene, and this was the only SNP identified for this locus, *MCM8* is the most likely gene causing this signal.

For the locus on chromosome 19, 3 candidate SNPs (rs10417628, rs12462821, rs7247495) were identified. The lead SNP in this locus (rs10417628) is also a missense variant, located in exon 5 of the *AMH* gene, making this the most likely causal gene at this locus. The other 2 variants were located in intronic and intronic non-coding RNA regions. Based on the used parameter settings, FUMA mapped 8 genes to the *AMH* locus (Supplemental Table 6), of which 4 were highlighted by eQTL mapping (*AMH*, *C19orf35*, *SPPL2B* and *LSM7*) and 1 through chromatin interactions (*ABHD17A*) (Figure 2B).

Most of the candidate SNPs were identified for the locus on chromosome 2 near *TEX41*. All 77 variants were located in either intronic or exonic long noncoding RNA regions. Of the 15 genes mapped to this locus (Supplemental Table 6), no genes were prioritized based on eQTL mapping, but several genes were prioritized based on chromosome interactions, including *ZEB2-AS1* (Figure 2C). In the other locus on chromosome 2, for which *CDCA7* is the nearest gene, no additional candidate SNPs were identified and no genes were mapped to this locus.

## Pathway analysis using DEPICT

Using the DEPICT tool, 188 suggestive associated SNPs ( $p < 5 \times 10^{-6}$ ) were clumped at LD  $r^2 < 0.1$  and a physical distance of 500 kb, resulting in 24 clumps as input for the enrichment analyses (Table S7). The top three prioritized gene sets were “URI1 PPI subnetwork”, “NFYB PPI subnetwork” and “nuclear inner membrane” (Table S8). “Induced Pluripotent Stem Cells” was identified as the highest prioritized cell type (Table S9). However, none of these enrichments were statistically significant (FDR > 0.05).

DEPICT prioritized nine genes at FDR < 0.05 as most likely causal genes (Table S10). Of the genome-wide associated loci, only *MCM8* (rs16991615), and *CDCA7* (rs11683493) were prioritized at this FDR threshold. *AMH*, *BMP4*, and *GAB2* were also prioritized by DEPICT, but FDR values were > 0.20. Of the genes significant in the gene-based MAGMA analysis, *EIF4EBP1* was prioritized at FDR < 0.05.

## LD Score Regression

We used LD score regression implemented in LD Hub to calculate SNP heritability for AMH based on the meta-analysis summary statistics. Total SNP heritability ( $h_g^2$ ) on the observed scale was estimated to be 15% (se = 7%). We additionally performed genetic correlations analyses between AMH and 345 traits on LD Hub. After correction for multiple testing, AMH was only significantly correlated with age at menopause ( $r_g = 0.82$ , se = 0.19, FDR = 0.003) (Table S11).

## MR analyses

IVW MR estimates did not indicate a causal effect of circulating AMH on breast cancer risk (OR<sub>IVW</sub> = 1.00, 95%CI: 0.74 – 1.36; Table 3). Results from the single SNP analysis including the variant in the *AMH* locus also did not support a causal association with breast cancer (OR<sub>IVW</sub> = 0.99, 95%CI = 0.87 – 1.12), whereas

analyses for the remaining variants suggested a risk decreasing effect of the SNPs in the *TEX41* and *CDCA7* loci and a risk increasing effect of the variant in the *MCM8* locus (Table 3). In agreement with these findings, a formal heterogeneity test for the IVW estimate indicated heterogeneity in causal effects amongst the four genetic variants ( $2.13 \times 10^{-11}$ ), although the interpretation of this heterogeneity p-value is limited due to the small number of included SNPs. Leave-one-out sensitivity analyses supported the outlying effect of rs16991615 (*MCM8* locus) (Figure S11).

For PCOS, the IVW MR estimate suggested that higher genetically predicted AMH levels are potentially associated with an increased risk of PCOS, but confidence intervals were wide and included the null ( $OR_{IVW} = 1.29$ , 95%CI = 0.85 – 1.95; Table 3). Single SNP analyses resulted in a similar effect estimate for the variant in the *AMH* locus ( $OR_{IVW} = 1.27$ , 95%CI = 0.64 – 2.56), risk increasing effects for the SNPs in the *TEX41* and *MCM8* loci, and a risk decreasing effect of the variant in the *CDCA7* locus (Table 3). The heterogeneity test did not suggest heterogeneous effects of the individual SNPs ( $p = 0.30$ ), most likely because of the high uncertainty in individual SNP estimates, but again interpretation of this p-value is limited with only four SNPs. Leave-one-out sensitivity analyses indicated rs11683493 (*CDCA7* locus) affected the IVW estimate most, and that exclusion of this variant resulted in a positive association ( $OR_{IVW} = 1.53$ , 95%CI = 1.01 – 2.33) (Figure S13).

## Discussion

We identified four loci for circulating AMH levels in women of European ancestry. In addition to confirming a previously reported signal in the *MCM8* locus, we discovered three new signals in and near the *AMH*, *TEX41* and *CDCA7* genes. In total, 35 genes were prioritized for these loci based on physical position, eQTL mapping and chromatin interactions, but pathway analyses did not reveal enrichments of gene-sets, tissues or cell types for genes annotated to suggestive associated SNPs. Genetic correlation analyses supported a shared genetic architecture between AMH levels and age at menopause. Exploratory MR analyses did not provide strong evidence of a causal effect of circulating AMH on breast cancer and PCOS.

We confirmed the association between rs16991615 and circulating AMH levels, previously reported by Ruth et al.<sup>8</sup>. This SNP is a missense variant located in exon 9 of the *MCM8* gene, rendering *MCM8* the most likely causal gene at this locus. In humans, *MCM8* plays a role in in homologous recombination, which is critical for DNA repair.<sup>42</sup> Previous studies have linked *MCM8* deficiency to premature ovarian failure and infertility, but also to cancer development.<sup>39; 43; 44</sup> Associations between rs16991615 and age at menopause<sup>45</sup> and number of ovarian follicles<sup>46</sup> have also been reported, which suggests that this locus is associated with circulating AMH levels because of its influence on the number of antral follicles.

Our GWAS study is the first AMH GWAS that identified a missense SNP (rs10417628) in the *AMH* gene in women. A previous AMH GWAS including adolescents from ALSPAC identified three SNPs in the *AMH* gene that were only significantly associated with AMH levels in male adolescents, and of which one (rs2385821) was in moderate LD with our lead SNP rs10417628 ( $R^2 = 0.55$ ).<sup>6</sup> However, approximate conditional and joint analyses suggested that these variants represent the same signal at the *AMH* locus. Although identification of a genetic variant in the gene encoding for AMH itself suggests that we reveal an actual signal for circulating AMH concentrations, a recent case report suggests that the amino acid

substitution corresponding to rs10417628 reduces AMH detection by the picoAMH assay from Ansh Labs without influencing AMH bioactivity.<sup>47</sup> We sought to verify this finding in a subsample of the Doetinchem Cohort Study, for which AMH was measured using both the picoAMH and the less sensitive Gen II assay from Beckman Coulter. For the only woman who was estimated to be homozygous for the T allele (dosage T allele = 1.9, age at measurement = 28.3 years), AMH levels were indeed undetectable using the picoAMH assay, whereas circulating AMH levels were detected using the Gen II assay (318 pg/mL). In addition, median AMH levels measured using the Gen II assay were less different between women homozygous for the reference allele and heterozygous women (median AMH levels<sub>homozygousrefallele</sub> = 953.0 pg/mL, IQR: 428.0 - 1999.0; median AMH levels<sub>heterozygous</sub> = 848.0 pg/mL, IQR: 509.0 - 1310.0), compared to AMH levels measured using the picoAMH assay (median AMH levels<sub>homozygousrefallele</sub> = 1485.9 pg/mL, IQR: 704.4 - 3150.0; median AMH levels<sub>heterozygous</sub> = 811.0 pg/mL, IQR: 462.3 - 1480.9). For ALSPAC, which also used the Gen II assay to measure AMH, the distribution of AMH levels was similar across adult women homozygous for the reference allele and heterozygous women as well. However, in the ALSPAC daughters median AMH levels were clearly higher in adolescents homozygous for the reference allele compared to heterozygous adolescents. Among the ALSPAC participants, only one adolescent was homozygous for the T allele, but her AMH levels could not be shared due to disclosure risk. Because of the lack of publicly available information on the antibodies and conformational epitopes of the Gen II assay, and the limited and inconsistent evidence in the current study, we do not want to draw any definite conclusions about this yet.

For the associated loci on chromosome 2 it is more challenging to assign possible causal genes, as *TEX41* is a long noncoding RNA and the SNP in the *CDCA7* locus was located in an intergenic region. Gene mapping based on chromatin interactions with *TEX41* highlighted several genes, including the long non-coding RNA *ZEB2-AS1* (*ZEB2* antisense RNA 1). *ZEB2-AS1* up-regulates expression of the protein *ZEB2*.<sup>48</sup> *ZEB2* (also known as *SIP1*) inhibits signal transduction in TGF- $\beta$  and BMP signalling through interaction



with ligand-activated SMAD proteins.<sup>49; 50</sup> Among other BMP proteins, BMP4 has been reported to regulate AMH expression through activation of SMAD proteins.<sup>51; 52</sup> Based on identification of *BMP4* in our gene-based association analysis and its prioritization by DEPICT, we hypothesize that BMP4 induced AMH expression may be regulated by ZEB2 interaction. However, fundamental laboratory research is needed to prove this.

*CDCA7* (also known as *JPO1*) is a direct target gene of the transcription factor *MYC* and is involved in apoptosis.<sup>53</sup> Functional annotation did not map any genes to this locus and thus the mechanism through which this locus affects circulating AMH levels remains to be elucidated. Also for *EIF4EBP1*, and *GAB2*, which were significantly associated with AMH levels in the gene-based association analysis, it is not yet clear if and how these genes influence AMH expression. A target gene of *EIF4EBP1* has been previously highlighted as candidate gene for age at menopause<sup>54</sup>, whereas *GAB2* is involved in FSH signalling<sup>55</sup> and has been identified as a potential causal gene for age at menarche.<sup>56</sup> Altogether these results suggest that these genes play a role in reproductive functioning and reproductive aging, potentially by affecting the number of antral follicles, which are the main producers of AMH in women.<sup>2</sup> Ideally, future studies should explore whether the observed genetic associations may merely reflect the size of the ovarian reserve, through adjusting analyses for antral follicle count. Such analyses would also show if we can actually use the identified variants as instruments for circulating AMH levels itself or for the quantity of antral follicles in MR analyses.

We did not find support for a causal effect of circulating AMH levels on breast cancer and PCOS risk in our exploratory MR analyses. To be valid genetic instruments for MR, SNPs have to fulfil the following three criteria<sup>57</sup>: (1) SNPs have to be associated with the circulating AMH levels; (2) SNPs cannot be associated with confounders of the studied AMH-outcome associations, and (3) SNPs cannot influence the outcomes through mechanisms that do not involve circulating AMH levels. Because rs10417628 in the *AMH* gene potentially reflects AMH detection instead of AMH expression, analyses including this

variant should be interpreted with caution. However, leave-one-out analyses excluding this variant did not affect IVW MR estimates. Based on the function of genes mapped to the loci on chromosomes 20 and 2, it is likely that these variants affect breast cancer and PCOS risk through mechanisms independent of AMH (e.g. DNA replication and apoptosis), in particular the *MCM8* locus, which has also been identified in breast cancer GWAS.<sup>39</sup> Due to the limited number of identified lead SNPs it was not possible to assess if our results were indeed biased by horizontal pleiotropy. Furthermore, weak instrument bias may still have biased MR results towards the null, since the F statistics may be overestimated in this GWAS (853.9 for breast cancer, and 422.4 for PCOS). Consequently, we should be cautious about excluding a causal effect of AMH on the studied outcomes.

Previous research suggests that AMH levels in females rise during puberty, until the mid to late twenties, and after that decrease until menopause.<sup>11; 58</sup> Based on these observations and the differences in both age and AMH distributions between the ALSPAC adolescents and other participants, we performed a sensitivity analysis excluding the adolescents from ALSPAC. This analysis revealed an additional locus on chromosome 5 (rs116090962, nearest gene: *CTB-99A3.1*), although we could not find clues for its association with circulating AMH levels in adult women only, nor with AMH levels in general. Study-specific betas revealed an opposite effect for the *MCM8* locus and a minimal effect for the *CDCA7* locus in adolescents compared to effects in adult women. A larger GWAS including older adolescents and a larger proportion of females aged 20 to 40, would be required to reveal potential gene-age interactions explain variation in AMH expression.

The main strengths of the current GWAS meta-analysis are its size, which is twice the size of the previous GWAS meta-analysis, and its larger proportion of women of early-reproductive age. Given AMH's function in ovarian follicle development, circulating levels and variation in AMH levels decrease with age. As a result, statistical power to identify genetic variants for circulating AMH increases if younger women are included. Still, our sample size remains relatively small for a GWAS, and future

larger studies may lead to the detection of additional variants for circulating AMH levels. This is supported by our chip heritability estimate of 15% (se = 7%), which indicates that there are likely more SNPs that contribute to variability in AMH levels. Identification of additional genetic variants will also facilitate increased power to identify pathways and tissues enriched for genes involved in AMH expression. A second limitation of this study is potential overlap in participants between the current AMH GWAS and the GWAS for breast cancer<sup>39</sup> (maximum n = 1,459; 20.7% of current study) and PCOS<sup>40</sup> (maximum n = 225; 3.2% of current study). Overlap in participants in two-sample MR analyses may bias effect estimates and inflate Type 1 error rates.<sup>59</sup>

In conclusion, we replicated the previously reported association with the *MCM8* locus and identified 3 novel loci for circulating AMH levels in women, including the *AMH* locus. The strongest signal in this locus possibly affects AMH detection by specific assays rather than AMH bioactivity, but further research is required to confirm this hypothesis. Genes mapped to the *MCM8*, *TEX41* and *CDCA7* loci are involved in the cell cycle and processes like DNA replication and apoptosis. The mechanism underlying their associations with AMH may affect the size of the ovarian follicle pool. MR analyses did not support a causal effect of AMH on breast cancer and PCOS, but these findings should be interpreted with caution because we could not robustly explore how valid the instruments were and weak instrument bias may have biased estimates towards the null.

## Supplemental Data

Document S1: Supplemental Methods.

Document S2: Figures S1 – S13.

Document S3: Tables S1 – S11.

## Declarations of Interest

FJMB has received fees and grant support from Merck Serono, Gedeon Richter, Ferring BV, and Roche.

DAL has received financial support from several national and international government and charitable funders as well as from Medtronic Ltd and Roche Diagnostics for research that is unrelated to this study.

NS is scientific consultant for Ansh Laboratories. The other authors declare no competing interests.

## Acknowledgments

The Generations Study 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 staff and data providers who have contributed to the Study. ICR acknowledges NHS 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.

For the Doetinchem Cohort Study we thank dr HSJ Picavet, A Blokstra MSc and Petra Vissink for their crucial role in data collection and data management.

For the ALSPAC Cohort, we are extremely grateful to all the families who took part in this 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.

SWAN thanks the study staff at each site and all the women who participated in SWAN. Clinical

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

Office: National Institute on Aging, Bethesda, MD – Chhanda Dutta 2016- present; Winifred Rossi 2012– 2016; Sherry Sherman 1994 – 2012; Marcia Ory 1994 – 2001; National Institute of Nursing Research, Bethesda, MD – Program Officers. Central Laboratory: University of Michigan, Ann Arbor – Daniel McConnell (Central Ligand Assay Satellite Services). SWAN Repository: University of Michigan, Ann Arbor – Siobán Harlow 2013 - Present; Dan McConnell 2011 - 2013; MaryFran Sowers 2000 – 2011. Coordinating Center: University of Pittsburgh, Pittsburgh, PA – Maria Mori Brooks, PI 2012 - present; Kim Sutton-Tyrrell, PI 2001 – 2012; New England Research Institutes, Watertown, MA - Sonja McKinlay, PI 1995 – 2001. Steering Committee: Susan Johnson, Current Chair, Chris Gallagher, Former Chair.

The Sister Study acknowledges the contributions of Jack A. Taylor, M.D., Ph.D., Molecular & Genetic Epidemiology Group, Epidemiology Branch, National Institute of Environmental Health Sciences, Durham, NC who secured initial funding for the Sister Study GWAS, and Hazel B. Nichols. Ph.D., Associate Professor, Department of Epidemiology, Gillings School of Global Public Health who was the PI

of the grant that allowed the AMH assays for the case-control study of AMH and breast cancer in premenopausal women.

## Funding

Nurses' Health Study and Nurses' Health Study II were supported by research grants from the National Institutes of Health (CA172726, CA186107, CA50385, CA87969, CA49449, CA67262, CA178949). The UK Medical Research Council and Wellcome (Grant ref: 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. This publication is the work of the listed authors, who will serve as guarantors for the contents of this paper. A comprehensive list of grants funding is available on the ALSPAC website (<http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf>). Funding for the collection of genotype and phenotype data used here was provided by the British Heart Foundation (SP/07/008/24066), Wellcome (WT092830M and WT08806) and UK Medical Research Council (G1001357). MCB, ALS and DAL work in a Unit that is funded by the University of Bristol and UK Medical Research Council (MC\_UU\_00011/6). MCB is funded by a UK Medical Research Council Skills Development Fellowship (MR/P014054/1) and DAL is a National Institute of Health Research Senior Investigator (NF-0616-10102). The Doetinchem Cohort Study is financially supported by the Ministry of Health, Welfare and Sports of the Netherlands. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Ansh Labs performed the AMH measurements for the Doetinchem Cohort Study free of charge. Ansh Labs was not involved in the data analysis, interpretation, or reporting, nor was it financially involved in any aspect of the study. RMGV is funded by the Honours Track of MSc Epidemiology, University Medical Center Utrecht with a grant from the Netherlands Organization for Scientific Research (NWO) (Grant number: 022.005.021). The Study of Women's Health Across the Nation (SWAN) has grant support from the National Institutes of Health

(NIH), DHHS, through the National Institute on Aging (NIA), the National Institute of Nursing Research (NINR) and the NIH Office of Research on Women's Health (ORWH) (Grants U01NR004061; U01AG012505, U01AG012535, U01AG012531, U01AG012539, U01AG012546, U01AG012553, U01AG012554, U01AG012495). The SWAN Genomic Analyses and SWAN Legacy has grant support from the NIA (U01AG017719). The Generations Study is funded by Breast Cancer Now and the Institute of Cancer Research (ICR). The ICR acknowledges NHS funding to the NIHR Biomedical Research Centre. The content of this manuscript is solely the responsibility of the authors and does not necessarily represent official views of the funders. The Sister Study was funded by the Intramural Research Program of the National Institutes of Health (NIH), National Institute of Environmental Health Sciences (Z01-ES044005 to DP Sandler); the AMH assays were supported by the Avon Foundation (02-2012-065) to H.B. Nichols and D.P. Sandler). The breast cancer genome-wide association analyses were supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the 'Ministère de l'Économie, de la Science et de l'Innovation du Québec' through Genome Québec and grant PSR-SIIRI-701, The National Institutes of Health (U19 CA148065, X01HG007492), Cancer Research UK (C1287/A10118, C1287/A16563, C1287/A10710) and The European Union (HEALTH-F2-2009-223175 and H2020 633784 and 634935). All studies and funders are listed in Michailidou et al (Nature, 2017).

## Web Resources

ALSPAC: <http://www.bristol.ac.uk/alspac/>

SWAN: <https://www.swanstudy.org/>

EASYQC: <https://www.uni-regensburg.de/medizin/epidemiologie-praeventivmedizin/genetische-epidemiologie/software/>

METAL: <http://csg.sph.umich.edu/abecasis/metal/>

GCTA: <https://cnsgenomics.com/software/gcta/>

FUMA: <https://fuma.ctglab.nl/>

DEPICT: <https://github.com/perslab/depict>

LD Hub: <http://ldsc.broadinstitute.org/>

## **Data and Code Availability**

Code generated for this study is available at Github

[[https://github.com/reneemgverdiesen/AMH\\_GWAS](https://github.com/reneemgverdiesen/AMH_GWAS)]. The full AMH GWAS summary statistics will be made available through the GWAS catalog (<https://www.ebi.ac.uk/gwas/>)



## References

1. Jost, A. (1947). The age factor in the castration of male rabbit fetuses. *Proceedings of the Society for Experimental Biology and Medicine Society for Experimental Biology and Medicine (New York, NY)* 66, 302.
2. Weenen, C., Laven, J.S., von Bergh, A.R., Cranfield, M., Groome, N.P., Visser, J.A., Kramer, P., Fauser, B.C., and Themmen, A.P. (2004). Anti-Müllerian hormone expression pattern in the human ovary: potential implications for initial and cyclic follicle recruitment. *MHR: Basic science of reproductive medicine* 10, 77-83.
3. Finkelstein, J.S., Lee, H., Karlamangla, A., Neer, R.M., Sluss, P.M., Burnett-Bowie, S.M., Darakananda, K., Donahoe, P.K., Harlow, S.D., Prizand, S.H., et al. (2020). Antimüllerian Hormone and Impending Menopause in Late Reproductive Age: The Study of Women's Health Across the Nation. *J Clin Endocrinol Metab* 105.
4. 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* 142, 2215-2226.
5. Homburg, R., and Crawford, G. (2014). The role of AMH in anovulation associated with PCOS: a hypothesis. *Human reproduction (Oxford, England)* 29, 1117-1121.
6. Perry, J.R., 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.

7. Schuh-Huerta, S.M., Johnson, N.A., Rosen, M.P., Sternfeld, B., Cedars, M.I., and Reijo Pera, R.A. (2012). Genetic variants and environmental factors associated with hormonal markers of ovarian reserve in Caucasian and African American women. *Hum Reprod* 27, 594-608.
8. 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-Mullerian hormone levels in pre-menopausal women of late reproductive age and relationship with genetic determinants of reproductive lifespan. *Hum Mol Genet* 28, 1392-1401.
9. He, C., Kraft, P., Chen, C., Buring, J.E., Paré, G., Hankinson, S.E., Chanock, S.J., Ridker, P.M., Hunter, D.J., and Chasman, D.I. (2009). Genome-wide association studies identify loci associated with age at menarche and age at natural menopause. *Nat Genet* 41, 724-728.
10. Stolk, L., Zhai, G., van Meurs, J.B., Verbiest, M.M., 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.
11. de Kat, A.C., van der Schouw, Y.T., Eijkemans, M.J., Herber-Gast, G.C., Visser, J.A., Verschuren, W.M., and Broekmans, F.J. (2016). Back to the basics of ovarian aging: a population-based study on longitudinal anti-Mullerian hormone decline. *BMC medicine* 14, 151.
12. Picavet, H.S.J., Blokstra, A., Spijkerman, A.M.W., and Verschuren, W.M.M. (2017). Cohort Profile Update: The Doetinchem Cohort Study 1987-2017: lifestyle, health and chronic diseases in a life course and ageing perspective. *Int J Epidemiol* 46, 1751-1751g.
13. Verschuren, W.M., Blokstra, A., Picavet, H.S., and Smit, H.A. (2008). Cohort profile: the Doetinchem Cohort Study. *Int J Epidemiol* 37, 1236-1241.
14. Sowers, M.F.R., Crawford, S.L., Sternfeld, B., Morganstein, D., Gold, E.B., Greendale, G.A., Evans, D.A., Neer, R., Matthews, K.A., and Sherman, S. (2000). SWAN: a multicenter, multiethnic,

- community-based cohort study of women and the menopausal transition. In *Menopause: Biology and Pathobiology*, Lobo RA, Kelsey J, and M. R, eds. (San Diego, Academic Press), pp 175-188.
15. Boyd, A., Golding, J., Macleod, J., Lawlor, D.A., Fraser, A., Henderson, J., Molloy, L., Ness, A., Ring, S., and Davey Smith, G. (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.
  16. 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.
  17. Nichols, H.B., Baird, D.D., Stanczyk, F.Z., Steiner, A.Z., Troester, M.A., Whitworth, K.W., and Sandler, D.P. (2015). Anti-Mullerian hormone concentrations in premenopausal women and breast cancer risk. *Cancer Prev Res (Phila)* 8, 528-534.
  18. Hankinson, S.E., Willett, W.C., Manson, J.E., Hunter, D.J., Colditz, G.A., Stampfer, M.J., Longcope, C., and Speizer, F.E. (1995). Alcohol, height, and adiposity in relation to estrogen and prolactin levels in postmenopausal women. *J Natl Cancer Inst* 87, 1297-1302.
  19. Eliassen, A.H., Missmer, S.A., Tworoger, S.S., Spiegelman, D., Barbieri, R.L., Dowsett, M., and Hankinson, S.E. (2006). Endogenous steroid hormone concentrations and risk of breast cancer among premenopausal women. *J Natl Cancer Inst* 98, 1406-1415.
  20. 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.

21. Winkler, T.W., Day, F.R., Croteau-Chonka, D.C., Wood, A.R., Locke, A.E., Magi, R., Ferreira, T., Fall, T., Graff, M., Justice, A.E., et al. (2014). Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc* 9, 1192-1212.
22. Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., Weedon, M.N., Loos, R.J., et al. (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* 44, 369-375, s361-363.
23. 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.
24. Beulens, J.W., Monninkhof, E.M., Verschuren, W.M., van der Schouw, Y.T., Smit, J., Ocke, M.C., Jansen, E.H., van Dieren, S., Grobbee, D.E., Peeters, P.H., et al. (2010). Cohort profile: the EPIC-NL study. *Int J Epidemiol* 39, 1170-1178.
25. 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.
26. Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* 8, 1826.
27. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 46, 310-315.
28. 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.

29. Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al. (2015). Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317-330.
30. Pers, T.H., Karjalainen, J.M., Chan, Y., Westra, H.J., Wood, A.R., Yang, J., Lui, J.C., Vedantam, S., Gustafsson, S., Esko, T., et al. (2015). Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* 6, 5890.
31. Zheng, J., Erzurumluoglu, A.M., Elsworth, B.L., Kemp, J.P., Howe, L., Haycock, P.C., Hemani, G., Tansey, K., Laurin, C., Pourcain, B.S., et al. (2017). LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* 33, 272-279.
32. Bulik-Sullivan, B.K., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 47, 291-295.
33. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.R., Duncan, L., Perry, J.R., Patterson, N., Robinson, E.B., et al. (2015). An atlas of genetic correlations across human diseases and traits. *Nat Genet* 47, 1236-1241.
34. Team, R.C. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
35. Laven, J.S., Mulders, A.G., Visser, J.A., Themmen, A.P., De Jong, F.H., and Fauser, B.C. (2004). Anti-Mullerian hormone serum concentrations in normoovulatory and anovulatory women of reproductive age. *J Clin Endocrinol Metab* 89, 318-323.
36. Davey Smith, G., and Ebrahim, S. (2005). What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? *Bmj* 330, 1076-1079.

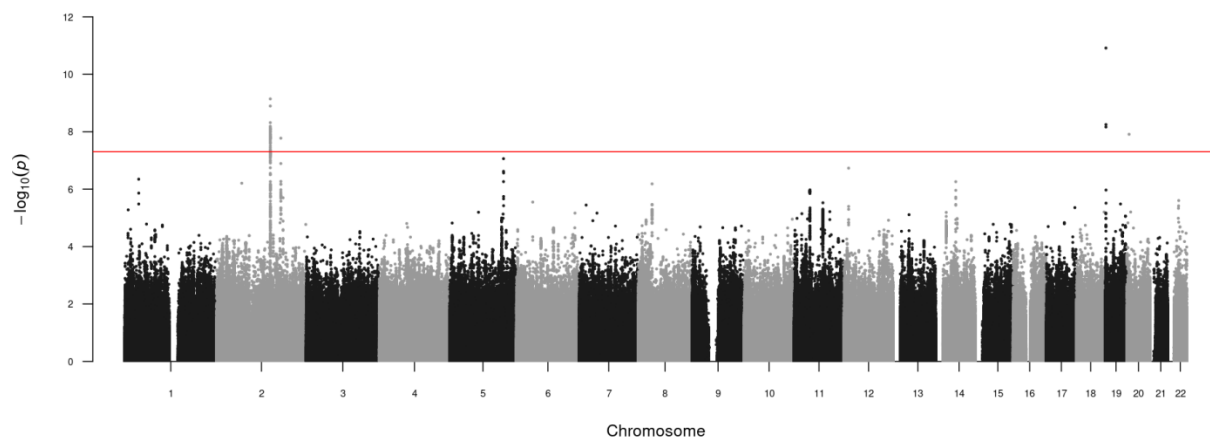
37. Pierce, B.L., and Burgess, S. (2013). Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. *Am J Epidemiol* 178, 1177-1184.
38. Hemani, G., Zheng, J., Elsworth, B., Wade, K.H., Haberland, V., Baird, D., Laurin, C., Burgess, S., Bowden, J., Langdon, R., et al. (2018). The MR-Base platform supports systematic causal inference across the human phenome. *Elife* 7.
39. Michailidou, K., Lindstrom, S., Dennis, J., Beesley, J., Hui, S., Kar, S., Lemacon, A., Soucy, P., Glubb, D., Rostamianfar, A., et al. (2017). Association analysis identifies 65 new breast cancer risk loci. *Nature* 551, 92-94.
40. Day, F., Karaderi, T., Jones, M.R., Meun, C., He, C., Drong, A., Kraft, P., Lin, N., Huang, H., Broer, L., et al. (2018). Large-scale genome-wide meta-analysis of polycystic ovary syndrome suggests shared genetic architecture for different diagnosis criteria. *PLoS Genet* 14, e1007813.
41. Burgess, S., and Thompson, S.G. (2015). Mendelian randomization: methods for using genetic variants in causal estimation. (CRC Press).
42. Park, J., Long, D.T., Lee, K.Y., Abbas, T., Shibata, E., Negishi, M., Luo, Y., Schimenti, J.C., Gambus, A., Walter, J.C., et al. (2013). The MCM8-MCM9 complex promotes RAD51 recruitment at DNA damage sites to facilitate homologous recombination. *Mol Cell Biol* 33, 1632-1644.
43. Lutzmann, M., Bernex, F., da Costa de Jesus, C., Hodroj, D., Marty, C., Plo, I., Vainchenker, W., Tosolini, M., Forichon, L., Bret, C., et al. (2019). MCM8- and MCM9 Deficiencies Cause Lifelong Increased Hematopoietic DNA Damage Driving p53-Dependent Myeloid Tumors. *Cell Rep* 28, 2851-2865.e2854.
44. Griffin, W.C., and Trakselis, M.A. (2019). The MCM8/9 complex: A recent recruit to the roster of helicases involved in genome maintenance. *DNA Repair (Amst)* 76, 1-10.

45. 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.
46. Schuh-Huerta, S.M., Johnson, N.A., Rosen, M.P., Sternfeld, B., Cedars, M.I., and Reijo Pera, R.A. (2012). Genetic markers of ovarian follicle number and menopause in women of multiple ethnicities. *Hum Genet* 131, 1709-1724.
47. Hoyos, L.R., Visser, J.A., McLuskey, A., Chazenbalk, G.D., Grogan, T.R., and Dumesic, D.A. (2020). Loss of anti-Müllerian hormone (AMH) immunoactivity due to a homozygous AMH gene variant rs10417628 in a woman with classical polycystic ovary syndrome (PCOS). *Hum Reprod* 35, 2294-2302.
48. Beltran, M., Puig, I., Pena, C., Garcia, J.M., Alvarez, A.B., Pena, R., Bonilla, F., and de Herreros, A.G. (2008). A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. *Genes Dev* 22, 756-769.
49. Conidi, A., Cazzola, S., Beets, K., Coddens, K., Collart, C., Cornelis, F., Cox, L., Joke, D., Dobрева, M.P., Dries, R., et al. (2011). Few Smad proteins and many Smad-interacting proteins yield multiple functions and action modes in TGFbeta/BMP signaling in vivo. *Cytokine Growth Factor Rev* 22, 287-300.
50. Postigo, A.A. (2003). Opposing functions of ZEB proteins in the regulation of the TGFbeta/BMP signaling pathway. *Embo j* 22, 2443-2452.
51. Estienne, A., Pierre, A., di Clemente, N., Picard, J.Y., Jarrier, P., Mansanet, C., Monniaux, D., and Fabre, S. (2015). Anti-Müllerian hormone regulation by the bone morphogenetic proteins in the sheep ovary: deciphering a direct regulatory pathway. *Endocrinology* 156, 301-313.

52. Pierre, A., Estienne, A., Racine, C., Picard, J.Y., Fanchin, R., Lahoz, B., Alabart, J.L., Folch, J., Jarrier, P., Fabre, S., et al. (2016). The Bone Morphogenetic Protein 15 Up-Regulates the Anti-Mullerian Hormone Receptor Expression in Granulosa Cells. *J Clin Endocrinol Metab* 101, 2602-2611.
53. Gill, R.M., Gabor, T.V., Couzens, A.L., and Scheid, M.P. (2013). The MYC-associated protein CDCA7 is phosphorylated by AKT to regulate MYC-dependent apoptosis and transformation. *Mol Cell Biol* 33, 498-513.
54. Horikoshi, M., Day, F.R., Akiyama, M., Hirata, M., Kamatani, Y., Matsuda, K., Ishigaki, K., Kanai, M., Wright, H., Toro, C.A., et al. (2018). Elucidating the genetic architecture of reproductive ageing in the Japanese population. *Nat Commun* 9, 1977.
55. Hunzicker-Dunn, M.E., Lopez-Biladeau, B., Law, N.C., Fiedler, S.E., Carr, D.W., and Maizels, E.T. (2012). PKA and GAB2 play central roles in the FSH signaling pathway to PI3K and AKT in ovarian granulosa cells. *Proc Natl Acad Sci U S A* 109, E2979-2988.
56. Elks, C.E., Perry, J.R., Sulem, P., Chasman, D.I., Franceschini, N., He, C., Lunetta, K.L., Visser, J.A., Byrne, E.M., Cousminer, D.L., et al. (2010). Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. *Nat Genet* 42, 1077-1085.
57. Lawlor, D.A., Harbord, R.M., Sterne, J.A., Timpson, N., and Davey Smith, G. (2008). Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med* 27, 1133-1163.
58. Kelsey, T.W., Wright, P., Nelson, S.M., Anderson, R.A., and Wallace, W.H. (2011). A validated model of serum anti-mullerian hormone from conception to menopause. *PLoS One* 6, e22024.
59. Burgess, S., Davies, N.M., and Thompson, S.G. (2016). Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol* 40, 597-608.



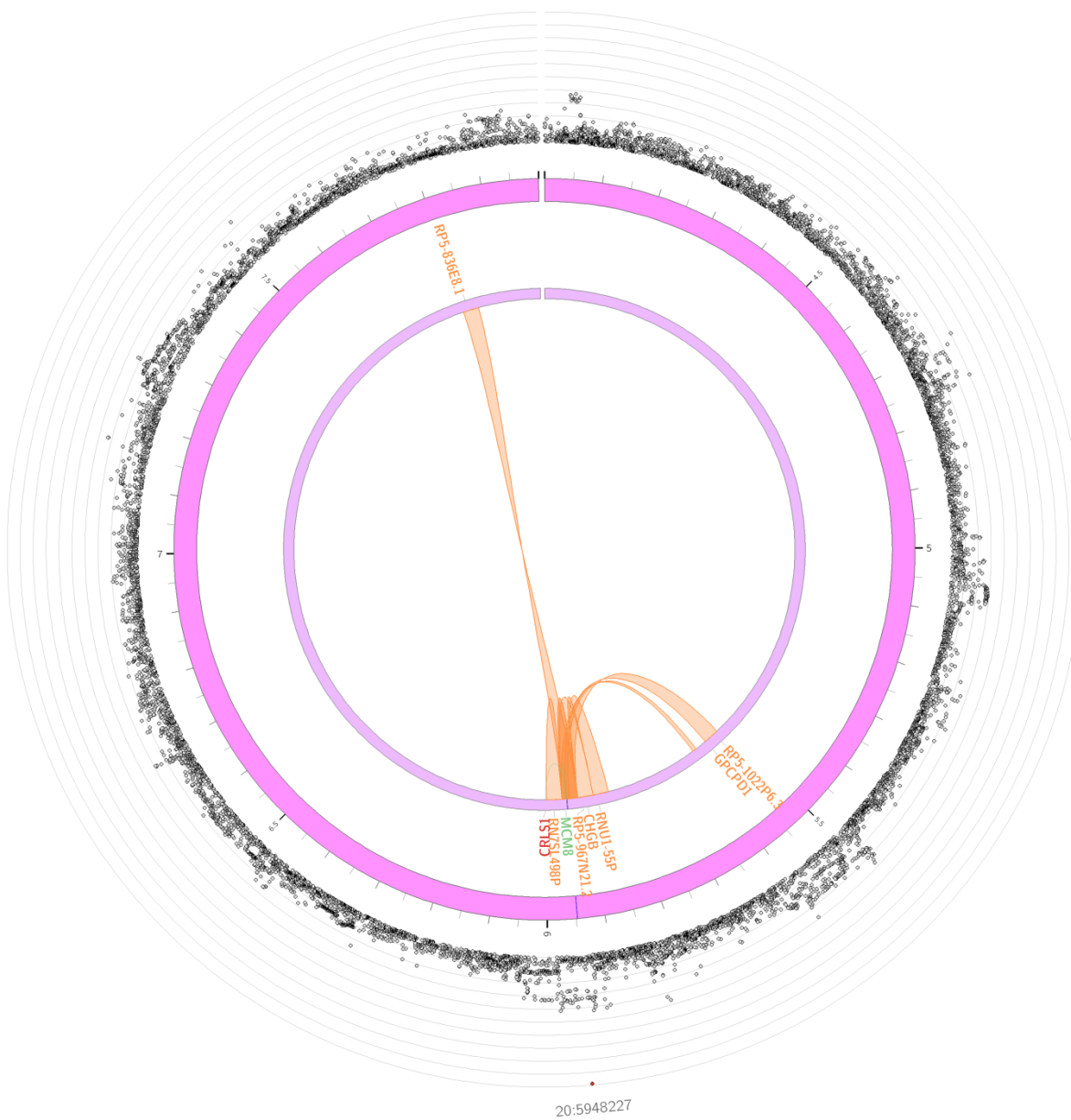
## Figures



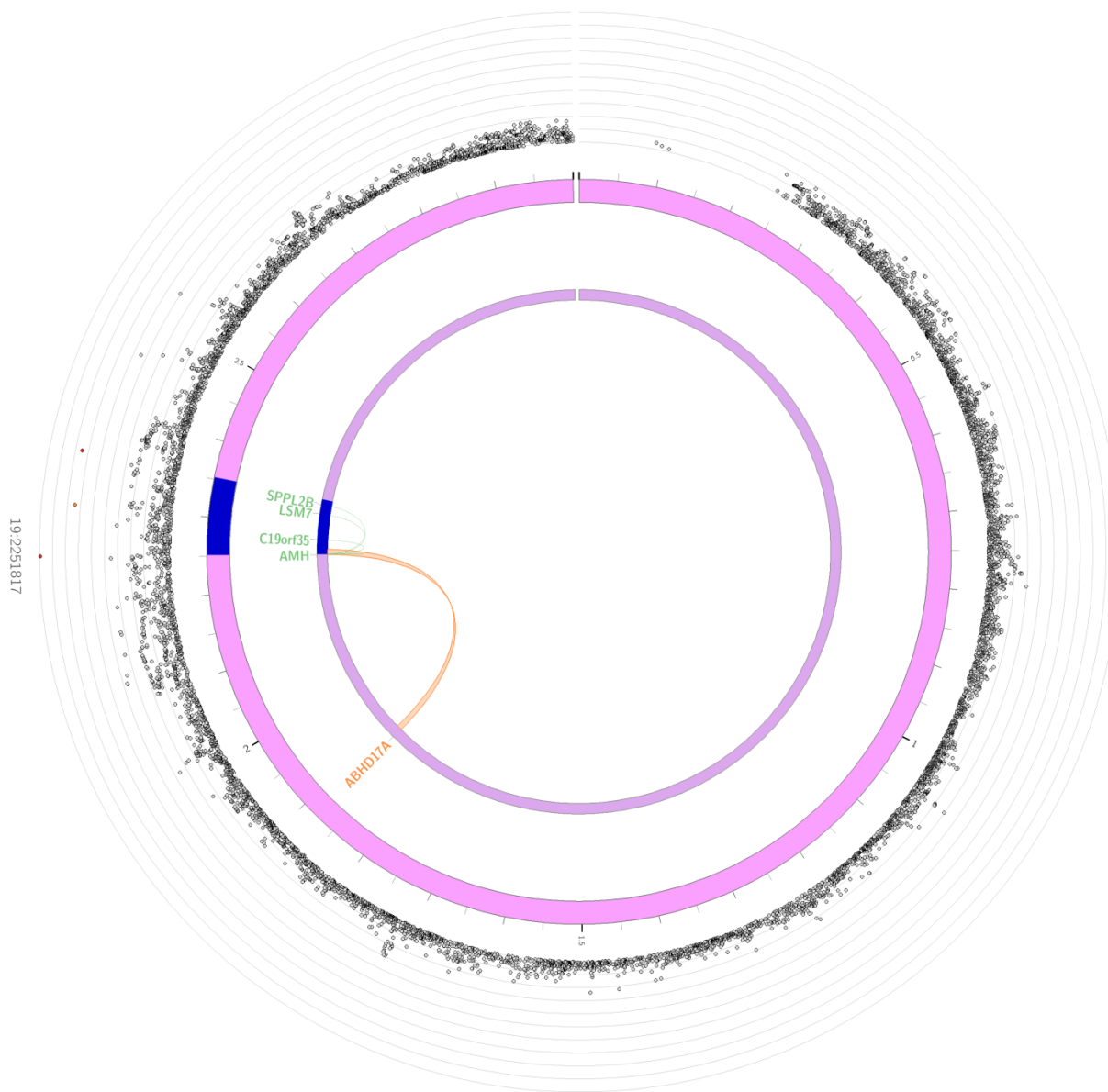
**Figure 1: Manhattan plot of genome-wide association results for inverse normally transformed AMH in women.**

Results are association results from meta-analysis of inverse normally transformed AMH in 7,049 women of European ancestry. Individual studies adjusted analyses for age at AMH measurement and population stratification (through kinship matrix or 10 principal components).

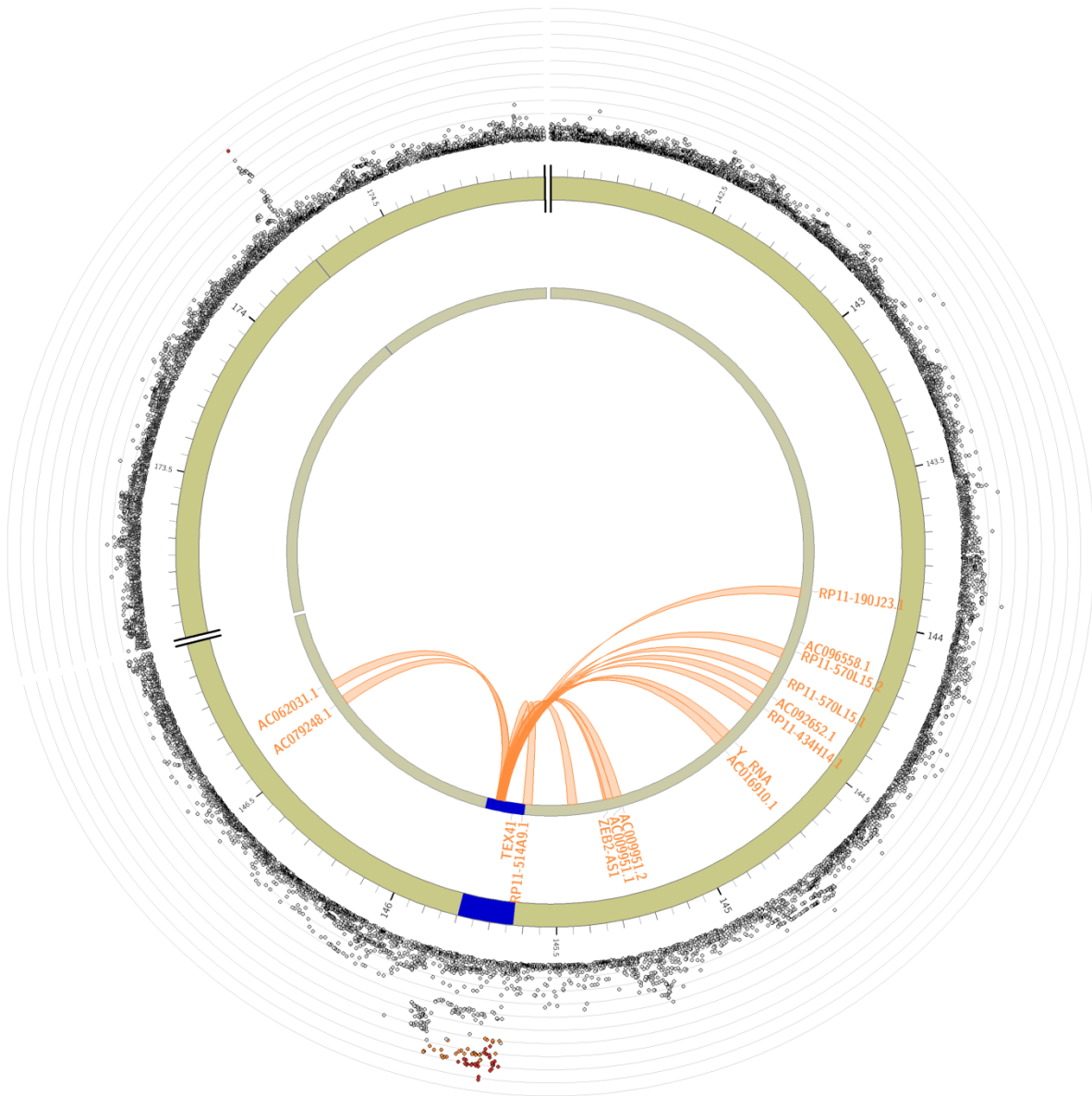
A



## B



C



**Figure 2: Circos plots for the genome-wide significant loci for inverse normally transformed AMH in women.**

Circos plots are presented for each of the identified loci: *MCM8* (panel A), *AMH* (panel B), *TEX41* and *CDCA7* (panel C). The outer layer represents the Manhattan plot. The second (including genomic positions) and third layers represents the chromosome ring, genomic risk loci are depicted in blue. Only genes mapped by either eQTLs or

chromatin mapping are plotted. Genes only mapped by eQTLs are green, genes only mapped by chromatin interactions are orange, and genes mapped by both have a red colour. Orange coloured lines represent chromatin interactions, green coloured lines are eQTL links. Plots were created using the FUMA platform<sup>26</sup>.

## Tables

**Table 1: Distributions of AMH and age per participating study.**

<b>Study</b>	<b>N</b>	<b>AMH, pmol/L (median (IQR))</b>	<b>Age at blood collection, years (median (IQR))</b>
Studies contributing to summary statistics GWAS Ruth et al. <sup>8*</sup>			
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	642	6.1 (2.0, 13.9)	44 (41, 47)
Additional studies			
Doetinchem Cohort Study	2,084	10.9 (2.9, 25.6)	37.2 (31.2, 42.9)
ALSPAC mothers	1,885	2.0 (0.4, 5.2)	46 (44, 49)
ALSPAC daughters	1,196	26.1 (18.2, 39.8)	15.3 (15.3, 15.5)
SWAN	425	1.1 (0.2, 3.3)	47.3 (45.3, 49.3)
<b>Total</b>	<b>7,049</b>		

\* In the original study ALSPAC mothers were included as well, but in the current analyses summary statistics from ALSPAC were included separately to assess potential genomic inflation due to inclusion of both ALSPAC mothers and daughters. Therefore, we treated the ALSPAC mothers as individual study.

**Table 2: Loci significantly associated ( $p < 5 \times 10^{-8}$ ) with inverse normally transformed AMH in women.**

Nearest gene	SNP	Chr	Pos	EA	OA	EAF	N	Imputation quality	Effect (SE)	P	Direction	P <sub>het</sub>	Percentage of variance in AMH explained
<i>AMH</i>	rs10417628	19	2,251,817	T	C	0.02	7049	0.83	-0.34 (0.05)	$1.2 \times 10^{-11}$	-----	0.14	0.50 %
<i>TEX41</i>	rs13009019	2	145,670,572	A	G	0.69	7049	0.95	-0.09 (0.01)	$7.2 \times 10^{-10}$	-----	0.24	0.35%
<i>MCM8</i>	rs16991615	20	5,948,227	A	G	0.07	7049	0.99	0.16 (0.03)	$1.2 \times 10^{-8}$	++++	0.0009	0.30%
<i>CDCA7</i>	rs11683493	2	174,259,325	T	C	0.57	7049	0.97	-0.08 (0.01)	$1.7 \times 10^{-8}$	-----	0.03	0.32%

Definition of columns: nearest gene, nearest gene identified using DEPICT tool (Subjects and Methods); SNP, genetic variant identified as lead SNP; Chr, chromosome; Pos, base pair position genomic build GRCh37; EA, effect allele; OA, other allele; EAF, effect allele frequency; N, number of samples contributing to estimate; Imputation quality, mean imputation quality over the included studies; Effect (SE); effect size and corresponding standard error; P, p-value; Direction, direction of effect for previous GWAS, ALSPAC mothers, ALSPAC daughters, Doetinchem Cohort Study and SWAN, respectively; P<sub>het</sub>, p-value for heterogeneity of effect across studies.

**Table 3: Mendelian randomization estimates for causal effects of circulating AMH on breast cancer and PCOS risk.**

Outcome	Method	Odds Ratio	95% CI	p
Breast Cancer	IVW	1.00	0.74 - 1.36	0.98
	Wald ratio estimate for rs10417628 (AMH)	0.99	0.87 - 1.12	0.85
	Wald ratio estimate for rs13009019 (TEX41)	0.84	0.72 - 0.97	0.02
	Wald ratio estimate for rs16991615 (MCM8)	1.60	1.37 - 1.87	1.79 x 10 <sup>-9</sup>
	Wald ratio estimate for rs11683493 (CDCA7)	0.76	0.65 - 0.89	9.41 x 10 <sup>-4</sup>
PCOS	IVW	1.29	0.85 - 1.95	0.23
	Wald ratio estimate for rs10417628 (AMH)	1.27	0.64 - 2.56	0.49
	Wald ratio estimate for rs13009019 (TEX41)	1.66	0.80 - 3.45	0.18
	Wald ratio estimate for rs16991615 (MCM8)	1.75	0.83 - 3.69	0.14
	Wald ratio estimate for rs11683493 (CDCA7)	0.66	0.29 - 1.50	0.32

Odds ratio and 95%CI are per 1 unit increase in inverse normally transformed AMH.

AMH, anti-Müllerian hormone; PCOS, polycystic ovary syndrome; IVW, inverse variance weighted; MR, Mendelian randomization.