

1 **Circulating anti-Müllerian hormone levels in pre-menopausal women: novel genetic**
2 **insights from a GWAS meta-analysis**

3 Natàlia Pujol-Gualdo#(1,2), Minna K. Karjalainen (3,4), Urmo Vösa (1), Riikka K. Arffman (2),
4 Reedik Mägi (1), Justiina Ronkainen(3), Triin Laisk*(1), Terhi T. Piltonen*(2)

5 #first and corresponding author / *co-last authorship

6 1. Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia

7 2. Department of Obstetrics and Gynecology, Research Unit of Clinical Medicine, University
8 of Oulu, Oulu, Finland

9 3. Research Unit of Population Health, Faculty of Medicine, University of Oulu, Oulu, Finland

10 4. Northern Finland Birth Cohorts, Arctic Biobank, Infrastructure for Population Studies,
11 Faculty of Medicine, University of Oulu, Oulu, Finland

12 **Abstract**

13 **Study question:** Can a genome-wide association study (GWAS) meta-analysis, including a
14 large sample of young premenopausal women from a founder population from Northern
15 Finland, identify novel genetic variants for circulating anti-Müllerian hormone (AMH) levels
16 and provide insights into biological pathways and tissues involved in AMH regulation?

17 **Summary answer:** We identified six loci associated with AMH levels at $P < 5 \times 10^{-8}$,
18 including the previously reported *MCM8*, *AMH* and *TEX41* loci, and three novel signals in or
19 near *CHEK2*, *BMP4* and *EIF4EBP1*. Gene set enrichment analysis highlighted significant
20 enrichment in renal system vasculature morphogenesis and tissue enrichment analysis
21 ranks the pituitary gland as a top associated tissue.

22 **What is known already:** AMH is expressed by preantral and small antral stage ovarian
23 follicles in women, and variation in age-specific circulating AMH levels has been associated
24 with several health conditions. However, the biological mechanisms underlying the
25 association between health conditions and AMH levels are not yet fully understood. Previous
26 GWAS have identified loci associated with AMH levels in pre-menopausal women, but they
27 were limited by small sample sizes or focused mostly on older pre-menopausal women.

28 **Study design, size, duration:** We performed a GWAS meta-analysis for AMH level
29 measurements in 9,668 pre-menopausal women.

30 **Participants/materials, setting, methods:** We performed a GWAS meta-analysis in which
31 we combined 2,619 AMH measurements (at age 31 years old) from a prospective founder
32 population cohort (Northern Finland Birth Cohort 1966, NFBC1966) with a previous GWAS
33 meta-analysis that included 7,049 pre-menopausal women (spanning age range 15-48).
34 NFBC1966 AMH measurements were quantified using an automated assay (Elecsys® AMH
35 Plus (Roche)). We annotated the genetic variants, combined different data layers to prioritise
36 potential candidate genes, described significant pathways and tissues enriched by the

37 GWAS signals, identified plausible regulatory roles using colocalization analysis and
38 leveraged publicly available summary statistics to assess genetic and phenotypic
39 correlations with multiple traits.

40 **Main results and the role of chance:** Three novel genome-wide significant loci were
41 identified. One of these is in complete linkage disequilibrium with c.1100delC in *CHEK2*,
42 which is found to be 4-fold enriched in the Finnish population compared to other European
43 populations. We propose a plausible regulatory effect of some of the GWAS variants linked
44 to AMH, as they colocalise with GWAS signals associated with gene expression levels of
45 *BMP4*, *TEX41* and *EIFBP41*. Gene set analysis highlighted significant enrichment in renal
46 system vasculature morphogenesis and tissue enrichment analysis ranked the pituitary
47 gland as the top association.

48 **Large scale data:** The GWAS meta-analysis summary statistics will be available for
49 download from the GWAS Catalog. Accession numbers will be provided upon publication.

50 **Limitations, reasons for caution:** This study only included women of European ancestry
51 and the unavailability of sufficiently sized relevant tissue data in gene expression datasets
52 hinders the assessment of potential regulatory effects in reproductive tissues.

53 **Wider implications of the findings:** Our results highlight the increased power of founder
54 populations and larger sample sizes to boost the discovery of novel trait-associated variants
55 underlying variation in AMH levels, which aided to characterise novel biological pathways
56 and plausible genetic regulatory effects linked with AMH levels variation for the first time.

57 **Study funding / competing interest(s):** This work has received funding from the European
58 Union's Horizon 2020 research and innovation programme under the MATER Marie
59 Sklodowska-Curie grant agreement No. 813707 and Oulu university scholarship foundation
60 (N.P.-G.), Academy of Finland, Sigrid Jusélius Foundation, Novo Nordisk, University of Oulu,
61 Roche Diagnostics (T.T.P). This work was supported by the Estonian Research Council
62 grant 1911 (R.M.). J.R. was supported by the European Union's Horizon 2020 research and
63 innovation program under grant agreements No. 874739 (LongITools), 824989 (EUCAN-
64 Connect), 848158 (EarlyCause) and 733206 (LifeCycle). U.V. was supported by the
65 Estonian Research Council grant PRG (PRG1291). The NFBC1966 received financial
66 support from University of Oulu Grant no. 24000692, Oulu University Hospital Grant no.
67 24301140, ERDF European Regional Development Fund Grant no. 539/2010 A31592.

68 **Keywords:** AMH; anti-Mullerian hormone, genome-wide association study, low-frequency
69 variants, reproductive ageing

70 INTRODUCTION

71 Anti-Müllerian hormone (AMH) is a member of the transforming growth factor β (TGF- β)
72 superfamily, which includes the bone morphogenic proteins (BMPs), growth differentiation
73 factors, activins and inhibins. Despite owing its name to its classical role in male sexual
74 differentiation, AMH is also expressed by the ovarian granulosa cells during the primary to
75 small antral stage of follicle development (Weenen *et al.*, 2004). In adult women, serum
76 AMH levels decrease with age, with undetectable levels following menopause, signalling
77 depletion of ovarian reserve (Finkelstei *et al.*, 2020). As a result, AMH is primarily known as

78 a serum marker for ovarian reserve. Previous studies have indicated that variations in age-
79 specific circulating AMH levels are linked with several health conditions, including breast
80 cancer (Ge *et al.*, 2018) and polycystic ovary syndrome (PCOS) (Homburg and Crawford,
81 2014). Therefore, identifying genetic determinants of inter-individual variation in AMH
82 measurements may offer valuable insight into their biological mechanistic effects and impact
83 in health and disease beyond reproductive ageing.

84 Three previous genome-wide association studies have identified a few loci associated with
85 AMH levels in pre-menopausal women (Schuh-Huerta *et al.*, 2012; Ruth *et al.*, 2019;
86 Verdiesen *et al.*, 2022). However, these studies were limited by small sample size (Schuh-
87 Huerta *et al.*, 2012) or focusing only or mostly on older pre-menopausal women (Ruth *et al.*,
88 2019; Verdiesen *et al.*, 2022). Since AMH decreases with age also the AMH variability
89 decreases in older ages, and therefore there is less power to detect associations in older
90 pre-menopausal women compared to younger pre-menopausal women. To address these
91 limitations, our current work analyses data from a single time point measured in relatively
92 young pre-menopausal women and combines this dataset with a previous meta-analysis
93 (Verdiesen *et al.*, 2022), to identify and characterise additional loci associated with AMH
94 levels in pre-menopausal women.

95 The studies conducted so far have focused on European ancestry, mainly powered to
96 discover associations with common genetic variants. Nonetheless, founder populations such
97 as those found in northern Finland, represent a powerful resource to accelerate discovery of
98 new biological mechanisms driven by low-frequency alleles that have risen to higher
99 frequency due to unique demographic history (Kurki *et al.*, 2023).

100 In the present study, we added 2,619 AMH measurements (at age 31 years old) from a
101 prospective founder population cohort (Northern Finland Birth Cohort 1966, NFBC1966) to a
102 previous GWAS meta-analysis of 7,049 pre-menopausal women (spanning age range 15-
103 48) (Verdiesen *et al.*, 2022), reaching the largest sample size for assessing AMH variation in
104 women from reproductive age to date (N=9,668). Beyond detecting three out of the four
105 previously identified signals near *TEX41*, *MCM8* and *AMH*, we identified three novel signals
106 near *EIF4EBP1*, *BMP4*, and *CHEK2*. In the *CHEK2* locus, the c.1100delC variant is enriched
107 in the Finnish population and in complete linkage disequilibrium ($r^2=1$) with the lead variant
108 identified in that locus. Additionally, we tested regulatory effects of GWAS variants on
109 specific genes and gene transcript expression in multiple tissues using colocalization
110 analysis, evaluated gene set and tissue set enrichments, and estimated genetic correlations
111 and phenotypic associations across multiple phenotypes, by leveraging publicly available
112 summary statistics.

113 **METHODS**

114 **Study design and setting**

115

116 This study is based on the prospective population-based NFBC1966 study (University of
117 Oulu, 1966, <http://urn.fi/urn:nbn:fi:att:bc1e5408-980e-4a62-b899-43bec3755243>). During
118 1966, 12231 children (of them 5,889 females) were born in the two northernmost provinces
119 of Finland (covering 48% of Finnish territory). Originally, the study was set to evaluate early
120 life factors on long-term health and work ability. The cohort population has been followed up
121 at four different time points: 1, 14, 31 and 46 years of age (set by the cohort centre).
122 Comprehensive questionnaires on female health and clinical examinations with biological
123 data collection were performed at ages 31 and 46 years.

124

125 The detailed cohort description and follow up protocol has been previously published
126 (Nordström *et al.*, 2022). Briefly, in 1997 (the 31-year follow-up) women living in the Northern
127 Finland area or in the Helsinki metropolitan area (n=4074 women) were invited to a clinical
128 examination, in which 3,127 (77%) women participated. The present study includes 2,619
129 women with AMH and genetic data available at 31 years old (median=31.1, interquartile
130 range (IQR)=30.9-31.4).

131
132 For the meta-analysis, publicly available summary statistics from an independent second
133 study evaluating AMH variations across 7,049 pre-menopausal women were used
134 (Verdiesen *et al.*, 2022). Publicly available summary statistics were downloaded from the
135 GWAS catalog (<https://www.ebi.ac.uk/gwas/publications/35274129>) under accession
136 number GCST90104596. This study included data from the AMH GWAS meta-analysis by
137 Ruth *et al.* (n=3,334) (Ruth *et al.*, 2019), which meta-analysed four studies, and three
138 cohorts which were additionally analysed and joined to the latter, conforming the most recent
139 meta-analysis (n=7,049) (Verdiesen *et al.*, 2022). The cohort details from the current meta-
140 analysis (n=9,668) can be found in Supplementary Table 1. More details of cohort
141 description of the studies included can be found in (Verdiesen *et al.*, 2022). All studies had
142 received ethical approval from institutional ethics committees.

143

144

145 **AMH measurement in NFBC1966**

146 Serum samples were drawn in 1997 (the 31-year follow up), sealed and stored at -20°C
147 since then. Serum AMH concentrations were measured using the automated Elecsys®
148 electrochemiluminescence immunoassay on a Cobas E411 analyser according to the
149 manufacturer's instructions (Roche Diagnostics, Germany). More detailed information on
150 AMH measurement has been published previously (Piltonen *et al.*, 2023).

151 The assay limits of detection and quantitation were 0.01 ng/ml and 0.03 ng/ml for AMH.
152 Intra- and inter-assay coefficients of variance were 1.0–1.8% and 2.9–4.4% for AMH. Limits
153 above the measuring ranges were 23 ng/ml for AMH.

154 AMH concentrations were converted to pmol/l using 1 ng/ml = 7.14 pmol/l. As AMH levels
155 are not normally distributed, for the following analysis AMH measurements were transformed
156 using rank-based inverse normal transformation, as done previously (Ruth *et al.*, 2019). For
157 rank-based inverse normal transformation we used the package *RNOmni* in R v3.6.3.

158 **Genotyping details and association analysis**

159 Genotyping of the NFBC1966 samples was performed with the Illumina Infinium
160 HumanCNV370-Duo array. Samples were excluded based on call rate < 95%, gender
161 mismatch, relatedness (identity by descent (IBD) estimate $\hat{p}_{ihat} > 0.2$) and outlying
162 heterozygosity. SNPs were excluded based on missingness rate >5%, HWE $p < 0.000001$
163 and MAF <1%. Genotypes were imputed with Beagle 4.1 using the SiSu v3 imputation
164 reference panel, which consisted of 3775 individuals of Finnish ancestry with sequenced
165 whole genomes.

166

167 Association analysis was performed using SNPTTEST v2.5.4-beta1, adjusted by the first 10
168 genetic principal components. Markers with INFO score >0.4 were kept for the meta-
169 analysis. Positions were converted to build hg37 to harmonise the data with the summary

170 statistics available from Verdiesen *et al.* before running the meta-analysis, using LiftOver
171 (Kent *et al.*, 2002).
172

173 **GWAS meta-analysis**

174 We conducted an inverse variance weighted fixed-effects meta-analysis with single genomic
175 control correction using GWAMA software (v2.2.2) (Mägi and Morris, 2010). A total of
176 13,903,812 variants were included in the meta-analysis of 9,688 AMH measurements. After
177 meta-analysis we kept variants present in both studies (totalling 7,900,839 variants) for
178 downstream analysis. Lead SNPs were identified as SNVs independent from each other
179 with P -value less than or equal to 5×10^{-8} . The maximum distance between LD blocks of
180 independent SNPs to merge into a single genomic locus was set to 500 kb.

181 **Heritability analysis**

182 The heritability (h^2) was estimated by single-trait LD score regression using the meta-
183 analysis summary statistics and HapMap 3 LD-scores using LDSC v1.0.1
184 (<https://github.com/bulik/ldsc>) (Bulik-Sullivan *et al.*, 2015).

185 **Annotation of GWAS signals**

186 We used FUMA (v1.4.0) to annotate the GWAS signals (Watanabe *et al.*, 2017). FUMA is an
187 online platform that performs annotation of GWAS signals using data from several
188 databases. First, FUMA identifies lead SNPs (p -value $\leq 5 \times 10^{-8}$ and SNP pairwise LD
189 $r^2 < 0.1$, based on 1000G European reference) and independent significant SNPs and each
190 risk locus (p -value $< 5 \times 10^{-8}$ and LD $r^2 < 0.6$) (Supplementary Table 2). Then FUMA
191 identifies potential candidate SNPs that are in LD with any of the identified independent
192 significant SNPs and annotates them via linking with several databases (ANNOVAR,
193 RegulomeDB, CADD scores etc.), which gives information on their location, functional
194 impact, and potential regulatory effects.

195

196

197 **Colocalization analysis**

198

199 We used HyPrColoc (Foley *et al.*, 2021), a colocalisation method for identifying the overlap
200 between our GWAS meta-analysis signals and cis-quantitative trait loci (cis-QTL) signals
201 from different tissues and cell types (expression QTLs, transcript QTLs, exon QTLs and
202 exon usage QTLs available in the eQTL Catalogue (Kerimov *et al.*, 2021)). We lifted the
203 GWAS summary statistics over to hg38 build to match the eQTL Catalogue using binary
204 liftOver tool (https://genome.sph.umich.edu/wiki/LiftOver#Binary_liftOver_tool).

205 For the genome-wide significant ($p < 5 \times 10^{-8}$) GWAS loci identified we extracted the +/-
206 500kb of its top hit from QTL datasets and ran the colocalization analysis against eQTL
207 Catalogue traits. For each eQTL Catalogue dataset we included all the QTL features which
208 shared at least 80% of tested variants with the variants present in our GWAS region. We
209 used the default settings for HyPrColoc analyses and did not specify any sample overlap
210 argument, because HyPrColoc paper (Foley *et al.*, 2021) demonstrates that assuming trait
211 independence gives reasonable results. HyPrColoc outputs the posterior probability that
212 genetic association signals for those traits are colocalising (we considered two or more

213 signals to colocalize if the posterior probability for a shared causal variant (PP4) was 0.8 or
214 higher). All results with a PP4 > 0.8 can be found in Supplementary Table 3.

215 **Gene prioritization**

216 In order to nominate plausible candidate genes in each locus, we prioritized genes according
217 to different levels of evidence: (1) genes containing lead variants that are coding variants or
218 coding variants in high LD ($r^2 \geq 0.6$) with lead variants; (2) genes whose expression was
219 regulated by eQTL variants that showed significant (posterior probability >0.8) colocalization
220 with our GWAS signals; (3) genes which showed a plausible biological role as defined
221 previously in the research literature.

222 **Gene set and tissue set enrichment analysis**

223 Gene set enrichment analysis and tissue set enrichment analysis were performed using
224 MAGMA v1.08 implemented in FUMA v1.4.0. Gene sets were obtained from Msigcbv7.0 for
225 'Curated gene sets' and 'GO terms'. A total of 15,485 gene terms were queried. The results
226 of this analysis are presented in Supplementary Table 4 and Supplementary Figure 2.
227 Tissue expression analysis was performed for 53 tissue types using MAGMA. The results of
228 this analysis are presented in Supplementary Table 5 and Supplementary Figure 2.

229 **Genetic correlation analysis**

230 The Complex Traits Genetics Virtual Lab (CTG-VL, <https://vl.genoma.io/>) was used to
231 calculate genetic correlations between the AMH meta-analysis and 1335 traits. We applied a
232 multiple testing correction (Benjamini-Hochberg FDR < 5%) to determine statistical
233 significance using the *p.adjust* function in R 3.6.3. Results of the genetic correlation analysis
234 are presented in Supplementary Figure 3 and Supplementary Table 6.

235

236 **Phenome-wide associations**

237

238 We evaluated the association of our GWAS lead variants and variants in high LD with these
239 by using human health traits and phenotypes available in the GWAS Catalog (e0_r2022-11-
240 29) as implemented in FUMA. We also queried the lead variants of this study from the
241 FinnGen study (data freeze 9, total n=377,277), which combines samples collected from
242 Finnish biobanks to digital health care data (Kurki *et al.*, 2023). The results obtained are
243 found in Supplementary Table 7 and 8 and Supplementary Figure 4.

244

245

246

247

248

249

250

251 **RESULTS**

252

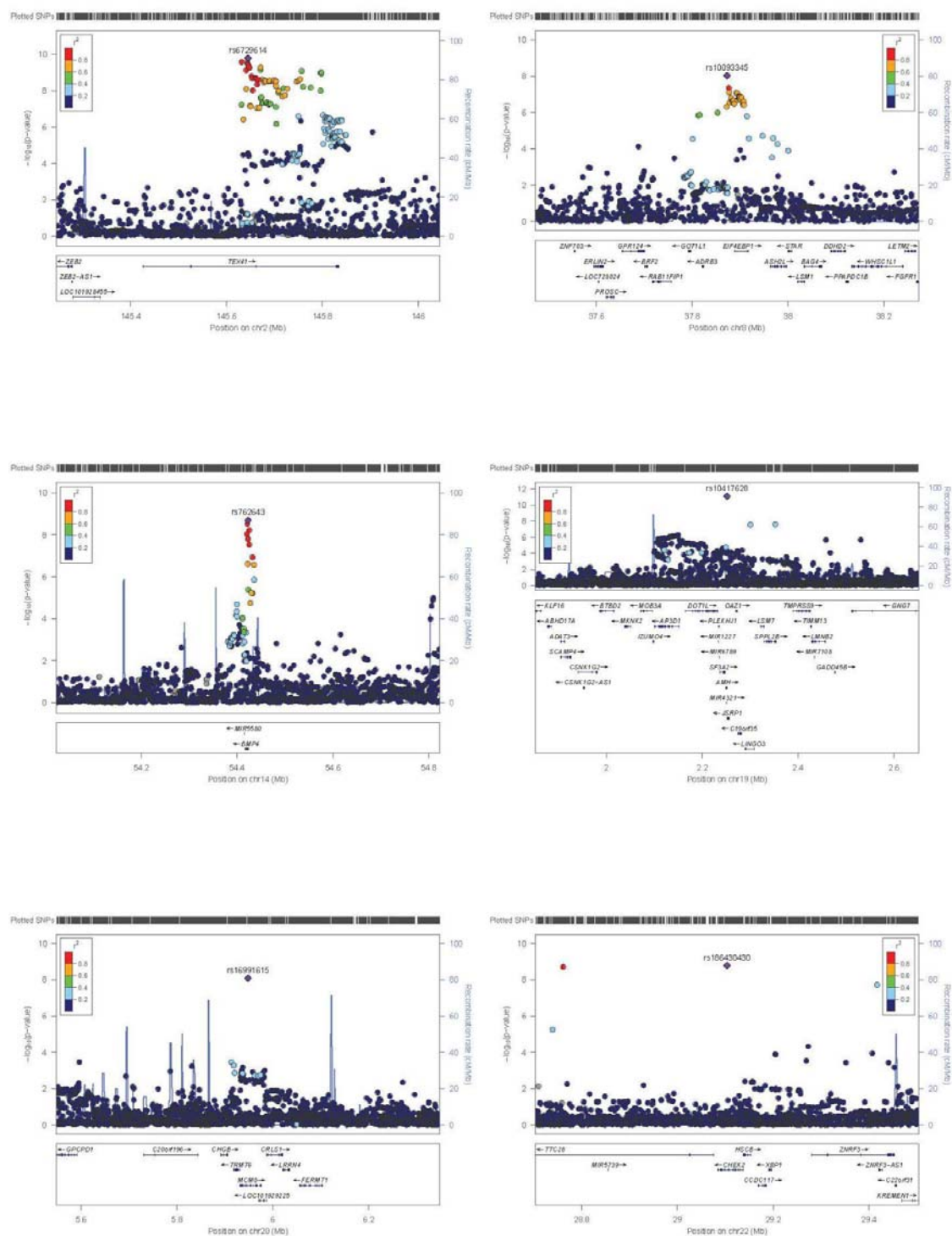
253 **Genome-wide association study for AMH levels**

254 To detect genetic factors associated with circulating AMH levels, we performed a GWAS
255 meta-analysis with data from two studies (NFBC1966 GWAS and summary statistics publicly

256 available from Verdiesen *et al.*, 2022 including a total of 9,668 AMH measurements
257 quantified from women of pre-menopausal age and European ancestry.

258 The meta-analysis identified a total of six associated loci, with six independent signals
259 significantly associated with AMH levels ($P \leq 5 \times 10^{-8}$) (Table 1, Supplementary Figure
260 1). A total of 70 SNPs reached genome-wide significance ($P \leq 5 \times 10^{-8}$).

261



262
263
264
265

Supplementary Figure 1. Regional plots showing the six independent loci significantly associated with AMH levels. Plots were created using the LocusZoom tool (locuszoom.org). The y axis represents $-\log_{10}(P\text{-values})$ for association of variants with AMH

266 levels and the x axis represents the chromosomal positions of the variants. Each lead variant
 267 is shown as a purple diamond in each of the six genetic loci defined.

268 **Table 1. Genome-wide significant signals ($p \leq 5 \times 10^{-8}$) in the meta-analysis for**
 269 **inverse normally transformed anti-Müllerian hormone levels in women. SNPs showing**
 270 **the most significant associations at each locus are shown, and novel genome-wide**
 271 **significant signals are highlighted in bold. EUR; European (non-Finnish), FINN:**
 272 **European (Finnish)**

273

rsID (lead variant)	Chromosome	Position (based on build 37)	Effect allele	Effect allele frequency (EUR/FINN)	Beta (SE)	P-value	Candidate gene	Source of support for candidate gene prioritisation
rs6729614	2	145644874	G	0.26/0.25	0.08 (0.01)	5.56×10^{-11}	<i>ZEB2/TEX41</i>	Colocalisation analysis and biological plausibility
rs10093345	8	37872776	T	0.72/0.68	-0.08 (0.01)	5.05×10^{-09}	<i>EIF4EBP1</i>	Colocalisation analysis and biological plausibility
rs762643	14	54422767	T	0.44/0.41	-0.07 (0.01)	3.99×10^{-09}	<i>BMP4</i>	Colocalisation analysis and biological plausibility

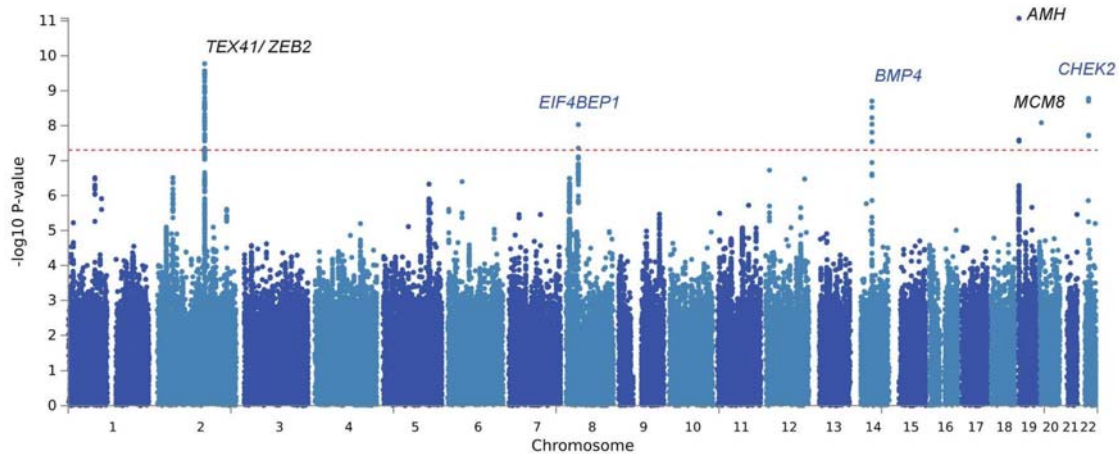
rs10417628	19	2251817	C	0.97/0.99	0.32 (0.04)	9.56 $\times 10^{-12}$	<i>AMH</i>	Coding variant and biological plausibility
rs16991615	20	5948227	A	0.06/0.02	0.16 (0.02)	4.68 $\times 10^{-09}$	<i>MCM8</i>	Coding variant and biological plausibility
rs186430430	22	29103598	C	0.002/0.008	0.79 (0.12)	9.69 $\times 10^{-11}$	<i>CHEK2</i>	Frameshift variant c.1100delC in high LD with the lead variant

274

275

276 Three of the associated loci were previously reported and were located in or near *AMH*
 277 (rs10417628, $p=9.34 \times 10^{-09}$), *TEX41* (rs6729614, $p=1.68 \times 10^{-10}$) and *MCM8* (rs16991615,
 278 $p=4.68 \times 10^{-09}$) (Ruth *et al.*, 2019; Verdiesen *et al.*, 2022), and three of the associated loci
 279 were novel. The novel loci included associations near *EIF4EBP1* (rs10093345, $p=9.34 \times 10^{-09}$),
 280 *BMP4* (rs762643, $p=1.98 \times 10^{-09}$) and *CHEK2* (rs186430430, $p=9.69 \times 10^{-11}$) (see Figure
 281 1, Table 1). All lead variants were present in the two datasets analysed, and the directions of
 282 effects were consistent between the datasets for all six lead variants (Figure 2).

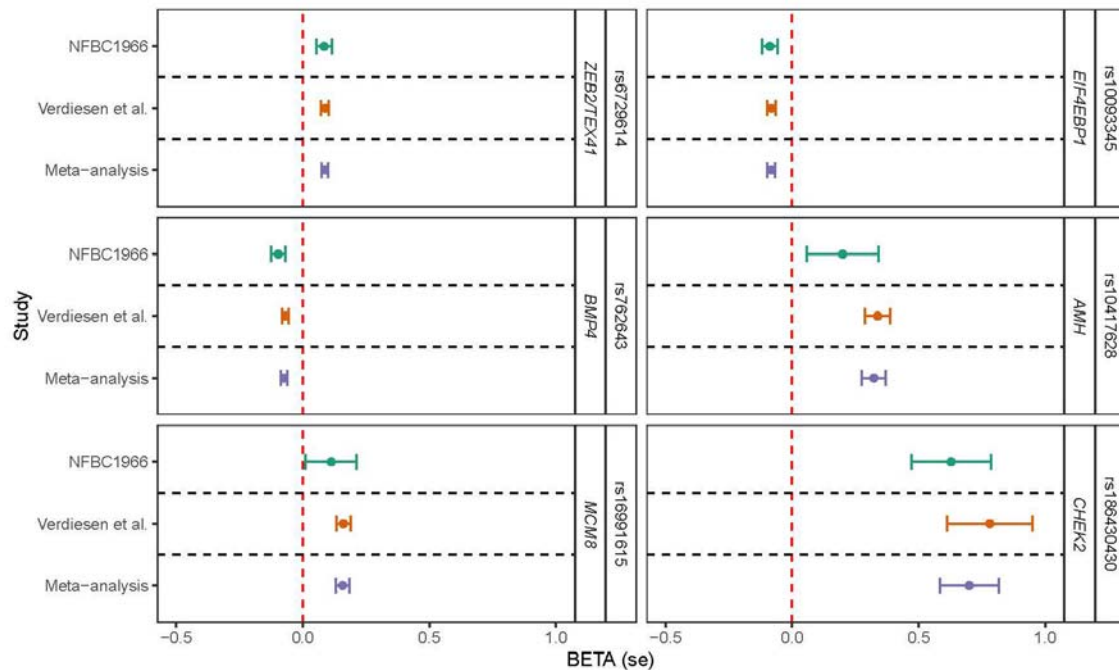
283 There was no evidence of excessive genomic inflation ($\lambda_{GC} = 1.01$) in the GWAS meta-
 284 analysis (LDSC intercept= 1.0039 (s.e. 0.0076)). The observed scale SNP heritability
 285 estimate was 0.13 (s.e. 0.05).



286

287 **Figure 1. Manhattan plot for GWAS meta-analysis for AMH levels in pre-menopausal**
 288 **women (n=9,668).** The novel loci are highlighted in blue. The y axis represents $-\log_{10}(P$ -
 289 values) for association of variants with AMH levels and the x axis represents the
 290 chromosomal positions of the variants. The red horizontal dashed line represents the
 291 threshold for genome-wide significance ($P < 5 \times 10^{-8}$).

292



293

294 **Figure 2. Forest plot of effect estimates for the 6 lead variants associated with AMH**
 295 **levels across datasets meta-analysed.** The betas and standard errors are shown for the
 296 two studies meta-analysed (NFBC1966 (green dots, $n=2,619$), Verdiesen *et al.* (orange dots,
 297 $n=7,049$), and the presented largest meta-analysis (Meta-analysis, purple dots, $n=9,668$).
 298 Candidate genes are also shown next to each genetic lead signal. Betas represent the effect
 299 sizes of each genetic variant association to phenotype, assessing the same risk allele across
 300 cohorts (Effect allele and effect allele frequency are reported in Table 1).

301

302 Characterisation of GWAS signals

303 Two out of the six lead variants that were detected are non-synonymous, making them
304 notably easier to interpret and establishing a more solid foundation for pinpointing potential
305 candidate genes. These non-synonymous variants are found in *AMH* (Anti-Mullerian
306 Hormone) (rs10417628 in chromosome 19, $p=9.56 \times 10^{-12}$) and *MCM8* (Minichromosome
307 Maintenance 8 Homologous Recombination Repair Factor) (rs16991615 in chromosome 20,
308 $p=4.68 \times 10^{-09}$) and were previously described to be associated with AMH levels (Ruth *et al.*,
309 2019; Verdiesen *et al.*, 2022).

310 In the novel locus identified on chromosome 22, the lead variant is an intronic variant near
311 *CHEK2* (Checkpoint Kinase 2) (rs186430430, $p=9.69 \times 10^{-11}$). Interestingly, this variant is in
312 complete linkage disequilibrium ($r^2=1$) with the frameshift variant c.1100delC in *CHEK2*
313 (rs555607708). Despite complete linkage disequilibrium, this rare variant is only found in the
314 association analysis of the NFBC1966 GWAS, showing a p-value of 6.69×10^{-05} and beta of
315 0.79 (s.e.=0.12) in this cohort only and not found in the summary statistics results from
316 Verdiesen *et al.*, likely explained by the fact that this variant is enriched in the Finnish
317 population (MAF= 0.008) compared to other European populations (MAF=0.002), and thus
318 better powered to be captured in a GWAS setting including the former population in the
319 analysis.

320 For the other three loci identified on chromosomes 2, 8 and 14, we defined for the first time
321 plausible shared causal variants between GWAS signals and transcript and gene expression
322 in specific tissues by colocalization analysis. This resulted in the prioritisation of three
323 plausible candidate genes, respectively: *TEX41* (Testis Expressed 41), *EIF4EBP1*
324 (Eukaryotic Translation Initiation Factor 4E Binding Protein 1) and *BMP4* (Bone
325 Morphogenetic Protein 4) (Supplementary Table 3).

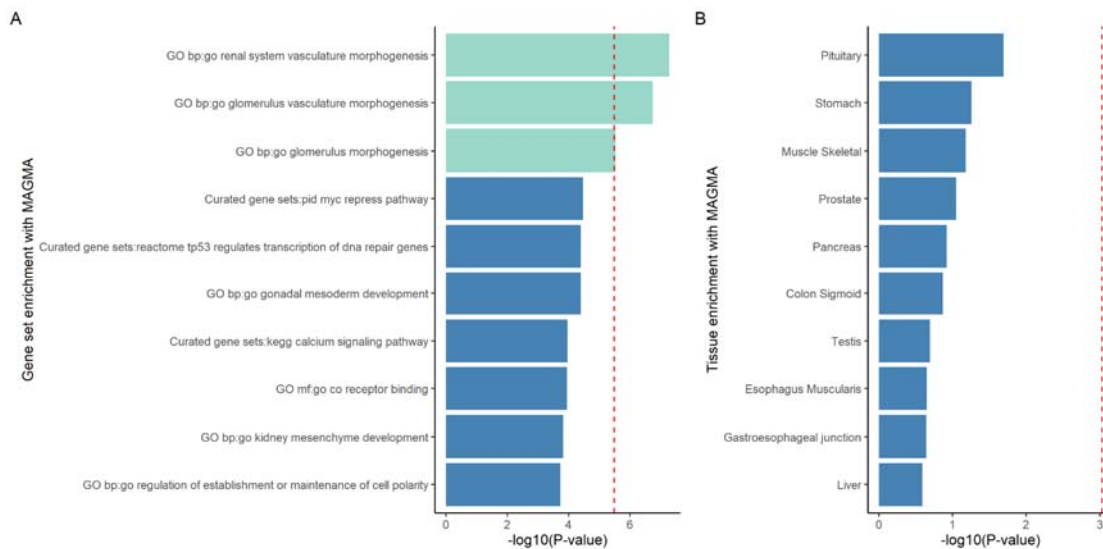
326 For instance, the association signal in chromosome 2, in an intronic region of *TEX41*,
327 colocalized with eQTL signals for *TEX1* transcripts ENST00000414256 in the testis
328 (PP4=0.99) and ENST00000445791 in the skin (PP4=1). *EIF4EBP1* was also prioritised
329 based on colocalization evidence in three tissues, with high colocalization probability
330 (PP4=0.90) between the GWAS signal and cis-eQTLs for *EIF4EBP1* expression in brain
331 frontal cortex tissue, and between the GWAS signal and cis-eQTLs for *EIF4EBP1*
332 expression in left heart ventricle. Additionally, the GWAS association signal in chromosome
333 8 showed colocalization with cis-eQTLs for *BMP4* transcripts ENST00000245451 and
334 ENST00000558489 in colon transverse (PP4=0.84) and *BMP4* transcript
335 ENST00000558489 in tibial nerve.

336 From colocalization results, only in the locus of chromosome 2, we observed that the
337 associations might be explained by a single variant. Namely, we observed that the non-
338 coding transcript (*TEX41*) exon variant rs17407477 and the intronic variant rs786244 were
339 explaining most of the shared association, for the testis transcript expression association
340 (posterior inclusion probability 0.90) and the skin transcript expression association (posterior
341 inclusion probability 0.94), respectively.

342 Although we identified three out of the four signals that were previously reported in
343 Verdiesen *et al.*, 2022, we were unable to detect the signal near *CDCA7* described in the
344 past work (rs11683493 (T), $p=1.7 \times 10^{-8}$). In our analysis using the NFBC66 GWAS data,
345 this variant showed a p-value of 0.09 and therefore, the resulting p-value from the meta-
346 analysis remained below the threshold of significance ($p=1.02 \times 10^{-5}$). Although there might
347 be various sources of variability which explains this discrepancy in our findings, we
348 hypothesise the difference in sample size to be the primary reason.

349 Gene set and tissue set enrichment analysis

350 Gene set enrichment analysis highlighted significant enrichment of renal system vasculature
351 morphogenesis ($p=5.2 \times 10^{-8}$), glomerulus vasculature morphogenesis ($p=1.77 \times 10^{-7}$) and
352 glomerulus morphogenesis ($p=2.99 \times 10^{-6}$) (see Supplementary Table 4 and Supplementary
353 Figure 2). Tissue expression analysis yielded the strongest enrichment in the pituitary gland,
354 even though not reaching significance after multiple testing correction ($p_{\text{non-adj}}=0.02$) (see
355 Supplementary Figure 2 and Supplementary Table 5 and Supplementary Figure 2).



356

357

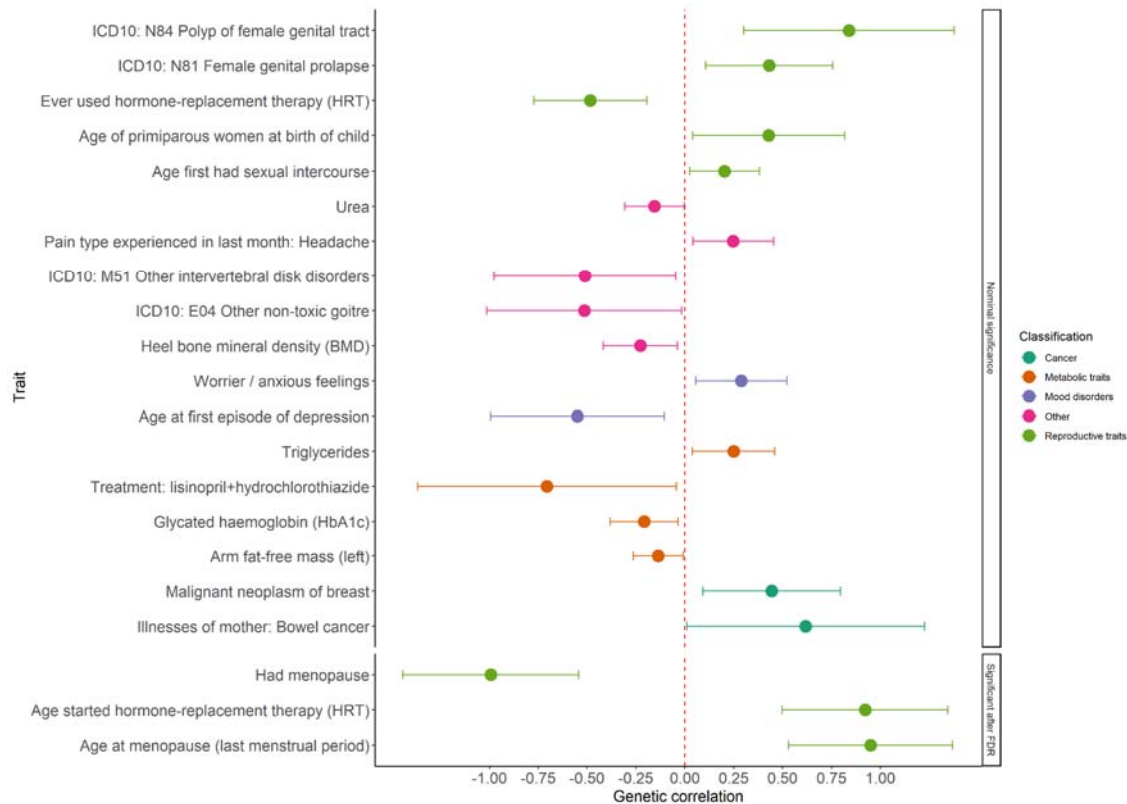
358 **Supplementary Figure 2. Gene set and tissue set enrichment results. A) Gene set**
359 **enrichment analyses with MAGMA.** Green bars and the red dashed line indicate
360 Bonferroni threshold, set to $p=0.05/15485=3.22 \times 10^{-6}$. **B) Tissue enrichment analyses**
361 **with MAGMA.** Red dashed line indicates Bonferroni threshold, set to $p=0.05/54=0.0009$.

362

363 Genetic correlations of AMH levels with diseases and traits

364 From genetic correlation analysis assessing the relationship between AMH levels and 1,335
365 traits (summary statistics for those traits are found in <http://www.nealelab.is/uk-biobank/>),
366 three traits reflecting menopausal timing remained significant after FDR correction, all
367 displaying very high genetic correlations with AMH levels: 'Age at menopause' ($r_g=0.95$,
368 $se=0.21$, $p=8.85 \times 10^{-6}$), 'Had menopause' (referring to having already undergone
369 menopause) ($r_g=-0.99$, $se=0.22$, $p=1.56 \times 10^{-5}$) and 'Age started hormone-replacement
370 therapy (HRT)' (age for initiation of hormone replacement therapy) ($r_g=0.92$, $se=0.21$,
371 $p=2.05 \times 10^{-5}$). Additionally, we observed a range of nominally significant associations
372 ($p<0.05$) with traits spanning reproductive/hormonal traits (such as polyp of female genital
373 tract and female genital prolapse), metabolic traits such as inverse relationship with HbA1c

374 and positive correlations with neoplasms such as breast and bowel cancer. (Supplementary
375 Figure 3 and full results and details can be found in Supplementary Table 6).



376

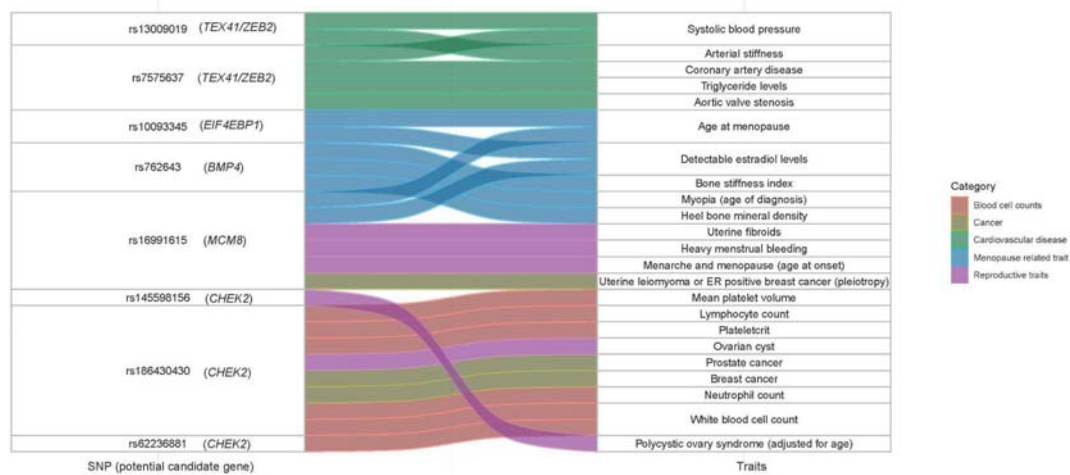
377 **Supplementary Figure 3. Genetic correlation results between AMH summary statistics**
378 **and publicly available summary statistics for other traits.** Out of 1,335 associations, the
379 lower section depicts significant associations after FDR correction ($p\text{-adj}<0.05$) and the
380 upper section depicts nominal significant correlations ($p<0.05$). Dots show the estimated
381 genetic correlation (r_g) and error bars indicate 95% confidence limits. Dotted red line
382 indicates no genetic correlation. ICD10: International Classifications of Diseases 10th
383 Revision.

384 Phenome-wide association study

385 To get more insights into phenotypic manifestations of the AMH-associated variants, we
386 performed a phenome-wide association study using the GWAS Catalog. We detected that
387 associated variants in the six loci identified had associations across several traits, spanning
388 from cardiovascular traits, age at menopause, estradiol levels, heel bone mineral density,
389 breast cancer, uterine fibroids, PCOS to blood cell counts (Supplementary Table 7 and
390 Supplementary Figure 3).

391 We performed a further look-up of the lead variants in the FinnGen study (data freeze 9, total
392 $n=377,277$), which combines samples collected from Finnish biobanks to digital health care
393 data. We detected that two of the lead variants had associations ($p<1 \times 10^{-5}$) with outcomes
394 in FinnGen, rs762643 (near *BMP4*) associated with cervical disc disorders and hernia, while
395 rs186430430 (near *CHEK2*) associated with a variety of outcomes including diseases of the

396 genitourinary tract, different neoplasms (e.g. leiomyoma of uterus, neoplasms of ovary), and
 397 breast-cancer-related outcomes (full results can be found in Supplementary Table 8).



398

399 **Supplementary Figure 4. River plot showing the association between lead variants (or**
 400 **variants in high LD with these) identified in our AMH GWAS (left column) and**
 401 **independent traits which show significant association results for respective variants**
 402 **in the GWAS Catalog (right column).**

403

404 DISCUSSION

405 Our GWAS meta-analysis of circulating AMH measurements in 9,668 pre-menopausal
 406 women, including 2,619 measurements from women at age 31 years old from the
 407 NFBC1966, identified three novel association signals near *EIF4EBP1*, *BMP4* and *CHEK2*.
 408 We also detected three previously identified signals near *TEX41*, *MCM8*, and *AMH*
 409 (Verdiesen *et al.*, 2022).

410 *CHEK2* was one of the most interesting associated loci. Of note, the lead variant near
 411 *CHEK2* (rs186430430, $p=9.69 \times 10^{-11}$, MAF= 0.002) is in complete linkage disequilibrium
 412 ($r^2=1$) with a frameshift variant, c.1100delC (rs555607708, found in NFBC GWAS only,
 413 $p=6.69 \times 10^{-05}$, OR=1.87, 95% CI= 1.37-2.55), as it is enriched in the Finnish population
 414 (MAF=0.008) compared to other non-Finnish and non-Estonian European populations
 415 (MAF=0.002) (Tyrmi *et al.*, 2022). Apart from being known as a moderate risk gene for
 416 breast cancer, different GWA studies highlighted loss of function alleles in *CHEK2* to be
 417 associated with firstly, PCOS (Tyrmi *et al.*, 2021) and secondly, with ovarian ageing
 418 (assessed as age at natural menopause) (Ruth *et al.*, 2021). Furthermore, recent research
 419 done with exome data has also identified an association between *CHEK2* truncating variants
 420 (excluding c.1110delC) and later age at menarche (Kentistou *et al.*, 2023).

421 *CHEK2* encodes a checkpoint kinase 2, which induces cell cycle arrest and apoptosis in
 422 response to DNA damage (Ahn *et al.*, 2004), also in oocytes with unrepaired DNA damage
 423 (Ruth *et al.*, 2021). In the work from Ruth *et al.*, 2021 Chek2^{-/-} female mice showed reduced
 424 follicular atresia around reproductive senescence, increased follicular response to

425 gonadotrophin stimulation, and also elevated AMH levels around reproductive senescence.
426 Our finding goes in line with this research and supports the hypothesis that loss of function in
427 *CHEK2* might result in decreased follicular atresia and higher AMH levels also in young
428 premenopausal women.

429 In our general meta-analysis, we also detected previously identified associations of a
430 missense variant in *MCM8* (Ruth *et al.*, 2019; Verdiesen *et al.*, 2022), a known DNA repair
431 gene (Park *et al.*, 2013). This missense variant has been associated previously with
432 premature ovarian failure, infertility, age at menopause (Day *et al.*, 2015) and cancer
433 (Michailidou *et al.*, 2017; Griffin and Trakselis, 2019; Lutzmann *et al.*, 2019). However, this
434 association was the weakest in NFBC66 alone. While this could support an age-specific
435 effect of the missense variant, as supported in Verdiesen *et al.* upon identifying a negative
436 effect direction for the cohort including adolescents versus older cohorts, this discrepancy
437 might be as well due to the variant's lower frequency in the Finnish population (MAF=0.02
438 compared to MAF=0.06 in other European populations) and a smaller sample size in the
439 NFBC1966 study. Further research including age-stratified groups with AMH levels available
440 and genetic data might inform potential age-specific effects of genetic signals.

441

442 Regarding the missense variant in the *AMH* gene also reported in Verdiesen *et al.*, the
443 association between the missense variant rs10417628 in the *AMH* gene and AMH levels
444 remains unclear in the current analysis. A case report from 2020 suggested that this variant
445 reduces AMH detection without affecting its bioactivity (Hoyos *et al.*, 2020). Previous GWAS
446 meta-analysis also provided inconclusive results due to inconsistent evidence from different
447 assay data. In the NFBC66 study using an automated assay (Elecsys® AMH Plus), the
448 missense variant showed a p-value of 0.15 (n=2,619), supporting the hypothesis of a
449 detection issue rather than an actual difference in AMH bioactivity. However, comparing
450 AMH values from different assays remains problematic, and recent evidence suggests lower
451 values in automated assays compared to Gen II and Ansh Labs assay (Moolhuijsen and
452 Visser, 2020). Further research is needed to validate the association between rs10417628
453 and AMH levels using different assays.

454 Colocalization analysis supports regulatory effects of GWAS variants associated with AMH
455 with genetic variants modifying expression levels of *TEX41*, *BMP4* and *EIF4EBP1* and thus
456 provides a refinement for the characterization of possible regulatory effects of the genetic
457 variants on different transcripts and genes.

458 Interestingly, we observed high posterior probability (PP4=0.99) of colocalization between
459 our GWAS signal and eQTL signal for the *TEX41* (Testis expressed 41) transcript
460 ENST00000414256 in the testis, prioritising this long non-coding RNA gene for the first time
461 in association with AMH levels based on colocalization analysis. In the same locus, *ZEB2*
462 shows biological plausibility as a candidate gene since it plays a role as inhibitor of signal
463 transduction in TGF- β and BMP signalling through interaction with ligand-activated SMAD
464 proteins (Postigo *et al.*, 2003; Conidi *et al.*, 2011).

465 Colocalization analysis also supported the nomination of *EIF4EBP1* in the novel locus
466 identified in chromosome 8. This gene encodes a translation repressor protein that
467 competitively binds to eukaryotic translation initiation factor 4E (EIF4E) (Gingras *et al.*, 1999;
468 Harris and Lawrence, 2003) and is a major substrate of mTOR and a key player in mTOR
469 signalling pathway. Studies show that both EIF4EBP1 and EIF4E are involved in cancer
470 development and progression where up-regulated EIF4E plays an oncogenic role in

471 carcinogenesis (Heikkinen *et al.*, 2013; Cha *et al.*, 2015). The same intergenic lead variant in
472 this locus has been associated with age at menopause, and other menopause-related traits
473 (ever had menopause and ever used hormone-replacement therapy), supporting a plausible
474 association with AMH levels.

475 We observed colocalization between our GWAS signal and variants modulating gene
476 expression of *BMP4*. *BMP4* has a known regulatory role on AMH expression through
477 activation of the SMAD proteins (Estienne *et al.*, 2015; Pierre *et al.*, 2016). Smad interacting
478 protein 1 (also known as *ZEB2*) is one of the plausible candidate genes in chromosome 2,
479 which would as well interact with SMAD proteins and participate in AMH regulation (Postigo
480 *et al.*, 2003).

481 Our GWAS of AMH measurements revealed significant enrichment in renal system vascular
482 morphogenesis, glomerulus vascular morphogenesis, and glomerulus morphogenesis gene
483 set analysis. This supports the close connection between urinary and reproductive system
484 development, originating from a common embryological origin, the intermediate mesoderm.
485 Additionally, in our gene set enrichment results, we observed a nominal significance
486 association with the gonadal mesoderm gene set ($p=4.2 \times 10^{-5}$), further supporting the
487 interconnectedness of the urinary and reproductive systems during development. We
488 propose that this observation may be linked to the identification of *BMP4* locus associated
489 with AMH levels. The *BMP4* signalling pathway is crucial during kidney development,
490 including ureteric bud outgrowth (Grinspon and Rey, 2014; Nishinakamura and Sakaguchi,
491 2014; Oxburgh *et al.*, 2014). This observation suggests that genetic variants affecting the
492 renal system's development and function could influence AMH levels in women later in life.
493 To gain further insight, studying AMH levels in younger cohorts and conducting sex-stratified
494 analyses may be valuable in understanding the association between urogenital development
495 and AMH levels in postnatal stages.

496 Tissue expression analysis also identified enrichment in the pituitary gland, although this did
497 not reach significance after multiple testing correction. This result goes in line with research
498 from recent years showing that AMH has versatile actions in different levels of the
499 hypothalamus-pituitary-gonadal axis (Silva and Giacobini, 2021), further supporting the
500 developmental alterations of neuroendocrine circuits regulating fertility.

501 We identified three significant genetic correlations which align with the observations in
502 Verdiesen *et al.*, indicating a strong positive genetic correlation between AMH and traits
503 reflecting age at menopause, suggesting shared underlying genetic risk factors and
504 supporting AMH as a proxy for ovarian reserve. We also found interesting nominal significant
505 associations, pointing towards a shared genetic background between AMH levels and breast
506 cancer, consistent with epidemiological studies (Ge *et al.*, 2018). We hypothesise both
507 *MCM8* and *CHEK2* loci may play a role in this relationship and Mendelian randomisation
508 studies with independent samples could further investigate this association's direction.

509 Upon querying the AMH GWAS lead variants and variants in high linkage disequilibrium with
510 those in other traits, we found shared significant signals amongst known epidemiological
511 associations (Homburg and Crawford, 2014; De Kat *et al.*, 2017; Ge *et al.*, 2018;
512 Moolhuijsen and Visser, 2020) with AMH such as PCOS, breast cancer, postmenopausal
513 status (age at menopause, estradiol levels, heel bone mineral density) and cardiovascular
514 disease. Additionally, we observed novel associations with other traits in European ancestry
515 studies, including uterine fibroids (coming from signal in *MCM8*) and blood cell counts
516 (coming from signals near *CHEK2*, which has been also associated recently with clonal
517 hematopoiesis (Kar *et al.*, 2022). Further epidemiological studies assessing the potential

518 causal relationships between these traits and AMH levels are warranted to better understand
519 the mechanisms underlying these associations.

520 Our study has several strengths, including a large sample size and inclusion of a founder
521 population, which allowed us to identify novel rare variants associated with AMH levels.
522 Furthermore, a younger cohort and equal age of measurement from 2,619 women increases
523 the variability of AMH levels and boosts the power to detect associations. We also performed
524 colocalization analysis for the first time, which provides a refinement for the characterization
525 of possible regulatory effects of the genetic variants on different transcripts and genes and
526 detect significant gene set enrichment. Our study has some limitations to consider. First, we
527 only included women of European ancestry, and our findings may not be generalizable to
528 other populations. Second, as with other reproductive phenotypes, the lack of sufficiently
529 sized datasets from relevant tissue in commonly used gene expression databases hinders a
530 more reliable assessment of the mechanisms underlying regulatory effects in reproductive
531 tissues.

532 In conclusion, our study expands our understanding of the genetic determinants of serum
533 AMH levels in pre-menopausal women by identifying new loci associated with serum AMH
534 concentration. Our results highlight the increased power of founder populations and larger
535 sample sizes to boost the discovery of novel trait-associated variants underlying variation in
536 AMH levels and to explore plausible genetic regulatory effects of the variants identified.
537 Further studies are needed to validate our findings and to explore the biological mechanisms
538 underlying the identified associations.

539

540 **Acknowledgments**

541 We want to thank the participants and investigators of the FinnGen study and the NFBC1966
542 study. Part of the computations were performed in the High-Performance Computing Center
543 of University of Tartu.

544

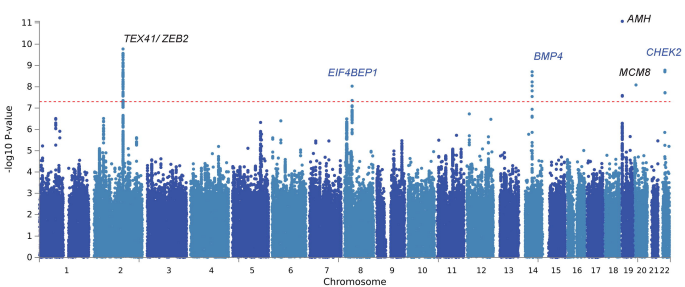
545 **Bibliography**

- 546 Ahn J, Urist M, Prives C. The Chk2 protein kinase. *DNA Repair (Amst)* [Internet]
547 2004;**3**:1039–1047. DNA Repair (Amst).
- 548 Bulik-Sullivan B, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, Daly MJ, Price AL,
549 Neale BM, Corvin A, *et al.* LD score regression distinguishes confounding from
550 polygenicity in genome-wide association studies. *Nat Genet* 2015;**47**:291–295. Nature
551 Publishing Group.
- 552 Cha YL, Li PD, Yuan LJ, Zhang MY, Zhang YJ, Rao HL, Zhang HZ, Zheng XFS, Wang HY.
553 EIF4EBP1 overexpression is associated with poor survival and disease progression in
554 patients with hepatocellular carcinoma. *PLoS One* [Internet] 2015;**10**:. PLoS One.
- 555 Conidi A, Cazzola S, Beets K, Coddens K, Collart C, Cornelis F, Cox L, Joke D, Dobрева
556 MP, Dries R, *et al.* Few Smad proteins and many Smad-interacting proteins yield
557 multiple functions and action modes in TGF β /BMP signaling in vivo. *Cytokine Growth*
558 *Factor Rev* [Internet] 2011;**22**:287–300. Cytokine Growth Factor Rev.
- 559 Day FR, Ruth KS, Thompson DJ, Lunetta KL, Pervjakova N, Chasman DI, Stolk L, Finucane
560 HK, Sulem P, Bulik-Sullivan B, *et al.* Large-scale genomic analyses link reproductive
561 aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated
562 DNA repair. *Nat Genet* [Internet] 2015;**47**:1294–1303. Nat Genet.
- 563 Estienne A, Pierre A, Clemente N Di, Picard JY, Jarrier P, Mansanet C, Monniaux D, Fabre

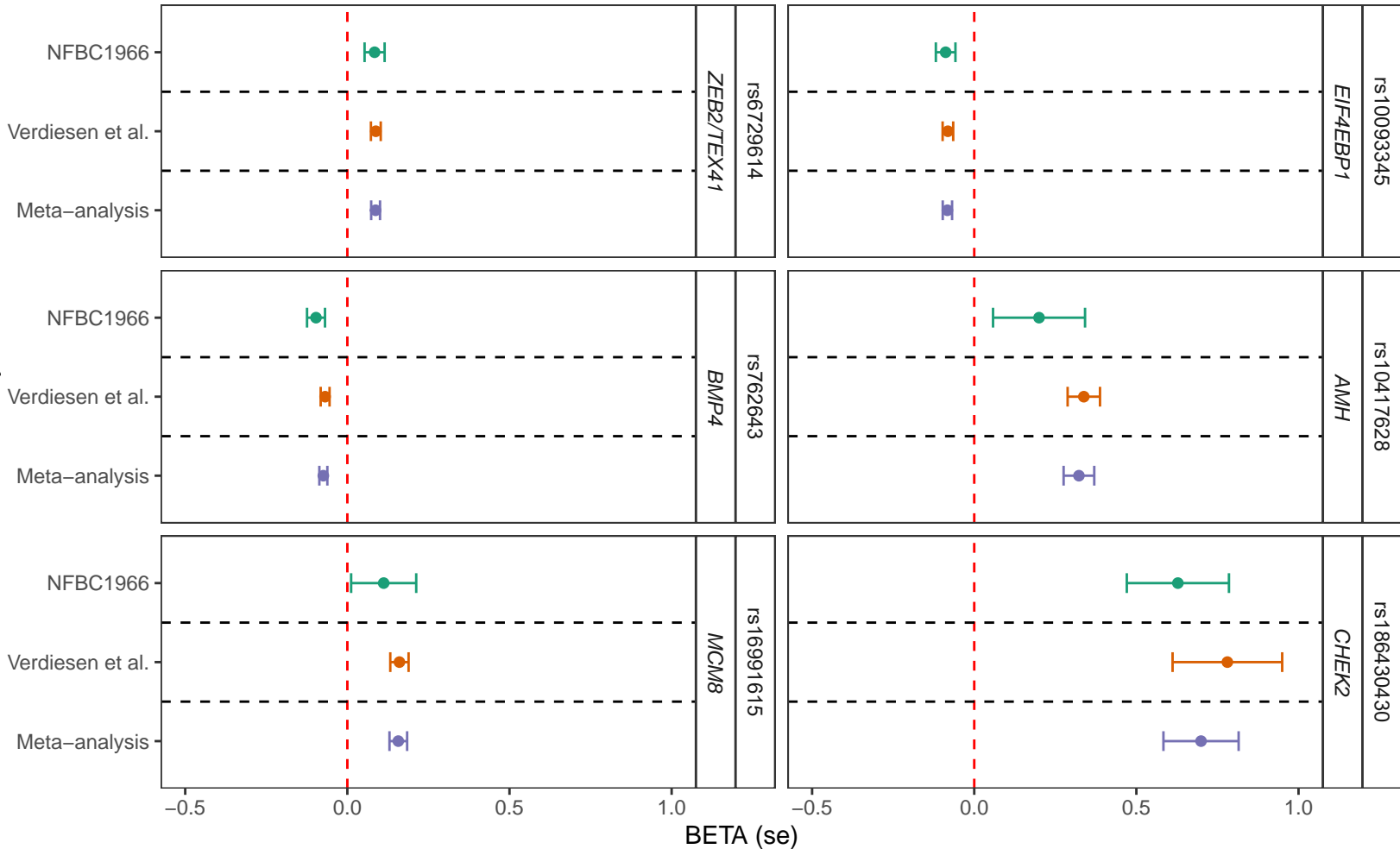
- 564 S. Anti-Müllerian hormone regulation by the bone morphogenetic proteins in the sheep
565 ovary: deciphering a direct regulatory pathway. *Endocrinology* [Internet] 2015;**156**:301–
566 313. *Endocrinology*.
- 567 Finkelstein JS, Lee H, Karlamangla A, Nee RM, Slus PM, Burnett-Bowie SAM, Darakananda
568 K, Donaho PK, Harlo SD, Prizan SH, *et al.* Antimüllerian Hormone and Impending
569 Menopause in Late Reproductive Age: The Study of Women’s Health Across the
570 Nation. *J Clin Endocrinol Metab* [Internet] 2020;**105**:E1862–E1871. *J Clin Endocrinol*
571 *Metab*.
- 572 Foley CN, Staley JR, Breen PG, Sun BB, Kirk PDW, Burgess S, Howson JMM. A fast and
573 efficient colocalization algorithm for identifying shared genetic risk factors across
574 multiple traits. *Nat Commun* [Internet] 2021;**12**:. *Nat Commun*.
- 575 Ge W, Clendenen T V., Afanasyeva Y, Koenig KL, Agnoli C, Brinton LA, Dorgan JF,
576 Eliassen AH, Falk RT, Hallmans G, *et al.* Circulating anti-Müllerian hormone and breast
577 cancer risk: A study in ten prospective cohorts. *Int J cancer* [Internet] 2018;**142**:2215–
578 2226. *Int J Cancer*.
- 579 Gingras AC, Raught B, Sonenberg N. eIF4 initiation factors: effectors of mRNA recruitment
580 to ribosomes and regulators of translation. *Annu Rev Biochem* 1999;**68**:913–963.
581 United States.
- 582 Griffin WC, Trakselis MA. The MCM8/9 complex: A recent recruit to the roster of helicases
583 involved in genome maintenance. *DNA Repair (Amst)* [Internet] 2019;**76**:1–10. *DNA*
584 *Repair (Amst)*.
- 585 Grinspon RP, Rey RA. When hormone defects cannot explain it: malformative disorders of
586 sex development. *Birth Defects Res C Embryo Today* [Internet] 2014;**102**:359–373.
587 *Birth Defects Res C Embryo Today*.
- 588 Harris TE, Lawrence JCJ. TOR signaling. *Sci STKE* 2003;**2003**:re15. United States.
- 589 Heikkinen T, Korpela T, Fagerholm R, Khan S, Aittomäki K, Heikkilä P, Blomqvist C, Carpén
590 O, Nevanlinna H. Eukaryotic translation initiation factor 4E (eIF4E) expression is
591 associated with breast cancer tumor phenotype and predicts survival after anthracycline
592 chemotherapy treatment. *Breast Cancer Res Treat* [Internet] 2013;**141**:79–88. *Breast*
593 *Cancer Res Treat*.
- 594 Homburg R, Crawford G. The role of AMH in anovulation associated with PCOS: a
595 hypothesis. *Hum Reprod* [Internet] 2014;**29**:1117–1121. *Hum Reprod*.
- 596 Hoyos LR, Visser JA, McLuskey A, Chazenbalk GD, Grogan TR, Dumesic DA. Loss of anti-
597 Müllerian hormone (AMH) immunoactivity due to a homozygous AMH gene variant
598 rs10417628 in a woman with classical polycystic ovary syndrome (PCOS). *Hum Reprod*
599 [Internet] 2020;**35**:2294–2302. *Hum Reprod*.
- 600 Kar SP, Quiros PM, Gu M, Jiang T, Mitchell J, Langdon R, Iyer V, Barcena C, Vijayabaskar
601 MS, Fabre MA, *et al.* Genome-wide analyses of 200,453 individuals yield new insights
602 into the causes and consequences of clonal hematopoiesis. *Nat Genet* [Internet]
603 2022;**54**:1155–1166. *Nat Genet*.
- 604 Kat AC De, Monique Verschuren W, Eijkemans MJC, Broekmans FJM, Schouw YT Van Der.
605 Anti-Müllerian Hormone Trajectories Are Associated With Cardiovascular Disease in
606 Women: Results From the Doetinchem Cohort Study. *Circulation* [Internet]
607 2017;**135**:556–565. *Circulation*.
- 608 Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler and D. The
609 human genome browser at UCSC. *Genome Res* [Internet] 2002;**12**:996–1006. *Genome*
610 *Res*.
- 611 Kentistou KA, Kaisinger LR, Stankovic S, Vaudel M, Oliveira EM de, Messina A, Walters RG,
612 Liu X, Busch AS, Helgason H, *et al.* Understanding the genetic complexity of puberty
613 timing across the allele frequency spectrum. *medRxiv* [Internet] 2023; Cold Spring
614 Harbor Laboratory Press Available from:
615 <https://www.medrxiv.org/content/early/2023/06/20/2023.06.14.23291322>.
- 616 Kerimov N, Hayhurst JD, Peikova K, Manning JR, Walter P, Kolberg L, Samoviča M,
617 Sakthivel MP, Kuzmin I, Trevanion SJ, *et al.* A compendium of uniformly processed
618 human gene expression and splicing quantitative trait loci. *Nat Genet* [Internet]

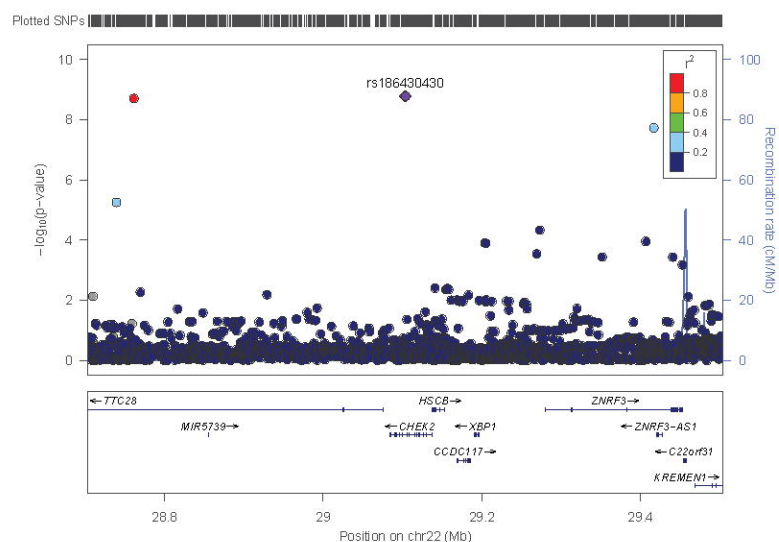
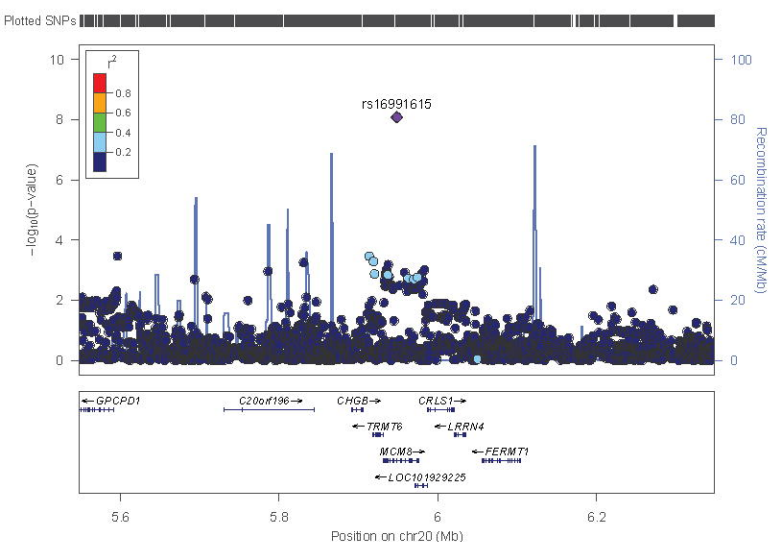
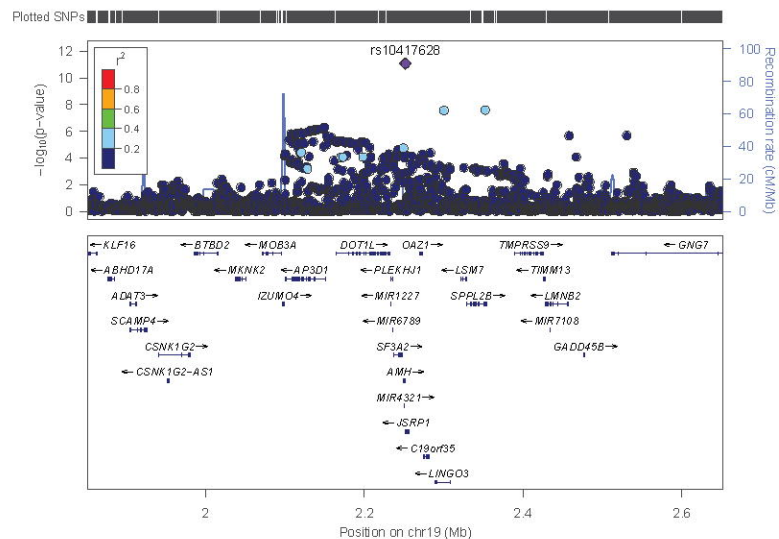
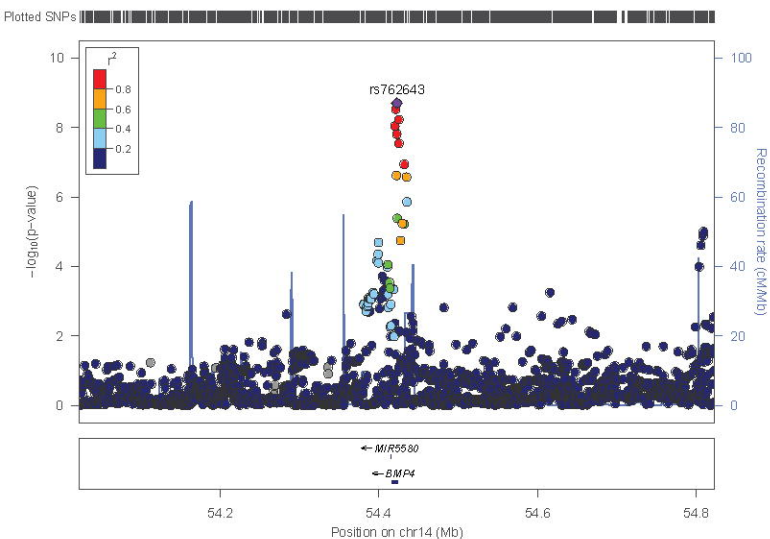
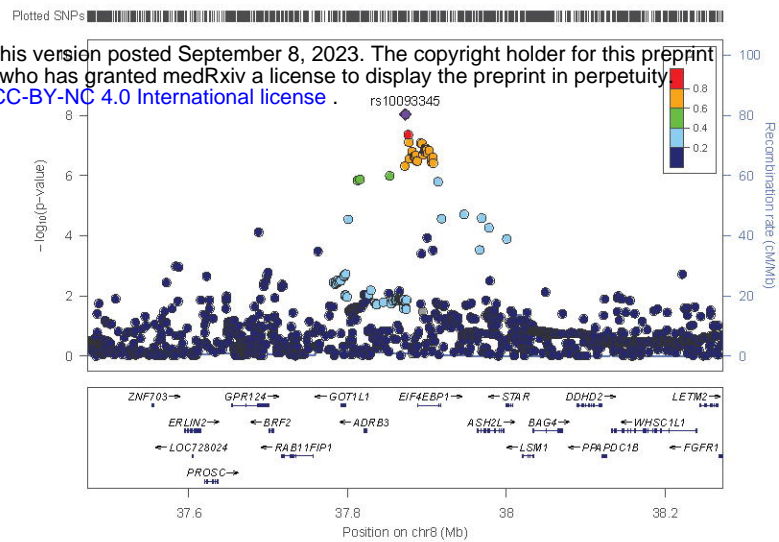
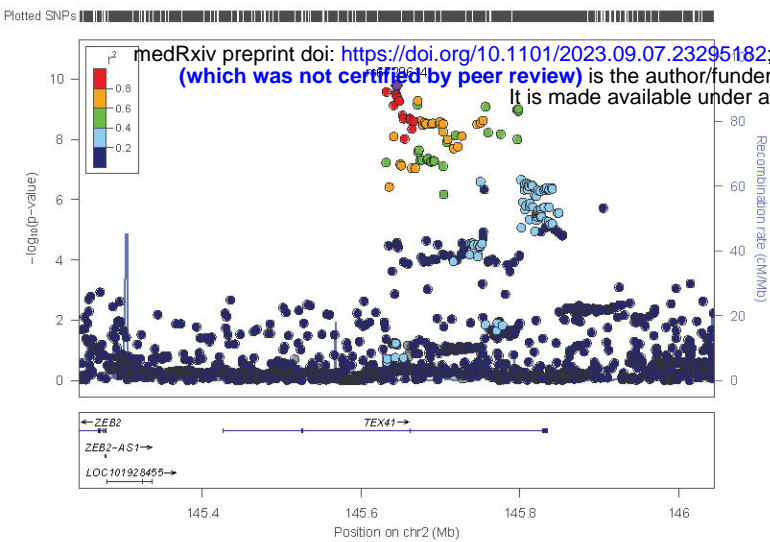
- 619 2021;**53**:1290–1299.
- 620 Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner KM, Reeve MP, Laivuori
621 H, Aavikko M, Kaunisto MA, *et al.* FinnGen provides genetic insights from a well-
622 phenotyped isolated population. *Nature* [Internet] 2023;**613**:508–518. *Nature*.
- 623 Lutzmann M, Bernex F, Costa de Jesus C da, Hodroj D, Marty C, Plo I, Vainchenker W,
624 Tosolini M, Forichon L, Bret C, *et al.* MCM8- and MCM9 Deficiencies Cause Lifelong
625 Increased Hematopoietic DNA Damage Driving p53-Dependent Myeloid Tumors. *Cell*
626 *Rep* [Internet] 2019;**28**:2851-2865.e4. *Cell Rep*.
- 627 Mägi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. *BMC*
628 *Bioinformatics* 2010;**11**:288.
- 629 Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, Lemaçon A, Soucy P, Glubb
630 D, Rostamianfar A, *et al.* Association analysis identifies 65 new breast cancer risk loci.
631 *Nature* [Internet] 2017;**551**:92–94. *Nature*.
- 632 Moolhuijsen LME, Visser JA. Anti-Müllerian Hormone and Ovarian Reserve: Update on
633 Assessing Ovarian Function. *J Clin Endocrinol Metab* [Internet] 2020;**105**:. *J Clin*
634 *Endocrinol Metab*.
- 635 Nishinakamura R, Sakaguchi M. BMP signaling and its modifiers in kidney development.
636 *Pediatr Nephrol* [Internet] 2014;**29**:681–686. *Pediatr Nephrol*.
- 637 Nordström T, Miettunen J, Auvinen J, Ala-Mursula L, Keinänen-Kiukaanniemi S, Veijola J,
638 Järvelin M-R, Sebert S, Männikkö M. Cohort Profile: 46 years of follow-up of the
639 Northern Finland Birth Cohort 1966 (NFBC1966). *Int J Epidemiol* 2022;**50**:1786-1787].
640 England.
- 641 Oxburgh L, Brown AC, Muthukrishnan SD, Fetting JL. Bone morphogenetic protein signaling
642 in nephron progenitor cells. *Pediatr Nephrol* [Internet] 2014;**29**:531–536. *Pediatr*
643 *Nephrol*.
- 644 Park J, Long DT, Lee KY, Abbas T, Shibata E, Negishi M, Luo Y, Schimenti JC, Gambus A,
645 Walter JC, *et al.* The MCM8-MCM9 complex promotes RAD51 recruitment at DNA
646 damage sites to facilitate homologous recombination. *Mol Cell Biol* [Internet]
647 2013;**33**:1632–1644. *Mol Cell Biol*.
- 648 Pierre A, Estienne A, Racine C, Picard JY, Fanchin R, Lahoz B, Alabart JL, Folch J, Jarrier
649 P, Fabre S, *et al.* The Bone Morphogenetic Protein 15 Up-Regulates the Anti-Müllerian
650 Hormone Receptor Expression in Granulosa Cells. *J Clin Endocrinol Metab* [Internet]
651 2016;**101**:2602–2611. *J Clin Endocrinol Metab*.
- 652 Piltonen TT, Komsa E, Morin-Papunen LC, Korhonen E, Franks S, Järvelin M-R, Arffman RK,
653 Ollila M-M. AMH as part of the diagnostic PCOS workup in large epidemiological
654 studies. *Eur J Endocrinol* 2023;**188**:547–554. England.
- 655 Postigo AA, Depp JL, Taylor JJ, Kroll KL. Regulation of Smad signaling through a differential
656 recruitment of coactivators and corepressors by ZEB proteins. *EMBO J* [Internet]
657 2003;**22**:2453–2462. *EMBO J*.
- 658 Ruth KS, Day FR, Hussain J, Martínez-Marchal A, Aiken CE, Azad A, Thompson DJ,
659 Knoblochova L, Abe H, Tarry-Adkins JL, *et al.* Genetic insights into biological
660 mechanisms governing human ovarian ageing. *Nature* [Internet] 2021;**596**:393–397.
661 *Nature*.
- 662 Ruth KS, Soares ALG, Borges MC, Eliassen AH, Hankinson SE, Jones ME, Kraft P, Nichols
663 HB, Sandler DP, Schoemaker MJ, *et al.* Genome-wide association study of anti-
664 Müllerian hormone levels in pre-menopausal women of late reproductive age and
665 relationship with genetic determinants of reproductive lifespan. *Hum Mol Genet*
666 [Internet] 2019;**28**:1392–1401. *Hum Mol Genet*.
- 667 Schuh-Huerta SM, Johnson NA, Rosen MP, Sternfeld B, Cedars MI, Reijo Pera RA. Genetic
668 markers of ovarian follicle number and menopause in women of multiple ethnicities.
669 *Hum Genet* [Internet] 2012;**131**:1709–1724. *Hum Genet*.
- 670 Silva MSB, Giacobini P. New insights into anti-Müllerian hormone role in the hypothalamic-
671 pituitary-gonadal axis and neuroendocrine development. *Cell Mol Life Sci* [Internet]
672 2021;**78**:. *Cell Mol Life Sci*.
- 673 Tyrmi JS, Arffman RK, Pujol-Gualdo N, Kurra V, Morin-Papunen L, Sliz E, FinnGen, Team

674 EBR, Piltonen TT, Laisk T, *et al.* Leveraging Northern European population history;
675 novel low frequency variants for polycystic ovary syndrome. *medRxiv* [Internet]
676 2021;2021.05.20.21257510 Available from:
677 <http://medrxiv.org/content/early/2021/05/24/2021.05.20.21257510.abstract>.
678 Tyrmi JS, Arffman RK, Pujol-Gualdo N, Kurra V, Morin-Papunen L, Sliz E, Piltonen TT, Laisk
679 T, Kettunen J, Laivuori H. Leveraging Northern European population history: novel low-
680 frequency variants for polycystic ovary syndrome. *Hum Reprod* [Internet] 2022;**37**:352–
681 365. Hum Reprod.
682 Verdiesen RMG, Schouw YT Van Der, Gils CH Van, Verschuren WMM, Broekmans FJM,
683 Borges MC, Gonçalves Soares AL, Lawlor DA, Eliassen AH, Kraft P, *et al.* Genome-
684 wide association study meta-analysis identifies three novel loci for circulating anti-
685 Müllerian hormone levels in women. *Hum Reprod* [Internet] 2022;**37**:1069–1082.
686 Oxford University Press.
687 Watanabe K, Taskesen E, Bochoven A van, Posthuma D. Functional mapping and
688 annotation of genetic associations with FUMA. *Nat Commun* 2017;**8**:1826.
689 Weenen C, Laven JSE, Bergh ARM von, Cranfield M, Groome NP, Visser JA, Kramer P,
690 Fauser BCJM, Themmen APN. Anti-Müllerian hormone expression pattern in the
691 human ovary: potential implications for initial and cyclic follicle recruitment. *Mol Hum*
692 *Reprod* [Internet] 2004;**10**:77–83. Mol Hum Reprod.
693



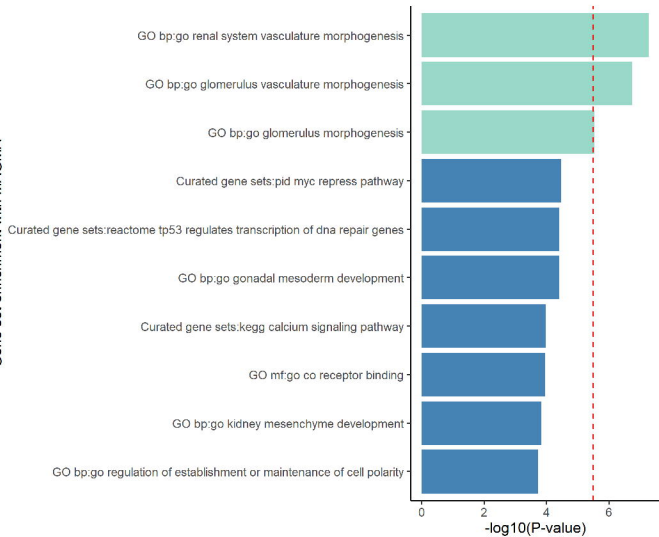
Study





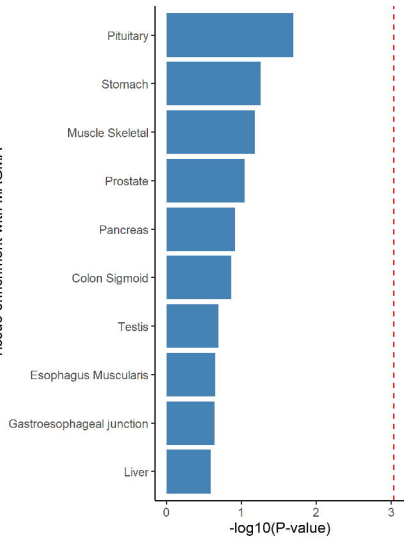
A

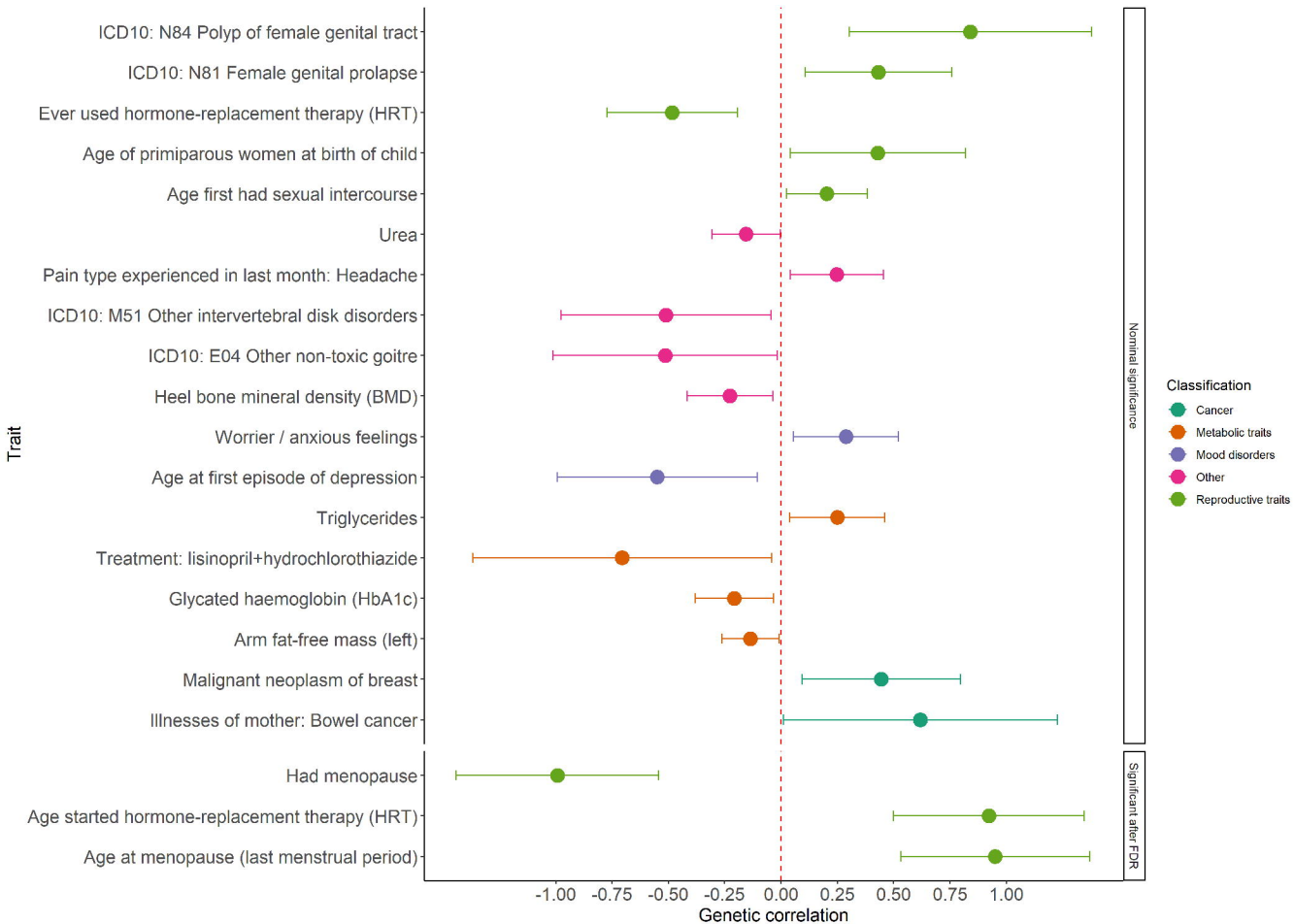
Gene set enrichment with MAGMA

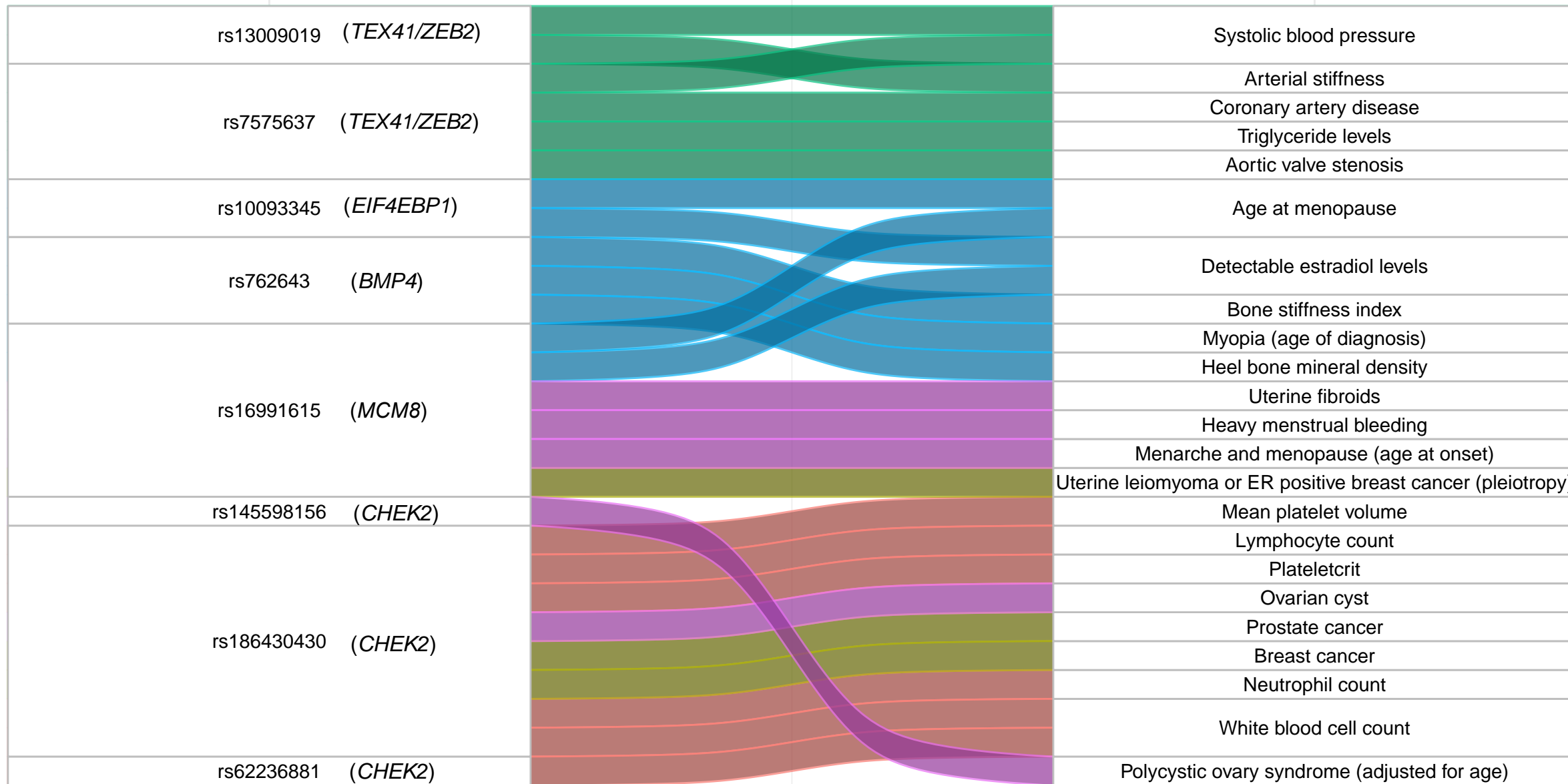


B

Tissue enrichment with MAGMA







Category

- Blood cell counts
- Cancer
- Cardiovascular disease
- Menopause related trait
- Reproductive traits

SNP (potential candidate gene)

Traits