

# 1 Genome-Wide Association Study of 2 Estradiol Levels, and the Causal Effect of 3 Estradiol on Bone Mineral Density

4 Daniel Schmitz<sup>1</sup>, Weronica E. Ek<sup>1</sup>, Elin Berggren<sup>1</sup>, Julia Höglund<sup>1</sup>, Torgny Karlsson<sup>1</sup>, Åsa  
5 Johansson<sup>1</sup>

6 <sup>1</sup>Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala  
7 University, Uppsala, Sweden.

8 Short title: GWAS of Estradiol Levels and Causal Effect on BMD

9 Keywords: GWAS, Estrogen, Bone mineral density, Mendelian Randomization

10 Corresponding Authors: Daniel Schmitz <[daniel.schmitz@igp.uu.se](mailto:daniel.schmitz@igp.uu.se)> or Åsa Johansson  
11 <[asa.johansson@igp.uu.se](mailto:asa.johansson@igp.uu.se)>

12 Reprint Requests: No

13 Grants/Fellowships supporting paper: The research was funded by the Swedish Research  
14 Council (Å.J), the cancer foundation (Å.J), the brain foundation (Å.J), the Å Wiberg (W.E.E),  
15 M Borgström (W.E.E), Hedströms, K och O F (W.E.E), A and M Rudbergs (W.E.E)  
16 foundations.

17 Disclosure summary: Nothing to declare

## 18 Abstract

19 **Context.** Estradiol is the primary female sex hormone and plays an important role for skeletal  
20 health in both sexes. Several enzymes are involved in estradiol metabolism but few genome-  
21 wide association studies (GWAS) have been performed to characterize the genetic  
22 contribution to variation in estrogen levels.

23 **Objective.** Identify genetic loci affecting estradiol levels and estimate causal effect of  
24 estradiol on bone mineral density (BMD).

25 **Design.** We performed GWAS for estradiol in males (N = 147,690) and females  
26 (N = 163,985) from UK Biobank (UKB). Estradiol was analyzed as a binary phenotype;  
27 above/below detection limit (175 pmol/L). We further estimated the causal effect of estradiol  
28 on BMD using Mendelian randomization.

29 **Results.** We identified 14 independent loci associated ( $P < 5 \times 10^{-8}$ ) with estradiol levels in  
30 males, of which one (*CYP3A7*) was genome-wide and seven nominally ( $P < 0.05$ ) significant  
31 in females. In addition, one female specific locus was identified. Most loci contain  
32 functionally relevant genes that have not been discussed in relation to estradiol levels in  
33 previous GWAS. For example, *SRD5A2*, which encodes a steroid 5-alpha reductase that is  
34 involved in processing androgens, and *UGT3A1* and *UGT2B7* which encode enzymes likely  
35 to be involved in estradiol elimination. The allele that tags the O blood group, at the *ABO*  
36 locus, was associated with higher estradiol levels. We identified a causal effect of high  
37 estradiol levels on increased BMD in both males ( $P = 1.58 \times 10^{-11}$ ) and females ( $P = 7.48 \times 10^{-6}$ ).

38 **Conclusion.** Our findings further support the importance of the body's own estrogen to  
39 maintain skeletal health in males and in females.

40

## 41 Introduction

42 Estrogen, the primary female sex hormone, is responsible for the development of the female  
43 reproductive system and secondary sex characteristics. Estrogen acts primarily as a growth  
44 hormone for the reproductive organs and regulates the menstrual cycle in women<sup>1</sup>.

45 Additionally, it plays a critical role in male sexual function<sup>2</sup>. There are three major forms of  
46 estrogen: estrone, estradiol and estriol. Estradiol is the major and most potent form<sup>3</sup>. In  
47 females, high estrogen levels have been associated with deep vein thrombosis<sup>4</sup> and increased  
48 risk for certain cancers, including breast cancer<sup>5</sup> in women. In males, higher levels have been  
49 suggested to be associated with reduced risk of type 2 diabetes (T2D)<sup>6</sup>. Estrogen is also an  
50 important regulator of bone metabolism in both females and males<sup>7</sup>. When estrogen levels  
51 drop during menopause, the cells that synthesize bone (osteoblasts) are unable to effectively  
52 produce bone mass<sup>8</sup>. The loss of ovarian estrogens after menopause has been associated with  
53 reduced bone mineral density (BMD) and higher risk of osteoporosis<sup>9,10</sup>.

54 Previously published genome-wide association studies (GWAS) for estrogen have been  
55 performed in cohorts of up to 11,000 individuals, mostly of European descent, both male and  
56 female and stratified by sex<sup>11-15</sup>. In addition, a GWAS was recently performed for sex  
57 hormone levels in UK biobank (UKB), a study which mainly focused on testosterone<sup>16</sup>. They  
58 identified strong sex-specific genetic effects on serum testosterone levels, but did not  
59 consider the results from estradiol in women, due to the strong link between estradiol levels  
60 and menopausal status/age of menopause. Previous studies have also addressed the causal  
61 effect of estradiol and testosterone levels in relation to disease risk using Mendelian  
62 randomization (MR)<sup>15-17</sup>. MR uses genetic data to evaluate a potential causal relationship  
63 between an exposure, for example hormones, and disease risk. The underlying assumption in  
64 MR is that if a hormone is involved in a disease process, then the genetic factors influencing

65 the levels of the hormone should only affect disease risk through their effects on hormone  
66 levels, and not through any alternative pathways. MR is conceptually similar to randomized  
67 controlled trials because genotypes are assigned randomly already at conception. MR can  
68 therefore be used to overcome some of the limitations of observational studies, such as bias  
69 due to unmeasured or unknown confounding. Likewise, reverse causation is also avoided due  
70 to genotypes being assigned before the disease manifests. For testosterone, many independent  
71 genetic variants were identified and included in the previous MR. They concluded that  
72 testosterone increases the risk for polycystic ovary syndrome (PCOS) and type 2 diabetes in  
73 women but decreases the risk for type 2 diabetes in men<sup>16</sup>. Even though the number of  
74 GWAS significant findings for estradiol has been quite limited in previous studies<sup>11–15</sup>, those  
75 variants have been used in MR studies to address the causal effect of estradiol on disease risk.  
76 Two recent MR studies, using five and two SNPs respectively, identified a beneficial effect  
77 of estradiol on bone mineral density (BMD) in males<sup>15,17</sup>, but no study has yet been  
78 performed in females.

79 By using the largest cohort with estradiol levels measured; almost half a million participants,  
80 we have been able to further highlight novel genes, important for regulating estradiol levels.  
81 Most assays for measuring estradiol levels are not very sensitive, and the methods used in  
82 UKB had a detection limit of 175 pmol/L. Normal ranges of serum estradiol levels are 36.7-  
83 183.6 pmol/L in men, 73.4 – 2753.5 pmol/L in premenopausal women and 0 – 73.4 pmol/L in  
84 postmenopausal women<sup>18</sup>. This results in a large fraction of the samples, especially men and  
85 postmenopausal women, having levels below detection limit. Similar to a previous GWAS  
86 for estradiol<sup>16</sup>, we have therefore analyzed estradiol as a binary phenotype (above or below  
87 detection limit), instead of removing all individuals below detection limit. We have stratified  
88 the cohort by sex and menopausal status, to enable comparing genetic effects across strata.  
89 Thanks to the larger sample size in UKB compared to previous studies, we were able to

90 identify enough genetic variants associated with estradiol levels to permit causal effect  
91 estimation of estradiol on bone mineral density, through MR analysis, also in females.

## 92 Material and Methods

### 93 Sample

94 UKB is a cross sectional cohort of about 500,000 participants, born between 1939 and 1970,  
95 that were recruited 2006—2010. Detailed information was collected on lifestyle and  
96 anthropometric traits, and on female specific factors, that are likely to effect hormone levels,  
97 including use of birth control pills, hormone replacement treatments, pregnancies and  
98 menopause. Participants also provided blood samples that were used to genotype  
99 approximately 800,000 single nucleotide polymorphisms (SNPs) across the genome and to  
100 measure levels of estradiol. The application to use data from UKB has been approved (under  
101 project number 41143). UKB has an ethics permit from the National Research Ethics  
102 Committee (REC reference 11 / NW / 0382). Phenotype data for this project was extracted  
103 from the UKB database April 2019. All analyses performed in this study are based on the  
104 samples and information collected at the participants initial visit to the assessment center.  
105 Ethical approval for the analyses performed in this study has also been approved by the  
106 Swedish Ethical Review Authority (Dnr: 2020-04415)

107

### 108 Genotype data

109 A total of 438,417 participants had been genotyped using the UKB Axiom array, and another  
110 49,994 participants had been genotyped using the similar (95% common marker content) UK  
111 BiLEVE array. SNPs had been imputed using UK10K and 1000 genomes phase 3, as  
112 reference panels<sup>19,20</sup>. For this project we used imputed data from the third release which

113 contained a total of 93,093,070 SNPs. We only analyzed SNPs with minor allele frequency  
114 (MAF)  $> 0.01$ , and removed SNPs deviating from Hardy-Weinberg (P-value  $< 1 \times 10^{-20}$ ) and  
115 markers with more than 5% missing genotype data, and with imputation quality  $< 0.3$ . After  
116 quality control (QC), a total of 7,651,231 autosomal SNPs, and 220,486 SNPs on the X-  
117 chromosome remained. We included Caucasian participants clustering with regards to their  
118 genetic principal components. We also excluded first- and second-degree relatives (genetic  
119 relationship  $> 0.044$ ), using kinship data. Samples with sex discordance, high  
120 heterozygosity/missingness, and with more than 5% missing genotypes were also excluded.  
121 After QC and exclusion, 361,975 unrelated Caucasian participants remained (167,168 males  
122 and 194,807 females).

### 123 Estradiol measurements

124 Estradiol had been measured by two step competitive analysis on a Beckman Coulter Unicel  
125 Dxl 800. We only included measurements from the blood samples that were collected during  
126 the participants initial visit to the assessment center. Among all participants, only 76,668 had  
127 estradiol measurements above detection limit (175 pmol/L). As the majority of the  
128 participants had measurements below detection limit, estradiol levels were analyzed as a  
129 binary phenotype (above/below detection limit).

### 130 GWAS and sensitivity analyses

131 GWAS was performed in males and females separately, using logistic regression with an  
132 additive genetic model implemented in PLINK version 2. The following covariates were  
133 included in both male and female analyses: age, body mass index (BMI), the first ten genetic  
134 principal components (to adjust for population structures and ethnic origins), and a binary  
135 indicator variable for UKB Axiom versus UK BiLEVE genotyping array, to adjust for any  
136 array differences. For females, hormone-replacement therapy (HRT) (current, ever or never),

137 oral contraceptive use (current, ever or never), number of live births, menopausal status and  
138 whether they have undergone hysterectomy, were also included. Women who were unsure  
139 about menopausal status were excluded. For this study, we applied a P-value threshold of  
140  $5 \times 10^{-8}$ , which is commonly used in GWAS when using a MAF cutoff of 1%<sup>21</sup>, and  
141 corresponds to a correction of 1 million independent tests (SNPs). Manhattan plots were  
142 generated using the R package hudson. We applied the clump-function in PLINK 1.90 to  
143 define start and end positions for each locus with the main parameters that determines the  
144 clumping being set to:  $R^2=0.1$ ,  $p_1=5 \times 10^{-8}$  and  $p_2=1 \times 10^{-4}$ . We subsequently performed  
145 conditional analyses for each locus, conditioning on the SNP with the lowest P-value from  
146 each locus. If there was any significant conditional SNP, this was repeated by also  
147 conditioning on the conditional SNP from the previous analyses, until no additional  
148 significant associations remained.

149 Several sensitivity analyses were performed for the GWAS. First, we stratified females into  
150 being pre- and postmenopausal. Secondly, we excluded all participants with a cancer  
151 diagnosis prior to assessment (when blood samples were taken). Finally, we also analyzed  
152 males and females with testosterone and sex hormone-binding globulin (SHBG) levels  
153 included as covariates, in addition to the other covariates and with age at menopause as a  
154 covariate in the female postmenopausal stratum.

#### 155 [Annotation of GWAS lead SNPs](#)

156 To identify likely target genes for the associated variants, we first identified genes containing  
157 the lead SNPs, or the closest gene(s) for each individual lead SNP. We also used HaploReg  
158 v4.1<sup>22</sup> to identify potential functional effects by the individual lead SNP, or any variant in  
159 linkage disequilibrium (LD,  $R^2 > 0.8$ ) with the lead SNP. We then checked for overlap  
160 between the lead SNPs and expression quantitative trait loci (eQTL) using data from the

161 Genotype-Tissue Expression (GTEx) project<sup>23</sup>. We considered our lead SNP to overlap with  
162 an eQTL if our lead SNP was in LD ( $R^2>0.8$ ) with a lead SNP for an eQTL. To identify  
163 overlap with other phenotypes from previously published GWAS we downloaded summary  
164 statistics from the GWAS catalog (Version downloaded: 1.0.3, date: 2020-11-13). We  
165 defined signals to be overlapping if a lead SNP from a previous GWAS was in LD ( $R^2>0.8$ )  
166 with any of our lead SNPs.

### 167 Mendelian Randomization

168 Firstly, we undertook a one-sample MR approach to investigate the causal effect of estradiol  
169 on BMD in UKB, in males and females separately. BMD was measured as heel BMD, based  
170 on an ultrasound measurement of the calcaneus (heel bone), using the Sahara Clinical Bone  
171 Sonometer. We used T-score, automated (data-field 78 in UKB). The T-score is the number  
172 of standard deviations a person's BMD differs from the mean BMD of their respective sex.  
173 The main MR analyses were performed with the R package gsmr (version 1.0.8)<sup>24</sup>. The  
174 GSMR approach allows both for heteroskedastic and correlated data via the variance-  
175 covariance matrix, and identifies and removes pleiotropic outliers prior to causal effect  
176 estimation. The default LD threshold was used ( $R^2=0.05$ ) and the significance threshold for  
177 an SNP being identified as an outlier, and therefore not belonging to the set of valid  
178 instruments, was set to the default value of  $\alpha=0.01$ . GSMR assumes a fixed effects model  
179 with the test statistic following a  $\chi^2$  distribution with 1 degree of freedom. We further  
180 performed sensitivity analyses, applying inverse variance weighted, weighted median, and  
181 the MR-Egger method, using the "MendelianRandomization" tool implemented in R. The  
182 MR-Egger intercept test was used to assess the possible presence of directional pleiotropy in  
183 the set of genetic instruments, where a deviation from zero intercept indicates pleiotropy<sup>25</sup>.  
184 Both the inverse variance weighted and the MR-Egger methods assume a random effects  
185 model, with underdispersion not being allowed. The standard error of the estimate for the

186 weighted median method is estimated using bootstrapping. If heterogeneity is present in the  
187 data, these methods would tend to show higher p-values as compared GSMR, as GSMR is  
188 expected to remove outliers. Secondly, the MR results were further replicated by two-sample  
189 MR, using the GWAS summary statistics for BMD from the GEFOS study<sup>26</sup>. This was done  
190 in order to avoid any weak-instrument bias towards the confounded association, potentially  
191 present in one-sample MR analyses. We used GWAS summary statistics for lumbar spine  
192 BMD (LSBMD) from the 2012 data release on [www.gefos.org](http://www.gefos.org) (downloaded: 2020-11-13),  
193 which was the only dataset with summary statistics for males and females separately that did  
194 not include samples from UKB. The two sets (for males and females) of genetic variants that  
195 had been selected in UKB were tested independently using GEFOS BMD summary statistics  
196 for males and females respectively. The overlap, with regards to SNPs, was not perfect  
197 between UKB and GEFOS. We therefore revised the set of SNPs for the MR in GEFOS  
198 slightly. For each of our GWAS locus, we selected the most significant SNP in LD with our  
199 lead SNP ( $R^2 > 0.3$ ) for which summary statistics were also available in the GEFOS data as  
200 a proxy. However, to avoid too weak instruments in the MR, only proxies that were  
201 associated with estradiol levels ( $P < 0.0001$ ) were included in the MR. Finally, we note that  
202 any winner's-curse bias in our effect estimates of the genetic variants associated with  
203 estradiol levels, should, on average, drive the estimates of the causal effect of estrogen on  
204 BMD towards the null.

## 205 Results

206 A total of 147,690 male participants had passed our genotype and estradiol QC (Figure 1,  
207 Table 1). Of these, 134,323 had estradiol levels below detection limit (175 pmol/L), while  
208 13,367 individuals had measured estradiol levels above, or equal to 175 pmol/L (Table 1).  
209 The number of females ( $N = 163,985$ ) was slightly higher compared to males, with 37,461

210 having estradiol levels above detection limit, and 126,524 below. Dramatically smaller  
211 fractions of males (9.1%) and postmenopausal women (7.9%) had detectable estrogen levels  
212 (Table 1), in comparison to the fraction of premenopausal women (71.9%) with detectable  
213 estrogen levels. Number of live births, age, and BMI were significantly lower among women  
214 with detectable estradiol in both post and premenopausal strata ( $P < 0.05$ ).

## 215 GWAS results

216 We identified 15 loci, on 14 different chromosomes, to be significantly associated ( $P < 5 \times 10^{-8}$ )  
217 with estradiol in males or females (Figure 2, Table 2). Of these, 13 were genome-wide  
218 significant only in males, one only in females (*MCM8*) and one in both males and females  
219 (*CYP3A7*). On chromosome 2, and 15, one conditional SNP was identified per locus, located  
220 in the vicinity to the primary lead SNPs (Table 2). Among the lead SNPs identified in males,  
221 some were located in or close to genes with a well-known role in estradiol metabolism.  
222 Among the most significant loci, we found *FAM9A* and *CYP19A1*, which have previously  
223 been identified in GWAS for estradiol in males<sup>12,15</sup>. One lead SNP, rs45446698, was  
224 genome-wide significant in both males and females, mapped to *CYP3A7*. It is also the most  
225 significant eQTL for *CYP3A7* in adrenal gland in GTEx ( $P = 1.2 \times 10^{-12}$ ), suggesting that the  
226 effect is mediated by the expression level of *CYP3A7*. The minor (G) allele increases the  
227 level of *CYP3A7* expression, but decreases the estradiol levels. There was also one lead SNP  
228 in *SHBG*, which codes for SHBG that had been linked to measurements of several other sex  
229 hormones<sup>16</sup>. Another lead SNP was located on the X chromosome, within the *AR* gene that  
230 encodes an androgen receptor. We also found a very strong association on chromosome 2,  
231 that overlaps with *SRD5A2*, which codes for 3-oxo-5- $\alpha$ -steroid 4-dehydrogenase 2,  
232 known to catalyze the conversion of testosterone into androgen or dihydrotestosterone<sup>27</sup>. Our  
233 most significant SNP is in LD ( $R^2 = 0.75$ ) with rs9282858, which is a missense variant that is

234 predicted to be deleterious (SIFT=0.03). Here, the minor allele was associated with higher  
235 estradiol levels (Table 2). At the same locus on chromosome two, there was also one strong  
236 conditional signal, further upstream of the primary signal, closer to *MEMO1*.

237 There was also an association to rs657152 in *ABO*, and SNP that is in LD ( $R^2=0.98$ ) with  
238 rs8176719, which tags the O blood group<sup>28</sup>. The A allele of rs657152 (Table 2) tags the non-  
239 O blood groups and we found rs657152-A to be associated with lower levels of estradiol,  
240 indicating that individuals with a non-O blood group have lower levels of estradiol compared  
241 to individuals with the O blood group. Two associations on chromosome 4 and 5  
242 respectively, mapped to *UGT2B7* (rs7662029) and *UGT3A1* (rs1073548). *UGT3A1* and  
243 *UGT2B7* are glucuronosyltransferases, which are enzymes that catalyze the glucosidation of  
244 lipophilic chemicals, facilitating their elimination<sup>29</sup>. Another lead SNP mapped to *SULT2A1*,  
245 which encodes a sulfotransferase which is believed to be important for sex-steroid  
246 biosynthesis by sulfating hydroxylated steroid hormones and bile acids<sup>30</sup>. Another lead SNP  
247 on chromosome 14 is a missense variant in *FKBP4*, which encodes the protein FKBP52 that  
248 has been suggested to be involved in the formation of steroid hormone nuclear receptors that  
249 regulate hormone levels<sup>31</sup>.

250 The remaining lead SNPs from the GWAS in males were mapped to loci with less clear links  
251 to estradiol metabolism. These include one lead SNP (and one conditional SNP) in a locus  
252 with a cluster of genes, including *IGHV3-7* and *IGHV6-1*, that encodes parts of  
253 immunoglobulin heavy chains that participates in the antigen recognition. Another lead SNP  
254 is located in *RBBP8* which encodes Retinoblastoma-binding protein 8 (RBBP8, CtIP), an  
255 endonuclease involved in the repair of double-stranded DNA breaks through homologous  
256 recombination as well as regulation of G2/M cell-cycle checkpoints as part of the BRCA1-  
257 RBBP8 complex<sup>32</sup>. Variation in *RBBP8* genotype and regulation has been reported in

258 multiple types of cancer<sup>33,34</sup>. A third SNP was a missense variant in *LINC01324*, a lncRNA  
259 which has been established as a predictor for melanoma progression<sup>35</sup>.

260 The lead SNP that was only significant in females (rs16991615, Table 2), is a missense  
261 variant located in *MCM8*. *MCM8* codes for a DNA helicase, which is involved in the repair  
262 of double-stranded DNA breaks and recombination<sup>36</sup>. Even though very few genome-wide  
263 significant SNPs were identified in women, eight out of 14 of the loci that were genome-wide  
264 significant in males, were at least nominally associated in females as well ( $P < 0.05$ ), and  
265 with the same direction of effect (Figure 3). Especially the genes with functions related to  
266 hormone metabolism; *UGT2B7*, *UGT3A1*, *CYP3A7*, *CYP19A1*, *SRD5A2*, and *SHBG*  
267 replicated in females, but also *IGHV3-7* and *ABO*.

#### 268 Sensitivity analysis

269 Even though the primary aim of the sensitivity analyses was to further verify the associations  
270 from the primary GWAS, we did perform GWAS with the different sets of covariates/strata  
271 in the sensitivity analyses as well. However, no additional genome-wide significant locus was  
272 identified. The effects of the findings from the primary analyses changed somewhat. When  
273 stratifying females into pre- and post-menopause there was one locus, *CYP3A7*, with a  
274 significant difference in the OR between the two strata (Supplementary Table S1<sup>37</sup>, Figure 3).  
275 For *CYP3A7*, the largest effect was seen in postmenopausal women OR = 0.703 (0.642 -  
276 0.770), followed by males OR = 0.811 (0.758 - 0.868), and then premenopausal OR = 0.907  
277 (0.837 - 0.984).

278 After removing all participants that had a prior cancer diagnosis at the time of assessment,  
279 there was no significant difference in OR compared to the primary GWAS, and all lead SNPs  
280 were still genome-wide significant, indicating that a prior cancer diagnosis did not influence  
281 our primary results (Supplementary Table S1<sup>37</sup>). However, when adjusting the estradiol

282 GWAS for testosterone and SHBG levels, the lead SNPs at *SHBG* and *FKBP4* were no  
283 longer genome-wide significant in males, and the corresponding effects were significantly  
284 lower than the previous signals. Also, for *UGT3A1* and *AR*, the P-values were below the  
285 threshold for genome-wide significance for estradiol when adjusting for testosterone and  
286 SHBG levels, but the ORs were not significantly different from the primary analyses. This  
287 indicates that the signals at *SHBG* and *FKBP4* might have been completely driven by SHBG  
288 and testosterone levels, whereas the other loci are more likely to be estradiol specific loci. We  
289 observed the same effect for the female-specific locus *MCM8*. There was no significant  
290 difference in OR between the primary GWAS in postmenopausal women with and without  
291 adjustment for age at menopause, suggesting that the results in females are not strongly  
292 confounded by age at menopause as has been discussed previously<sup>16</sup>.

### 293 GWAS overlap with estradiol and related phenotypes

294 We did a systematic check for overlap with previous GWAS results that were available in the  
295 GWAS catalog (Supplementary Table S2<sup>37</sup>). For lead SNPs at *UGT3A1*, *LINC01324* and  
296 *RBBP8*, we found no overlap with previous GWAS signals. Several SNPs at the *AR* locus  
297 have previously been associated with male pattern baldness<sup>38</sup>, but none of those are in strong  
298 LD with our lead SNP ( $R^2 < 0.3$ ). The lead SNP at the other ten loci, all overlapped with at  
299 least one previous GWAS. There were several overlapping signals with phenotypes that are  
300 related to sex hormones. For example, the lead SNPs at *SRD5A2* and *FAM9A* overlap with  
301 male pattern baldness, as well. The lead SNPs at *CYP3A7* and *SULT2A1* overlap with  
302 GWAS-hits for several sex hormones and other metabolites, *UGT2B7* overlaps with serum  
303 and urinary metabolites, while *SHBG* overlaps with testosterone levels. Our lead SNPs in  
304 *FAM9A*, *CYP19A1*, and *CYP3A7* all overlap with previous GWAS for BMD. Other loci  
305 overlap with GWAS for more distal traits. For example, the lead SNP in *ABO* overlaps with  
306 results from over 100 GWAS, with associations to metabolites, liver enzymes and proteins as

307 well as cardiovascular diseases, cancer, and type 2 diabetes. The lead SNP at *FKBP4*  
308 overlaps with waist to hip ratio and lung function. *MCM8*, the female specific locus, overlaps  
309 with traits related to age at menopause. This agrees partly with our sensitivity analyses where  
310 the OR decreased slightly from OR = 1.139 (1.087 – 1.193) in the primary GWAS to OR =  
311 1.126 (1.035 – 1.225) when adjusted for age at menopause. Still the OR was significantly  
312 different from unity in the sensitivity analyses suggesting that the locus is probably  
313 associated with both age at menopause and estradiol levels.

314

### 315 Mendelian randomization

316 A total of 209,043 participants (96,432 males and 112,611 females) had BMD measurements  
317 and covariate information available, and were included in the MR analysis. Mean value for T-  
318 scores in males was -0.085, and for females -0.58. The average T-scores below zero agrees  
319 with the average age in UKB participant being higher than the general population, and  
320 thereby the BMD being lower compared to the mean BMD of respective sex. The instrument  
321 variables for the MR analyses were selected from the lead SNPs in the primary GWAS for  
322 males and females, respectively. For males, all significant lead SNPs and conditional SNPs  
323 were selected, resulting in 16 instruments (Table 3). The estimated effects on estradiol level  
324 (Supplementary Table S3<sup>37</sup>) were taken from the primary GWAS for all instruments, also for  
325 the SNPs that were identified in the conditional analyses. Hence, the estimates for the  
326 conditional SNPs differ slightly from the ones in Table 2. Using GSMR, three instruments  
327 were identified as pleiotropic and were excluded from the primary analysis in males  
328 (Supplementary Table S3<sup>37</sup>). We identified a causal effect of high estradiol levels on BMD in  
329 males (Table 3, Figure 4a). As estradiol was analyzed as a binary phenotype, the unit for the  
330 estimate from the MR can be interpreted as standard units increase in T-score associated with

331 having estradiol levels above detection limit. Similar MR results were obtained using  
332 weighted median and MR-Egger, which do not remove pleiotropic SNPs. The MR-Egger  
333 intercept was not significantly different from zero, indicating that these results are not  
334 strongly influenced by directional pleiotropy.

335 To get a reasonable number of instruments to be used for the MR analysis in females, we  
336 used a less strict P-value cutoff ( $P < 1 \times 10^{-7}$ ), resulting in a total of 4 instrumental variables  
337 (Supplementary Table S3<sup>37</sup>) that were included in the analysis. In the GSMR analyses, the  
338 HEIDI-outlier flag was disabled in females, as there were fewer ( $< 10$ ) SNPs than  
339 recommended for this approach to work properly<sup>24</sup>. We identified a causal effect of high  
340 estradiol levels on BMD also in women ( $\beta = 0.03$ ,  $P = 7.48 \times 10^{-6}$ , Figure 4b). All three  
341 alternative MR methods showed the same direction of effect, although the null hypothesis of  
342 no effect could not be rejected, and the MR Egger intercept was not significantly different  
343 from zero (Table 3).

344 To further validate the MR results we also did a two-sample MR for males and females  
345 respectively using the instruments selected for estradiol in UKB and the GWAS summary  
346 statistics from the BMD GWAS in GEFOS<sup>26</sup>. A slightly different set of instrumental  
347 variables (Supplementary Table S3<sup>37</sup>) had to be selected, since considerably fewer SNPs were  
348 analyzed in GEFOS due to their use of the HapMap reference panel for imputations. We  
349 included 8 instruments for the replication in males, and four in females. The MR estimates  
350 were in the same direction as the one-sample MR (Figure 4c, d), but with a larger effect in  
351 both males ( $\beta = 0.18$ ,  $P = 0.00012$ ) and females ( $\beta = 0.41$ ,  $P = 3.49 \times 10^{-5}$ )

352 Previous concerns that age at menopause influences estradiol levels, could potentially render  
353 some of our selected instruments invalid since being affected by confounding (correlated  
354 horizontal) pleiotropy. We therefore also performed a third set of MR-analyses in females.

355 Here, we selected the instruments for estradiol from the GWAS in males, which should not  
356 be influenced by age at menopause. To reduce the risk for male-specific effects on estradiol  
357 levels, we filtered for SNPs that were significant also in females, after adjusting for the 16  
358 instruments tested ( $P < 0.0031$ ), resulting in six instruments and used the estradiol estimates  
359 for females in UKB (Supplementary Table S3<sup>37</sup>). Using the GSMR approach, we could  
360 identify a significant effect on BMD in females, similar to the effect estimated in males  
361 ( $\beta = 0.16$ ,  $P = 4.97 \times 10^{-6}$  (Table 3, Figure 4e). These results indicate that the effect in females is  
362 not driven by age at menopause.

## 363 Discussion

364 We have presented results from a large GWAS for estradiol levels in males and females.  
365 Many of the significant SNPs are annotated to genes with functions that are biologically  
366 relevant to estrogen metabolism. For example, by converting and transporting androgen and  
367 endogens, synthesis or elimination of steroid hormones. In males, we identified 14 loci, to be  
368 associated with serum estradiol levels, and another two independent SNPs on chromosome 2  
369 and 15. In females, we identified two significant loci. One locus, within the *CYP3A7* gene,  
370 overlaps with our findings in males, but the other locus, mapping to *MCM8*, was specific to  
371 females. However, a majority (eight) of the male loci, was at least nominally significant in  
372 females, even if the estimated effect size was lower. These results suggest that genetic effects  
373 on estradiol levels might differ between males and females. However, in premenopausal  
374 women, estradiol levels vary during the menstrual cycle, limiting our ability to estimate the  
375 genetic effects on estradiol in females. In postmenopausal women, on the other hand,  
376 estradiol levels drop rapidly, and it has previously been suggested that the estradiol levels in  
377 females are more likely to be determined by the number of years since menopause<sup>39</sup>. For  
378 those reasons, we performed sensitivity analyses, stratifying females by menopausal status,

379 and also adjusting for age at menopause in the postmenopausal strata. The only locus being  
380 significantly different between pre- and postmenopausal women was *CYP3A7*. The largest  
381 effect was seen in postmenopausal women, followed by males, and then premenopausal  
382 women. *CYP3A7* encodes cytochrome P450 CYP3A7, which metabolizes  
383 dehydroepiandrosterone (DHEA), a precursor of both androgen and estrogen synthesis, and  
384 shows catalytic activity for conversion of estrone to hydroxyestrogens<sup>40,41</sup>. As much as 75%  
385 of the circulating estrogens in premenopausal women and 50% of circulating androgens in  
386 men are known to be derived from DHEA. After menopause, DHEA is thought to be the  
387 major precursor of androgens and estrogens in females<sup>42</sup>. Interestingly, the effect of the  
388 *CYP3A7* SNP disappeared when adjusting for SHBG and testosterone levels in females, but  
389 the effect was unaffected in males. However, *CYP3A7* loses genome-wide significance when  
390 adjusting for testosterone levels. This suggests that *CYP3A7* might have different functions in  
391 males and in females, as well as in pre- and in postmenopausal women. In postmenopausal  
392 women *CYP3A7* appears to mainly influence estradiol levels by regulating the amount of  
393 available testosterone as a precursor for estrogen.

394 Several other loci, with previous links to estrogen metabolism were also identified. *SULT2A1*  
395 encodes a sulfotransferase which are active on hydroxylated steroid hormones and bile  
396 acids<sup>30</sup>. Sulfotransferase 2A1, encoded by *SULT2A1*, mediates the sulfation of a wide range  
397 of steroids and sterols, including DHEA, a precursor of estrogen and testosterone. *SULT2A1*  
398 was only associated with estradiol in males, and this association was similar (slightly  
399 stronger) when adjusting for testosterone and SHBG levels. Two of the most well-known  
400 loci, that has previously also been associated with estradiol in males in GWAS, are *FAM9A*  
401 and *CYP19A1*<sup>12,15</sup>. Aromatase, encoded by *CYP19A1*, is a key enzyme in the synthesis of  
402 estradiol and estrone from testosterone and androstenedione, respectively<sup>43</sup>. The association  
403 to *CYP19A1* was at least nominally significant, in both the pre- and postmenopausal women,

404 as well as in all sensitivity analyses. The P-values were even numerically slightly lower when  
405 adjusting for SHBG and testosterone, which agrees with the function of the protein. *FAM9A*  
406 is exclusively expressed in testis<sup>59</sup> and it is therefore not surprising that there was no  
407 association to *FAM9A* in females. Previously, *FAM9A* has also been associated with  
408 testosterone levels<sup>15</sup>, and our analyses indicates that the association was partly driven by  
409 associations to testosterone or SHBG levels. We could therefore conclude that the SNP at  
410 *FAM9A* is likely to affect both testosterone and estrogen metabolism. This signal also  
411 overlaps with a GWAS signal for male pattern baldness<sup>38</sup>, which further supports the  
412 importance of this locus. However, *FAM9A*, encodes the protein FAM9A, that has a rather  
413 unknown function. Also, the SNP is located between *FAM9A* and *FAM9B* and the annotation  
414 of this locus should therefore be interpreted with care.

415 We identified a strong association between estradiol levels and the *SHBG* locus. The *SHBG*  
416 gene encodes SHBG that has been linked to measurements of several other sex hormones  
417 previously<sup>16</sup>. SHBG is a glycoprotein that binds to sex hormones, including testosterone and  
418 estrogen, in the bloodstream, and thereby regulates the amount of bioavailable hormones<sup>44</sup>.  
419 Only 1–2% of sex hormones are unbound and therefore bioavailable<sup>45–47</sup>. The association to  
420 SHBG in males, completely disappeared after adjusting for SHBG levels. In females, there  
421 was a nominally significant association to SHBG in the primary GWAS, with the same  
422 direction of effect as in males. Interestingly, after adjusting for SHBG and testosterone levels,  
423 there was still an association (P=0.00033), but the effect was in the opposite direction.  
424 Similarly, an opposite effect between males and females has previously been seen for total  
425 testosterone, where there is a negative genetic correlation between total testosterone and  
426 SHBG in females, but a positive correlation in males<sup>16</sup>. This clearly illustrates the  
427 heterogenous effects between males and females for sex hormones. We also identified two  
428 associations to the X chromosome in males, one mapping to the *AR* gene, which encodes the

429 androgen receptor. In mammals, the action of androgens acts through this single androgen  
430 receptor. In different tissues, the androgen receptor is activated by testosterone, or  
431 testosterone is first converted into dihydrotestosterone which is then the activating  
432 molecule<sup>48</sup>. Therefore, the androgen receptor plays an important role in sex differentiation,  
433 mainly through regulating gene expression<sup>49</sup>. Interestingly no association ( $P>0.05$ ) was seen  
434 in females which might be due to the much lower levels of androgens in females. We also  
435 found an association to *SRD5A2* which encodes 3-oxo-5- $\alpha$ -steroid 4-dehydrogenase 2 and is  
436 located on chromosome 2. This enzyme catalyzes the conversion of testosterone to 5- $\alpha$ -  
437 dihydrotestosterone<sup>27</sup>. Thus, the variant in *SRD5A2* might influence estradiol levels in  
438 men by affecting the amount of testosterone available for conversion to estradiol. The SNP in  
439 *SRD5A2* is nominally significant in females, an association that is more pronounced in  
440 postmenopausal women. There is also a conditional signal in males, 200 kb upstream, close  
441 to *MEMO1* which encodes the mediator of ErbB2-driven cell motility 1 (MEMO1). In  
442 previous studies, it has been shown that MEMO1 regulates the extranuclear function of the  
443 estradiol receptor<sup>50</sup>, which is a possible link to estradiol levels. It is therefore possible that the  
444 GWAS signal on chromosome two, indeed represent associations with two different genes.  
445 Another interesting association is with *FKBP4*, which encodes FKBP52 (FK506-binding  
446 protein 4). FKBP52 has been show to regulate the androgen, glucocorticoid and progesterone  
447 receptor signaling pathways<sup>31</sup>, by playing an important role in the formation of steroid  
448 hormone nuclear receptors<sup>51</sup>. *FKBP4* has been shown to regulate androgen receptor  
449 transactivation activity and also been linked to testosterone levels<sup>52</sup>. When adjusting for  
450 testosterone and SHBG levels, the *FKBP4* association was weaker, and no longer significant.  
451  
452 Two loci, mapping to *UGT2B7* and *UGT3A1* were also identified. These two loci encode  
453 UDP-glycosyltransferases, which are responsible for the addition of sugars to a broad range

454 of lipophilic molecules to improve the water solubility and thereby play a major role in  
455 elimination of potentially toxic xenobiotics and endogenous compounds. Both UGT3A1 and  
456 UGT2B7 have been shown to metabolize estrogens by glucuronidation, i.e., the transferring  
457 of glucuronic acid to the metabolites<sup>29,53,54</sup>. This suggests that UDP-glycosyltransferases  
458 might play a major role in estradiol metabolisms. Also, *ABO* encodes a glycosyltransferase  
459 that is known to modify carbohydrates on the red blood cell antigens. *ABO* is the gene that is  
460 responsible for the ABO blood groups<sup>28</sup>. The effect allele of our study (the A allele) is in LD  
461 with the non-O blood groups, indicating that men with a non-O blood type have lower  
462 estrogen levels, or the other way around, men with blood group O have higher levels. Since  
463 *ABO* and its histo-blood group properties have been linked to a vast number of diseases and  
464 functions, it is not unlikely that our finding is indicative of an association between ABO  
465 antigens and estradiol levels. The most well-known association for ABO, is that individuals  
466 with the O blood group have lower risk of cardiovascular diseases<sup>55</sup>, probably due to lower  
467 coagulation activity. There are very limited findings of a clear link between ABO and  
468 estrogen levels. However, one study found an interaction between ABO and HRT in relation  
469 to coagulation activity<sup>56</sup>. ABO antigens are present on most epithelial and endothelial cells as  
470 well as on T cells, B cells and platelets and are also detectable in most body fluids such as  
471 saliva<sup>57</sup>. Recently, ABO was linked to Covid-19, where the susceptibility to SARS-CoV-2  
472 has been suggested to be explained by modulation of sialic acid-containing receptors  
473 distributed on host cell surface<sup>58</sup>. Covid-19 has a higher mortality in males, which has been  
474 suggested to be linked to their lower estrogen levels<sup>59,60</sup>. We also identified several loci with  
475 less clear links to estrogen metabolism including *LINC01324*, *IGHV3-7*, *RBBP8*, and *MCM8*.  
476 *MCM8* is the female-specific locus, where the lead SNP is a missense variant. Genetic  
477 variants in *MCM8* have previously been associated with premature ovarian failure<sup>61,62</sup>, a  
478 condition that is characterized by low estrogen levels<sup>51</sup>. According to GTEx, *MCM8* is highly

479 expressed in both female and male reproductive organs, despite only being significant in the  
480 female GWAS cohort, which further support this locus as being involved in sex hormone  
481 metabolism.

482 We also performed MR analysis and estimated a causal effect of estradiol on BMD in males,  
483 and to our knowledge, for the first time also in females. Compared to previous studies in  
484 males, with not more than five SNPs as instruments for estradiol<sup>17</sup>, we included as many as  
485 16 SNPs in males and four in females. Interestingly, the effect estimate was higher for  
486 females than males in the two-sample MR, which could agree with the fact that BMD  
487 decreases more rapidly in females by age. Estrogen can be used as a preventative treatment in  
488 postmenopausal women with low BMD<sup>63,64</sup>, to prevent osteoporosis, and our results were  
489 therefore not surprising. However, one concern that has been raised previously<sup>16</sup>, is that  
490 estradiol levels are driven by number of years till/from menopause. To reduce potential bias  
491 in our results due to instruments being associated with age at menopause, we applied a  
492 second approach where we used instruments from males, to minimize our instruments'  
493 association to age at menopause. Using these instruments, we could replicate the MR results  
494 in women which suggests that age at menopause is not dramatically influencing our results.

495

496 One of the major strengths of this study is the large cohort from UKB with estradiol  
497 measurements available. However, only a subset of the participants had estradiol levels above  
498 detection limit. To increase the number of participants to analyze, we created a binary  
499 phenotype. Since normal levels of estradiol in males and postmenopausal women have a  
500 range below or just above the detection limit, we were not able to capture the full spectrum of  
501 individual variation in the cohort. Also, very few loci were identified in females. This could  
502 be due to hormone levels fluctuating during the menstrual cycle lowering the power to

503 identify true genetic effects behind estradiol levels in premenopausal women. However, we  
504 also conducted a GWAS including only women that had entered menopause, which should  
505 give a more stable phenotype. However, no additional significant GWAS hits were identified,  
506 even if the number of participants in the GWAS was comparable to the GWAS in males, and  
507 the range of estradiol levels was comparable. This suggest that there is an underlying  
508 difference in genetic effects between sexes.

509 In summary, we have identified biologically relevant genetic loci associated with variation in  
510 estradiol levels in males and females and found differences in genetic effects between sexes.  
511 In addition, we used an MR approach to show that estradiol levels have a causal effect on  
512 good bone health in both males in females, confirming that estrogen replacement therapy may  
513 be a good treatment to prevent osteoporosis and prevent fractures in elderly of both sexes.  
514 Our findings support previous research on the synthesis of estradiol and its effects on bone  
515 mineral density, as well as provide new insights into the genetic components of estrogen  
516 regulation and metabolism.

## 517 Acknowledgements

518 This research was conducted using the UKB Resource under  
519 applicationnumber 41143, following the restrictions on data availability set up by  
520 the UKB. We acknowledge all participants and staff involved in UKB for their  
521 valuable contribution. The computations were performed on resources provided  
522 by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science  
523 (UPPMAX) under project sens2017538. The research was funded by the Swedish Research  
524 Council (Å.J), and the cancer (Å.J), the brain (Å.J), the heart and lung (Å.J), the  
525 Å Wiberg (W.E.E), M Borgström (W.E.E), Hedströms, K och O F (W.E.E), A and  
526 M Rudbergs (W.E.E) foundations.

## 527 Data Availability

528 The data used for this study is available for bona fide researchers from the UK Biobank  
529 Resource (<http://www.ukbiobank.ac.uk/about-biobank-uk/>), and can be accessed by an  
530 application to the UK Biobank. GEFOS data included in this study can be downloaded from  
531 <http://www.gefos.org>.

## 532 Author contributions

533 W.E.E and Å.J designed the study; E.B, W.E.E, D.S performed the data analysis; D.S  
534 generated the figures; W.E.E, D.S and Å.J wrote the manuscript; W.E.E, T.K, D.S, E.B, Å.J,  
535 J.H interpreted the data, contributed to and reviewed the manuscript.

## 536 Conflict of interest

537 The authors declare no conflict of interests. The funding sources had no influence an took no  
538 part in the design or conduct of this research.

539

## 540 References

- 541 1. Bates GW, Bowling M. Physiology of the female reproductive axis. *Periodontol 2000*.  
542 2013;61(1):89-102. doi:10.1111/j.1600-0757.2011.00409.x
- 543 2. Hess RA, Bunick D, Lee K-H, et al. A role for oestrogens in the male reproductive  
544 system. *Nature*. 1997;390(6659):509-512. doi:10.1038/37352
- 545 3. Thomas MP, Potter BVL. The structural biology of oestrogen metabolism. *J Steroid*  
546 *Biochem Mol Biol*. 2013;137:27-49. doi:10.1016/j.jsbmb.2012.12.014
- 547 4. Cauley JA, Lucas FL, Kuller LH, Stone K, Browner W, Cummings SR. Elevated  
548 serum estradiol and testosterone concentrations are associated with a high risk for  
549 breast cancer. *Ann Intern Med*. 1999;130(4 Pt 1):270-277. doi:10.7326/0003-4819-130-  
550 4\_part\_1-199902160-00004
- 551 5. Rosendaal FR, Van Hylckama Vlieg A, Tanis BC, Helmerhorst FM. Estrogens,  
552 progestogens and thrombosis. *J Thromb Haemost*. 2003;1(7):1371-1380.  
553 doi:10.1046/j.1538-7836.2003.00264.x
- 554 6. Vikan T, Schirmer H, Njølstad I, Svartberg J. Low testosterone and sex hormone-  
555 binding globulin levels and high estradiol levels are independent predictors of type 2  
556 diabetes in men. *Eur J Endocrinol*. 2010;162(4):747-754. doi:10.1530/EJE-09-0943
- 557 7. Manolagas SC, O'Brien CA, Almeida M. The role of estrogen and androgen receptors  
558 in bone health and disease. *Nat Rev Endocrinol*. Published online 2013.  
559 doi:10.1038/nrendo.2013.179
- 560 8. Väänänen HK, Härkönen PL. Estrogen and bone metabolism. *Maturitas*. Published  
561 online 1996. doi:10.1016/0378-5122(96)01015-8

- 562 9. Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J. *Harrison's*  
563 *Principles of Internal Medicine*. Vol 2012. Mcgraw-hill New York; 2012.
- 564 10. Riggs BL, Khosla S, Melton LJ. A Unitary Model for Involutional Osteoporosis:  
565 Estrogen Deficiency Causes Both Type I and Type II Osteoporosis in Postmenopausal  
566 Women and Contributes to Bone Loss in Aging Men. *J Bone Miner Res*.  
567 1998;13(5):763-773. doi:10.1359/jbmr.1998.13.5.763
- 568 11. Pott J, Bae YJ, Horn K, et al. Genetic Association Study of Eight Steroid Hormones  
569 and Implications for Sexual Dimorphism of Coronary Artery Disease. *J Clin*  
570 *Endocrinol Metab*. 2019;104(11):5008-5023. doi:10.1210/jc.2019-00757
- 571 12. Chen Z, Tao S, Gao Y, et al. Genome-wide association study of sex hormones,  
572 gonadotropins and sex hormone-binding protein in Chinese men. *J Med Genet*.  
573 2013;50(12):794-801. doi:10.1136/jmedgenet-2013-101705
- 574 13. Liu M, Ingle JN, Fridley BL, et al. TSPYL5 SNPs: association with plasma estradiol  
575 concentrations and aromatase expression. *Mol Endocrinol*. 2013;27(4):657-670.  
576 doi:10.1210/me.2012-1397
- 577 14. Prescott J, Thompson DJ, Kraft P, et al. Genome-wide association study of circulating  
578 estradiol, testosterone, and sex hormone-binding globulin in postmenopausal women.  
579 *PLoS One*. 2012;7(6):e37815. doi:10.1371/journal.pone.0037815
- 580 15. Eriksson AL, Perry JRB, Coviello AD, et al. Genetic determinants of circulating  
581 estrogen levels and evidence of a causal effect of estradiol on bone density in men. *J*  
582 *Clin Endocrinol Metab*. 2018;103(3):991-1004. doi:10.1210/jc.2017-02060
- 583 16. Ruth KS, Day FR, Tyrrell J, et al. Using human genetics to understand the disease

- 584 impacts of testosterone in men and women. *Nat Med.* 2020;26(2):252-258.  
585 doi:10.1038/s41591-020-0751-5
- 586 17. Nethander M, Vandenput L, Eriksson AL, Windahl S, Funck-Brentano T, Ohlsson C.  
587 Evidence of a Causal Effect of Estradiol on Fracture Risk in Men. *J Clin Endocrinol*  
588 *Metab.* 2019;104(2):433-442. doi:10.1210/jc.2018-00934
- 589 18. Nakamoto J, Salameh WA, Carlton E. Endocrine Testing. In: Jameson JL, De Groot  
590 LJBT-E (Sixth E, eds. *Endocrinology*. Elsevier; 2010:2802-2834. doi:10.1016/B978-1-  
591 4160-5583-9.00155-6
- 592 19. Huang J, Ellinghaus D, Franke A, Howie B, Li Y. 1000 Genomes-based imputation  
593 identifies novel and refined associations for the Wellcome Trust Case Control  
594 Consortium phase 1 Data. *Eur J Hum Genet.* 2012;20(7):801-805.  
595 doi:10.1038/ejhg.2012.3
- 596 20. Walter K, Min JL, Huang J, et al. The UK10K project identifies rare variants in health  
597 and disease. *Nature.* 2015;526(7571):82-89. doi:10.1038/nature14962
- 598 21. Xu C, Tachmazidou I, Walter K, Ciampi A, Zeggini E, Greenwood CMT. Estimating  
599 genome-wide significance for whole-genome sequencing studies. *Genet Epidemiol.*  
600 2014;38(4):281-290. doi:10.1002/gepi.21797
- 601 22. Ward LD, Kellis M. HaploReg: A resource for exploring chromatin states,  
602 conservation, and regulatory motif alterations within sets of genetically linked variants.  
603 *Nucleic Acids Res.* 2012;40(D1):D930-D934. doi:10.1093/nar/gkr917
- 604 23. Carithers LJ, Moore HM. The Genotype-Tissue Expression (GTEx) Project.  
605 *Biopreserv Biobank.* 2015;13(5):307-308. doi:10.1089/bio.2015.29031.hmm

- 606 24. Zhu Z, Zheng Z, Zhang F, et al. Causal associations between risk factors and common  
607 diseases inferred from GWAS summary data. *Nat Commun.* 2018;9(1):224.  
608 doi:10.1038/s41467-017-02317-2
- 609 25. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using  
610 the MR-Egger method. *Eur J Epidemiol.* 2017;32(5):377-389.
- 611 26. Estrada K, Styrkarsdottir U, Evangelou E, et al. Genome-wide meta-analysis identifies  
612 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nat*  
613 *Genet.* 2012;44(5):491-501. doi:10.1038/ng.2249
- 614 27. Makridakis NM, Di Salle E, Reichardt JKV. Biochemical and pharmacogenetic  
615 dissection of human steroid 5 $\alpha$ -reductase type II. *Pharmacogenetics.* 2000;10(5):407-  
616 413. doi:10.1097/00008571-200007000-00004
- 617 28. Ogasawara K, Bannai M, Saitou N, et al. Extensive polymorphism of ABO blood  
618 group gene: Three major lineages of the alleles for the common ABO phenotypes.  
619 *Hum Genet.* 1996;97(6):777-783. doi:10.1007/BF02346189
- 620 29. Mackenzie PI, Rogers A, Treloar J, Jorgensen BR, Miners JO, Meech R. Identification  
621 of UDP Glycosyltransferase 3A1 as a UDP N -Acetylglucosaminyltransferase. *J Biol*  
622 *Chem.* 2008;283(52):36205-36210. doi:10.1074/jbc.M807961200
- 623 30. Riches Z, Stanley EL, Bloomer JC, Coughtrie MWH. Quantitative evaluation of the  
624 expression and activity of five major sulfotransferases (SULTs) in human tissues: the  
625 SULT &quot;pie&quot;. *Drug Metab Dispos.* 2009;37(11):2255-2261.  
626 doi:10.1124/dmd.109.028399
- 627 31. Sivils JC, Storer CL, Galigniana MD, Cox MB. Regulation of Steroid Hormone

- 628 Receptor Function By the 52-kDa FK506-Binding Protein (FKBP52). *Curr Opin*  
629 *Pharmacol.* 2011;11(4):314. doi:10.1016/J.COPH.2011.03.010
- 630 32. Yu X, Baer R. Nuclear localization and cell cycle-specific expression of CtIP, a  
631 protein that associates with the BRCA1 tumor suppressor. *J Biol Chem.*  
632 2000;275(24):18541-18549. doi:10.1074/jbc.M909494199
- 633 33. Wong AKC, Ormonde PA, Pero R, et al. Characterization of a carboxy-terminal  
634 BRCA1 interacting protein. *Oncogene.* 1998;17(18):2279-2285.  
635 doi:10.1038/sj.onc.1202150
- 636 34. Wu M, Soler DR, Abba MC, et al. CtIP silencing as a novel mechanism of tamoxifen  
637 resistance in breast cancer. *Mol Cancer Res.* 2007;5(12):1285-1295. doi:10.1158/1541-  
638 7786.MCR-07-0126
- 639 35. Yang S, Xu J, Zeng X. A six-long non-coding RNA signature predicts prognosis in  
640 melanoma patients. *Int J Oncol.* 2018;52(4):1178-1188. doi:10.3892/ijo.2018.4268
- 641 36. Lee KY, Im JS, Shibata E, et al. MCM8-9 complex promotes resection of double-  
642 strand break ends by MRE11-RAD50-NBS1 complex. *Nat Commun.* 2015;6.  
643 doi:10.1038/ncomms8744
- 644 37. Schmitz D, Ek WE, Berggren E, Höglund J, Karlsson T, Johansson Å. Genome-Wide  
645 Association Study of Estradiol Levels, and the Causal Effect of Estradiol on Bone  
646 Mineral Density, Supplementary Data. *Zenodo.* Published online March 3, 2021.  
647 doi:10.5281/ZENODO.4575528
- 648 38. Hagensars SP, Hill WD, Harris SE, et al. Genetic prediction of male pattern baldness.  
649 Noethen MM, ed. *PLOS Genet.* 2017;13(2):e1006594.

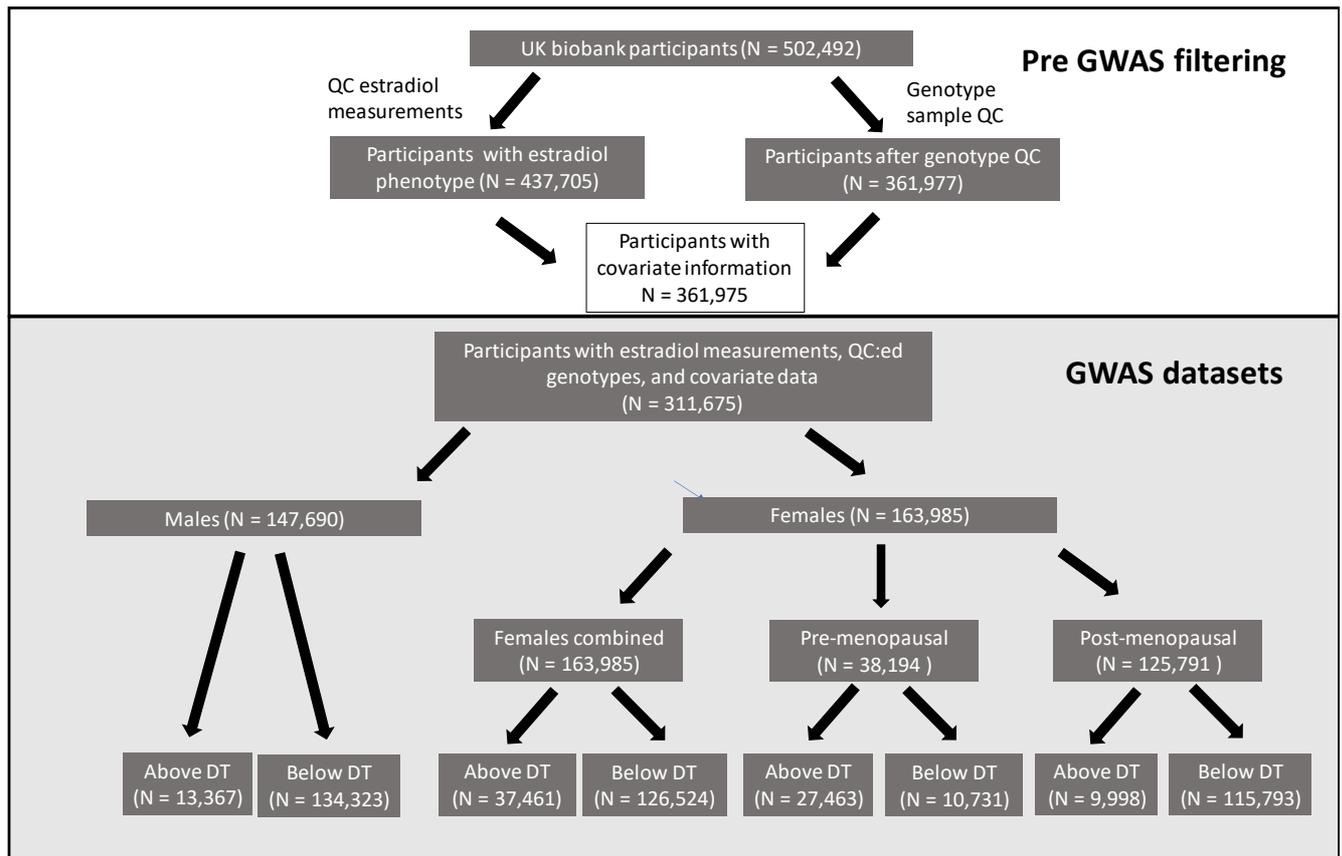
- 650           doi:10.1371/journal.pgen.1006594
- 651   39.   Richardson H, Ho V, Pasquet R, et al. Baseline estrogen levels in postmenopausal  
652       women participating in the MAP.3 breast cancer chemoprevention trial. In:  
653       *Menopause*. Vol 27. Lippincott Williams and Wilkins; 2020:693-700.  
654       doi:10.1097/GME.0000000000001568
- 655   40.   Lee AJ, Conney AH, Zhu BT. Human Cytochrome P450 3A7 Has a Distinct High  
656       Catalytic Activity for the 16 $\alpha$ -Hydroxylation of Estrone but not 17 $\beta$ -Estradiol. *Cancer*  
657       *Res*. 2003;63(19):6532 LP - 6536.  
658       <http://cancerres.aacrjournals.org/content/63/19/6532.abstract>
- 659   41.   Ohmori S, Nakasa H, Asanome K, et al. Differential catalytic properties in metabolism  
660       of endogenous and exogenous substrates among CYP3A enzymes expressed in COS-7  
661       cells. *Biochim Biophys Acta - Gen Subj*. 1998;1380(3):297-304. doi:10.1016/S0304-  
662       4165(97)00156-6
- 663   42.   Simpson ER, Davis SR. Minireview: Aromatase and the Regulation of Estrogen  
664       Biosynthesis—Some New Perspectives. *Endocrinology*. 2001;142(11):4589-4594.  
665       doi:10.1210/endo.142.11.8547
- 666   43.   Sohl CD, Guengerich FP. Kinetic analysis of the three-step steroid aromatase reaction  
667       of human cytochrome P450 19A1. *J Biol Chem*. 2010;285(23):17734-17743.  
668       doi:10.1074/jbc.M110.123711
- 669   44.   Hammond GL. Access of reproductive steroids to target tissues. *Obstet Gynecol Clin*  
670       *North Am*. 2002;29(3):411-423. doi:10.1016/S0889-8545(02)00008-6
- 671   45.   Coviello AD, Haring R, Wellons M, et al. A genome-wide association meta-analysis of

- 672 circulating sex hormone-binding globulin reveals multiple loci implicated in sex  
673 steroid hormone regulation. *PLoS Genet.* 2012;8(7). doi:10.1371/journal.pgen.1002805
- 674 46. Somboonporn W, Davis SR. Testosterone Effects on the Breast: Implications for  
675 Testosterone Therapy for Women. *Endocr Rev.* 2004;25(3):374-388.  
676 doi:10.1210/er.2003-0016
- 677 47. Dunn JF, Nisula BC, Rodbard D. Transport of steroid hormones: Binding of 21  
678 endogenous steroids to both testosterone-binding globulin and corticosteroid-binding  
679 globulin in human plasma. *J Clin Endocrinol Metab.* 1981;53(1):58-68.  
680 doi:10.1210/jcem-53-1-58
- 681 48. SL D, R B. Androgen physiology. *Semin Reprod Med.* 2006;24(2). doi:10.1055/S-  
682 2006-939565
- 683 49. Mooradian AD, Morley JE, Korenman SG. Biological actions of androgens. *Endocr*  
684 *Rev.* 1987;8(1):1-28. doi:10.1210/edrv-8-1-1
- 685 50. Jiang K, Yang Z, Cheng L, et al. Mediator of ERBB2-driven cell motility (MEMO)  
686 promotes extranuclear estrogen receptor signaling involving the growth factor  
687 receptors IGF1R and ERBB2. *J Biol Chem.* 2013;288(34):24590-24599.  
688 doi:10.1074/jbc.M113.467837
- 689 51. Hou Q, Gorski J, Daikoku T, et al. Estrogen receptor and progesterone receptor genes  
690 are expressed differentially in mouse embryos during preimplantation development.  
691 *Proc Natl Acad Sci U S A.* 1993;90(20):9460-9464. doi:10.1073/pnas.90.20.9460
- 692 52. Chen H, Yong W, Hinds TD, et al. Fkbp52 regulates androgen receptor transactivation  
693 activity and male urethra morphogenesis. *J Biol Chem.* 2010;285(36):27776-27784.

- 694 doi:10.1074/jbc.M110.156091
- 695 53. Dong M, Owens IS, Sheen YY. Cloning and expression of human liver UDP-  
696 glucuronosyltransferase cDNA, UDPGTh2. *Arch Pharm Res.* 1997;20(5):459-464.  
697 doi:10.1007/BF02973940
- 698 54. Kallionpää RA, Järvinen E, Finel M. Glucuronidation of estrone and 16 $\alpha$ -  
699 hydroxyestrone by human UGT enzymes: The key roles of UGT1A10 and UGT2B7. *J*  
700 *Steroid Biochem Mol Biol.* 2015;154:104-111. doi:10.1016/j.jsbmb.2015.07.013
- 701 55. Zhang H, Mooney CJ, Reilly MP. ABO Blood Groups and Cardiovascular Diseases.  
702 *Int J Vasc Med.* 2012;2012:641917. doi:10.1155/2012/641917
- 703 56. Guimarães DAM, dos Santos MS, Gomes KB, et al. Interaction between oral estrogen  
704 plus progestogen therapy and ABO blood groups on coagulation activation in  
705 postmenopausal women. *Menopause.* 2012;19(3):339-345.  
706 doi:10.1097/gme.0b013e31822b721f
- 707 57. Schachter H, Michaels MA. A quantitative difference in the activity of blood group A-  
708 specific N-acetylgalactosaminyltransferase in serum from A 1 and A 2 human subjects.  
709 *Biochem Biophys Res Commun.* 1971;45(4):1011-1018.
- 710 58. Silva-Filho JC, Melo CGF de, Oliveira JL de. The influence of ABO blood groups on  
711 COVID-19 susceptibility and severity: A molecular hypothesis based on carbohydrate-  
712 carbohydrate interactions. *Med Hypotheses.* 2020;144:110155.  
713 doi:10.1016/J.MEHY.2020.110155
- 714 59. Suba Z. Prevention and therapy of covid-19 via exogenous estrogen treatment for both  
715 male and female patients; An opinion paper. *J Pharm Pharm Sci.* 2020;23(1):75-85.

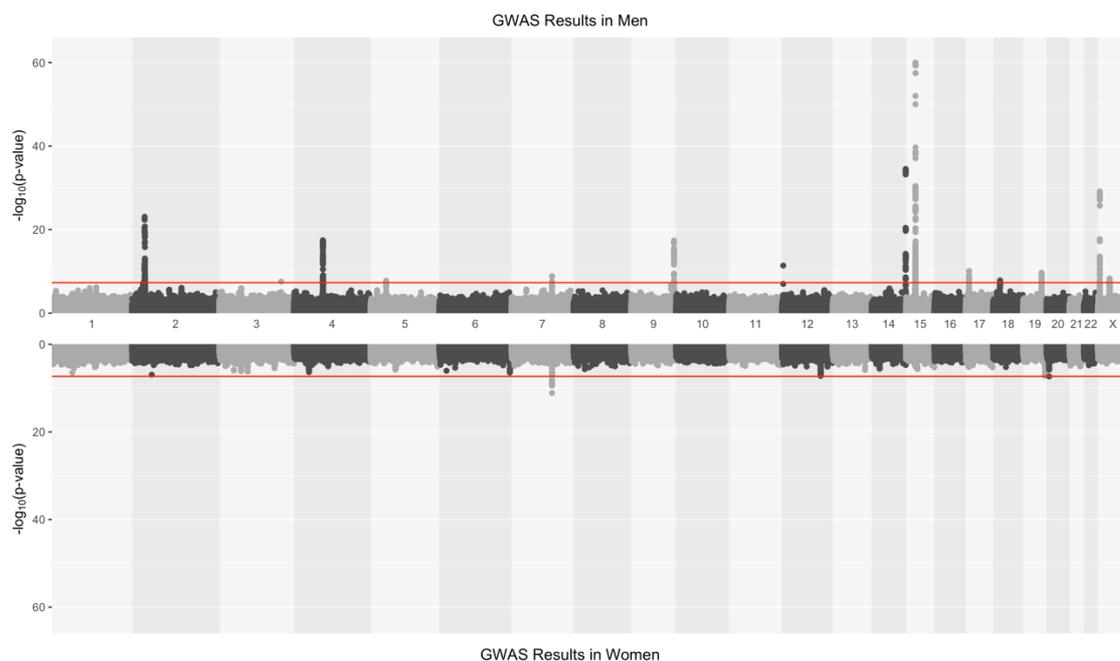
716 doi:10.18433/JPPS31069

- 717 60. Li Y, Jerkic M, Slutsky AS, Zhang H. Molecular mechanisms of sex bias differences  
718 in COVID-19 mortality. *Crit Care*. 2020;24(1):405. doi:10.1186/s13054-020-03118-8
- 719 61. AlAsiri S, Basit S, Wood-Trageser MA, et al. Exome sequencing reveals MCM8  
720 mutation underlies ovarian failure and chromosomal instability. *J Clin Invest*.  
721 2015;125(1):258-262. doi:10.1172/JCI78473
- 722 62. Dou X, Guo T, Li G, Zhou LG, Qin Y, Chen ZJ. Minichromosome maintenance  
723 complex component 8 mutations cause primary ovarian insufficiency. *Fertil Steril*.  
724 2016;106(6):1485-1489.e2. doi:10.1016/j.fertnstert.2016.08.018
- 725 63. Lufkin EG, Wahner HW, O'Fallon WM, et al. Treatment of postmenopausal  
726 osteoporosis with transdermal estrogen. *Ann Intern Med*. 1992;117(1):1-9.  
727 doi:10.7326/0003-4819-117-1-1
- 728 64. Abdi F, Mobedi H, Bayat F, Mosaffa N, Dolatian M, Tehrani FR. The effects of  
729 transdermal estrogen delivery on bone mineral density in postmenopausal women: A  
730 meta-analysis. *Iran J Pharm Res*. 2017;16(1):380-389. doi:10.22037/ijpr.2017.1967



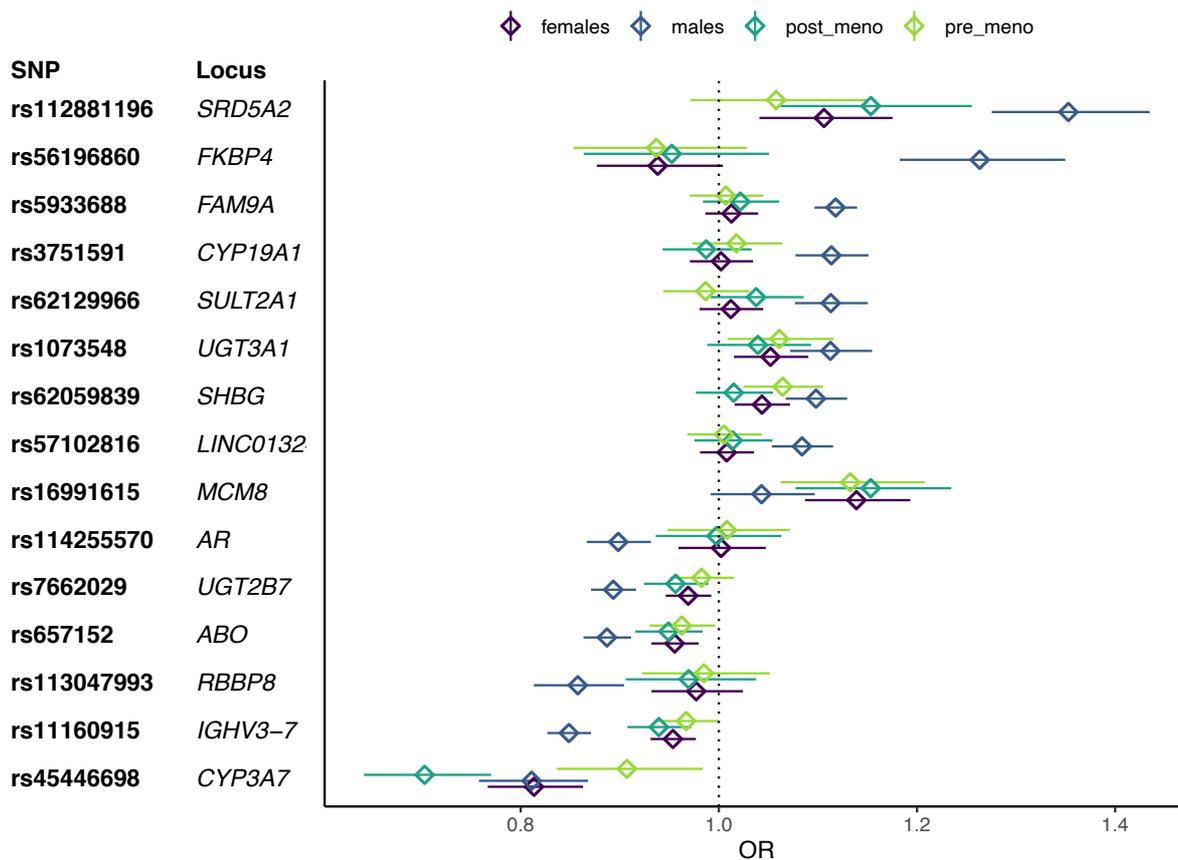
731

732 *Figure 1. Flowchart of UK Biobank participants included in the four GWAS. QC = quality control, DT = detection limit (for*  
 733 *estradiol measurement).*

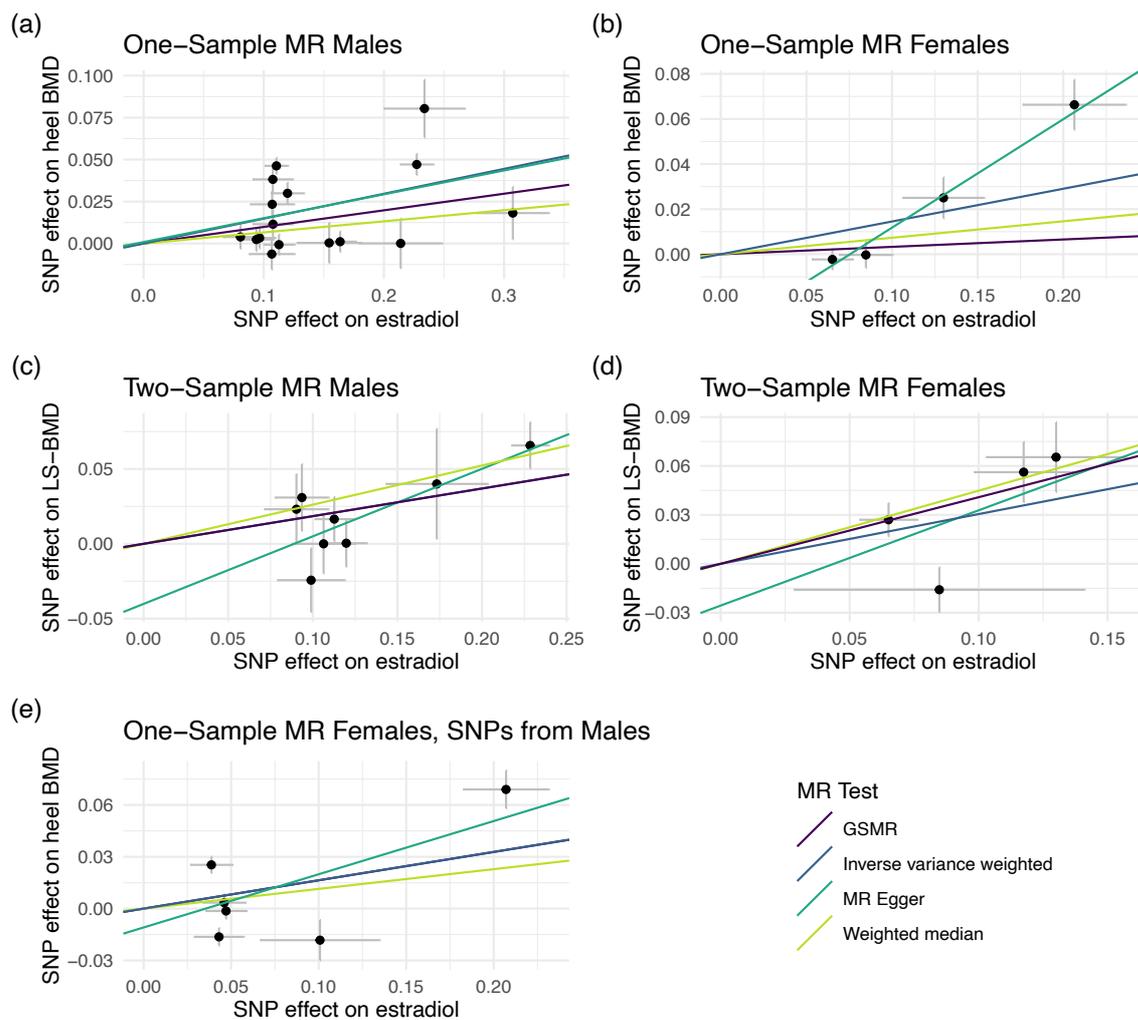


734

735 *Figure 2. Hudson plot for the estradiol GWAS in males (upper) and females (lower). The red horizontal line indicated the*  
 736 *genome-wide significance threshold ( $5 \times 10^{-8}$ ).*



737  
 738 *Figure 3. OR and 95% CIs for the lead SNPs in the male and female strata, as well as the two female subgroups: pre- and*  
 739 *post-menopause. The lead SNP for each locus that reached genome-wide significance in males or females in the primary*  
 740 *analyses are included. Non-overlapping confidence intervals typically indicate a significant difference in ORs well below*  
 741 *p = 0.05.*



742

743 *Figure 4. Mendelian randomization analyses to estimate the causal effect of estradiol on bone mineral*  
744 *density (BMD). (a) One-sample MR in UKB males, (b) One-sample MR in UKB females. (c) Two-sample*  
745 *MR with GEFOS males, and (d) Two-sample MR in GEFOS females. (e) One-Sample MR analysis with*  
746 *estradiol instruments selected in the UKB males, and MR performed in UKB females.*

747 *Table 1. Baseline characteristics in the UKB participants included in our GWAS*

	Males		Postmenopausal Females		Premenopausal Females	
	Above DT	Below DT	Above DT	Below DT	Above DT	Below DT
Participants, N (%)***	13499 (9.1)	135200 (90.1)	10090 (7.9)	117158 (92.1)	27672 (71.9)	10807 (28.1)
Participants with all covariates, N***	13367	134323	9998	115793	27463	10731
Estradiol levels (pmol/L), mean (Q1 – Q3)	203.8 (188.8 - 230.8)	NA	302.2 (221.2 - 475.2)	NA	433.7 (291.7 - 686.5)	NA
BMI*, mean (Q1 – Q3)	<b>28.4 (25.2 - 30.8)</b>	<b>27.8 (25.0 - 30.0)</b>	<b>27.5 (23.6 - 30.2)</b>	<b>27.2 (23.7 - 29.8)</b>	<b>26.2 (22.7 - 28.7)</b>	<b>26.6 (22.9 - 29.2)</b>
Age*, mean (Q1 – Q3)	57.1 (51 - 64)	57.1 (51 - 64)	<b>55.0 (50 - 60)</b>	<b>60.6 (57 - 65)</b>	<b>45.9 (43 - 48)</b>	<b>47.5 (44 - 50)</b>
Smoking-Current, N (%)	1179 (8.73)	11424 (8.45)	<b>784 (7.77)</b>	<b>7445 (6.35)</b>	<b>1862 (6.73)</b>	<b>809 (7.49)</b>
Smoking-Never, N (%)	6662 (49.3)	65993 (48.81)	5635 (55.85)	67537 (57.65)	17920 (64.76)	6909 (63.93)
Smoking-Occasional, N (%)	423 (3.1)	4424 (3.27)	<b>243 (2.41)</b>	<b>1884 (1.61)</b>	755 (2.73)	281 (2.60)
Smoking-Previous, N (%)	5175 (38.3)	52905 (39.13)	3399 (33.69)	39842 (34.01)	7083 (25.60)	2788 (25.80)
Had hysterectomy, N (%)**	NA	NA	<b>4868 (48.25)</b>	<b>27841 (23.76)</b>	<b>79 (0.29)</b>	<b>76 (0.70)</b>
Hormone replacement therapy, N (never/current/previous)*	NA	NA	<b>3474 / 4586 / 2024</b>	<b>58112 / 4606 / 54169</b>	<b>26797 / 403 / 424</b>	<b>10166 / 157 / 463</b>
Oral contraceptives, N (never/current/previous)*	NA	NA	<b>1252 / 28 / 8798</b>	<b>25005 / 87 / 91874</b>	<b>2442 / 1090 / 24098</b>	<b>883 / 1476 / 8430</b>
Number of live births, median (Q1 – Q3)**	NA	NA	<b>1.79 (1 - 2)</b>	<b>1.9 (1 - 3)</b>	1.58 (0 - 2)	1.56 (0 - 2)
Age at menarche, median (Q1 – Q3)*	NA	NA	<b>12.62 (12 - 14)</b>	<b>12.54 (12 - 14)</b>	<b>12.68 (12 - 14)</b>	<b>12.59 (12 - 14)</b>
Age at menopause, median (Q1 – Q3)*	NA	NA	<b>44.62 (44 - 52)</b>	<b>46.53 (46 - 53)</b>	NA	NA

748

749 DT: detection limit.

750 Q1: 1<sup>st</sup> quartile.

751 Q3: 3<sup>rd</sup> quartile.

752 Bold cells are significantly ( $P < 0.05$ ) associated with binary estradiol phenotype.

753 \*: Tested with Student's t-test.

754 \*\*: Tested with Fisher's exact test.

755 \*\*\* These are the actual number of participants analyzed in the GWAS

Primary/ Conditional	Lead SNP	Chr:position	GWAS strata	Gene	Eff allele	MAF	Males		Females	
							OR (95% CI)	P-Value	OR (95% CI)	P-Value
Primary	rs112881196 <sup>a</sup>	2:31982811	males	<i>SRD5A2</i>	G	0.039	1.353 (1.275 - 1.435)	8.52E-24	1.106 (1.041 - 1.175) <sup>f</sup>	0.00113
Conditional	rs62142080	2:32182528	males	<i>MEMO1/SRD5A2</i>	C	0.396	0.927 (0.903 - 0.952)	2.14E-08	0.989 (0.965 - 1.013) <sup>f</sup>	0.384
Primary	rs57102816	3:164506246	males	<i>LINC01324</i>	TA	0.257	1.084 (1.054 - 1.115)	2.87E-08	1.008 (0.981 - 1.036) <sup>f</sup>	5.66E-01
Primary	rs7662029	4:69961912	males	<i>UGT2B7</i>	G	0.454	0.893 (0.871 - 0.916)	3.61E-18	0.969 (0.946 - 0.992) <sup>f</sup>	0.00945
Primary	rs1073548	5:36012617	males	<i>UGT3A1</i>	G	0.126	1.113 (1.072 - 1.155)	1.75E-08	1.052 (1.015 - 1.090)	0.00528
Primary	rs45446698 <sup>b</sup>	7:99332948	males/females	<i>CYP3A7</i>	G	0.043	0.811 (0.758 - 0.868)	1.38E-09	0.813 (0.767 - 0.863)	7.62E-12
Primary	rs657152 <sup>c</sup>	9:136139265	males	<i>ABO</i>	A	0.340	0.887 (0.863 - 0.911)	4.14E-18	0.955 (0.932 - 0.980) <sup>f</sup>	0.000332
Primary	rs56196860 <sup>d</sup>	12:2908330	males	<i>FKBP4</i>	A	0.031	1.263 (1.183 - 1.350)	4.11E-12	0.938 (0.877 - 1.004) <sup>f</sup>	0.0656
Primary	rs11160915	14:106514653	males	<i>IGHV3-7</i>	A	0.444	0.849 (0.827 - 0.871)	3.08E-35	0.954 (0.931 - 0.977) <sup>f</sup>	0.000106
Primary	rs28892005	15:51519945	males	<i>CYP19A1</i>	A	0.350	0.795 (0.774 - 0.817)	1.01E-60	0.962 (0.939 - 0.986) <sup>f</sup>	0.00227
Conditional	rs3751591	15:51606710	males	<i>CYP19A1</i>	G	0.167	1.125 (1.088 - 1.163)	3.22E-12	1.004 (0.972 - 1.036) <sup>f</sup>	0.811789
Primary	rs62059839	17:7533015	males	<i>SHBG</i>	T	0.262	1.098 (1.067 - 1.130)	8.53E-11	1.044 (1.016 - 1.072)	0.00184
Primary	rs113047993	18:20585399	males	<i>RBBP8</i>	T	0.068	0.858 (0.813 - 0.904)	1.47E-08	0.977 (0.932 - 1.024) <sup>f</sup>	0.337
Primary	rs62129966	19:48374950	males	<i>SULT2A1</i>	A	0.164	1.113 (1.077 - 1.150)	2.18E-10	1.012 (0.981 - 1.045) <sup>f</sup>	0.456
Primary	rs16991615	20:5948227	females	<i>MCM8</i>	A	0.065	1.043 (0.992 - 1.097)	0.101	1.139 (1.087 - 1.193)	4.67E-08
Primary	rs5933688 <sup>e</sup>	X:8880680	males	<i>FAM9A</i>	G	0.275	1.118 (1.097 - 1.140)	7.49E-30	1.013 (0.986 - 1.040) <sup>f</sup>	0.348
Primary	rs114255570	X:67005508	males	<i>AR</i>	A	0.077	0.899 (0.867 - 0.931)	5.6E-09	1.002 (0.959 - 1.047) <sup>f</sup>	0.919

757 <sup>a</sup> LD with rs9282858, which is missense variant in *SRD5A2* (SIFT=0.03 deleterious, PolyPhen2=0.337 benign).<sup>7</sup>

758 <sup>b</sup> eQTL for *CYP3A7*.

759 <sup>c</sup> rs657152 tags the O blood group.

760 <sup>d</sup> rs56196860 is a missense variant in *FKBP4*.

761 <sup>e</sup> rs16991615 is a missense variant in *MCM8*.

762 <sup>f</sup> Significant (P<0.05) difference in OR between males and females.

MR method	Males (one-sample MR: UKB)		Males (two-sample MR: UKB - GEFOS)		Females (one-sample MR: UKB)		Females (two-sample MR: UKB - GEFOS)		Females, with SNPs from males (one-sample MR: UKB)	
	Estimate	P-value	Estimate	P-value	Estimate	P-value	Estimate	P-value	Estimate	P-value
<b>GSMR</b>	9.89E-02	1.57E-11	1.85E-01	1.20E-04	3.27E-02	7.48E-06	4.08E-01	3.49E-05	1.64E-01	4.97E-06
<b>Weighted median</b>	6.61E-02	2.82E-02	2.62E-01	1.01E-04	7.31E-02	1.26E-01	4.48E-01	1.25E-04	1.14E-01	1.49E-01
<b>IVW</b>	1.48E-01	5.49E-05	1.84E-01	1.06E-03	1.45E-01	9.98E-02	3.05E-01	6.11E-02	1.64E-01	1.93E-01
<b>MR-Egger</b>	1.42E-01	2.23E-01	4.51E-01	1.67E-02	4.81E-01	2.56E-02	5.84E-01	4.98E-01	3.08E-01	2.73E-01
<b>MR-Egger intercept</b>	8.87E-04	6.95E-01	-4.02E-02	8.36E-02	-3.63E-02	1.71E-01	-2.56E-02	7.23E-01	-1.09E-02	5.17E-01