

# 1 Genome-wide analyses identify 21 2 infertility loci and over 400 reproductive 3 hormone loci across the allele frequency 4 spectrum

5 Samvida S. Venkatesh<sup>1,2</sup>, Laura B. L. Wittemans<sup>3,4</sup>, Duncan S. Palmer<sup>1,5</sup>, Nikolas A. Baya<sup>1,2</sup>,  
6 Teresa Ferreira<sup>1</sup>, Barney Hill<sup>1,5</sup>, Frederik Heymann Lassen<sup>1,2</sup>, Melody J. Parker<sup>1,6</sup>, Saskia  
7 Reibe<sup>1,5</sup>, Ahmed Elhakeem<sup>7,8</sup>, Karina Banasik<sup>9,10</sup>, Mie T. Bruun<sup>11</sup>, Christian Erikstrup<sup>12,13</sup>, Bitten  
8 A. Jensen<sup>14</sup>, Anders Juul<sup>15,16</sup>, Christina Mikkelsen<sup>17,18</sup>, Henriette S. Nielsen<sup>19,20</sup>, Sisse R.  
9 Ostrowski<sup>17,20</sup>, Ole B. Pedersen<sup>20,21</sup>, Palle D. Rohde<sup>22</sup>, Erik Sorensen<sup>17</sup>, Henrik Ullum<sup>23</sup>, David  
10 Westergaard<sup>9,10</sup>, Asgeir Haraldsson<sup>24,25</sup>, Hilma Holm<sup>26</sup>, Ingileif Jonsdottir<sup>24,26</sup>, Isleifur Olafsson<sup>27</sup>,  
11 Thora Steingrimsdottir<sup>24,28</sup>, Valgerdur Steinthorsdottir<sup>26</sup>, Gudmar Thorleifsson<sup>26</sup>, Jessica  
12 Figueredo<sup>29</sup>, Minna K. Karjalainen<sup>30,31,32</sup>, Anu Pasanen<sup>33</sup>, Benjamin M. Jacobs<sup>34</sup>, Nikki  
13 Hubers<sup>35,36</sup>, Genes & Health Research Team, Estonian Biobank Research Team, Estonian  
14 Health Informatics Research Team, DBDS Genomic Consortium, FinnGen, Margaret  
15 Lippincott<sup>37,38</sup>, Abigail Fraser<sup>7,8</sup>, Deborah A. Lawlor<sup>7,8</sup>, Nicholas J. Timpson<sup>7,8</sup>, Mette Nyegaard<sup>22</sup>,  
16 Kari Stefansson<sup>24,26</sup>, Reedik Magi<sup>29</sup>, Hannele Laivuori<sup>30,39,40,41</sup>, David A. van Heel<sup>42</sup>, Dorret I.  
17 Boomsma<sup>35,36</sup>, Ravikumar Balasubramanian<sup>37,38</sup>, Stephanie B. Seminara<sup>37,38</sup>, Yee-Ming  
18 Chan<sup>38,43</sup>, Triin Laisk<sup>29</sup>, Cecilia M. Lindgren<sup>1,2,4,44</sup>

- 19  
20 1. Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of  
21 Oxford, Oxford OX3 7LF, United Kingdom  
22 2. Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford,  
23 Oxford OX3 7BN, United Kingdom  
24 3. Novo Nordisk Research Centre Oxford, Oxford, United Kingdom  
25 4. Nuffield Department of Women's and Reproductive Health, Medical Sciences Division,  
26 University of Oxford, United Kingdom  
27 5. Nuffield Department of Population Health, Medical Sciences Division, University of Oxford,  
28 Oxford, United Kingdom  
29 6. Nuffield Department of Clinical Medicine, University of Oxford, John Radcliffe Hospital,  
30 Oxford, United Kingdom  
31 7. MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, United Kingdom  
32 8. Population Health Science, Bristol Medical School, University of Bristol, Bristol, United  
33 Kingdom  
34 9. Novo Nordisk Foundation Center for Protein Research, University of Copenhagen,  
35 Copenhagen, Denmark  
36 10. Department of Obstetrics and Gynecology, Copenhagen University Hospital, Hvidovre,  
37 Copenhagen, Denmark  
38 11. Department of Clinical Immunology, Odense University Hospital, Odense, Denmark  
39 12. Department of Clinical Immunology, Aarhus University Hospital, Aarhus, Denmark

- 40 13. Department of Clinical Medicine, Health, Aarhus University, Aarhus, Denmark
- 41 14. Department of Clinical Immunology, Aalborg University Hospital, Aalborg, Denmark
- 42 15. Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of  
43 Copenhagen; Copenhagen, Denmark
- 44 16. Department of Growth and Reproduction, Copenhagen University Hospital-Rigshospitalet,  
45 Copenhagen, Denmark
- 46 17. Department of Clinical Immunology, Copenhagen University Hospital, Rigshospitalet,  
47 Copenhagen, Denmark
- 48 18. Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and  
49 Medical Science, Copenhagen University, Copenhagen, Denmark
- 50 19. Department of Obstetrics and Gynecology, The Fertility Clinic, Hvidovre University Hospital,  
51 Copenhagen, Denmark
- 52 20. Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of  
53 Copenhagen, Copenhagen, Denmark
- 54 21. Department of Clinical Immunology, Zealand University Hospital, Kge, Denmark
- 55 22. Genomic Medicine, Department of Health Science and Technology, Aalborg University,  
56 Aalborg, Denmark
- 57 23. Statens Serum Institut, Copenhagen, Denmark
- 58 24. Faculty of Medicine, University of Iceland, Reykjavik, Iceland
- 59 25. Children's Hospital Iceland, Landspítali University Hospital, Reykjavik, Iceland
- 60 26. deCODE genetics/Amgen, Inc., Reykjavik, Iceland
- 61 27. Department of Clinical Biochemistry, Landspítali University Hospital, Reykjavik, Iceland
- 62 28. Department of Obstetrics and Gynecology, Landspítali University Hospital, Reykjavik,  
63 Iceland
- 64 29. Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia
- 65 30. Institute for Molecular Medicine Finland, Helsinki Institute of Life Science, University of  
66 Helsinki, Helsinki, Finland
- 67 31. Research Unit of Population Health, Faculty of Medicine, University of Oulu, Finland
- 68 32. Northern Finland Birth Cohorts, Arctic Biobank, Infrastructure for Population Studies, Faculty  
69 of Medicine, University of Oulu, Oulu, Finland
- 70 33. Research Unit of Clinical Medicine, Medical Research Center Oulu, University of Oulu, and  
71 Department of Children and Adolescents, Oulu University Hospital, Oulu, Finland
- 72 34. Centre for Preventive Neurology, Wolfson Institute of Population Health, Queen Mary  
73 University London, London, EC1M 6BQ, United Kingdom
- 74 35. Department of Biological Psychology, Netherlands Twin Register, Vrije Universiteit,  
75 Amsterdam, The Netherlands
- 76 36. Amsterdam Reproduction and Development Institute, Amsterdam, The Netherlands
- 77 37. Harvard Reproductive Sciences Center and Reproductive Endocrine Unit, Massachusetts  
78 General Hospital, Boston, Massachusetts, United States of America
- 79 38. Harvard Medical School, Boston, Massachusetts, United States of America
- 80 39. Medical and Clinical Genetics, University of Helsinki and Helsinki University Hospital,  
81 Helsinki, Finland
- 82 40. Department of Obstetrics and Gynecology, Tampere University Hospital, Finland

- 83 41. Center for Child, Adolescent, and Maternal Health Research, Faculty of Medicine and  
84 Health Technology, Tampere University, Finland  
85 42. Blizard Institute, Queen Mary University London, London, E1 2AT, United Kingdom  
86 43. Division of Endocrinology, Department of Pediatrics, Boston Children's Hospital, Boston,  
87 Massachusetts, United States of America  
88 44. Broad Institute of Harvard and MIT, Cambridge, Massachusetts, United States of America

## 89 Abstract

90 Genome-wide association studies (GWASs) may help inform treatments for infertility, whose  
91 causes remain unknown in many cases. Here we present GWAS meta-analyses across six  
92 cohorts for male and female infertility in up to 41,200 cases and 687,005 controls. We identified  
93 21 genetic risk loci for infertility ( $P \leq 5E-08$ ), of which 12 have not been reported for any  
94 reproductive condition. We found positive genetic correlations between endometriosis and all-  
95 cause female infertility ( $r_g = 0.585$ ,  $P = 8.98E-14$ ), and between polycystic ovary syndrome and  
96 anovulatory infertility ( $r_g = 0.403$ ,  $P = 2.16E-03$ ). The evolutionary persistence of female infertility-  
97 risk alleles in *EBAG9* may be explained by recent directional selection. We additionally identified  
98 up to 269 genetic loci associated with follicle-stimulating hormone (FSH), luteinising hormone,  
99 oestradiol, and testosterone through sex-specific GWAS meta-analyses (N=6,095-246,862).  
100 While hormone-associated variants near *FSHB* and *ARL14EP* colocalised with signals for  
101 anovulatory infertility, we found no  $r_g$  between female infertility and reproductive hormones  
102 ( $P > 0.05$ ). Exome sequencing analyses in the UK Biobank (N=197,340) revealed that women  
103 carrying testosterone-lowering rare variants in *GPC2* were at higher risk of infertility (OR=2.63,  
104  $P = 1.25E-03$ ). Taken together, our results suggest that while individual genes associated with  
105 hormone regulation may be relevant for fertility, there is limited genetic evidence for correlation  
106 between reproductive hormones and infertility at the population level. We provide the first  
107 comprehensive view of the genetic architecture of infertility across multiple diagnostic criteria in  
108 men and women, and characterise its relationship to other health conditions.

## 109 Introduction

110 Infertility, defined as the inability to achieve pregnancy within 12 months of regular unprotected  
111 sexual intercourse, affects one in six couples across the globe<sup>1</sup>. A range of demographic,  
112 environmental, and genetic factors may drive infertility, including the age-related decline of sperm  
113 and oocyte quality and quantity<sup>2,3</sup>, infectious diseases<sup>4-6</sup>, and rare Mendelian disorders such as  
114 cystic fibrosis<sup>7,8</sup>. However, the exact cause remains undetermined in up to 28% of couples and  
115 40% of women with infertility<sup>9,10</sup>. Given that current treatments such as *in vitro* fertilisation pose  
116 physical, emotional, and financial burdens on couples and healthcare systems<sup>11-14</sup>, a richer  
117 understanding of the biology and pathophysiology of infertility is urgently necessary.

118  
119 Heritable women's reproductive health diseases, particularly endometriosis<sup>15</sup> and polycystic ovary  
120 syndrome (PCOS)<sup>16</sup>, are thought to be responsible for a considerable proportion of female  
121 infertility, with PCOS in particular accounting for up to 80% of cases of anovulatory infertility<sup>17</sup>. It  
122 is hypothesised that sex-hormone dysregulation<sup>18,19</sup> and obesity<sup>20</sup>, which often accompany  
123 reproductive diseases, may be involved in the aetiology of infertility. Yet little is known about the  
124 genetic basis of reproductive hormones and infertility, which are not well-phenotyped in men or  
125 women in large studies<sup>21,22</sup>. Moreover, negative selection against infertility naturally limits the  
126 frequency of risk alleles in the population<sup>23</sup>. Genome-wide association studies (GWASs) have  
127 thus typically queried proxy measures of fertility such as childlessness<sup>24,25</sup>, which may partly arise  
128 from socio-economic and behavioural factors.

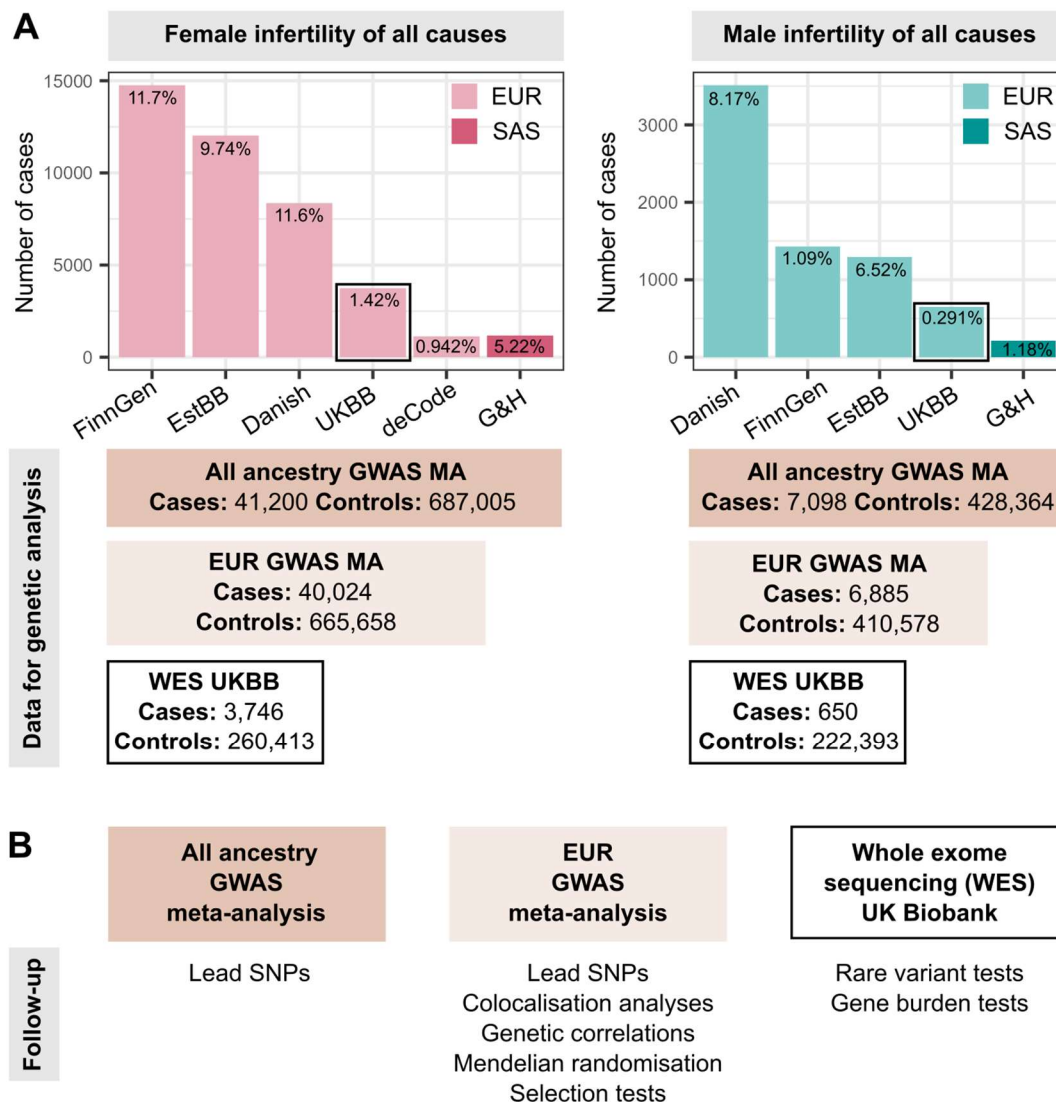
129  
130 We aggregated data from a range of sources, including primary care and hospital electronic health  
131 records (EHRs) and self-report, across six cohorts with over 1 million participants, to perform the  
132 first reported GWAS meta-analyses for male infertility and five categories of female infertility. In  
133 addition, we report results from the largest sex-specific GWASs to date for five reproductive  
134 hormones. By aggregating this data with complementary rare variant genetic association testing  
135 from the UK Biobank, we catalogue the common and rare genetic contributions to infertility and  
136 reproductive hormone levels, quantify the extent of shared genetic architecture between these  
137 traits, and prioritise genes and cell types for further functional investigation of the hormonal and  
138 non-hormonal drivers of infertility.

## 139 Results

### 140 Genome-wide meta-analyses identify novel genetic loci for female 141 and male infertility

142 We identified female infertility of all causes (F-ALL), anatomical causes (F-ANAT), anovulation  
143 (F-ANOV), unknown causes, i.e., idiopathic infertility as defined by exclusion of known causes of  
144 infertility (anatomical or anovulatory causes, PCOS, endometriosis, or uterine leiomyomas) (F-  
145 EXCL), or idiopathic infertility defined by inclusion of diagnostic codes for idiopathic infertility (F-  
146 INCL), as well as male infertility of all causes (M-ALL) in six cohorts, primarily of European

147 ancestry (Figure 1 and Supp. Tables 1 and 2). The case-control ratio of all-cause female infertility  
 148 ranged from 0.9% in the deCODE Genetics dataset<sup>26</sup> to 11.7% in FinnGen<sup>27</sup>, whereas the case-  
 149 control ratio of male infertility was between 0.3% (UKBB) and 8.2% (Danish Biobank) (Figure 1  
 150 and Supp. Table 2). Anatomical female infertility was the least common cause of infertility in three  
 151 of six cohorts (prevalence in UKBB=0.01%, EstBB=2.0%, FinnGen=0.8%). Due to varying sample  
 152 ascertainment, the case-control ratio does not necessarily reflect the population prevalence of  
 153 infertility.  
 154



155  
 156 **Figure 1. Overview of study cohorts and analyses presented for infertility genetic association**  
 157 **studies.** (A) Case numbers in each cohort contributing cases to genome-wide association study (GWAS)  
 158 meta-analyses (MA) for female (left) and male (right) infertility. The prevalence of all-cause infertility in each  
 159 cohort (%) is noted on the barplots. EUR=European ancestry, SAS=South Asian ancestry. EstBB=Estonian  
 160 Biobank, Danish=Danish Blood Donor Study/Copenhagen Hospital Biobank, UKBB=UK Biobank,  
 161 G&H=Genes and Health cohort. Total case and control counts for each type of genetic analysis: all ancestry  
 162 GWAS meta-analysis (dark rectangles), EUR-only GWAS meta-analysis (light rectangles), and UK Biobank  
 163 whole exome sequencing (WES) analyses (black outlined rectangles) are displayed. Male infertility in



164 deCode, with <100 cases, was excluded from GWAS MA. Note the different Y-axis scales in each subplot.  
165 (B) Downstream analyses performed for each type of genetic analysis: lead variants were identified via  
166 distance-based pruning for all-ancestry and EUR-only GWAS meta-analyses; colocalisation, genetic  
167 correlation, and selection analyses were only performed for EUR meta-analyses due to the need for  
168 ancestry-matched linkage disequilibrium (LD) information; rare variant and gene burden tests were  
169 performed with WES data for the UK Biobank EUR-ancestry subset.

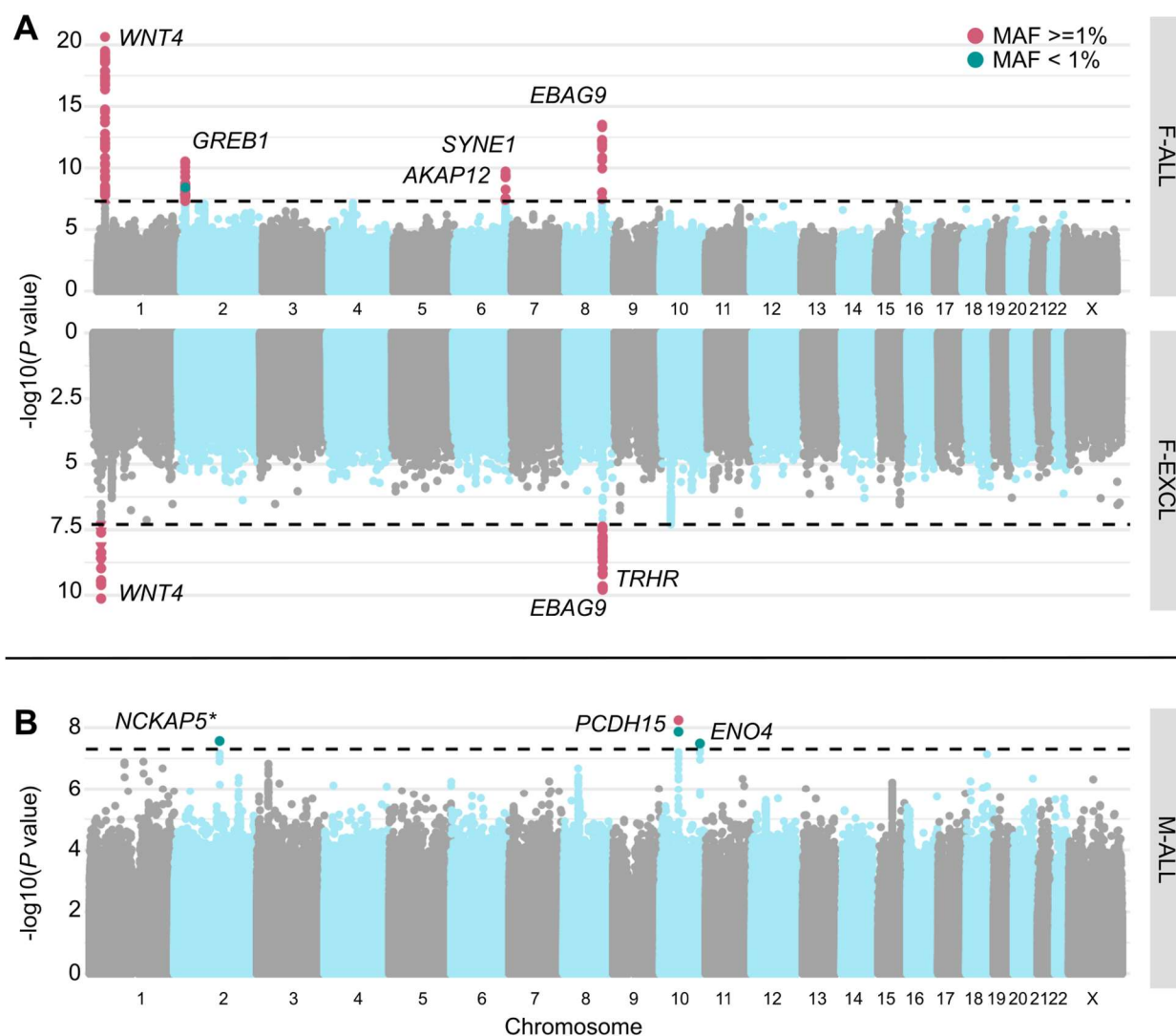
## 170 Novel genetic loci for infertility

171 We performed GWAS meta-analyses, testing up to 28.4 million genetic variants for associations  
172 with each of the above categories of infertility, in up to 41,200 cases/687,005 controls in women,  
173 and 7,098 cases/428,364 controls in men (Figure 1 and Supp. Table 2). We identified 19 unique  
174 genome-wide significant (GWS,  $P < 5E-8$ ) loci associated with at least one category of female  
175 infertility and two loci for male infertility (minor allele frequency (MAF) range 0.24%-46%, lead  
176 variants reported in at least two cohorts) (Figure 2, Table 1, and Supp. Figure 1). There was no  
177 evidence for heterogeneity in lead variant effects across cohorts (Supp. Text).

178  
179 Among the variants associated with multiple subtypes of female infertility is rs1964514, an intronic  
180 variant in *PKHD1L1* (OR (95% CI) for F-ALL=1.13 (1.09-1.16), F-EXCL=1.13 (1.09-1.17), F-  
181 INCL=1.18 (1.11-1.25)). This variant is 76 kb upstream of *EBAG9*, an oestrogen-responsive gene  
182 previously reported to have a recessive association with female infertility<sup>28</sup> and thought to  
183 suppress maternal immune response during pregnancy<sup>29,30</sup>. We also identified an intronic variant  
184 in *WNT4*, rs61768001, associated with three categories of female infertility (F-ALL=0.909 (0.891-  
185 0.927), F-EXCL=0.923 (0.902-0.946), F-INCL=0.870 (0.839-0.903)). *WNT4* is highly pleiotropic  
186 for female reproductive traits, as it is reported to associate with gestational length<sup>31</sup>, uterine  
187 fibroids<sup>32,33</sup>, endometriosis<sup>34,35</sup>, female genital prolapse<sup>27</sup>, and bilateral oophorectomy<sup>27</sup>. Such  
188 pleiotropy is expected, as *WNT4* is a key regulator of female reproductive organ development in  
189 embryogenesis<sup>36-38</sup>.

190  
191 The nearest gene to the idiopathic infertility-associated variant rs111597692 (F-EXCL OR=1.16  
192 (1.10-1.22)) is *TRHR*, which encodes the thyrotropin-releasing hormone receptor. Mice with  
193 *TRHR* knockouts display a phenotype similar to primary ovarian insufficiency<sup>39,40</sup>. The F-ANOV  
194 associated variant rs72827480 (OR=0.905 (0.873-0.938)) colocalises with a testis-eQTL for  
195 *INHBB* in the GTEx Project<sup>41</sup> (posterior probability (PP) of shared causal variant=91.6%) (Supp.  
196 Table 4). *INHBB* encodes the beta subunit of inhibin B, which regulates hypothalamic, pituitary,  
197 and gonadal hormone secretion<sup>42</sup>, and ovarian follicle and oocyte development<sup>43</sup>.

198  
199 Finally, an intronic variant in *ENO4*, which is expressed in the testis and may play a role in sperm  
200 motility<sup>44</sup>, is associated with male infertility (rs139862664, OR=0.388 (0.277-0.543)). Male mice  
201 with *ENO4* knockouts display infertility, abnormal sperm morphology and physiology, and  
202 decreased testis weight, among other altered male reproductive tract phenotypes<sup>45</sup>.



203  
204 **Figure 2. Miami and Manhattan plots for selected infertility meta-analyses.** (A) Genetic variants  
205 associated with female infertility of all causes (F-ALL) (top) and idiopathic infertility (unknown causes)  
206 defined by exclusion of known causes such as anatomical or anovulatory causes, PCOS, endometriosis,  
207 or uterine leiomyomas (bottom). (B) Genetic variants associated with male infertility of all causes (M-ALL).  
208 Each point depicts a single SNP, with genome-wide significant (GWS) SNPs ( $P < 5E-08$ , dashed line)  
209 coloured in pink for common variants with minor allele frequency (MAF)  $\geq$  1% and green for those with  
210 MAF < 1%. SNPs are annotated with the mapped gene. \* indicates that lead variant is reported in only one  
211 cohort.

**Table 1. Lead variants associated with infertility in GWAS meta-analyses.** A1 is the effect allele. \*lead variant is reported in only one cohort.

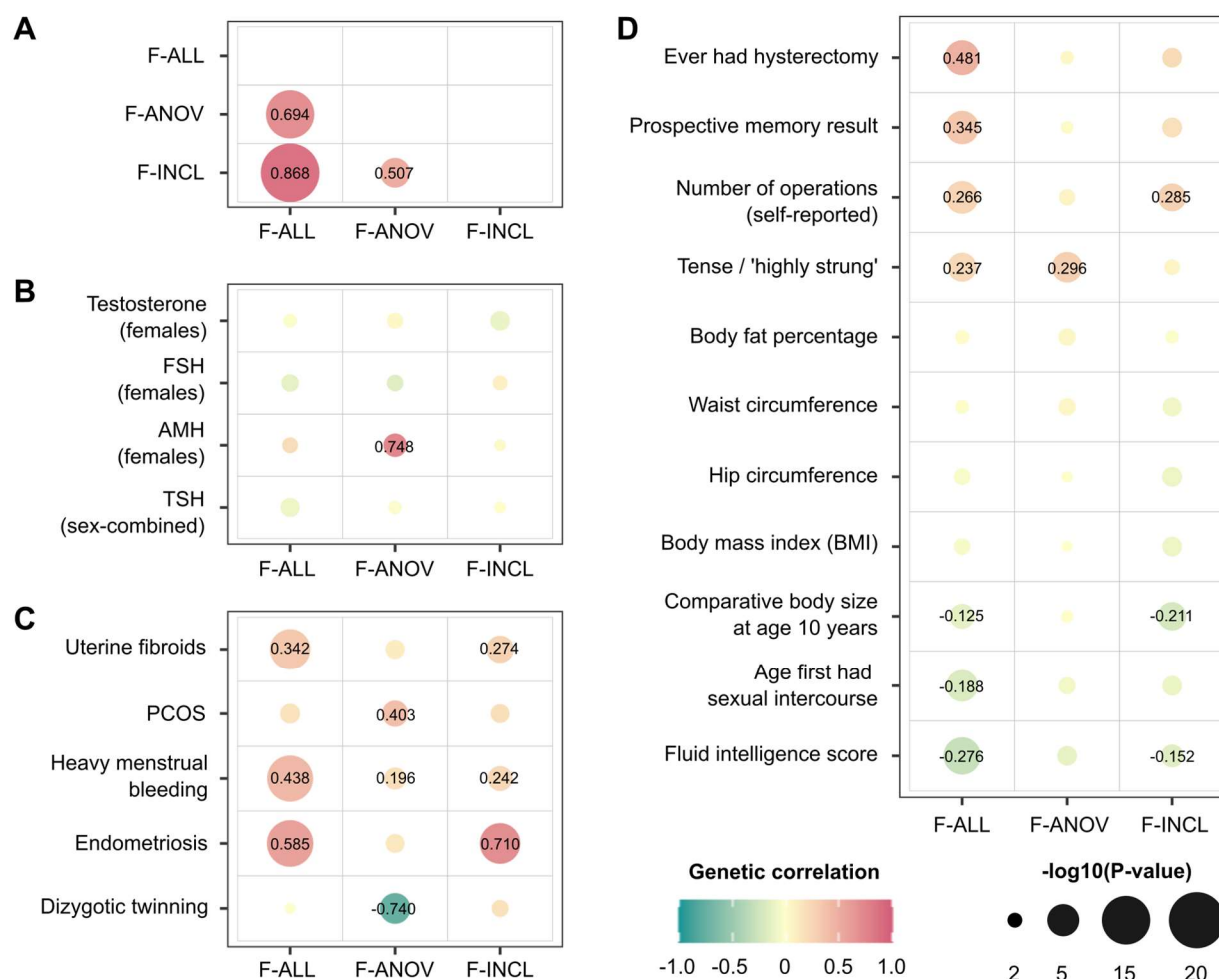
RSID	chr:pos:A1:A2 (hg38)	Mapped gene	All ancestries			EUR only		
			Average MAF	OR (95% CI)	P-value	Average MAF	OR (95% CI)	P-value
<b>Female infertility of all causes (F-ALL)</b>								
rs61768001	chr1:22139327:T:C	<i>WNT4</i>	0.166	0.909 (0.891-0.927)	2.25E-21	0.163	0.911 (0.893-0.93)	1.24E-19
rs10200851	chr2:11581956:T:C	<i>GREB1</i>	0.458	0.951 (0.937-0.965)	2.90E-11	0.456	0.951 (0.936-0.965)	5.84E-11
rs6938404	chr6:151222906:T:C	<i>AKAP12</i>	0.453	0.958 (0.943-0.973)	3.88E-08	0.453	0.958 (0.943-0.973)	3.88E-08
rs17803970	chr6:152232583:A:T	<i>SYNE1</i>	0.0836	1.09 (1.06-1.12)	1.91E-10	0.0836	1.10 (1.07-1.13)	7.50E-11
rs1964514	chr8:109463457:C:G	<i>EBAG9</i>	0.0595	1.13 (1.09-1.16)	3.01E-14	0.0597	1.13 (1.09-1.16)	6.68E-14
<b>Anatomical female infertility (F-ANAT)</b>								
rs340879	chr1:213983171:T:C	<i>PROX1</i>	0.418	0.906 (0.874-0.939)	4.95E-08	0.418	0.902 (0.869-0.936)	5.06E-08
<b>Anovulatory female infertility (F-ANOV)</b>								
rs72665317	chr1:22040580:T:G	<i>CDC42</i>	0.190	0.875 (0.839-0.913)	7.76E-10	0.18	0.886 (0.847-0.927)	1.45E-07
rs72827480	chr2:120388925:T:C	<i>INHBB</i>	0.401	0.905 (0.873-0.938)	4.20E-08	0.401	0.905 (0.873-0.938)	4.20E-08
rs1852684	chr2:145068818:T:G	<i>ZEB2</i>	0.367	1.12 (1.08-1.16)	9.25E-10	0.35	1.12 (1.08-1.17)	3.44E-10
rs552953683	chr8:102898586:T:C	<i>AZIN1</i>	0.0024	0.341 (0.234-0.498)	2.54E-08	0.0024	0.341 (0.234-0.498)	2.54E-08
rs9696009	chr9:123856954:A:G	<i>DENND1A</i>	0.0777	1.21 (1.14-1.29)	6.87E-10	0.0695	1.24 (1.16-1.32)	2.40E-10
rs9902027	chr17:7537667:T:C	<i>TNFSF12</i>	0.255	0.895 (0.86-0.931)	4.06E-08	0.255	0.895 (0.86-0.931)	4.06E-08
rs143459581	chr22:28068862:T:C	<i>PITPNB</i>	0.0419	1.30 (1.19-1.43)	1.21E-08	0.0419	1.30 (1.19-1.43)	1.21E-08
rs17879961	chr22:28725099:A:G	<i>CHEK2</i>	0.0389	0.739 (0.673-0.811)	1.55E-10	0.0389	0.739 (0.673-0.811)	1.55E-10
<b>Idiopathic female infertility, exclusion definition (F-EXCL)</b>								
rs61768001	chr1:22139327:T:C	<i>WNT4</i>	0.165	0.923 (0.902-0.946)	7.49E-11	0.162	0.928 (0.906-0.951)	2.48E-09
rs111597692	chr8:109039973:T:C	<i>TRHR</i>	0.0323	1.16 (1.10-1.22)	1.51E-08	0.0323	1.16 (1.1-1.22)	1.51E-08
rs17378154	chr8:109568721:A:G	<i>EBAG9</i>	0.059	1.13 (1.09-1.17)	1.64E-10	0.0593	1.13 (1.09-1.17)	3.36E-10
<b>Idiopathic female infertility, inclusion definition (F-INCL)</b>								
rs61768001	chr1:22139327:T:C	<i>WNT4</i>	0.170	0.87 (0.839-0.903)	6.87E-14	0.165	0.872 (0.840-0.905)	8.96E-13
rs11692588	chr2:11544358:A:G	<i>GREB1</i>	0.358	0.919 (0.892-0.947)	2.98E-08	0.358	0.919 (0.892-0.947)	2.98E-08
rs190290095	chr4:39786858:A:G	<i>UBE2K</i>	0.0022	0.227 (0.137-0.375)	7.60E-09	0.0022	0.227 (0.137-0.375)	7.60E-09
rs851982	chr6:151703850:T:C	<i>ESR1</i>	0.428	0.921 (0.895-0.947)	7.60E-09	0.437	0.922 (0.896-0.949)	2.86E-08
rs17378154	chr8:109568721:A:G	<i>EBAG9</i>	0.0565	1.18 (1.11-1.25)	2.47E-08	0.0569	1.18 (1.11-1.25)	4.97E-08
rs74156208	chr10:61509370:A:G	<i>TMEM26</i>	0.184	1.10 (1.06-1.14)	4.96E-08	0.187	1.10 (1.07-1.15)	5.44E-08
<b>Male infertility of all causes (M-ALL)</b>								
rs1228269928*	chr2:132923776:A:T	<i>NCKAP5</i>	0.0006	0.0995 (0.0441-0.224)	2.72E-08	0.0006	0.0995 (0.0441-0.224)	2.72E-08
rs150639836	chr10:53879806:T:C	<i>PCDH15</i>	0.0109	0.505 (0.402-0.636)	5.72E-09	0.0109	0.505 (0.402-0.636)	5.72E-09
rs139862664	chr10:116879589:C:G	<i>ENO4</i>	0.0072	0.388 (0.277-0.543)	3.29E-08	0.0072	0.388 (0.277-0.543)	3.29E-08



## 215 Genetic relationships between infertility and female reproductive conditions

216 Genome-wide, we observed positive genetic correlation between endometriosis and F-ALL ( $r_g$   
 217 (SE)=0.585 (0.0785),  $P=8.98E-14$ ) and F-INCL ( $r_g=0.710$  (0.115),  $P=5.94E-10$ ). We also  
 218 observed positive correlation between F-ANOV and PCOS, the most common cause of  
 219 anovulatory infertility ( $r_g=0.403$  (0.131),  $P=2.20E-3$ ), and negative correlation between F-ANOV  
 220 and spontaneous dizygotic twinning, a heritable metric of female fecundity that captures the  
 221 propensity for multiple ovulation<sup>46</sup> ( $r_g=-0.740$  (0.182),  $P=4.93E-05$ ).  
 222

223 Two loci associated with both endometriosis and female infertility - *WNT4* and *ESR1* - may share  
 224 the same putative causal variant (PP>93.6%, Supp. Table 5). Variants in both these genes have  
 225 previously been associated with endometriosis-related infertility<sup>47-50</sup>. While *GREB1* and *SYNE1*  
 226 also contain overlapping signals for infertility and endometriosis, there is strong evidence against  
 227 shared causal variants (PP>75%, Supp. Table 5). Finally, three of eight loci for anovulatory  
 228 infertility - *INHBB*, *PITPNB*, and *CHEK2* - may share a causal variant with PCOS (PP>89.2%,  
 229 Supp. Table 5).  
 230



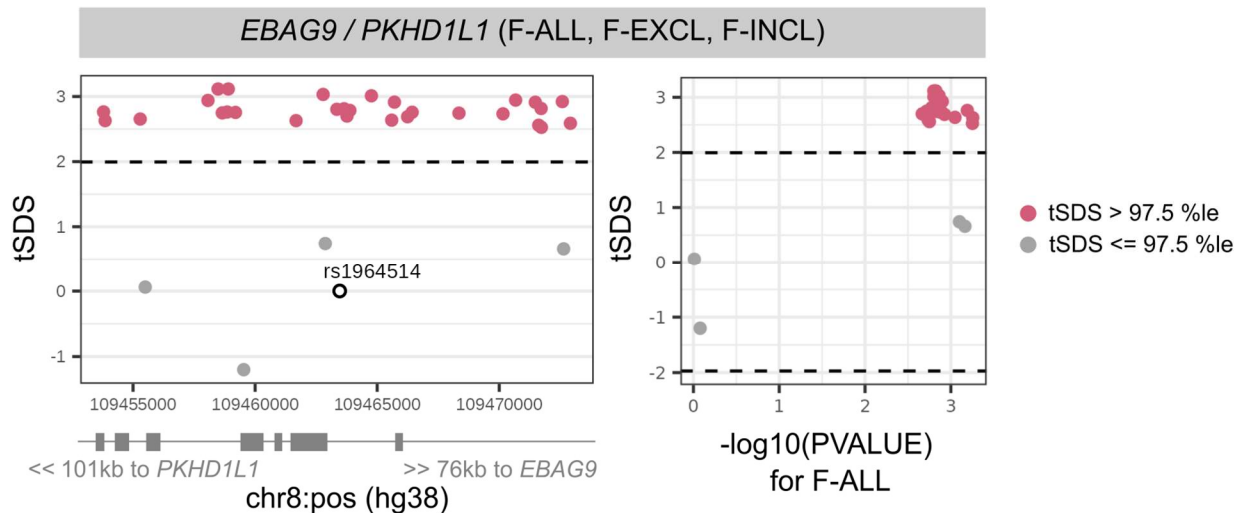
231

232 **Figure 3. Genetic correlations between female infertility and other phenotypes.** SNP-based genetic  
233 correlations ( $r_g$ ) between significantly heritable phenotypes ( $Z>4$ ) were estimated using LD-score  
234 regression, performed using the LDSC software<sup>51</sup> on a subset of 1 million HapMap3 SNPs<sup>52</sup>. Points are  
235 coloured by  $r_g$  estimate, scaled by significance ( $-\log_{10}(P)$ ), and labelled with the associated  $r_g$  estimate if  
236 nominally significant without correction for multiple testing ( $P<0.05$ ). (A) Genetic correlations among the  
237 three significantly heritable definitions of female infertility (all cause=F-ALL, anovulatory=F-ANOV, and  
238 idiopathic infertility defined by inclusion=F-INCL). (B) Genetic correlations between female infertility traits  
239 and reproductive hormones: testosterone, follicle stimulating hormone (FSH), and anti-Mullerian hormone  
240 (AMH, publicly available summary statistics) in female-specific analyses, and thyroid stimulating hormone  
241 (TSH, publicly available summary statistics) from sex-combined analysis. (C) Genetic correlations between  
242 female infertility traits and female reproductive conditions, with summary statistics generated from the  
243 largest available European-ancestry studies for each trait (see Methods). PCOS=polycystic ovary  
244 syndrome. (D) Genetic correlations between female infertility traits and selected heritable phenotypes ( $Z>4$ )  
245 in the UK Biobank, as generated by the Neale lab<sup>53</sup>. Correlations with all heritable phenotypes can be found  
246 in Supp. Table 12.

## 247 Selection pressure may explain the persistence of some infertility- 248 associated variants in the population

249 The genome-wide SNP heritability estimates (on the liability scale, accounting for disease  
250 prevalence<sup>54</sup>) for all categories of infertility are  $<10\%$  (lowest for M-ALL at 1.12% (SE=0.93) and  
251 highest for F-ANOV at 9.54% (2.16)) (Supp. Table 6). This is lower than heritability estimates of  
252 two-thirds of all heritable binary phenotypes in the UK Biobank with population prevalence similar  
253 to that of infertility (64 phenotypes with  $Z>4$  and prevalence  $<5\%$ )<sup>53</sup>. We hypothesised that  
254 infertility risk-increasing alleles are subject to negative selection<sup>55</sup>, so we tested whether there  
255 was evidence for: (i) variants associated with infertility in loci under historical or recent directional  
256 selection<sup>56-58</sup>, or (ii) recent directional selection (over the last 2,000 to 3,000 years) measured by  
257 singleton density scores (SDSs)<sup>56</sup> and balancing selection measured by standardised BetaScan2  
258 scores (StdB2)<sup>59</sup> at infertility loci.

259  
260 While we found no genome-wide signature of directional selection against infertility (Supp. Text),  
261 we observed extreme SDSs (in the highest 99.75<sup>th</sup> percentile (%ile) of SNPs within 10kb of a  
262 GWAS Catalog variant) at the *EBAG9* locus associated with female infertility, indicating recent  
263 positive selection (Figure 4 and Supp. Table 7). *EBAG9* is associated with infectious response  
264 phenotypes, suggesting that the locus may be under selection for its effects on the immune  
265 system. We additionally observed signatures of balancing selection, which maintains multiple  
266 alleles in the population through mechanisms such as heterozygote advantage or time-varying  
267 fitness<sup>60,61</sup>, at the female infertility loci *GREB1* (StdB2 in the 98.6<sup>th</sup>-99.4<sup>th</sup> %ile of SNPs within 10kb  
268 of a GWAS Catalog variant) and *INHBB* (98.5<sup>th</sup> %ile), and the male infertility locus *PCDH15* (98.7<sup>th</sup>  
269 %ile); however, variants at these loci with high probability of association with infertility did not  
270 have high balancing selection scores (Supp. Figure 2 and Supp. Table 7).

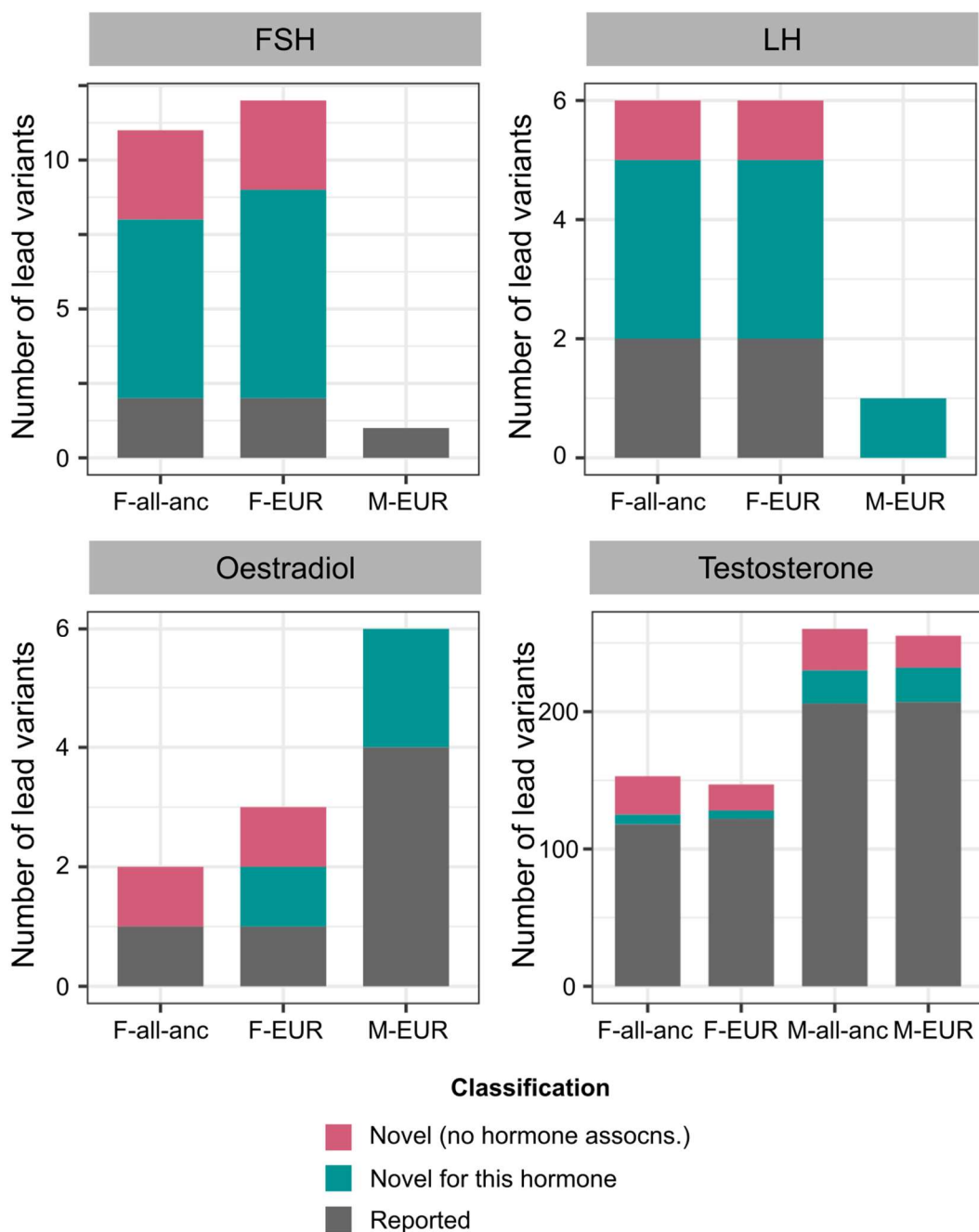


271  
 272 **Figure 4. Directional selection scores at infertility-associated *EBAG9* locus.** Recent directional  
 273 selection, as measured by trait-aligned Singleton Density Scores (tSDSs) at the *EBAG9* locus. The window  
 274 of +/- 10 kb around the lead variant associated with female infertility of all causes (F-ALL) is displayed,  
 275 along with the location of nearest gene transcription start sites (TSSs). The tSDSs are aligned to the  
 276 infertility-risk increasing allele, wherein a positive tSDS indicates positive selection for infertility-risk  
 277 increasing allele at the locus. Dashed lines indicate 2.5th percentile (%ile) and 97.5th %ile of SDSs, and  
 278 variants below or above this threshold respectively are coloured in pink. Left: Locus plots depicting genomic  
 279 position on the x-axis and tSDS on the y-axis. The lead variant rs1964514 (open circle) is not present in  
 280 the tSDS dataset and thus assigned a score of 0. Right: Scatter plots depicting relationship between  $-\log_{10}(\text{PVALUE})$   
 281 of the GWAS p-value for the variant association with F-ALL on the x-axis and tSDS on the y-axis.

## 282 Genetic determinants of reproductive hormone levels

### 283 Identification of novel reproductive hormone loci

284 As hormone dysregulation is central to many infertility diagnoses<sup>18,19</sup>, we conducted sex-specific  
 285 GWAS meta-analyses of five reproductive hormones - follicle-stimulating hormone (FSH)  
 286 ( $N_{\text{female}}=57,890$ ,  $N_{\text{male}}=6,095$ ), luteinising hormone (LH) ( $N_{\text{female}}=47,986$ ,  $N_{\text{male}}=6,769$ ), oestradiol  
 287 ( $N_{\text{female}}=97,887$ ,  $N_{\text{male}}=39,165$ ), progesterone ( $N_{\text{female}}=18,368$ ), and total testosterone  
 288 ( $N_{\text{female}}=246,862$ ,  $N_{\text{male}}=243,951$ ) - collected at assessment centre visits or identified through  
 289 EHRs, in six cohorts and publicly available summary statistics (Supp. Table 9). We identified GWS  
 290 loci associated with FSH (9 novel/2 previously known in females (F) and 0/1 in males (M)), LH  
 291 (4/2 in F and 1/0 in M), oestradiol (1/1 in F and 2/4 in M), and testosterone (35/118 in F and 55/206  
 292 in M), but found no genetic variants associated with progesterone (Figure 5, Supp. Figure 3, and  
 293 Supp. Figure 4). Several of the reported signals we replicated are near genes encoding the  
 294 hormone-specific subunits themselves, such as *FSHB* for FSH and *LHB* for LH, or enzymes for  
 295 steroid hormone metabolism, such as *CYP3A7* for oestradiol and *HSD17B13* for testosterone  
 296 (Supp. Text).



297  
 298 **Figure 5. Number of novel and reported reproductive hormone associations.** Each panel displays a  
 299 different hormone (FSH=follicle-stimulating hormone, LH=luteinising hormone). Lead variants in each  
 300 analysis stratum (F=female-specific, M=male-specific, all-anc=all ancestry meta-analysis, EUR=European-  
 301 only meta-analysis) are classified as: (1) novel (no hormone associations) if they are not in LD ( $r^2 < 0.1$ ) with,  
 302 and conditionally independent of (conditional  $P$ -value  $P_{cond} < 0.05$ ), any variants within a 1Mb window of the  
 303 lead variant that are associated with 28 reproductive hormones in the GWAS Catalog<sup>62</sup>, plotted in pink, (2)  
 304 novel for this hormone if they are not in LD ( $r^2 < 0.1$ ) with, and conditionally independent of ( $P_{cond} < 0.05$ ), the  
 305 respective hormone-associated variants within a 1Mb window of the lead variant, plotted in green, and (3)  
 306 reported otherwise, plotted in grey. Note the different Y-axis scales in each subplot. assocns.=associations.

307 We further classified lead variants as entirely novel hormone associations by defining a protocol  
308 based on linkage disequilibrium (LD) and conditional independence from published SNPs  
309 associated with any of 28 reproductive hormones in the GWAS Catalog<sup>62</sup> (see Methods for  
310 detailed classification protocol).

311  
312 We found 39 novel variants for testosterone in men, including those near *SPOCK1* (rs1073917:  
313  $\beta$  (SE)=-0.0160 (0.0029),  $P=4.69E-08$ ), which is a target for the androgen receptor<sup>63</sup>, *NR4A3*  
314 (rs10988865:  $\beta=-0.0161$  (0.0029),  $P=4.33E-08$ ), which coordinates the cellular response to  
315 corticotropin-hormone and thyrotropin-hormone releasing stimuli<sup>64,65</sup> and regulates  
316 adipogenesis<sup>66</sup>, and obesity-associated genes *ANKS1B* (rs144998814:  $\beta=0.133$  (0.0162),  
317  $P=2.34E-16$ ) and *ANO10* (rs6809522:  $\beta=0.016$  (0.0029),  $P=3.00E-08$ )<sup>67</sup> (Supp. Table 10). The 28  
318 novel reproductive hormone variants associated with testosterone in women include those near  
319 *LAMTOR4* (rs17250196:  $\beta=-0.131$  (0.0067),  $P=4.02E-86$ ), associated with hyperthyroidism<sup>39</sup> and  
320 age at menarche and menopause<sup>68</sup>, obesity-associated *CCDC146* (rs138240474:  $\beta=-0.116$   
321 (0.0207),  $P=2.03E-08$ )<sup>67</sup>, which is also expressed in the fallopian tubes and endometrium<sup>69,70</sup>, and  
322 *SLC8A1* (rs12611602:  $\beta=0.0163$  (0.003),  $P=3.79E-08$ ), which causes increased pancreatic beta  
323 cell proliferation and insulin secretion in knockout mouse models<sup>71</sup>. Finally, we report lead SNPs  
324 independent of previously published hormone variants in the *HELQ* locus for FSH-F (rs4235062:  
325  $\beta=-0.046$  (0.0065),  $P=1.50E-12$ ), *TMEM150B* locus for FSH-F (rs28875253:  $\beta=-0.0599$  (0.0061),  
326  $P=9.90E-23$ ) and LH-F (rs11668309:  $\beta=0.0519$  (0.0071),  $P=3.91E-13$ ), and in the *SOX15-SAT2*  
327 locus for oestradiol-F (rs3933469:  $\beta=0.0363$  (0.0051),  $P=1.02E-12$ ) (Supp. Table 10).

328  
329 Our results were robust to the inclusion of summary statistics from publicly available datasets,  
330 and there was no evidence for heterogeneity in variant effects across cohorts (Supp. Text).

### 331 Sex-specific genetic architecture of testosterone

332 Only 9.80% (of 153 total) lead variants for testosterone in females and 5.75% (of 261 total) lead  
333 variants for testosterone in males reach GWS in both sexes; and 45.9% of variants have opposing  
334 directions of effect in men and women (Supp. Figure 6). Indeed, we found no significant genetic  
335 correlation between testosterone in men and women ( $r_g$  (SE)=0.0361 (0.0428),  $P=0.399$ ). The  
336 heritability of testosterone in women is enriched in the adrenal gland ( $P=1.03E-03$ ) and  
337 hepatocytes ( $P=9.36E-04$ ); but only the latter is enriched for the heritability of testosterone in men  
338 ( $P=3.61E-04$ ), as is the liver more broadly ( $P=1.16E-06$ ) (Supp. Figure 10, stratified LD-score  
339 regression performed across 205 tissues and cell-types from the Genotype Tissue Expression  
340 (GTEx) Project database<sup>41</sup> and the Franke lab single-cell database<sup>72</sup>). Finally, although  
341 testosterone regulates several traits hypothesised to be under sexual selection and may be under  
342 selection itself<sup>73</sup>, we do not find significant genome-wide directional selection for testosterone in  
343 men or women (mean genome-wide trait-SDS is not significantly different from 0, both  $P>0.05$ )  
344 (Supp. Text).



## 345 Genetic relationships between female infertility, reproductive 346 hormones, and obesity

347 We observed no genome-wide genetic correlations between any category of female infertility and:  
348 (i) any reproductive hormone in this study, or (ii) thyroid stimulating hormone (TSH), or (iii) anti-  
349 Mullerian hormone (AMH), the latter two based on publicly available summary statistics<sup>74,75</sup> (all  
350  $P>0.05$ , Figure 3B). Mendelian randomisation (MR) analyses indicated a genetically causal  
351 protective effect of FSH on risk of F-ALL (OR (95% CI)=0.776 (0.678-0.888),  $P=2.15E-04$ ) and F-  
352 EXCL (0.716 (0.604-0.850),  $P=1.26E-04$ ) (Supp. Table 11).

353  
354 We found evidence for shared variants between hormones and infertility at the *FSHB* locus  
355 associated with FSH, LH, and testosterone (PP>84.8% for colocalisation with F-ANOV), and the  
356 *ARL14EP* locus associated with LH (PP=89.3% for colocalisation with F-ANOV) (Supp. Table 12).  
357 There was no evidence for colocalisation at any of the >300 other GWS loci associated with  
358 infertility or reproductive hormones in our study (Supp. Table 12). Our results suggest that while  
359 these traits are not significantly correlated at a genome-wide level, a small number of genes may  
360 drive infertility linked to hormone dysregulation.

361  
362 Across 703 heritable phenotypes in the UK Biobank, we found 15 traits to be genetically correlated  
363 with female infertility, which we broadly group into: female reproductive conditions (such as having  
364 had a hysterectomy,  $r_g$  (SE)=0.481 (0.0963)), general illness (such as number of operations,  
365  $r_g=0.266$  (0.0588)), and cognitive test results (overall prospective memory test  $r_g=0.345$  (0.0736),  
366 overall fluid intelligence  $r_g=-0.276$  (0.0502)) (Figure 3D and Supp. Table 13). 24 obesity-related  
367 traits, including body mass index (BMI), waist-to-hip ratio (WHR), and body fat percentage, are  
368 correlated with testosterone and FSH, but are not genetically correlated with any category of  
369 female infertility (all  $P>0.05$ , Figure 3D, Supp. Figure 7, and Supp. Table 13). However, MR  
370 analyses using genetic instruments for BMI, WHR, and WHR adjusted for BMI (WHRadjBMI)<sup>67</sup>  
371 indicated evidence for bi-directional causal relationships between infertility and abdominal obesity  
372 independent of overall obesity. While genetically predicted WHRadjBMI is a risk factor for F-ALL  
373 (OR (95% CI)=1.10 (1.05-1.16),  $P=1.71E-04$ ) and F-ANOV (1.29 (1.16-1.45),  $P=4.66E-06$ ), the  
374 latter is itself causal for increased WHRadjBMI ( $\beta$  (SE)=0.0547 (0.0133),  $P=3.74E-05$ ) (Supp.  
375 Table 11).

376  
377 Variants associated with all-cause female infertility are in genes enriched for expression in ovarian  
378 stromal cells (partitioned heritability  $P=2.52E-03$ ). We did not find significant enrichment of  
379 infertility heritability in any of the 205 tissues and cell-types from the GTEx project database<sup>41</sup> and  
380 the Franke lab single-cell database<sup>72</sup>.

## 381 Rare variant contribution to reproductive-hormone and infertility 382 genetics

383 We analysed the 450k UK Biobank exome sequencing dataset to characterise the association  
384 between rare coding variation (MAF<1%) and binary traits with >100 cases (F-ALL (3,746 cases,



385 260,413 controls), F-EXCL (3,012 cases, 261,147 controls), and M-ALL (650 cases, 222,393  
386 controls), and quantitative traits with >10,000 participants (FSH-F (N=20,800), LH-F (N=16,391),  
387 oestradiol-F (N=54,609), and testosterone (N<sub>female</sub>=197,038, N<sub>male</sub>=197,340) (Figure 1)). Gene-  
388 burden analyses implicate the *PLEKHG4* gene, which is highly expressed in the testis and ovary,  
389 for F-EXCL (burden test OR (95% CI)=1.04 (1.02-1.06) when aggregated across all variant  
390 annotations with MAF<1%, Cauchy  $P=1.37E-08$ ) (Supp. Table 14). Rare variants in *PLEKHG4*  
391 cause cerebellar ataxia<sup>76</sup>, which is a feature of some syndromes that also cause steroid hormone  
392 deficiency and hypogonadism<sup>77,78</sup>.

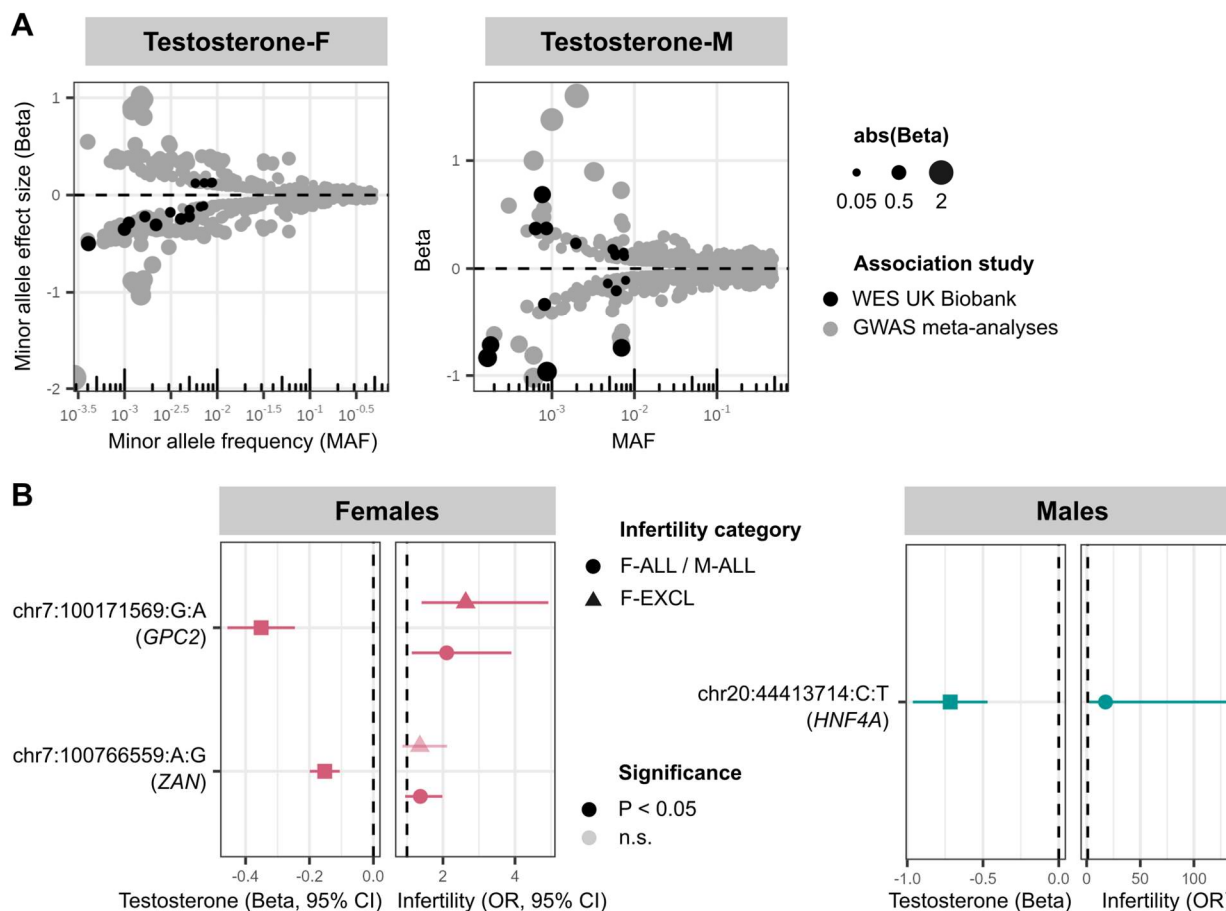
### 393 Novel genes for testosterone implicated by gene burden analyses

394 Gene-based analyses identify 27 genes associated with testosterone-F and 24 genes for  
395 testosterone-M ( $P<5E-06$ ), of which eleven have not previously been implicated in GWASs (Supp.  
396 Text). We report the first known association of *HSD11B1* with testosterone-F (burden test  
397  $P=1.93E-06$  when aggregated across missense variants with MAF<0.01%); pathogenic variants  
398 in this gene are reported to cause hyperandrogenism due to cortisone reductase deficiency<sup>79,80</sup>  
399 (Supp. Figure 11 and Supp. Table 14). We also report the association of testosterone-M with  
400 *HSD17B2* (burden test  $P=1.33E-11$  when aggregated across pLoF variants with MAF<0.1%),  
401 which encodes the enzyme hydroxysteroid 17 $\beta$ -dehydrogenase 2 that catalyses the oxidation of  
402 oestradiol, testosterone, and dihydrotestosterone to less active forms and thus regulates the  
403 biological potency of steroid hormones<sup>81,82</sup> (Supp. Figure 11 and Supp. Table 14).

### 404 Increased risk of infertility in individuals carrying rare testosterone- 405 associated variants

406 Two genes associated with testosterone in female UK Biobank participants are also associated  
407 with infertility risk ( $P<1.00E-03$ , Bonferroni adjustment for 50 unique genes): *TRIM4* (F-ALL,  
408 burden test OR=1.03 (1.01-1.05),  $P=4.05E-04$  across all variants with MAF<0.1%) and *CYP3A43*  
409 (F-EXCL, burden test OR=1.02 (1.01-1.03),  $P=4.84E-04$  across all variants with MAF<1%). The  
410 latter encodes the steroid hormone metabolic enzyme testosterone 6-beta-hydroxylase; but  
411 neither gene has previously been implicated in infertility.

412  
413 Finally, we identified 29 unique genes carrying rare variants (MAF<1%) associated with  
414 testosterone in male or female participants in the UK Biobank. Eighteen of the 29 genes also  
415 contain common testosterone-associated variants from GWASs (MAF>1%), but the rare variant  
416 has a larger absolute effect size in the majority (83%) of these (Figure 6A, Supp. Table 15, and  
417 Supp. Text).



418  
419

420 **Figure 6. Rare variants associated with testosterone and infertility in UK Biobank whole exome**  
 421 **sequencing (WES) analyses.** (A) Effect size versus allele frequency of genetic variants associated with  
 422 total testosterone. Variants discovered at genome-wide significance ( $P < 5E-08$ ) in GWAS meta-analyses  
 423 (coloured in grey) and exome-wide significance in the UK Biobank WES analyses (coloured in black) are  
 424 plotted, sized by the absolute value of their effect size. Effect sizes are aligned to the minor allele, plotted  
 425 against MAF on the log x-axis. (B) Effects of testosterone-associated rare variants on infertility in females  
 426 (left) and males (right). Per gene, the variant with lowest  $P$ -value of all variants that reach exome-wide  
 427 significance ( $P < 1E-07$ ) in UK Biobank WES analyses for testosterone is displayed, for all variants with  
 428 nominally significant effects on infertility. Effect sizes ( $\beta$  and 95% confidence intervals (CIs) for the variant  
 429 effect on testosterone are to the left of each plot, and effect sizes (odds ratios (ORs) and 95% CIs) for the  
 430 variant effect on infertility are to the right of each plot. Variants that reach nominal significance ( $P < 0.05$ ) are  
 431 coloured in solid shapes.

432  
 433 The eleven novel testosterone associations include a female testosterone-lowering missense  
 434 variant in *STAG3* (chr7:100204708:C:T,  $\beta = -0.284$ ,  $P = 2.31E-08$ ); *STAG3* is also associated with  
 435 primary ovarian insufficiency in women<sup>83,84</sup>, and induces female infertility through the absence of  
 436 oocytes in knockout mouse models<sup>39</sup>. While we did not find significant association between the  
 437 *STAG3* variant and female infertility in the UK Biobank ( $P > 0.05$ ), we observed increased risk of  
 438 idiopathic infertility in women carrying a novel testosterone-lowering variant in *GPC2*  
 439 (chr7:100171569:G:A, F-EXCL OR=2.63 (1.40-4.92),  $P = 1.25E-03$ ) (Figure 6B). *GPC2* is highly  
 440 expressed in the testis, and *GPC2*-knockout mouse models display reduced adrenal gland size<sup>39</sup>.

441 The gene has not previously been reported to be associated with testosterone or infertility. Taken  
442 together, our results indicate a potential role for infertility driven by rare hormone-disrupting  
443 variants.

## 444 Discussion

445 Our large-scale genetic investigation of infertility and related reproductive phenotypes in over 1  
446 million individuals identified 19 genetic loci associated with female infertility, two with male  
447 infertility, and novel variants for the reproductive hormones FSH (3 novel variants), LH (1),  
448 oestradiol (1), and total testosterone (28) in women and for total testosterone in men (39). Through  
449 rare-variant and gene-based analyses in the UK Biobank, we additionally identified *PLEKHG4*  
450 associated with female infertility and 50 genes for testosterone, including the first reported  
451 hormone-associated variants in some members of the hydroxysteroid dehydrogenase enzyme  
452 family. We found evidence at non-hormonal, pleiotropic, infertility loci for recent directional  
453 selection (*EBAG9*) and balancing selection (*GREB1*, *INHBB*, *PCDH15*). Although there was  
454 evidence for distinct genetic architectures of infertility and reproductive hormones, we showed  
455 that individual genes containing rare protein-coding variants associated with testosterone (*GPC2*,  
456 *CYP3A43*, *TRIM4*) were also associated with higher risk of infertility in the UK Biobank.

457  
458 Previous efforts to catalogue the genome-wide architecture of infertility have relied on proxy  
459 measures such as childlessness and number of children ever born<sup>24,25</sup>, which may be confounded  
460 by behavioural, socio-economic, and lifestyle factors. While we did find modest genetic correlation  
461 between female infertility and age at first sexual intercourse (-18.8%), indicating that the latter  
462 captures some shared biology with fertility, our meta-analyses did not replicate the associations  
463 of infertility proxy variables with putative behavioural loci for risk-taking<sup>85,86</sup> or educational  
464 attainment<sup>85,87-89</sup>. Instead, we nominate genes with putative roles in both male and female  
465 gonads, such as *TRHR* for ovarian insufficiency<sup>39,40</sup> and *ENO4* for sperm motility<sup>44</sup>.

466  
467 The strong genetic correlation of 71% between idiopathic infertility and endometriosis may  
468 indicate that some proportion of idiopathic cases are due to under-diagnosis of endometriosis,  
469 whose early treatment may prevent future infertility<sup>15,90</sup>. Our subtype-specific analyses highlight  
470 the value in dissecting heterogeneous causes of infertility. For example, PCOS is a heritable  
471 cause of up to 80% of anovulatory infertility cases that may be treated through induced  
472 ovulation<sup>17,91,92</sup>. However, as only three of eight loci for anovulatory infertility colocalise with known  
473 PCOS signals and the genetic correlation between these traits is only 40%, other hypothalamic-  
474 pituitary-ovarian disorders, endocrinopathies (hypothyroidism, hyperprolactinaemia, etc.) and  
475 ovarian insufficiency may also contribute significantly to the genetic aetiology of anovulatory  
476 infertility and require treatments different from those for PCOS-associated infertility<sup>93</sup>. Weight loss  
477 for overweight patients is often recommended as beneficial for fertility<sup>94,95</sup>, but we did not find  
478 substantial genetic correlation between obesity and infertility. Our findings add genetic support to  
479 evidence from randomised controlled trials demonstrating no fertility benefits from short-term  
480 weight loss in overweight and obese women<sup>96</sup>. Instead, we observed bi-directional causal  
481 relationships between abdominal obesity and anovulatory infertility, suggesting physiological  
482 feedback mechanisms whose complex interplay requires deeper study. Taken together, these

483 results suggest a critical need for a richer understanding of the genetic and non-genetic  
484 contributions to infertility.

485

486 The testes and ovaries were not significantly enriched for the heritability of infertility or  
487 testosterone, despite being reproductive organs that are major sites for testosterone  
488 production<sup>97,98</sup>. However, neither organ is disaggregated into tissues or cell types in the GTEx  
489 database, so gene expression profiles may not capture cell-type specific effects. Indeed, we found  
490 enrichment of testosterone heritability in the androgen-secreting thecal cells and androgen-  
491 responsive granulosa cells of the ovary<sup>99-101</sup>, and female infertility in ovarian stromal cells.  
492 Although there are several causal roles hypothesised for stromal dysfunction in infertility, such as  
493 impaired folliculogenesis<sup>102</sup>, restricted blood flow<sup>103</sup>, and ovarian scarring<sup>104</sup>, more work is needed  
494 to robustly replicate these findings. In general, more functional studies of gonadal cell types, in  
495 both men and women, are needed to enable a mechanistic understanding of the genetic variation  
496 associated with reproductive hormones and infertility.

497

498 We employed a broad search strategy to maximise sample sizes for cases of infertility and  
499 reproductive hormone levels in our meta-analyses. Diagnostic criteria for infertility vary by country  
500 and have changed over time<sup>1</sup>, which may explain the wide spread in the prevalence of infertility  
501 across cohorts. Reproductive hormone values in this study were assayed using different  
502 methodologies, in primary care or hospital EHRs, and at different ages and stages of the  
503 menstrual cycle in women. A majority of samples in our study were derived from the UK Biobank  
504 and measured during and post-menopause (ages 40-69), whereas infertility occurs pre-  
505 menopause, so we urge caution in interpreting the lack of correlation between these traits.  
506 Although we were able to adjust for covariates such as age, which can account for some of the  
507 effect of menopause on hormone levels, we did not have the data granularity to account for  
508 hormonal fluctuations during the menstrual cycle and pregnancy. In the future, longitudinal  
509 GWASs that can incorporate mean and variance of hormone levels over the menstrual cycle, or  
510 phenotypes that calculate ratios between various hormones over time, will likely reveal  
511 fundamental biology that is missed by the broad-stroke assessments in this study.

512

513 Our results indicate that balancing selection and recent positive selection at pleiotropic loci may  
514 explain the persistence of genetic factors for infertility. For example, the *EBAG9* locus associated  
515 with female infertility is under directional selection, perhaps because *EBAG9*, which is highly  
516 expressed in CD34-/CD41+/CD42+ megakaryocytes<sup>69,70</sup>, plays a role in T-cell mediated  
517 cytotoxicity as part of the adaptive immune memory response to infection<sup>105</sup>. However, a  
518 complementary role for *EBAG9* may be in the placenta during early pregnancy, where reduction  
519 of *EBAG9* levels is associated with inappropriate activation of the maternal immune system and  
520 results in foetal rejection<sup>106</sup>.

521

522 In conclusion, in this comprehensive large-scale investigation of the genetic determinants of  
523 infertility and reproductive hormones across men and women, we identified several genes  
524 associated with infertility and analysed their effects on reproductive disease and selection  
525 pressures. We did not find evidence that reproductive hormone dysregulation and obesity are  
526 strongly correlated with infertility at the population level, but instead nominate individual hormone-

527 associated genes with effects on fertility. Other genetic and non-genetic avenues must be  
528 explored to treat complex and heterogeneous fertility disorders that impact the physical,  
529 emotional, and financial well-being of millions of individuals across the globe.

## 530 Methods

### 531 Study populations and phenotype identification

#### 532 Binary traits (infertility)

533 Cases were identified in UK Biobank, Copenhagen Hospital Biobank and Danish Blood Donor  
534 Study, deCode, Estonian Biobank, FinnGen, and Genes and Health (Supp. Text). We defined five  
535 categories of female infertility: all causes (F-ALL), anovulatory (F-ANOV), anatomical (F-ANAT,  
536 including tubal, uterine, and cervical origins), idiopathic infertility by exclusion of known causes  
537 (anatomical and anovulatory infertility, PCOS, endometriosis, and uterine leiomyoma) (F-EXCL),  
538 and idiopathic infertility by inclusion of a diagnosis code for idiopathic infertility (F-INCL), and male  
539 infertility of all causes (M-ALL). Cases were identified through self-report (F-ALL, F-EXCL, M-ALL)  
540 and through primary- and secondary-care codes (Supp. Table 1). Within each subtype, sex-  
541 matched controls were defined as individuals not identified as cases for that subtype.

#### 542 Quantitative traits (reproductive hormones)

543 Hormones were included from UK Biobank, Avon Longitudinal Study of Parents and Children  
544 (ALSPAC), deCode, Estonian Biobank, and Genes and Health (Supp. Text). We extracted  
545 measurements of FSH, LH, oestradiol, progesterone, and testosterone from biobank assessment  
546 centres or primary- and secondary-care records (Supp. Table 16). If repeated measurements  
547 were available for an individual, we retained the recorded hormone value closest to the individual's  
548 median hormone value over time. Each hormone was regressed on *age*, *age*<sup>2</sup>, and cohort-specific  
549 covariates specified below; the residuals from this regression were rank-based inverse normally  
550 transformed (RINTed) prior to GWAS.

### 551 Meta-analysis of GWAS summary statistics

#### 552 Genome-wide association testing

553 Association analyses were performed separately within each ancestry and sex stratum for all  
554 strata with at least 100 cases (binary traits) or 1,000 individuals (quantitative traits). For binary  
555 traits, each variant passing QC was tested for association under an additive model using  
556 REGENIE<sup>107</sup> or SAIGE<sup>108</sup>, with adjustments for *age*, *age*<sup>2</sup>, and cohort-specific covariates, with the  
557 Firth correction applied to control for inflation at rare variants and traits with low case-control  
558 ratios<sup>107,108</sup>. For quantitative traits, the RINTed hormone value was tested for association under  
559 an additive model using REGENIE<sup>107</sup> or SAIGE<sup>108</sup>, with adjustments for cohort-specific genetic  
560 covariates. Any deviations from this GWAS protocol are noted in the Supplementary Text.



## 561 Meta-analysis

562 Prior to meta-analysis, summary statistics from all studies underwent thorough quality control to  
563 retain variants that met the following criteria: (1) on the autosomes or X chromosome, (2) with  
564 imputation information score >0.8 (where available), (3) bi-allelic variants with A, C, G, T alleles,  
565 (4) with standard errors <10 and  $P$ -values in [0,1], and (5) without duplicate entries. Fixed-effects  
566 inverse-variance weighted meta-analysis was performed using METAL<sup>109</sup>. We report results from  
567 European-ancestry and all-ancestry meta-analyses for each trait. Genome-wide significance was  
568 established at  $P<5E-08$ .

## 569 Identification and classification of lead variants

570 Distance-based pruning was used to identify lead variants as the SNP with the lowest  $P$ -value  
571 within each 1Mb window at all loci with at least one GWS variant with  $P<5E-08$ .

572  
573 Hormone-associated variants were classified based on conditional analysis as (1) previously  
574 reported for the hormone of interest, (2) previously reported for any of 28 reproductive hormones,  
575 or (3) novel, based on SNP associations published in the GWAS Catalog as of 27 March 2023<sup>62</sup>  
576 (Supp. Table 17). We adapted criteria developed by Benonisdottir *et al.* (2016)<sup>110</sup> to classify novel  
577 variants as those that are not in LD with ( $r^2<0.1$ ), and conditionally independent of  
578 ( $P_{conditional}<0.05$ ), all published hormone-associated variants within 1 Mb; all other variants are  
579 considered to be previously reported. Conditional analysis was performed in GCTA-COJO<sup>111</sup>, with  
580 LD information for European-ancestry individuals derived from the 1000 Genomes dataset<sup>112</sup>.

581  
582 For lead variants on the X chromosome and those from multi-ancestry analyses, for which  
583 estimating LD is more difficult due to differences in recombination rates and selection pressures  
584 between sexes and populations<sup>113–115</sup>, we did not use the above LD-based classification system.  
585 Instead, a lead SNP was considered novel if it was not within 1 Mb of a published hormone-  
586 associated variant or if its effect was independent of published variants within a 1 Mb window  
587 ( $P_{conditional}<0.05$ ), and reported if not.

## 588 SNP-based heritability

589 The following analyses, which rely on population-specific LD patterns, were restricted to  
590 European-ancestry summary statistics with pre-computed LD-scores based on European-  
591 ancestry individuals in the 1000 Genomes dataset<sup>112</sup>, restricted to HapMap3 SNPs<sup>52</sup>. We  
592 estimated the SNP-based heritability ( $h_G^2$ ) of a trait from GWAS summary statistics using LD-  
593 score regression as implemented in the LDSC software<sup>51</sup>. For infertility traits, the observed-scale  
594 heritability ( $h_{obs}^2$ ) was converted to liability-scale heritability ( $h_{liab}^2$ ), which accounts for the disease  
595 prevalence in the sample ( $k$ ) and population ( $K$ ), under the assumption that sample prevalence  
596 equals the population prevalence<sup>54</sup>.



## 597 Genetic correlations

598 LDSC was used to estimate genetic correlations between infertility traits, hormone levels, and a  
599 collection of other phenotypes in the UK Biobank in European-ancestry individuals. To simplify  
600 computation of  $r_g$  across a large number of traits, we used an extension of the LDSC software  
601 which allows for simultaneous estimation of multiple genetic correlations<sup>116</sup>.

602  
603 We estimated genetic correlations among the three categories of female infertility with significant  
604 heritability ( $Z > 4$ )<sup>51</sup>: F-ALL, F-ANOV, and F-INCL, as well as among heritable female reproductive  
605 hormones (FSH and testosterone in females). We additionally obtained summary statistics from  
606 GWASs of thyroid stimulating hormone (TSH)<sup>75</sup> (sex-combined analysis, N=247,107 participants)  
607 and anti-Mullerian hormone (N=7,049 pre-menopausal participants)<sup>74</sup> from the largest publicly  
608 available European-ancestry studies to date. We also tested for genetic correlations between  
609 infertility and reproductive hormones. Significant  $r_g$  after multiple testing was established at 2.38E-  
610 03 (FWER controlled at 5% across 21 tests using the Bonferroni method).

611  
612 We collated European-ancestry GWAS summary statistics for four female reproductive disorders:  
613 (1) endometriosis from Rahmioglu *et al.* (2023)<sup>35</sup>, (57,248 cases and 698,764 controls), (2) heavy  
614 menstrual bleeding by meta-analysing GWAS data from Gallagher *et al.* (2019)<sup>117</sup> and FinnGen  
615 data freeze 9<sup>27</sup> (31,309 cases and 318,510 controls), (3) PCOS by meta-analysing GWAS data  
616 from Tyrmi *et al.* (2022)<sup>92</sup> and a UKBB-based GWAS (14,467 cases and 430,267 controls), and  
617 (4) uterine fibroids by meta-analysing GWAS data generated by the Neale lab<sup>53</sup> and FinnGen data  
618 freeze 9<sup>27</sup>, (42,446 cases and 588,955 controls). We additionally obtained summary statistics  
619 from a GWAS of spontaneous dizygotic (DZ) twinning (8,265 cases (mothers of DZ twins) and  
620 264,567 controls; plus 26252 DZ twins and 417,433 additional controls) from Mbarek *et al.* (2024),  
621 the largest European-ancestry study of female fecundity to date<sup>46</sup>. Significant  $r_g$  after multiple  
622 testing was established at 2.00E-03 (FWER controlled at 5% across 25 tests using the Bonferroni  
623 method).

624  
625 We downloaded LD-score formatted summary statistics for European-ancestry individuals across  
626 703 heritable phenotypes ( $Z > 4$ ) from the Neale lab round 2 collection<sup>53</sup>. The number of effectively  
627 independent phenotypes estimated by the Neale lab ( $M_{eff}=340$ ) was used to establish significant  
628  $r_g$  after multiple testing at 2.45E-05 (FWER controlled at 5% across 2,040 tests using the  
629 Bonferroni method).

## 630 Mendelian randomisation

631 The following analyses were all performed with summary statistics from European-ancestry  
632 GWASs, using the TwoSampleMR v0.5.7 package<sup>118</sup>.

633  
634 We constructed genetic instruments for BMI, WHR, and WHRadjBMI with female-specific lead  
635 variants from a recent European-ancestry GWAS meta-analysis with a maximum sample size of  
636 434,785 female participants<sup>67</sup>. SNPs were weighted by their female-specific effect sizes. The  
637 mean F-statistic across all SNPs in each instrument indicated sufficient strength for MR

638 (BMI=61.3, WHR=74.8, WHRadjBMI=84.7, recommended $>10^{119}$ ). As the instrument GWASs  
639 included participants from UK Biobank, we conducted a sensitivity analysis to avoid bias from  
640 sample overlap between instrument and outcome GWASs by constructing obesity-trait  
641 instruments from an earlier release of summary statistics from the GIANT Consortium without  
642 UKBB participants<sup>120</sup> (Supp. Table 11). As the WHRadjBMI instrument may be confounded due  
643 to adjustment for a correlated variable<sup>121</sup>, i.e. adjustment for BMI in the WHR GWAS, we  
644 performed multivariable MR with a joint instrument for BMI and WHR to estimate the BMI-adjusted  
645 causal effect of WHR on reproductive outcomes. We found no difference in effect estimates from  
646 MR conducted using an instrument for WHRadjBMI and multivariable MR (Supp. Table 19).

647  
648 Hormone instruments were constructed for reproductive hormones in this study with F-statistic $>10$   
649 (FSH-F=38.7, testosterone-F=66.1), using GWAS summary statistics from European-ancestry  
650 GWASs excluding UK Biobank participants to avoid sample overlap with outcome GWASs.

651  
652 We also performed reciprocal MR to test the genetically predicted causal effects of infertility on  
653 obesity and reproductive hormone levels. Genetic instruments were constructed for subtypes of  
654 infertility with F-statistic $>10$  (F-ALL=51.0, F-ANOV=36.2), using GWAS summary statistics from  
655 European-ancestry GWASs excluding UK Biobank participants to avoid sample overlap with  
656 outcome GWASs. We assessed the causal direction between each pair of traits tested with  
657 Steiger filtering of instruments and the Steiger directionality test.

658  
659 We report results from the inverse-variance weighted (IVW) method, the MR-Egger method which  
660 is robust to horizontal pleiotropy<sup>122</sup>, and the weighted median method which protects against  
661 outlier variants<sup>123</sup> (Supp Table 11).

## 662 Colocalisation

663 The following analyses were all performed with summary statistics from European-ancestry  
664 GWASs, using the Bayesian framework implemented in the coloc v5.1.0 package<sup>124</sup> under a  
665 single causal variant assumption<sup>125</sup>. Only common variants (MAF $>1\%$ ) within windows of  $\pm 50$   
666 kb around each lead variant for an infertility or reproductive hormone trait were retained. For each  
667 pair of traits tested for colocalisation, we set the prior probabilities of variants in a locus being  
668 causally associated with trait 1 ( $p_1$ ) and trait 2 ( $p_2$ ) to  $1E-04$  (99% confidence in a true association),  
669 and the prior for joint association  $p_{12}$  to  $1E-06$  (assuming equal likelihood of shared and non-  
670 shared causal variants for each trait in a locus) as recommended by the developers of coloc<sup>125</sup>.  
671 We tested five hypotheses: H0=no association with either trait in region, H1=association with trait  
672 1 in region, but not trait 2, H2=association with trait 2 in region, but not trait 1, H3=association  
673 with both traits in region, but different causal variants, and H4=association with both traits in  
674 region, and a shared causal variant. A pair of traits were considered to colocalise if posterior  
675 probability of H4 $>50\%$  and the ratio of posterior probabilities of H4/H3 $>5^{124,126}$ .

676  
677 We tested for colocalisation between each female infertility category and each female-specific  
678 hormone (FSH, LH, oestradiol, and testosterone) at all genetic loci associated with at least one  
679 of the pair of traits tested. The single male infertility locus with common variants (MAF $>1\%$ ) in the

680 European-ancestry analysis did not contain enough significant associations (only 12 common  
681 variants with  $P < 1E-06$ ) for colocalisation analyses.

682  
683 Because we noticed that some lead variants for female infertility had previously been reported as  
684 associated with endometriosis and PCOS, we estimated the posterior probability (PP) of  
685 colocalisation of genetic signals between each category of female infertility and each of these two  
686 reproductive disorders. European-ancestry summary statistics for endometriosis and PCOS were  
687 obtained as described in the genetic correlations section above.

688  
689 We assessed colocalisation of genetic signals for female infertility with eQTLs for all proximal  
690 genes with transcription start sites (TSSs) within 1 Mb of an infertility lead variant. Publicly  
691 available eQTL data was downloaded from the GTEx project<sup>41</sup>.

## 692 Tissue and cell-type prioritisation

693 We estimated the polygenic contributions of genes with tissue-specific expression profiles to the  
694 heritability of infertility and hormones using stratified LD-score regression (partitioned heritability  
695 analyses)<sup>51</sup>. We restricted these analyses to traits with highly significant heritability in European-  
696 ancestry analyses ( $Z > 7$ ) (F-ALL, testosterone-F, and testosterone-M), as recommended by the  
697 developers, Finucane *et al.* (2015)<sup>127</sup>.

698  
699 Gene sets and LD scores for 205 tissues and cell-types from the GTEx Project database<sup>41</sup> and  
700 the Franke lab single-cell database<sup>72</sup> were downloaded from Finucane *et al.* (2018)<sup>128</sup>. We  
701 established tissue-wide significance at  $-\log_{10}(P) > 2.75$ , which corresponds to  $FDR < 5\%$ .

## 702 Ovarian cell types

703 As the ovary, a reproductive tissue of interest, is not well characterised in the GTEx project, we  
704 identified two publicly available single-cell gene expression datasets for ovarian cell types: (1)  
705 from Fan *et al.* (2019), who performed single-cell RNA sequencing on ovarian tissue from five  
706 adult women undergoing fertility preservation procedures with 20,676 cells across 19 identified  
707 cell types<sup>129</sup>, and (2) from Jin *et al.* (2022), who performed single-nucleus RNA sequencing on  
708 autopsy samples from four women (aged 49-54 years, with normal ovarian histology) with 42,568  
709 cells across 8 identified cell types<sup>130</sup>. The datasets were aligned and filtered using the QC  
710 pipelines provided by the authors of each study, and clustered with identical parameters to  
711 replicate the results of each individual study. Gene sets for each cluster were identified as  
712 recommended by Finucane *et al.* (2018)<sup>128</sup> - briefly, we identified differential expression between  
713 the cells in each cluster and all other clusters by using the Wilcoxon rank sum test implemented  
714 in Seurat v3.0<sup>131-133</sup>, and returned the top 10% of genes that are specifically expressed in each  
715 cluster (positive average log-fold-change values), ranked by differential expression  $P$ -value. We  
716 computed annotation-specific LD scores for these gene sets using hg38 coordinates for gene  
717 TSSs and TESs obtained from Ensembl<sup>134</sup>, across 1 million HapMap3 variants<sup>52</sup> with LD  
718 information from European-ancestry individuals in the 1000 Genomes phase 3 dataset<sup>112</sup>.

## 719 Overlaps with genetic regions under selection

720 To avoid confounding by population stratification, selection look-ups were restricted to GWAS  
721 summary statistics from European-ancestry individuals.

## 722 Directional selection

723 Following guidelines described by Mathieson *et al.* (2023)<sup>25</sup>, we identified 54 genomic regions  
724 under directional selection from three previously reported genome-wide scans: (1) 39 regions  
725 from the Composite of Multiple Signals (CMS) test, which infers historical selection on the order  
726 of the past 50,000 years<sup>58</sup>, (2) 12 regions from an ancient DNA scan that uses inferences of allele  
727 frequency from ancient genomes to determine selection over the past 10,000 years<sup>57</sup>, and (3)  
728 three regions from Singleton Density Scores (SDSs), which use the pattern of singleton variants  
729 to identify recent selection in the past 2,000 to 3,000 years<sup>56</sup>. For each genomic window under  
730 directional selection, we report the infertility-associated variants with the lowest *P*-value.

## 731 Singleton density scores

732 We downloaded publicly available SDSs for SNPs in the UK10K dataset<sup>56</sup> to report the highest  
733 SDS (positive selection of derived allele over ancestral allele in the past 2,000 to 3,000 years)  
734 and lowest SDS (negative selection) within the +/-10kb window around each infertility or hormone  
735 lead SNP. To calculate trait-SDS for each phenotype, we aligned each SDS to the trait-increasing  
736 allele rather than the derived allele<sup>56</sup>. For each lead variant window containing variants with  
737 extreme SDSs (top 97.5th %ile or bottom 2.5th %ile), we report the direction of selection with  
738 respect to the trait-increasing allele. Percentiles of SDSs were evaluated only on a subset of  
739 variants within 10kb of any variant reported in the GWAS Catalog to account for genomic context.  
740 Further, as variants that are sub-GWS for a trait may nonetheless be under selection, we  
741 calculated the genome-wide mean trait-SDS in each bin of 1000 variants, ranked by *P*-value for  
742 the trait association, following the protocol outlined by Field *et al.* (2016)<sup>56</sup>.

## 743 Balancing selection

744 We accessed publicly available standardised BetaScan2 scores, which detect balancing selection  
745 using polymorphism and substitution data, for all SNPs in the 1000 Genomes dataset<sup>59</sup>. We tested  
746 whether the +/-10kb window around each infertility or hormone lead variant contained SNPs with  
747 scores in the 99th %ile of standardised BetaScan2 scores. Percentiles of SDSs were evaluated  
748 only on a subset of variants within 10kb of any variant reported in the GWAS Catalog to account  
749 for genomic context. For each lead variant window, we report the highest standardised BetaScan2  
750 score and its percentile.

## 751 Whole exome sequencing analyses in the UK Biobank

### 752 Exome sequencing quality control

#### 753 Quality control outline

754 We first considered an initial set of “high quality” variants to evaluate the mean call rate and depth  
755 of coverage for each sample. We then ran a sample and variant level pre-filtering step and  
756 calculated sample-level QC metrics. Using these metrics, we removed sample outliers based on  
757 median absolute deviation (MAD) thresholds, and excluded sites which did not pass variant QC  
758 according to Karzcewski *et al.* (2022)<sup>135</sup>. We then applied a genotype-level filter using genotype  
759 quality (GQ), depth (DP), and heterozygote allele balance (AB). The resultant high-quality  
760 European call set consisted of 402,375 samples and 25,229,669 variants. For details see  
761 Supplementary Text.

#### 762 Variant annotation

763 We annotated variants using Variant Effect Predictor (VEP) v105 (corresponding to gencode  
764 v39)<sup>136</sup> with the LOFTEE v1.04\_GRCh38<sup>137</sup> and dbNSFP<sup>138</sup> plugins, annotating variants with  
765 CADD v1.6<sup>139</sup>, and REVEL using dbNSFP4.3<sup>140</sup> and loss of function confidence using LOFTEE.  
766 Complete instructions and code for this step are provided in our VEP\_105\_LOFTEE repository<sup>141</sup>,  
767 which contains a docker/singularity container to ensure reproducibility of annotations. We then  
768 ran SpliceAI v1.3<sup>142</sup> using the gencode v39 gene annotation file to ensure alignment between  
769 VEP and SpliceAI transcript annotations. We defined ‘canonical’ transcripts to be used for variant-  
770 specific annotations as follows: set MANE Select<sup>143</sup> as the canonical, where available, and if a  
771 MANE Select transcript is not present, set canonical and restrict to protein coding genes. Note  
772 that for VEP 105, this is equivalent to selecting the ‘canonical’ transcript in protein coding genes.  
773 Then, using the collection of missense, pLoF, splice metrics, and annotations of variant  
774 consequence on the ‘canonical’ transcript, we determine a set of variant categories for gene-  
775 based testing.

#### 776 Variant categories for gene-based tests

- 777 1. **High confidence pLoF:** high-confidence LoF variants, as defined by LOFTEE <sup>137</sup>  
778 (LOFTEE HC).
- 779 2. **Damaging missense/protein-altering:** at least one of:
  - 780 a. Variant annotated as missense/start-loss/stop-loss/in-frame indel and  
781 (REVEL $\geq$ 0.773 or CADD $\geq$ 28.1 (or both)).
  - 782 b. Any variant with SpliceAI delta score (DS) $\geq$ 0.2 where SpliceAI DS the maximum  
783 of the set {DS\_AG, DS\_AL, DS\_DG, DS\_DL} for each annotated variant (where  
784 DS\_AG, DS\_AL, DS\_DG and DS\_DL are delta score (acceptor gain), delta score  
785 (acceptor loss), delta score (donor gain), and delta score (donor loss),  
786 respectively).
  - 787 c. Low-confidence LoF variants, as defined by LOFTEE (LOFTEE LC)
- 788 3. **Other missense/protein-altering:**



789 Missense/start-loss/stop-loss/in-frame indel not categorised in (2) (Damaging  
790 missense/protein-altering).

791 4. **Synonymous**: synonymous variants with SpliceAI DS<0.2 in the gene (our 'control' set).

792  
793 REVEL and CADD score cut-offs are chosen to reflect the supporting level for pathogenicity (PP3)  
794 from the American College of Medical Genetics and Genomics and the Association for Molecular  
795 Pathology (ACMG/AMP) criteria<sup>144</sup>.

796  
797 Variant counts and average allele counts for each annotation, split by population label and binned  
798 by MAF are displayed in Supp. Figure 13 and Supp. Figure 14, respectively.

## 799 Genetic association testing

800 We carried out rare variant genetic association testing in the European-ancestry subset of the UK  
801 Biobank using Scalable and Accurate Implementation of GEneralized mixed model (SAIGE)<sup>108</sup>,  
802 a mixed model framework that accounts for sample relatedness and case-control imbalance  
803 through a saddle-point approximation in binary traits. All rare-variant analysis was carried out on  
804 the UK Biobank Research Analysis Platform (RAP) using SAIGE version wzhou88/saige:1.1.9<sup>108</sup>.  
805 In the sex-combined analyses, we account for *age*, *sex*, *age*<sup>2</sup>, *age* × *sex*, *age*<sup>2</sup> × *sex*, and the first  
806 10 genetic principal components as fixed effects; and *age*, *age*<sup>2</sup>, and the first 10 principal  
807 components in sex-specific analyses. All continuous traits were inverse rank normalised prior to  
808 association testing.

809  
810 For SAIGE step 0, we constructed a genetic relatedness matrix (GRM) using the UK Biobank  
811 genotyping array data. We LD pruned the genotyped data using PLINK (`--indep-pairwise`  
812 `50 5 0.05`)<sup>145</sup>, and created a sparse GRM using 5000 randomly selected markers, with  
813 relatedness cutoff of 0.05, using the `createSparseGRM.R` function within SAIGE. To generate  
814 a variance ratio file for subsequent steps in SAIGE, we extracted 1000 variants each with MAC<20  
815 and MAC>20, and combined these markers to define a PLINK file for the variance ratio  
816 determination.

817  
818 In SAIGE step 1 for each trait, the curated phenotype data and sparse GRM were used to fit a  
819 null model with no genetic contribution. All parameters were set at the defaults in SAIGE, except  
820 `--relatednessCutoff 0.05, --useSparseGRMtoFitNULL TRUE` and `--`  
821 `isCateVarianceRatio TRUE`. Tolerance for fitting the null generalised linear mixed model  
822 was set to 0.00001.

## 823 Rare variant and gene based testing

824 Following null model fitting, we carried out variant and gene-based testing in SAIGE step 2 using  
825 the variant categories described above, with the `--is_single_in_groupTest TRUE` flag. All  
826 other parameters were set to default, except `--`  
827 `maxMAF_in_groupTest=0.0001,0.001,0.01, --is_Firth_beta TRUE, --`  
828 `pCutoffforFirth=0.1, and --is_fastTest TRUE`. We included the following collection of



829 group tests, using the annotations defined in methods: variant annotation.

- 830
- 831 ● High confidence pLoF
  - 832 ● Damaging missense/protein-altering
  - 833 ● Other missense/protein-altering
  - 834 ● Synonymous
  - 835 ● High confidence pLoF or Damaging missense/protein-altering
  - 836 ● High confidence pLoF or Damaging missense/protein-altering or Other missense/protein-
  - 837 altering or Synonymous

838

839 We then carried out Cauchy combination tests<sup>146</sup> across these annotations for each gene.

## 840 Data and code availability

841 Cohorts may be contacted individually for access to raw data. Summary statistics for all  
842 phenotypes will be made available through the GWAS Catalog upon publication. All code used in  
843 this study will be made available through GitHub upon publication.

## 844 Acknowledgements

845 Thanks to Prof Pier Palamara and Prof Zoltan Kutalik for helpful discussions. We are grateful to  
846 the participants of all cohorts. This research has partly been conducted using the UK Biobank  
847 Resource under Application Number 11867. Genes & Health is/has recently been core-funded by  
848 Wellcome (WT102627, WT210561), the Medical Research Council (UK) (M009017,  
849 MR/X009777/1, MR/X009920/1), Higher Education Funding Council for England Catalyst, Barts  
850 Charity (845/1796), Health Data Research UK (for London substantive site), and research delivery  
851 support from the NHS National Institute for Health Research Clinical Research Network (North  
852 Thames). Genes & Health is/has recently been funded by Alnylam Pharmaceuticals, Genomics  
853 PLC; and a Life Sciences Industry Consortium of Astra Zeneca PLC, Bristol-Myers Squibb  
854 Company, GlaxoSmithKline Research and Development Limited, Maze Therapeutics Inc, Merck  
855 Sharp & Dohme LLC, Novo Nordisk A/S, Pfizer Inc, Takeda Development Centre Americas Inc.  
856 We thank Social Action for Health, Centre of The Cell, members of our Community Advisory  
857 Group, and staff who have recruited and collected data from volunteers. We thank the NIHR  
858 National Biosample Centre (UK Biocentre), the Social Genetic & Developmental Psychiatry  
859 Centre (King's College London), Wellcome Sanger Institute, and Broad Institute for sample  
860 processing, genotyping, sequencing and variant annotation. We thank: Barts Health NHS Trust,  
861 NHS Clinical Commissioning Groups (City and Hackney, Waltham Forest, Tower Hamlets,  
862 Newham, Redbridge, Havering, Barking and Dagenham), East London NHS Foundation Trust,  
863 Bradford Teaching Hospitals NHS Foundation Trust, Public Health England (especially David  
864 Wyllie), Discovery Data Service/Endeavour Health Charitable Trust (especially David Stables),  
865 Voror Health Technologies Ltd (especially Sophie Don), NHS England (for what was NHS Digital)  
866 - for GDPR-compliant data sharing backed by individual written informed consent. Most of all we  
867 thank all of the volunteers participating in Genes & Health. This study was funded by the European  
868 Union through the European Regional Development Fund Project No. 2014-2020.4.01.15-0012

869 GENTRANSMED. Data analysis was carried out in part in the High-Performance Computing  
870 Center of University of Tartu. This research has partly been conducted using the ALSPAC  
871 resource under Project Number B4568. The activities of the EstBB are regulated by the Human  
872 Genes Research Act, which was adopted in 2000 specifically for the operations of the EstBB.  
873 Individual level data analysis in the EstBB was carried out under ethical approval 1.1-12/624 from  
874 the Estonian Committee on Bioethics and Human Research (Estonian Ministry of Social Affairs),  
875 using data according to release application 3-10/GI/10790 from the Estonian Biobank. This study  
876 was funded by the Estonian Research Council grants PRG1911. We want to acknowledge the  
877 participants and investigators of the FinnGen study. The FinnGen project is funded by two grants  
878 from Business Finland (HUS 4685/31/2016 and UH 4386/31/2016) and the following industry  
879 partners: AbbVie Inc., AstraZeneca UK Ltd, Biogen MA Inc., Bristol Myers Squibb (and Celgene  
880 Corporation & Celgene International II Sàrl), Genentech Inc., Merck Sharp & Dohme LCC, Pfizer  
881 Inc., GlaxoSmithKline Intellectual Property Development Ltd., Sanofi US Services Inc., Maze  
882 Therapeutics Inc., Janssen Biotech Inc, Novartis Pharma AG, and Boehringer Ingelheim  
883 International GmbH. Following biobanks are acknowledged for delivering biobank samples to  
884 FinnGen: Auria Biobank ([www.auria.fi/biopankki](http://www.auria.fi/biopankki)), THL Biobank ([www.thl.fi/biobank](http://www.thl.fi/biobank)), Helsinki  
885 Biobank ([www.helsinginbiopankki.fi](http://www.helsinginbiopankki.fi)), Biobank Borealis of Northern Finland  
886 ([https://www.ppshp.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-](https://www.ppshp.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-English.aspx)  
887 [English.aspx](http://www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere)), Finnish Clinical Biobank Tampere ([www.tays.fi/en-](http://www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere)  
888 [US/Research\\_and\\_development/Finnish\\_Clinical\\_Biobank\\_Tampere](http://www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere)), Biobank of Eastern  
889 Finland ([www.ita-suomenbiopankki.fi/en](http://www.ita-suomenbiopankki.fi/en)), Central Finland Biobank ([www.ksshp.fi/fi-](http://www.ksshp.fi/fi-FI/Potilaalle/Biopankki)  
890 [FI/Potilaalle/Biopankki](http://www.ksshp.fi/fi-FI/Potilaalle/Biopankki)), Finnish Red Cross Blood Service Biobank  
891 ([www.veripalvelu.fi/verenluovutus/biopankkitoiminta](http://www.veripalvelu.fi/verenluovutus/biopankkitoiminta)), Terveystalo Biobank  
892 ([www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/](http://www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/)) and Arctic Biobank  
893 ([https://www oulu.fi/en/university/faculties-and-units/faculty-medicine/northern-finland-birth-](https://www oulu.fi/en/university/faculties-and-units/faculty-medicine/northern-finland-birth-cohorts-and-arctic-biobank)  
894 [cohorts-and-arctic-biobank](https://www oulu.fi/en/university/faculties-and-units/faculty-medicine/northern-finland-birth-cohorts-and-arctic-biobank)). All Finnish Biobanks are members of BBMRI.fi infrastructure  
895 ([www.bbmri.fi](http://www.bbmri.fi)). Finnish Biobank Cooperative -FINBB (<https://finbb.fi/>) is the coordinator of  
896 BBMRI-ERIC operations in Finland. The Finnish biobank data can be accessed through the  
897 Fingenious® services (<https://site.fingenious.fi/en/>) managed by FINBB. We are extremely  
898 grateful to all the families who took part in this study, the midwives for their help in recruiting them,  
899 and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians,  
900 clerical workers, research scientists, volunteers, managers, receptionists and nurses. We  
901 acknowledge the contribution of the participants in the deCODE study. Patients and control  
902 subjects in FinnGen provided informed consent for biobank research, based on the Finnish  
903 Biobank Act. Alternatively, separate research cohorts, collected prior the Finnish Biobank Act  
904 came into effect (in September 2013) and start of FinnGen (August 2017), were collected based  
905 on study-specific consents and later transferred to the Finnish biobanks after approval by Fimea  
906 (Finnish Medicines Agency), the National Supervisory Authority for Welfare and Health.  
907 Recruitment protocols followed the biobank protocols approved by Fimea. The Coordinating  
908 Ethics Committee of the Hospital District of Helsinki and Uusimaa (HUS) statement number for  
909 the FinnGen study is Nr HUS/990/2017. The FinnGen study is approved by Finnish Institute for  
910 Health and Welfare (permit numbers: THL/2031/6.02.00/2017, THL/1101/5.05.00/2017,  
911 THL/341/6.02.00/2018, THL/2222/6.02.00/2018, THL/283/6.02.00/2019, THL/1721/5.05.00/2019  
912 and THL/1524/5.05.00/2020), Digital and population data service agency (permit numbers:

913 VRK43431/2017-3, VRK/6909/2018-3, VRK/4415/2019-3), the Social Insurance Institution  
914 (permit numbers: KELA 58/522/2017, KELA 131/522/2018, KELA 70/522/2019, KELA  
915 98/522/2019, KELA 134/522/2019, KELA 138/522/2019, KELA 2/522/2020, KELA 16/522/2020),  
916 Findata permit numbers THL/2364/14.02/2020, THL/4055/14.06.00/2020,  
917 THL/3433/14.06.00/2020, THL/4432/14.06/2020, THL/5189/14.06/2020,  
918 THL/5894/14.06.00/2020, THL/6619/14.06.00/2020, THL/209/14.06.00/2021,  
919 THL/688/14.06.00/2021, THL/1284/14.06.00/2021, THL/1965/14.06.00/2021,  
920 THL/5546/14.02.00/2020, THL/2658/14.06.00/2021, THL/4235/14.06.00/2021, Statistics Finland  
921 (permit numbers: TK-53-1041-17 and TK/143/07.03.00/2020 (earlier TK-53-90-20)  
922 TK/1735/07.03.00/2021, TK/3112/07.03.00/2021) and Finnish Registry for Kidney Diseases  
923 permission/extract from the meeting minutes on 4th July 2019. The Biobank Access Decisions for  
924 FinnGen samples and data utilised in FinnGen Data Freeze 10 include: THL Biobank  
925 BB2017\_55, BB2017\_111, BB2018\_19, BB\_2018\_34, BB\_2018\_67, BB2018\_71, BB2019\_7,  
926 BB2019\_8, BB2019\_26, BB2020\_1, BB2021\_65, Finnish Red Cross Blood Service Biobank  
927 7.12.2017, Helsinki Biobank HUS/359/2017, HUS/248/2020, HUS/150/2022 § 12, §13, §14, §15,  
928 §16, §17, §18, and §23, Auria Biobank AB17-5154 and amendment #1 (August 17 2020) and  
929 amendments BB\_2021-0140, BB\_2021-0156 (August 26 2021, Feb 2 2022), BB\_2021-0169,  
930 BB\_2021-0179, BB\_2021-0161, AB20-5926 and amendment #1 (April 23 2020) and it's  
931 modification (Sep 22 2021), Biobank Borealis of Northern Finland\_2017\_1013, 2021\_5010,  
932 2021\_5018, 2021\_5015, 2021\_5023, 2021\_5017, 2022\_6001, Biobank of Eastern Finland  
933 1186/2018 and amendment 22 § /2020, 53§/2021, 13§/2022, 14§/2022, 15§/2022, Finnish  
934 Clinical Biobank Tampere MH0004 and amendments (21.02.2020 & 06.10.2020), §8/2021,  
935 §9/2022, §10/2022, §12/2022, §20/2022, §21/2022, §22/2022, §23/2022, Central Finland  
936 Biobank 1-2017, and Terveystalo Biobank STB 2018001 and amendment 25th Aug 2020, Finnish  
937 Hematological Registry and Clinical Biobank decision 18th June 2021, Arctic biobank P0844:  
938 ARC\_2021\_1001. We want to acknowledge the participants and investigators of FinnGen study.  
939 The FinnGen project is funded by two grants from Business Finland (HUS 4685/31/2016 and UH  
940 4386/31/2016) and the following industry partners: AbbVie Inc., AstraZeneca UK Ltd, Biogen MA  
941 Inc., Bristol Myers Squibb (and Celgene Corporation & Celgene International II Sàrl), Genentech  
942 Inc., Merck Sharp & Dohme LCC, Pfizer Inc., GlaxoSmithKline Intellectual Property Development  
943 Ltd., Sanofi US Services Inc., Maze Therapeutics Inc., Janssen Biotech Inc, Novartis Pharma AG,  
944 and Boehringer Ingelheim International GmbH. Following biobanks are acknowledged for  
945 delivering biobank samples to FinnGen: Auria Biobank ([www.auria.fi/biopankki](http://www.auria.fi/biopankki)), THL Biobank  
946 ([www.thl.fi/biobank](http://www.thl.fi/biobank)), Helsinki Biobank ([www.helsinginbiopankki.fi](http://www.helsinginbiopankki.fi)), Biobank Borealis of Northern  
947 Finland ([https://www.ppsHP.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-](https://www.ppsHP.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-English.aspx)  
948 [English.aspx](http://www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere)), Finnish Clinical Biobank Tampere ([www.tays.fi/en-](http://www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere)  
949 [US/Research\\_and\\_development/Finnish\\_Clinical\\_Biobank\\_Tampere](http://www.ita-suomenbiopankki.fi/en)), Biobank of Eastern  
950 Finland ([www.ita-suomenbiopankki.fi/en](http://www.ita-suomenbiopankki.fi/en)), Central Finland Biobank ([www.kssHP.fi/fi-](http://www.kssHP.fi/fi-FI/Potilaalle/Biopankki)  
951 [FI/Potilaalle/Biopankki](http://www.veripalvelu.fi/verenluovutus/biopankkitoiminta)), Finnish Red Cross Blood Service Biobank  
952 ([www.veripalvelu.fi/verenluovutus/biopankkitoiminta](http://www.veripalvelu.fi/verenluovutus/biopankkitoiminta)), Terveystalo Biobank  
953 ([www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/](http://www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/)) and Arctic Biobank  
954 ([https://www oulu.fi/en/university/faculties-and-units/faculty-medicine/northern-finland-birth-](https://www oulu.fi/en/university/faculties-and-units/faculty-medicine/northern-finland-birth-cohorts-and-arctic-biobank)  
955 [cohorts-and-arctic-biobank](http://www.bbmri.fi)). All Finnish Biobanks are members of BBMRI.fi infrastructure  
956 ([www.bbmri.fi](http://www.bbmri.fi)). Finnish Biobank Cooperative -FINBB (<https://finbb.fi/>) is the coordinator of

957 BBMRI-ERIC operations in Finland. The Finnish biobank data can be accessed through the  
958 Fingenious® services (<https://site.fingenious.fi/en/>) managed by FINBB. This study was based on  
959 the CHB reproduction protocol and DBDS (ethical approval NVK-1805807; NVK-1700407; SJ-  
960 740).

## 961 Funding Statement

962 S.S.V. is supported by the Rhodes Scholarships, Clarendon Fund, and the Medical Sciences  
963 Doctoral Training Centre at the University of Oxford. L.B.L.W. was supported by the Wellcome  
964 Trust. B.M.J. is funded by an Medical Research Council (MRC) Clinical  
965 Research Training Fellowship (CRTF) jointly supported by the UK MS Society (B.M.J.;  
966 grant reference: MR/V028766/1). A.P. is supported by Alma and K.A. Snellman Foundation.  
967 Genes & Health is/has recently been core-funded by Wellcome (WT102627, WT210561), the  
968 Medical Research Council (UK) (M009017, MR/X009777/1, MR/X009920/1), Higher Education  
969 Funding Council for England Catalyst, Barts Charity (845/1796), Health Data Research UK (for  
970 London substantive site), and research delivery support from the NHS National Institute for Health  
971 Research Clinical Research Network (North Thames). A.E. and D.A.L contributions are supported  
972 by the UK Medical Research Council (MC\_UU\_00032/05) and the European Research Council  
973 under the European Union's Horizon 2020 research and innovation program (grant agreements  
974 No 101021566). Genes & Health is/has recently been funded by Alnylam Pharmaceuticals,  
975 Genomics PLC; and a Life Sciences Industry Consortium of Astra Zeneca PLC, Bristol-Myers  
976 Squibb Company, GlaxoSmithKline Research and Development Limited, Maze Therapeutics Inc,  
977 Merck Sharp & Dohme LLC, Novo Nordisk A/S, Pfizer Inc, Takeda Development Centre Americas  
978 Inc. The UK Medical Research Council and Wellcome (Grant ref: 217065/Z/19/Z) and the  
979 University of Bristol provide core support for ALSPAC. A comprehensive list of grants funding is  
980 available on the ALSPAC website ([http://www.bristol.ac.uk/alspac/external/documents/grant-  
981 acknowledgements.pdf](http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf)). Genome-wide genotyping data was generated by Sample Logistics and  
982 Genotyping Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of  
983 America) using support from 23andMe. C.M.L. is supported by the Li Ka Shing Foundation, NIHR  
984 Oxford Biomedical Research Centre, Oxford, NIH (1P50HD104224-01), Gates Foundation (INV-  
985 024200), and a Wellcome Trust Investigator Award (221782/Z/20/Z). The research was supported  
986 by the Wellcome Trust Core Award Grant Number 203141/Z/16/Z with additional support from the  
987 NIHR Oxford BRC. The views expressed are those of the authors and not necessarily those of  
988 the NHS, the NIHR or the Department of Health.

## 989 Competing Interests Statement

990 L.B.L.W. is currently employed by Novo Nordisk Research Centre Oxford but, while she  
991 conducted the research described in this manuscript, was only affiliated to the University of  
992 Oxford. V.S., G.T., H.H., I.J., and K.S. are employees of deCODE genetics, a subsidiary of  
993 Amgen. C.M.L. reports grants from Bayer AG and Novo Nordisk and has a partner who works at  
994 Vertex. The other authors declare no conflicts of interest.

## 995 Supplementary Author List

- 996 Genes & Health Research Team: Shaheen Akhtar, Mohammad Anwar, Elena Arciero, Omar  
997 Asgar, Samina Ashraf, Saeed Bidi, Gerome Breen, James Broster, Raymond Chung, David  
998 Collier, Charles J Curtis, Shabana Chaudhary, Megan Clinch, Grainne Colligan, Panos Deloukas,  
999 Ceri Durham, Faiza Durrani, Fabiola Eto, Sarah Finer, Joseph Gafton, Ana Angel, Chris Griffiths,  
1000 Joanne Harvey, Teng Heng, Sam Hodgson, Qin Qin Huang, Matt Hurles, Karen A Hunt, Shapna  
1001 Hussain, Kamrul Islam, Vivek Iyer, Benjamin M Jacobs, Ahsan Khan, Claudia Langenberg, Cath  
1002 Lavery, Sang Hyuck Lee, Daniel MacArthur, Sidra Malik, Daniel Malawsky, Hilary Martin, Dan  
1003 Mason, Rohini Mathur, Mohammed Bodrul Mazid, John McDermott, Caroline Morton, Bill  
1004 Newman, Elizabeth Owor, Asma Qureshi, Shwetha Ramachandruppa, Mehru Raza, Jessry  
1005 Russell, Nishat Safa, Miriam Samuel, Michael Simpson, John Solly, Marie Spreckley. Daniel  
1006 Stow, Michael Taylor, Richard C Trembath, Karen Tricker, David A van Heel, Klaudia Walter,  
1007 Caroline Winckley, Suzanne Wood, John Wright, Ishevanhu Zengeya, Julia Zöllner.
- 1008 Estonian Biobank Research Team: Andres Metspalu, Lili Milani, Tõnu Esko, Mari Nelis and  
1009 Georgi Hudjashov.
- 1010 Estonian Health Information Research Team: Raivo Kolde, Sven Laur, Sulev Reisberg and Jaak  
1011 Vilo.
- 1012
- 1013 DBDS Genomic Consortium: Agnete Lundgaard; Alexander Pil Henriksen; Bertram Dalskov  
1014 Kjerulff; Bitten Aagaard Jensen; Bjarke Feenstra; Christian Erikstrup; Christina Mikkelsen; Daniel  
1015 Gudbjartsson; David Westergaard; Erik Sørensen; Frank Geller; Gregor Jemec; Henrik Hjalgrim;  
1016 Henrik Ullum; Hreinn Stefánsson; Ioanna Nissen; Ioannis Louloudis ; Jakob Bay; Jens Kjærgaard  
1017 Boldsen; Joseph Dowsett; Kari Stefansson; Karina Banasik; Katrine Kaspersen; Khoa Manh Dinh;  
1018 Klaus Rostgaard; Kristoffer Burgdorf; Lise Wegner Thørner; Lisette Kogelman; Lotte Hindhede;  
1019 Margit Anita Hørup Larsen; Maria Didriksen; Mette Nyegaard; Michael Schwinn; Mie Topholm  
1020 Bruun; Mona Ameri Chalmer; Ole Birger Pedersen; Palle Duun Rohde; Rikke Louise Jacobsen;  
1021 Sisse Rye Ostrowski; Søren Brunak; Susan Mikkelsen; Thomas Folkmann Hansen; Thomas  
1022 Werge; Thorsten Brodersen; Unnur Þorsteinsdóttir.
- 1023
- 1024 FinnGen: See separate document.



## 1025 References

- 1026 1. Infertility prevalence estimates, 1990–2021.  
1027 <https://www.who.int/publications/i/item/978920068315> (2023).
- 1028 2. Brandt, J. S., Cruz Ithier, M. A., Rosen, T. & Ashkinadze, E. Advanced paternal age,  
1029 infertility, and reproductive risks: A review of the literature. *Prenat. Diagn.* **39**, 81–87 (2019).
- 1030 3. Crawford, N. M. & Steiner, A. Z. Age-related infertility. *Obstet. Gynecol. Clin. North Am.* **42**,  
1031 15–25 (2015).
- 1032 4. Tsevat, D. G., Wiesenfeld, H. C., Parks, C. & Peipert, J. F. Sexually transmitted diseases  
1033 and infertility. *Am. J. Obstet. Gynecol.* **216**, 1–9 (2017).
- 1034 5. Passos, L. G. *et al.* The Correlation between Chlamydia Trachomatis and Female Infertility:  
1035 A Systematic Review. *Rev. Bras. Ginecol. Obstet.* **44**, 614–620 (2022).
- 1036 6. Tjahyadi, D. *et al.* Female Genital Tuberculosis: Clinical Presentation, Current Diagnosis,  
1037 and Treatment. *Infect. Dis. Obstet. Gynecol.* **2022**, 3548190 (2022).
- 1038 7. Shteinberg, M., Taylor-Cousar, J. L., Durieu, I. & Cohen-Cymbarknoh, M. Fertility and  
1039 Pregnancy in Cystic Fibrosis. *Chest* **160**, 2051–2060 (2021).
- 1040 8. Jarzabek, K. *et al.* Cystic fibrosis as a cause of infertility. *Reprod. Biol.* **4**, 119–129 (2004).
- 1041 9. Matzuk, M. M. & Lamb, D. J. The biology of infertility: research advances and clinical  
1042 challenges. *Nat. Med.* **14**, 1197–1213 (2008).
- 1043 10. Krausz, C. & Riera-Escamilla, A. Genetics of male infertility. *Nat. Rev. Urol.* **15**, 369–384  
1044 (2018).
- 1045 11. Payne, N. & van den Akker, O. Mental health and coping with fertility treatment cessation  
1046 during the COVID-19 pandemic in the UK. *J. Psychosom. Obstet. Gynaecol.* **43**, 550–556  
1047 (2022).
- 1048 12. Kwan, I., Wang, R., Pearce, E. & Bhattacharya, S. Pain relief for women undergoing oocyte  
1049 retrieval for assisted reproduction. *Cochrane Database Syst. Rev.* **5**, CD004829 (2018).
- 1050 13. Katz, P. *et al.* Costs of infertility treatment: results from an 18-month prospective cohort  
1051 study. *Fertil. Steril.* **95**, 915–921 (2011).
- 1052 14. Lai, J. D. *et al.* Unmet financial burden of infertility care and the impact of state insurance  
1053 mandates in the United States: analysis from a popular crowdfunding platform. *Fertil. Steril.*  
1054 **116**, 1119–1125 (2021).
- 1055 15. Bonavina, G. & Taylor, H. S. Endometriosis-associated infertility: From pathophysiology to  
1056 tailored treatment. *Front. Endocrinol.* **13**, 1020827 (2022).
- 1057 16. Hoeger, K. M., Dokras, A. & Piltonen, T. Update on PCOS: Consequences, Challenges,  
1058 and Guiding Treatment. *J. Clin. Endocrinol. Metab.* **106**, e1071–e1083 (2021).
- 1059 17. Thessaloniki ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group. Consensus  
1060 on infertility treatment related to polycystic ovary syndrome. *Hum. Reprod.* **23**, 462–477  
1061 (2008).
- 1062 18. Vannuccini, S. *et al.* Infertility and reproductive disorders: impact of hormonal and  
1063 inflammatory mechanisms on pregnancy outcome. *Hum. Reprod. Update* **22**, 104–115  
1064 (2016).
- 1065 19. Concepción-Zavaleta, M. *et al.* Assessment of hormonal status in male infertility. An  
1066 update. *Diabetes Metab. Syndr.* **16**, 102447 (2022).
- 1067 20. Silvestris, E., de Pergola, G., Rosania, R. & Loverro, G. Obesity as disruptor of the female  
1068 fertility. *Reprod. Biol. Endocrinol.* **16**, 22 (2018).
- 1069 21. A life-course approach to women’s health. *Nat. Med.* **30**, 1 (2024).
- 1070 22. Eisenberg, M. L. *et al.* Male infertility. *Nat Rev Dis Primers* **9**, 49 (2023).
- 1071 23. Gardner, D. K. & Sakkas, D. Making and selecting the best embryo in the laboratory. *Fertil.*  
1072 *Steril.* **120**, 457–466 (2023).
- 1073 24. Barban, N. *et al.* Genome-wide analysis identifies 12 loci influencing human reproductive



- 1074 behavior. *Nat. Genet.* **48**, 1462–1472 (2016).
- 1075 25. Mathieson, I. *et al.* Genome-wide analysis identifies genetic effects on reproductive  
1076 success and ongoing natural selection at the FADS locus. *Nat Hum Behav* **7**, 790–801  
1077 (2023).
- 1078 26. Hakonarson, H., Gulcher, J. R. & Stefansson, K. deCODE genetics, Inc.  
1079 *Pharmacogenomics* **4**, 209–215 (2003).
- 1080 27. Kurki, M. I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated  
1081 population. *Nature* **613**, 508–518 (2023).
- 1082 28. Heyne, H. O. *et al.* Mono- and biallelic variant effects on disease at biobank scale. *Nature*  
1083 **613**, 519–525 (2023).
- 1084 29. Wicherek, L. *et al.* The placental RCAS1 expression during stillbirth. *Reprod. Biol.*  
1085 *Endocrinol.* **3**, 24 (2005).
- 1086 30. Ohshima, K., Nakashima, M., Sonoda, K., Kikuchi, M. & Watanabe, T. Expression of  
1087 RCAS1 and FasL in human trophoblasts and uterine glands during pregnancy: the possible  
1088 role in immune privilege. *Clin. Exp. Immunol.* **123**, 481–486 (2001).
- 1089 31. Zhang, G. *et al.* Genetic Associations with Gestational Duration and Spontaneous Preterm  
1090 Birth. *N. Engl. J. Med.* **377**, 1156–1167 (2017).
- 1091 32. Välimäki, N. *et al.* Genetic predisposition to uterine leiomyoma is determined by loci for  
1092 genitourinary development and genome stability. *Elife* **7**, (2018).
- 1093 33. Rafnar, T. *et al.* Variants associating with uterine leiomyoma highlight genetic background  
1094 shared by various cancers and hormone-related traits. *Nat. Commun.* **9**, 3636 (2018).
- 1095 34. Luong, H. T. *et al.* Fine mapping of variants associated with endometriosis in the WNT4  
1096 region on chromosome 1p36. *Int. J. Mol. Epidemiol. Genet.* **4**, 193–206 (2013).
- 1097 35. Rahmioglu, N. *et al.* The genetic basis of endometriosis and comorbidity with other pain  
1098 and inflammatory conditions. *Nat. Genet.* **55**, 423–436 (2023).
- 1099 36. Pellegrino, M., Maiorino, R. & Schonauer, S. WNT4 signaling in female gonadal  
1100 development. *Endocr. Metab. Immune Disord. Drug Targets* **10**, 168–174 (2010).
- 1101 37. Bernard, P. & Harley, V. R. Wnt4 action in gonadal development and sex determination. *Int.*  
1102 *J. Biochem. Cell Biol.* **39**, 31–43 (2007).
- 1103 38. Pitzer, L. M., Moroney, M. R., Nokoff, N. J. & Sikora, M. J. WNT4 Balances Development vs  
1104 Disease in Gynecologic Tissues and Women’s Health. *Endocrinology* **162**, (2021).
- 1105 39. Groza, T. *et al.* The International Mouse Phenotyping Consortium: comprehensive knockout  
1106 phenotyping underpinning the study of human disease. *Nucleic Acids Res.* **51**, D1038–  
1107 D1045 (2023).
- 1108 40. Ochoa, D. *et al.* The next-generation Open Targets Platform: reimaged, redesigned,  
1109 rebuilt. *Nucleic Acids Res.* **51**, D1353–D1359 (2023).
- 1110 41. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–  
1111 585 (2013).
- 1112 42. Ying, S. Y. Inhibins and activins: chemical properties and biological activity. *Proc. Soc. Exp.*  
1113 *Biol. Med.* **186**, 253–264 (1987).
- 1114 43. Chada, M. *et al.* Inhibin B, follicle stimulating hormone, luteinizing hormone, and estradiol  
1115 and their relationship to the regulation of follicle development in girls during childhood and  
1116 puberty. *Physiol. Res.* **52**, 341–346 (2003).
- 1117 44. Nawaz, S. *et al.* A variant in sperm-specific glycolytic enzyme enolase 4 (ENO4) causes  
1118 human male infertility. *J. Gene Med.* e3583 (2023).
- 1119 45. Nakamura, N. *et al.* Disruption of a spermatogenic cell-specific mouse enolase 4 (eno4)  
1120 gene causes sperm structural defects and male infertility. *Biol. Reprod.* **88**, 90 (2013).
- 1121 46. Mbarek, H. *et al.* Genome-wide association study meta-analysis of dizygotic twinning  
1122 illuminates genetic regulation of female fecundity. *Hum. Reprod.* **39**, 240–257 (2024).
- 1123 47. Paskulin, D. D., Cunha-Filho, J. S., Paskulin, L. D., Souza, C. A. B. & Ashton-Prolla, P.  
1124 ESR1 rs9340799 is associated with endometriosis-related infertility and in vitro fertilization

- 1125 failure. *Dis. Markers* **35**, 907–913 (2013).
- 1126 48. Wang, W. *et al.* Association of an oestrogen receptor gene polymorphism in Chinese Han  
1127 women with endometriosis and endometriosis-related infertility. *Reprod. Biomed. Online* **26**,  
1128 93–98 (2013).
- 1129 49. Mafrà, F., Catto, M., Bianco, B., Barbosa, C. P. & Christofolini, D. Association of WNT4  
1130 polymorphisms with endometriosis in infertile patients. *J. Assist. Reprod. Genet.* **32**, 1359–  
1131 1364 (2015).
- 1132 50. Osiński, M. *et al.* The assessment of GWAS - identified polymorphisms associated with  
1133 infertility risk in Polish women with endometriosis. *Ginekol. Pol.* **89**, 304–310 (2018).
- 1134 51. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity  
1135 in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
- 1136 52. International HapMap 3 Consortium *et al.* Integrating common and rare genetic variation in  
1137 diverse human populations. *Nature* **467**, 52–58 (2010).
- 1138 53. Neale lab round 2 GWAS blogpost. <http://www.nealelab.is/blog/2019/10/24/updating-snp-heritability-results-from-4236-phenotypes-in-uk-biobank>.
- 1139 54. Dempster, E. R. & Lerner, I. M. Heritability of Threshold Characters. *Genetics* **35**, 212–236  
1140 (1950).
- 1141 55. Tropf, F. C. *et al.* Human fertility, molecular genetics, and natural selection in modern  
1142 societies. *PLoS One* **10**, e0126821 (2015).
- 1143 56. Field, Y. *et al.* Detection of human adaptation during the past 2000 years. *Science* **354**,  
1144 760–764 (2016).
- 1145 57. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*  
1146 **528**, 499–503 (2015).
- 1147 58. Grossman, S. R. *et al.* Identifying recent adaptations in large-scale genomic data. *Cell* **152**,  
1148 703–713 (2013).
- 1149 59. Siewert, K. M. & Voight, B. F. BetaScan2: Standardized Statistics to Detect Balancing  
1150 Selection Utilizing Substitution Data. *Genome Biol. Evol.* **12**, 3873–3877 (2020).
- 1151 60. Dobzhansky, T. A review of some fundamental concepts and problems of population  
1152 genetics. *Cold Spring Harb. Symp. Quant. Biol.* **20**, 1–15 (1955).
- 1153 61. Lewontin, R. C. The Units of Selection. *Annu. Rev. Ecol. Syst.* **1**, 1–18 (1970).
- 1154 62. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations.  
1155 *Nucleic Acids Res.* **42**, D1001–6 (2014).
- 1156 63. Rytinki, M. *et al.* Dynamic SUMOylation is linked to the activity cycles of androgen receptor  
1157 in the cell nucleus. *Mol. Cell. Biol.* **32**, 4195–4205 (2012).
- 1158 64. Terashima, R., Tani, T., Sakakibara, K., Kurusu, S. & Kawaminami, M. Thyrotropin-  
1159 releasing hormone stimulates NR4A3 expression in the pituitary thyrotrophs of proestrus  
1160 rats. *Endocr. J.* **70**, 805–814 (2023).
- 1161 65. Fernandez, P. M. *et al.* Nuclear receptors Nor1 and NGFI-B/Nur77 play similar, albeit  
1162 distinct, roles in the hypothalamo-pituitary-adrenal axis. *Endocrinology* **141**, 2392–2400  
1163 (2000).
- 1164 66. Zhao, Y. & Bruemmer, D. NR4A orphan nuclear receptors: transcriptional regulators of  
1165 gene expression in metabolism and vascular biology. *Arterioscler. Thromb. Vasc. Biol.* **30**,  
1166 1535–1541 (2010).
- 1167 67. Pulit, S. L. *et al.* Meta-analysis of genome-wide association studies for body fat distribution  
1168 in 694 649 individuals of European ancestry. *Hum. Mol. Genet.* **28**, 166–174 (2019).
- 1169 68. Kichaev, G. *et al.* Leveraging Polygenic Functional Enrichment to Improve GWAS Power.  
1170 *Am. J. Hum. Genet.* **104**, 65–75 (2019).
- 1171 69. Mountjoy, E. *et al.* An open approach to systematically prioritize causal variants and genes  
1172 at all published human GWAS trait-associated loci. *Nat. Genet.* **53**, 1527–1533 (2021).
- 1173 70. Ghossaini, M. *et al.* Open Targets Genetics: systematic identification of trait-associated  
1174 genes using large-scale genetics and functional genomics. *Nucleic Acids Res.* **49**, D1311–  
1175

- 1176 D1320 (2021).
- 1177 71. Nguidjoe, E. *et al.* Heterozygous inactivation of the Na/Ca exchanger increases glucose-  
1178 induced insulin release,  $\beta$ -cell proliferation, and mass. *Diabetes* **60**, 2076–2085 (2011).
- 1179 72. Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using  
1180 predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
- 1181 73. Martinez, G. & Garcia, C. Sexual selection and sperm diversity in primates. *Mol. Cell.*  
1182 *Endocrinol.* **518**, 110974 (2020).
- 1183 74. Verdiesen, R. M. G. *et al.* Genome-wide association study meta-analysis identifies three  
1184 novel loci for circulating anti-Müllerian hormone levels in women. *Hum. Reprod.* **37**, 1069–  
1185 1082 (2022).
- 1186 75. Williams, A. T. *et al.* Genome-wide association study of thyroid-stimulating hormone  
1187 highlights new genes, pathways and associations with thyroid disease. *Nat. Commun.* **14**,  
1188 6713 (2023).
- 1189 76. Ishikawa, K. *et al.* An autosomal dominant cerebellar ataxia linked to chromosome 16q22.1  
1190 is associated with a single-nucleotide substitution in the 5' untranslated region of the gene  
1191 encoding a protein with spectrin repeat and Rho guanine-nucleotide exchange-factor  
1192 domains. *Am. J. Hum. Genet.* **77**, 280–296 (2005).
- 1193 77. Holmes, G. AN ATTEMPT TO CLASSIFY CEREBELLAR DISEASE, WITH A NOTE ON  
1194 MARIE'S HEREDITARY CEREBELLAR ATAXIA. *Brain* **30**, 545–567 (1908).
- 1195 78. Seminara, S. B., Acierno, J. S., Jr, Abdulwahid, N. A., Crowley, W. F., Jr & Margolin, D. H.  
1196 Hypogonadotropic hypogonadism and cerebellar ataxia: detailed phenotypic  
1197 characterization of a large, extended kindred. *J. Clin. Endocrinol. Metab.* **87**, 1607–1612  
1198 (2002).
- 1199 79. Lawson, A. J. *et al.* Cortisone-reductase deficiency associated with heterozygous mutations  
1200 in 11beta-hydroxysteroid dehydrogenase type 1. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 4111–  
1201 4116 (2011).
- 1202 80. Draper, N. *et al.* Mutations in the genes encoding 11beta-hydroxysteroid dehydrogenase  
1203 type 1 and hexose-6-phosphate dehydrogenase interact to cause cortisone reductase  
1204 deficiency. *Nat. Genet.* **34**, 434–439 (2003).
- 1205 81. Durocher, F., Morissette, J., Labrie, Y., Labrie, F. & Simard, J. Mapping of the HSD17B2  
1206 gene encoding type II 17 beta-hydroxysteroid dehydrogenase close to D16S422 on  
1207 chromosome 16q24.1-q24.2. *Genomics* **25**, 724–726 (1995).
- 1208 82. Saloniemi, T., Jokela, H., Strauss, L., Pakarinen, P. & Poutanen, M. The diversity of sex  
1209 steroid action: novel functions of hydroxysteroid (17 $\beta$ ) dehydrogenases as revealed by  
1210 genetically modified mouse models. *J. Endocrinol.* **212**, 27–40 (2012).
- 1211 83. Le Quesne Stabej, P. *et al.* STAG3 truncating variant as the cause of primary ovarian  
1212 insufficiency. *Eur. J. Hum. Genet.* **24**, 135–138 (2016).
- 1213 84. Caburet, S. *et al.* Mutant cohesin in premature ovarian failure. *N. Engl. J. Med.* **370**, 943–  
1214 949 (2014).
- 1215 85. Day, F. R. *et al.* Physical and neurobehavioral determinants of reproductive onset and  
1216 success. *Nat. Genet.* **48**, 617–623 (2016).
- 1217 86. Pasman, J. A. *et al.* The CADM2 Gene and Behavior: A Phenome-Wide Scan in UK-  
1218 Biobank. *Behav. Genet.* **52**, 306–314 (2022).
- 1219 87. Mills, M. C. *et al.* Identification of 371 genetic variants for age at first sex and birth linked to  
1220 externalising behaviour. *Nat Hum Behav* **5**, 1717–1730 (2021).
- 1221 88. Gunnarsson, B. *et al.* A sequence variant associating with educational attainment also  
1222 affects childhood cognition. *Sci. Rep.* **6**, 36189 (2016).
- 1223 89. Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with  
1224 educational attainment. *Nature* **533**, 539–542 (2016).
- 1225 90. de Ziegler, D., Borghese, B. & Chapron, C. Endometriosis and infertility: pathophysiology  
1226 and management. *Lancet* **376**, 730–738 (2010).

- 1227 91. Day, F. *et al.* Large-scale genome-wide meta-analysis of polycystic ovary syndrome  
1228 suggests shared genetic architecture for different diagnosis criteria. *PLoS Genet.* **14**,  
1229 e1007813 (2018).
- 1230 92. Tyrmi, J. S. *et al.* Leveraging Northern European population history: novel low-frequency  
1231 variants for polycystic ovary syndrome. *Hum. Reprod.* **37**, 352–365 (2022).
- 1232 93. Crosignani, P. G., Bianchedi, D., Riccaboni, A. & Vegetti, W. Management of anovulatory  
1233 infertility. *Hum. Reprod.* **14 Suppl 1**, 108–119 (1999).
- 1234 94. Gambineri, A. *et al.* Female infertility: which role for obesity? *Int J Obes Suppl* **9**, 65–72  
1235 (2019).
- 1236 95. Maheshwari, A., Stofberg, L. & Bhattacharya, S. Effect of overweight and obesity on  
1237 assisted reproductive technology--a systematic review. *Hum. Reprod. Update* **13**, 433–444  
1238 (2007).
- 1239 96. Gaskins, A. J. Recent advances in understanding the relationship between long- and short-  
1240 term weight change and fertility. *F1000Res.* **7**, (2018).
- 1241 97. Zirkin, B. R. & Papadopoulos, V. Leydig cells: formation, function, and regulation. *Biol.*  
1242 *Reprod.* **99**, 101–111 (2018).
- 1243 98. Burger, H. G. Androgen production in women. *Fertil. Steril.* **77 Suppl 4**, S3–5 (2002).
- 1244 99. Fortune, J. E. & Armstrong, D. T. Androgen production by theca and granulosa isolated  
1245 from proestrous rat follicles. *Endocrinology* **100**, 1341–1347 (1977).
- 1246 100. Liu, Y. X. & Hsueh, A. J. Synergism between granulosa and theca-interstitial cells in  
1247 estrogen biosynthesis by gonadotropin-treated rat ovaries: studies on the two-cell, two-  
1248 gonadotropin hypothesis using steroid antisera. *Biol. Reprod.* **35**, 27–36 (1986).
- 1249 101. Wen, X., Li, D., Tozer, A. J., Docherty, S. M. & Iles, R. K. Estradiol, progesterone,  
1250 testosterone profiles in human follicular fluid and cultured granulosa cells from luteinized  
1251 pre-ovulatory follicles. *Reprod. Biol. Endocrinol.* **8**, 117 (2010).
- 1252 102. Kinnear, H. M. *et al.* The ovarian stroma as a new frontier. *Reproduction* **160**, R25–R39  
1253 (2020).
- 1254 103. Samimi, M. *et al.* The Effects of Synbiotic Supplementation on Metabolic Status in Women  
1255 With Polycystic Ovary Syndrome: a Randomized Double-Blind Clinical Trial. *Probiotics*  
1256 *Antimicrob. Proteins* **11**, 1355–1361 (2019).
- 1257 104. Zhou, Y., Ding, X. & Wei, H. Reproductive immune microenvironment. *J. Reprod. Immunol.*  
1258 **152**, 103654 (2022).
- 1259 105. Rüder, C. *et al.* The tumor-associated antigen EBAG9 negatively regulates the cytolytic  
1260 capacity of mouse CD8+ T cells. *J. Clin. Invest.* **119**, 2184–2203 (2009).
- 1261 106. Basta, P. *et al.* The immunohistochemical analysis of RCAS1, HLA-G, and B7H4-positive  
1262 macrophages in partial and complete hydatidiform mole in both applied therapeutic surgery  
1263 and surgery followed by chemotherapy. *Am. J. Reprod. Immunol.* **65**, 164–172 (2011).
- 1264 107. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and  
1265 binary traits. *Nat. Genet.* **53**, 1097–1103 (2021).
- 1266 108. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in  
1267 large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).
- 1268 109. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of  
1269 genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
- 1270 110. Benonisdotir, S. *et al.* Epigenetic and genetic components of height regulation. *Nat.*  
1271 *Commun.* **7**, 13490 (2016).
- 1272 111. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide  
1273 complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
- 1274 112. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation.  
1275 *Nature* **526**, 68–74 (2015).
- 1276 113. Schaffner, S. F. The X chromosome in population genetics. *Nat. Rev. Genet.* **5**, 43–51  
1277 (2004).



- 1278 114. Veeramah, K. R. & Novembre, J. Demographic events and evolutionary forces shaping  
1279 European genetic diversity. *Cold Spring Harb. Perspect. Biol.* **6**, a008516 (2014).
- 1280 115. Hinch, A. G., Altemose, N., Noor, N., Donnelly, P. & Myers, S. R. Recombination in the  
1281 human Pseudoautosomal region PAR1. *PLoS Genet.* **10**, e1004503 (2014).
- 1282 116. Creators Brendan Bulik-Sullivan hiliaryfinucane rkwalters Tim Poterba stevengazal  
1283 Oleksandr Frei1 Duncan Palmer Show affiliations 1. University of Oslo. *astheeggeggs/ldsc:*  
1284 *ldsc rg vector*. doi:10.5281/zenodo.10497662.
- 1285 117. Gallagher, C. S. *et al.* Genome-wide association and epidemiological analyses reveal  
1286 common genetic origins between uterine leiomyomata and endometriosis. *Nat. Commun.*  
1287 **10**, 4857 (2019).
- 1288 118. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the  
1289 human phenome. *Elife* **7**, (2018).
- 1290 119. Burgess, S., Thompson, S. G. & CRP CHD Genetics Collaboration. Avoiding bias from  
1291 weak instruments in Mendelian randomization studies. *Int. J. Epidemiol.* **40**, 755–764  
1292 (2011).
- 1293 120. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity  
1294 biology. *Nature* **518**, 197–206 (2015).
- 1295 121. Hartwig, F. P., Tilling, K., Davey Smith, G., Lawlor, D. A. & Borges, M. C. Bias in two-  
1296 sample Mendelian randomization when using heritable covariable-adjusted summary  
1297 associations. *Int. J. Epidemiol.* **50**, 1639–1650 (2021).
- 1298 122. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid  
1299 instruments: effect estimation and bias detection through Egger regression. *Int. J.*  
1300 *Epidemiol.* **44**, 512–525 (2015).
- 1301 123. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent Estimation in  
1302 Mendelian Randomization with Some Invalid Instruments Using a Weighted Median  
1303 Estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).
- 1304 124. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic  
1305 association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
- 1306 125. Wallace, C. Eliciting priors and relaxing the single causal variant assumption in  
1307 colocalisation analyses. *PLoS Genet.* **16**, e1008720 (2020).
- 1308 126. Schmiedel, B. J. *et al.* COVID-19 genetic risk variants are associated with expression of  
1309 multiple genes in diverse immune cell types. *Nat. Commun.* **12**, 6760 (2021).
- 1310 127. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide  
1311 association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
- 1312 128. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies  
1313 disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
- 1314 129. Fan, X. *et al.* Single-cell reconstruction of follicular remodeling in the human adult ovary.  
1315 *Nat. Commun.* **10**, 3164 (2019).
- 1316 130. Jin, C. *et al.* The regulatory landscapes of human ovarian ageing. *bioRxiv* (2022)  
1317 doi:10.1101/2022.05.18.492547.
- 1318 131. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21  
1319 (2019).
- 1320 132. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell  
1321 transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.*  
1322 **36**, 411–420 (2018).
- 1323 133. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of  
1324 single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
- 1325 134. Martin, F. J. *et al.* Ensembl 2023. *Nucleic Acids Res.* **51**, D933–D941 (2023).
- 1326 135. Karczewski, K. J. *et al.* Systematic single-variant and gene-based association testing of  
1327 thousands of phenotypes in 394,841 UK Biobank exomes. *Cell Genom* **2**, 100168 (2022).
- 1328 136. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).



- 1329 137. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in  
1330 141,456 humans. *Nature* **581**, 434–443 (2020).
- 1331 138. Liu, X., Li, C., Mou, C., Dong, Y. & Tu, Y. dbNSFP v4: a comprehensive database of  
1332 transcript-specific functional predictions and annotations for human nonsynonymous and  
1333 splice-site SNVs. *Genome Med.* **12**, 103 (2020).
- 1334 139. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the  
1335 deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–  
1336 D894 (2019).
- 1337 140. Ioannidis, N. M. *et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of  
1338 Rare Missense Variants. *Am. J. Hum. Genet.* **99**, 877–885 (2016).
- 1339 141. *vep105\_loftee*. (Github).
- 1340 142. Jaganathan, K. *et al.* Predicting Splicing from Primary Sequence with Deep Learning. *Cell*  
1341 **176**, 535–548.e24 (2019).
- 1342 143. Morales, J. *et al.* A joint NCBI and EMBL-EBI transcript set for clinical genomics and  
1343 research. *Nature* **604**, 310–315 (2022).
- 1344 144. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a  
1345 joint consensus recommendation of the American College of Medical Genetics and  
1346 Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).
- 1347 145. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer  
1348 datasets. *Gigascience* **4**, 7 (2015).
- 1349 146. Liu, Y. & Xie, J. Cauchy combination test: a powerful test with analytic p-value calculation  
1350 under arbitrary dependency structures. *J. Am. Stat. Assoc.* **115**, 393–402 (2020).
- 1351 147. Kivisild, T. *et al.* Patterns of genetic connectedness between modern and medieval  
1352 Estonian genomes reveal the origins of a major ancestry component of the Finnish  
1353 population. *Am. J. Hum. Genet.* **108**, 1792–1806 (2021).
- 1354 148. Su, W. *et al.* Role of HSD17B13 in the liver physiology and pathophysiology. *Mol. Cell.*  
1355 *Endocrinol.* **489**, 119–125 (2019).
- 1356 149. Tardif, S. *et al.* Zonadhesin is essential for species specificity of sperm adhesion to the egg  
1357 zona pellucida. *J. Biol. Chem.* **285**, 24863–24870 (2010).
- 1358 150. Fry, A. *et al.* Comparison of Sociodemographic and Health-Related Characteristics of UK  
1359 Biobank Participants With Those of the General Population. *Am. J. Epidemiol.* **186**, 1026–  
1360 1034 (2017).
- 1361 151. Fertility treatment 2021: preliminary trends and figures. [https://www.hfea.gov.uk/about-](https://www.hfea.gov.uk/about-us/publications/research-and-data/fertility-treatment-2021-preliminary-trends-and-figures/)  
1362 [us/publications/research-and-data/fertility-treatment-2021-preliminary-trends-and-figures/](https://www.hfea.gov.uk/about-us/publications/research-and-data/fertility-treatment-2021-preliminary-trends-and-figures/).
- 1363 152. Leitsalu, L. *et al.* Cohort Profile: Estonian Biobank of the Estonian Genome Center,  
1364 University of Tartu. *Int. J. Epidemiol.* **44**, 1137–1147 (2015).
- 1365 153. Sørensen, E. *et al.* Data Resource Profile: The Copenhagen Hospital Biobank (CHB). *Int. J.*  
1366 *Epidemiol.* **50**, 719–720e (2021).
- 1367 154. Nisbett, R. E. *et al.* Intelligence: new findings and theoretical developments. *Am. Psychol.*  
1368 **67**, 130–159 (2012).
- 1369 155. Rask-Andersen, M., Karlsson, T., Ek, W. E. & Johansson, Å. Modification of Heritability for  
1370 Educational Attainment and Fluid Intelligence by Socioeconomic Deprivation in the UK  
1371 Biobank. *Am. J. Psychiatry* **178**, 625–634 (2021).
- 1372 156. Balbo, N., Billari, F. C. & Mills, M. Fertility in Advanced Societies: A Review of Research: La  
1373 fécondité dans les sociétés avancées: un examen des recherches. *Eur. J. Popul.* **29**, 1–38  
1374 (2013).
- 1375 157. Petersen, B.-S., Fredrich, B., Hoepfner, M. P., Ellinghaus, D. & Franke, A. Opportunities  
1376 and challenges of whole-genome and -exome sequencing. *BMC Genet.* **18**, 14 (2017).
- 1377 158. Park, J.-H. *et al.* Distribution of allele frequencies and effect sizes and their  
1378 interrelationships for common genetic susceptibility variants. *Proc. Natl. Acad. Sci. U. S. A.*  
1379 **108**, 18026–18031 (2011).

- 1380 159. Ruth, K. S. *et al.* Using human genetics to understand the disease impacts of testosterone  
1381 in men and women. *Nat. Med.* **26**, 252–258 (2020).
- 1382 160. Tchernof, A. *et al.* Androgens and the Regulation of Adiposity and Body Fat Distribution in  
1383 Humans. *Compr. Physiol.* **8**, 1253–1290 (2018).
- 1384 161. Loh, N. Y. *et al.* Sex hormones, adiposity, and metabolic traits in men and women: a  
1385 Mendelian randomisation study. *Eur. J. Endocrinol.* **186**, 407–416 (2022).
- 1386 162. Palmer, B. F. & Clegg, D. J. The sexual dimorphism of obesity. *Mol. Cell. Endocrinol.* **402**,  
1387 113–119 (2015).
- 1388 163. Schulze, M. B. & Stefan, N. Genetic Predisposition to Abdominal Adiposity and  
1389 Cardiometabolic Risk. *JAMA: the journal of the American Medical Association* vol. 317  
1390 2334 (2017).
- 1391 164. Goossens, G. H., Jocken, J. W. E. & Blaak, E. E. Sexual dimorphism in cardiometabolic  
1392 health: the role of adipose tissue, muscle and liver. *Nat. Rev. Endocrinol.* **17**, 47–66 (2021).
- 1393 165. Agarwal, A., Mulgund, A., Hamada, A. & Chyatte, M. R. A unique view on male infertility  
1394 around the globe. *Reprod. Biol. Endocrinol.* **13**, 37 (2015).
- 1395 166. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide  
1396 range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
- 1397 167. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data.  
1398 *Nature* **562**, 203–209 (2018).
- 1399 168. Zhang, D., Dey, R. & Lee, S. Fast and robust ancestry prediction using principal component  
1400 analysis. *Bioinformatics* **36**, 3439–3446 (2020).
- 1401 169. Dey, R. & Lee, S. Asymptotic properties of principal component analysis and shrinkage-  
1402 bias adjustment under the generalized spiked population model. *J. Multivar. Anal.* **173**,  
1403 145–164 (2019).
- 1404 170. randomForest: Breiman and Cutler’s Random Forests for Classification and Regression.  
1405 *Comprehensive R Archive Network (CRAN)* [https://cran.r-](https://cran.r-project.org/web/packages/randomForest)  
1406 [project.org/web/packages/randomForest](https://cran.r-project.org/web/packages/randomForest).
- 1407 171. Fraser, A. *et al.* Cohort Profile: the Avon Longitudinal Study of Parents and Children:  
1408 ALSPAC mothers cohort. *Int. J. Epidemiol.* **42**, 97–110 (2013).
- 1409 172. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat.*  
1410 *Genet.* **48**, 1279–1283 (2016).
- 1411 173. Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in  
1412 large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
- 1413 174. Hansen, T. F. *et al.* DBDS Genomic Cohort, a prospective and comprehensive resource for  
1414 integrative and temporal analysis of genetic, environmental and lifestyle factors affecting  
1415 health of blood donors. *BMJ Open* **9**, e028401 (2019).
- 1416 175. Kong, A. *et al.* Detection of sharing by descent, long-range phasing and haplotype  
1417 imputation. *Nat. Genet.* **40**, 1068–1075 (2008).
- 1418 176. Gudbjartsson, D. F. *et al.* Large-scale whole-genome sequencing of the Icelandic  
1419 population. *Nat. Genet.* **47**, 435–444 (2015).
- 1420 177. Reisberg, S. *et al.* Translating genotype data of 44,000 biobank participants into clinical  
1421 pharmacogenetic recommendations: challenges and solutions. *Genet. Med.* **21**, 1345–1354  
1422 (2019).
- 1423 178. Finer, S. *et al.* Cohort Profile: East London Genes & Health (ELGH), a community-based  
1424 population genomics and health study in British Bangladeshi and British Pakistani people.  
1425 *Int. J. Epidemiol.* **49**, 20–21i (2020).
- 1426 179. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program.  
1427 *Nature* **590**, 290–299 (2021).
- 1428 180. Suhre, K. *et al.* Connecting genetic risk to disease end points through the human blood  
1429 plasma proteome. *Nat. Commun.* **8**, 14357 (2017).
- 1430 181. Dennis, J. K. *et al.* Clinical laboratory test-wide association scan of polygenic scores

- 1431 identifies biomarkers of complex disease. *Genome Med.* **13**, 6 (2021).  
1432 182. Thareja, G. *et al.* Differences and commonalities in the genetic architecture of protein  
1433 quantitative trait loci in European and Arab populations. *Hum. Mol. Genet.* **32**, 907–916  
1434 (2023).  
1435 183. Pott, J. *et al.* Genetic Association Study of Eight Steroid Hormones and Implications for  
1436 Sexual Dimorphism of Coronary Artery Disease. *J. Clin. Endocrinol. Metab.* **104**, 5008–  
1437 5023 (2019).  
1438 184. Prins, B. P. *et al.* Genome-wide analysis of health-related biomarkers in the UK Household  
1439 Longitudinal Study reveals novel associations. *Sci. Rep.* **7**, 11008 (2017).  
1440 185. Kuhn, R. M., Haussler, D. & Kent, W. J. The UCSC genome browser and associated tools.  
1441 *Brief. Bioinform.* **14**, 144–161 (2013).  
1442 186. hail 0.2 documentation. <https://hail.is/docs/0.2/>.