

1 Genome-wide analysis identifies 66 variants  
2 underlying anatomical variation in human  
3 neuroendocrine structures and reveals links  
4 to testosterone

5 Hannah Currant<sup>1,2,3†</sup>, Christopher Arthofer<sup>4†</sup>, Teresa  
6 Ferreira<sup>1,2</sup>, Gwenaelle Douaud<sup>4</sup>, Barney Hill<sup>1,2</sup>, Samvida S  
7 Venkatesh<sup>2,5</sup>, Nikolas A Baya<sup>2,5</sup>, Duncan S Palmer<sup>1,2</sup>, Saskia  
8 Reibe<sup>1,2</sup>, Anje Moltke-Prehn<sup>6</sup>, Tune H Pers<sup>6</sup>, Andreas  
9 Bartsch<sup>7</sup>, Jesper Andersson<sup>4</sup>, Margaret F Lippincott<sup>8,9</sup>, Yee-  
10 Ming Chan<sup>10</sup>, Stephanie B Seminara<sup>9,8</sup>, Thomas E  
11 Nichols<sup>2,4</sup>, Christoffer Nellaker<sup>2</sup>, Stephen Smith<sup>4</sup>, Søren  
12 Brunak<sup>3</sup>, Frederik J Lange<sup>4†</sup> and Cecilia M Lindgren<sup>2,5†</sup>

13 <sup>1</sup>Nuffield Department for Population Health, University of  
14 Oxford, Old Road Campus, Oxford, OX3 7LF, UK.

15 <sup>2</sup>Big Data Institute, University of Oxford, Old Road Campus,  
16 Oxford, OX3 7LF, UK.

17 <sup>3</sup>Novo Nordisk Foundation Center for Protein Research,  
18 University of Copenhagen, Blegdamsvej, Copenhagen, 2200,  
19 Denmark.

20 <sup>4</sup>Oxford Centre for Functional MRI of the Brain, University of  
21 Oxford, John Radcliffe Hospital, Oxford, OX3 9DU, UK.

22 <sup>5</sup>Wellcome Centre for Human Genetics, University of Oxford, Old  
23 Road Campus, Oxford, OX3 7LF, UK.

24 <sup>6</sup>Novo Nordisk Foundation Center for Basic Metabolic Research,  
25 University of Copenhagen, Blegdamsvej, Copenhagen, 2200,  
26 Denmark.

27 <sup>7</sup>Department of Neuroradiology, University of Heidelberg, Im  
28 Neuenheimer Feld, Heidelberg, 69120, Germany.

29 <sup>8</sup>Reproductive Endocrine Unit, Massachusetts General Hospital,  
30 Fruit Street Gray, Boston, 02114, MA, USA.

31 <sup>9</sup>Harvard Center for Reproductive Medicine, Massachusetts  
32 General Hospital, Fruit Street Gray, Boston, 02114, MA, USA.

33 <sup>10</sup>Division of Endocrinology, Boston's Children's Hospital,  
34 Longwood Avenue, Boston, 02115, MA, USA.

35 Contributing authors: [hannah.currant@ndph.ox.ac.uk](mailto:hannah.currant@ndph.ox.ac.uk);  
36 [christoph.arthofer@ndcn.ox.ac.uk](mailto:christoph.arthofer@ndcn.ox.ac.uk); [teresa.ferreira@bdi.ox.ac.uk](mailto:teresa.ferreira@bdi.ox.ac.uk);  
37 [gwenaelle.douaud@ndcn.ox.ac.uk](mailto:gwenaelle.douaud@ndcn.ox.ac.uk); [Barney.Hill@ndph.ox.ac.uk](mailto:Barney.Hill@ndph.ox.ac.uk);  
38 [samvida.venkatesh@univ.ox.ac.uk](mailto:samvida.venkatesh@univ.ox.ac.uk); [nikolas.baya@pmb.ox.ac.uk](mailto:nikolas.baya@pmb.ox.ac.uk);  
39 [Duncan.Palmer@ndph.ox.ac.uk](mailto:Duncan.Palmer@ndph.ox.ac.uk); [saskia.reibe-pal@bdi.ox.ac.uk](mailto:saskia.reibe-pal@bdi.ox.ac.uk);  
40 [anjamj@sund.ku.dk](mailto:anjamj@sund.ku.dk); [tune.pers@sund.ku.dk](mailto:tune.pers@sund.ku.dk);  
41 [bartsch@radvisory.net](mailto:bartsch@radvisory.net); [jesper.andersson@ndcn.ox.ac.uk](mailto:jesper.andersson@ndcn.ox.ac.uk);  
42 [MLIPPINCOTT@mgh.harvard.edu](mailto:MLIPPINCOTT@mgh.harvard.edu);  
43 [Yee-Ming.Chan@childrens.harvard.edu](mailto:Yee-Ming.Chan@childrens.harvard.edu);  
44 [SSEMINARA@mgh.harvard.edu](mailto:SSEMINARA@mgh.harvard.edu); [thomas.nichols@bdi.ox.ac.uk](mailto:thomas.nichols@bdi.ox.ac.uk);  
45 [christoffer.nellaker@wrh.ox.ac.uk](mailto:christoffer.nellaker@wrh.ox.ac.uk); [stephen.smith@ndcn.ox.ac.uk](mailto:stephen.smith@ndcn.ox.ac.uk);  
46 [soren.brunak@cpr.ku.dk](mailto:soren.brunak@cpr.ku.dk); [frederik.lange@dtc.ox.ac.uk](mailto:frederik.lange@dtc.ox.ac.uk);  
47 [cecilia.lindgren@bdi.ox.ac.uk](mailto:cecilia.lindgren@bdi.ox.ac.uk);

48 †These authors contributed equally to this work.

## 49 Abstract

50 The hypothalamus, pituitary gland and olfactory bulbs are neuroanatomical  
51 structures key to the regulation of the endocrine system. Variation  
52 in their anatomy can affect the function of the reproductive system.  
53 To investigate this relationship, we extracted four largely unexplored  
54 phenotypes from 34,834 individuals within UK Biobank by quantifying  
55 the volume of the hypothalamus, pituitary gland and olfactory bulbs  
56 using multi-modal magnetic resonance imaging. Genome-wide association  
57 studies of these phenotypes identified 66 independent common  
58 genetic associations with endocrine-related neuroanatomical volumes  
59 ( $P < 5 \times 10^{-8}$ ), five of which had a prior association to testosterone  
60 levels, representing enrichment of testosterone-associated SNPs  
61 over random chance ( $P$ -value =  $9.89 \times 10^{-12}$ ). Exome-wide rare  
62 variant burden analysis identified *STAB1* as being significantly associated  
63 with hypothalamus volume ( $P = 3.78 \times 10^{-7}$ ), with known  
64 associations to brain iron levels. Common variants associated with  
65 hypothalamic grey matter volume were also found to be associated  
66 with iron metabolism, in which testosterone plays a key role. These  
67 results provide initial evidence of common and rare genetic effects on  
68 both anatomical variation in neuroendocrine structures and their function  
69 in hormone production and regulation. Variants associated with  
70 pituitary gland volume were enriched for gene expression specific to

71 theca cells, responsible for testosterone production in ovaries, suggest-  
72 ing shared underlying genetic variation affecting both neuroanatomical  
73 and gonadal endocrine tissues. Cell-type expression enrichment analysis  
74 across hypothalamic cell types identified tanycytes to be associated ( $P =$   
75  $1.69 \times 10^{-3}$ ) with olfactory bulb volume associated genetic variants, a  
76 cell type involved in release of gonadotropin-releasing hormone into the  
77 bloodstream. Voxel-wise analysis highlighted associations between the  
78 variants associated with pituitary gland volume and areas of intracranial  
79 venous drainage involved in hormonal release into the blood circulation.  
80 Together, our results suggest a shared role of genetics impacting both  
81 the anatomy and function of neuroendocrine structures within the repro-  
82 ductive system in their production and release of reproductive hormones.

83 **Keywords:** endocrinology, brain anatomy, image-derived phenotypes, genetic  
84 association studies

## 85 Introduction

86 The endocrine system is integral to reproductive function, with its disruption  
87 impacting development, fertility and broader disease. The intricate task of reg-  
88 ulating the endocrine system is orchestrated by neuroanatomical structures.  
89 Three neuroanatomical structures are of key importance: (1) the hypothala-  
90 mus, which contains specialised neurons that produce gonadotropin releasing  
91 hormone (GnRH); (2) the pituitary gland, which responds to GnRH and aids  
92 regulation of hormone production [1]; and (3) the olfactory bulbs, along which  
93 the GnRH releasing neurons migrate during fetal development [2]. The pitu-  
94 itary gland, when stimulated by GnRH, produces luteinising hormone (LH)  
95 and follicle stimulating hormone (FSH) which, in turn, regulate the production  
96 of other sex hormones such as testosterone, progesterone and oestrogen [3].  
97 Collectively, these hormones control reproductive processes including puberty,  
98 fertility, menstruation and menopause.

99 Variation in the anatomy (specifically, the size/volume) of these three neu-  
100 roanatomical structures occurs within the population and has been reported to  
101 be associated with variation in their function within the reproductive system  
102 across a spectrum of severity. For example, population-level common varia-  
103 tion in the volume of the pituitary gland has been found to be correlated with  
104 sex-steroid concentration [4]. At the more pathological end of the spectrum,  
105 individuals with Kallmann syndrome exhibit hypoplastic or absent olfactory  
106 bulbs, alongside additional characteristics including delayed or absent onset of  
107 puberty, infertility and anosmia [5].

108 Magnetic resonance imaging (MRI) offers the opportunity to accurately  
109 characterise anatomical variation in brain structure. Further, biobanks con-  
110 taining such imaging data allow us to do so at a unprecedented scale. Novel  
111 and advancing methodologies enable quantification of anatomical variation at

112 such scale, offering opportunities for well-powered statistical analyses of sub-  
113 tle differences [6]. Image-derived phenotypes (IDPs), such as the volumes of  
114 anatomical structures, have proven a valuable resource for enabling genetic dis-  
115 covery across numerous biological systems including the cardiac [7], ophthalmic  
116 [8], and neurological [6] systems. By extracting low-dimensional phenotypes  
117 from high-dimensional data (MRI images), based on *a priori* knowledge of bio-  
118 logical relationships, there is the opportunity to increase statistical discovery  
119 power in genetic association studies.

120 Examining the genetic contributors to anatomical variation in these neuro-  
121 endocrine structures provides an opportunity to further our understanding  
122 of their underlying biology and their function. In doing so, we aim to eluci-  
123 date their connection to the broader endocrine and reproductive systems. To  
124 do this, we extracted quantitative IDPs measuring structure volume from MRI  
125 images of 34,834 individuals within UK Biobank (UKB), representative of the  
126 European subset (Methods) of individuals with MRI available. We performed  
127 common-variant genome- and rare-variant exome-wide association studies of  
128 these neuroendocrine IDPs. This is, to our knowledge, the first population-  
129 scale study of IDPs relating to these neuroendocrine structures and the largest  
130 study of genetic effect on their anatomy by far. We hypothesise that use of  
131 these IDPs at unprecedented scale will allow for discovery of novel genetic vari-  
132 ants affecting neuroanatomy, and additional understanding of links between  
133 the neurological and endocrine systems through common biological pathways.

## 134 Results

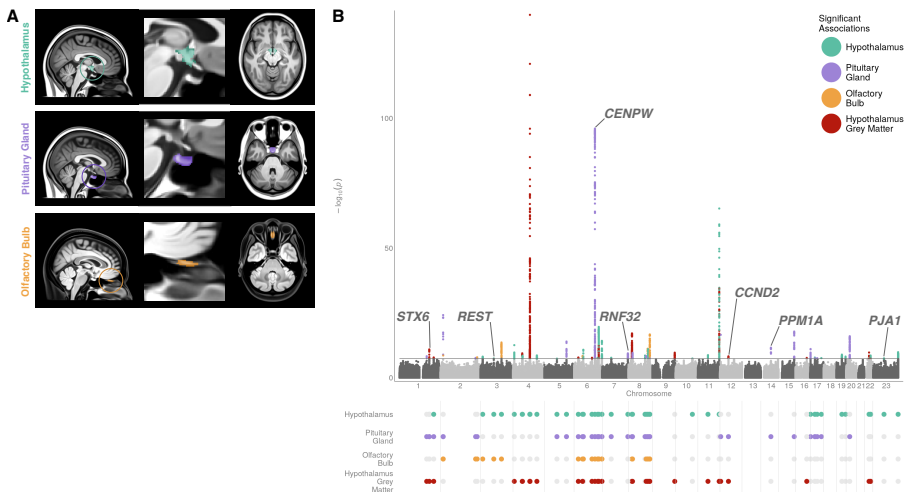
### 135 Extraction of phenotypes describing neuroendocrine 136 anatomy

137 The volumes of the hypothalamus, pituitary gland and olfactory bulbs (com-  
138 bined left and right) were extracted with a label propagation approach from the  
139 MRI images of 34,834 individuals following quality control (Fig. 1A). Hypotha-  
140 lamic grey matter volume was extracted directly from grey matter probability  
141 maps (see Methods for details of image-derived phenotype extraction). All  
142 measurements were taken as absolute volumes (without normalisation for  
143 brain size) while accounting for partial volumes after label propagation (see  
144 Methods 10). The hypothalamus is the largest of the four volumes (mean =  
145  $466.16\text{mm}^3 \pm 50.84$  (SD)), followed by the hypothalamic grey matter (mean  
146 =  $456.31\text{mm}^3 \pm 62.06$ ), pituitary gland (mean =  $268.02\text{mm}^3 \pm 64.15$ ) and olfac-  
147 tory bulbs (mean =  $62.68\text{mm}^3 \pm 17.55$ ). Across all four IDPs, males had higher  
148 volumes than females, in keeping with prior knowledge [9] (Fig. A1, Table  
149 A1). We adjusted the IDPs for a number of variables including age, sex and  
150 head size (for complete details see genome-wide association studies Methods)  
151 to carry forward for further analysis.



## 152 Genetic variation associated with anatomical variation in 153 neuroendocrine structures

154 We conducted the largest to-date genome-wide association studies (GWAS) of  
155 four under-explored IDPs describing the volume of the neuroendocrine struc-  
156 tures across ten million SNPs (minor allele frequency (MAF)  $>0.01$ ). We  
157 identified 66 independent lead variants significantly associated ( $P < 5 \times 10^{-8}$ )  
158 with one of the four neuroendocrine IDPs in a sex-combined sample: 26 asso-  
159 ciated with hypothalamus volume, 16 with pituitary gland volume, 6 with  
160 olfactory bulb volume, and 18 with hypothalamic grey matter volume (Fig.  
161 1B, Table A2). Application of LD-score regression (LDSC) [10] showed all  
162 GWAS were controlled for residual population structure (Fig. A2, Table A4,  
163 maximum LDSC intercept = 1.02, minimum LDSC intercept  $P = 0.02$ ).



**Fig. 1** GWAS of the volume of four neuroendocrine structures (hypothala-  
mus, pituitary gland, olfactory bulb, and hypothalamic grey matter volume)  
extracted from MRI images A) Atlas consisting of a standard brain template with the  
endocrine-associated neuroanatomical segmentations overlaid in colour, as circled in the left  
of the plot. B) The upper panel is a Manhattan plot presenting the result of four GWAS of  
hypothalamus (turquoise), pituitary gland (purple), mean of left and right olfactory bulbs  
(gold), and hypothalamic grey matter (red) volume, in a sex-combined sample. Each point  
represents a genetic variant with the  $x$ -axis representing the genomic position within each  
chromosome and the  $y$ -axis representing the  $-\log_{10}(P\text{-value})$  of the association in an additive  
linear model. Variants with a prior association to testosterone or menstrual timing are  
highlighted and their gene name listed. The lower panel is an upset plot representing the  
overlap of the signals across the GWAS of the four neuroendocrine traits. The  $x$ -axis  
represents the genomic position, aligned with the Manhattan plot. The presence of an overlap  
between the signal is shown with a coloured dot, no overlap is shown with a grey dot.

164 In five cases, using conditional joint analysis using GCTA-COJO [11], we  
165 identified two or more independent signals within the same locus. Although  
166 lead SNPs were unique to each phenotype, conditional analysis across GWAS

167 identified a number of loci with overlapping genetic association signal across  
168 the four IDPs (Fig. 1B). Genome-wide SNP-based heritability ( $h_{\text{SNP}}^2 \pm$  stan-  
169 dard error (SE)), was calculated [10] for hypothalamus ( $0.25 \pm 0.02$ ), pituitary  
170 gland ( $0.20 \pm 0.03$ ), olfactory bulb ( $0.11 \pm 0.02$ ), and hypothalamic grey matter  
171 volume ( $0.25 \pm 0.02$ ).

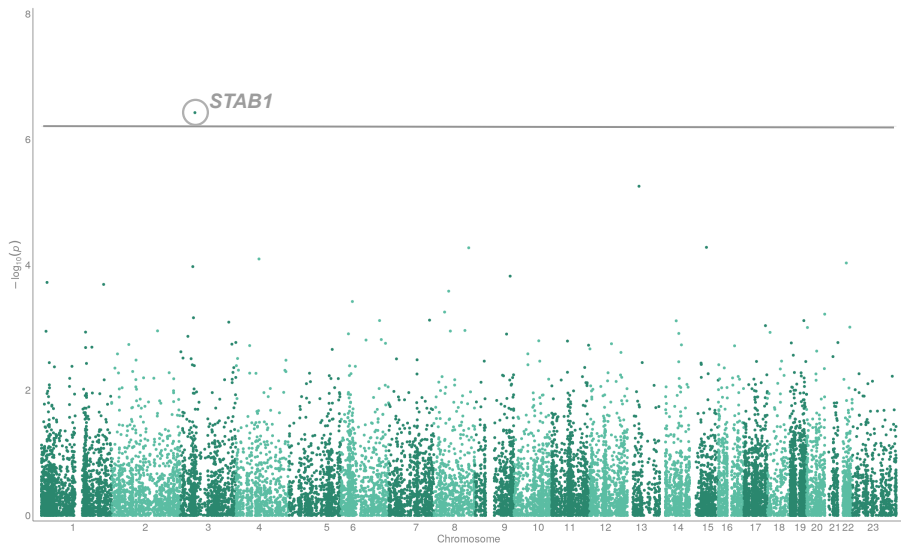
## 172 Sexual dimorphism in genetic effect

173 We observed sexual dimorphism in sex-stratified GWAS (18,487 females and  
174 16,347 males, Fig. A3): among lead variants significantly associated ( $P <$   
175  $5 \times 10^{-8}$ ) with each of the four neuroendocrine IDPs in the sex-combined,  
176 female or male populations, four variants displayed a statistically significant  
177 difference ( $P < 5.5 \times 10^{-4}$ ) in effect size between females and males (Fig. A4A,  
178 Table A5). These were: rs6544040, in an intronic region of *VIT* (hypothala-  
179 mous volume females:  $\beta = 0.07$  and SE = 0.01, males:  $\beta = -0.002$  and  
180 SE = 0.01; sex difference  $P = 1.31 \times 10^{-5}$ ); rs186990314, an intronic variant  
181 in *SMR3B* (pituitary gland volume females:  $\beta = 0.001$  and SE = 0.02, males:  
182  $\beta = -0.09$  and SE = 0.02; sex difference  $P = 4.22 \times 10^{-5}$ ); rs7749444, an  
183 intronic variant near *PRIM2BP* (olfactory bulb volume females:  $\beta = 0.01$  and  
184 SE = 0.01, males:  $\beta = 0.06$  and SE = 0.01; sex difference  $P = 1.17 \times 10^{-3}$ );  
185 and rs144968764, an intronic variant near *MND1* (hypothalamic grey matter  
186 volume females:  $\beta = -0.24$  and SE = 0.04, males:  $\beta = 0.01$  and SE = 0.05; sex  
187 difference  $P = 6.65 \times 10^{-5}$ ). Common genetic variants in *VIT* have been asso-  
188 ciated with cortical surface area asymmetry [12]. Additionally, comparison of  
189 the SNP-based heritability [10] of each of the traits across the sexes showed a  
190 nominally significant difference in the SNP-heritability of olfactory bulb vol-  
191 ume ( $t$ -test  $P = 0.04$ , Fig. A4B), with females having a higher heritability  
192 ( $h_{\text{SNP}}^2 = 0.16 \pm 0.03$ ) than males ( $h_{\text{SNP}}^2 = 0.07 \pm 0.03$ ).

## 193 Rare variant association testing across the exome

194 To assess the rare variant contribution to neuroendocrine volumes, we car-  
195 ried out rare-variant association studies (RVAS) across the exome at the  
196 variant- ( $N=32,537,048$ ) and gene-level ( $N=19,489$ ) for each of the four neu-  
197 roendocrine IDPs. Following multiple testing correction ( $P < 6.25 \times 10^{-7}$ ),  
198 rare damaging variation in a single gene, *STAB1*, was found to be associated  
199 (Cauchy combination test [13] of all damaging variants across all MAF thresh-  
200 olds,  $P = 3.78 \times 10^{-7}$ ) with hypothalamus volume (Fig. 2, Methods). Within  
201 *STAB1* we found a particularly strong association ( $P=7.73 \times 10^{-8}$ ) subset-  
202 ting to ultra-rare ( $MAF \leq 1 \times 10^{-4}$ ) pLoF or damaging missense variants  
203 with hypothalamus volume. No common variants ( $MAF > 0.01$ ) within this  
204 gene were significantly associated with hypothalamus volume in our GWAS.  
205 GWAS have previously found variants mapped to this gene to be associated  
206 with vertex-wise sulcal depth [14] and cortical thickness [15]. *STAB1* encodes  
207 stabilin 1, a type 1 transmembrane receptor [16] which has been associated  
208 with brain iron levels [17] and functional brain measurements [18]. Brain iron

209 levels and homeostasis are known to play a role in several neurological diseases  
210 including Parkinson's [19], for which *STAB1* is also a gene of interest [20].

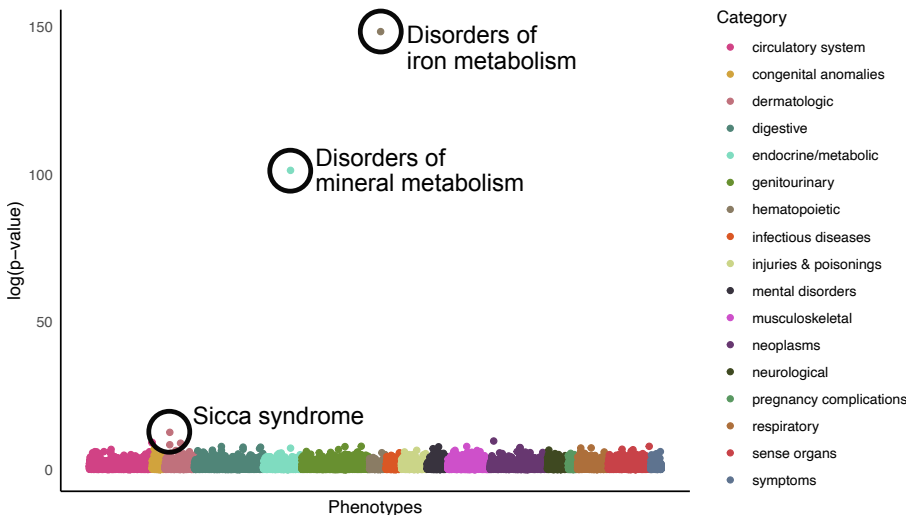


**Fig. 2 RVAS of hypothalamus volume.** A Manhattan plot of RVAS of hypothalamus volume in a sex-combined sample using a cauchy combination test using damaging gene burden tests. Variants in odd and even chromosomes are coloured in dark and light green. The  $y$ -axis is the  $-\log_{10}(P\text{-values})$  of association. The exome-wide significant ( $P < 6.25 \times 10^{-7}$ ) gene, *STAB1* association is highlighted.

## 211 Phenome-wide association analyses across clinical 212 phenotypes

213 To explore the overlap between the 66 lead variants identified in our GWAS  
214 and a range of clinical phenotypes, we carried out phenome-wide association  
215 studies (PheWAS) across 1839 phecodes in the UKB by mapping International  
216 Statistical Classification of Diseases and Related Health Problems 10th Revision  
217 (ICD10) codes (Methods). We chose to use phecodes to group ICD10 codes  
218 as they are designed to capture clinically meaningful concepts for research and  
219 tend to comprise a collection of ICD10 codes, resulting in larger case counts  
220 and so increased power compared to two-digit ICD10 codes (Fig. A5, Tables  
221 A6, A7, A8, A9) [21]. Following Benjamini-Hochberg (BH) [22] false discovery  
222 rate (FDR) multiple test correction, three phenotypes were significantly asso-  
223 ciated (BH FDR  $P < 0.05$ ) with hypothalamic grey matter associated SNPs  
224 (Fig. 3): disorders of iron metabolism ( $P = 3.27 \times 10^{-65}$ ), disorders of min-  
225 eral metabolism ( $P = 8.99 \times 10^{-45}$ ) and sicca syndrome ( $P = 3.48 \times 10^{-6}$ ).  
226 As highlighted following rare-variant association testing, iron homeostasis is  
227 integral to brain function with its dysfunction seen in several neurological  
228 diseases. Additionally, diseases of iron overload in males are associated with

229 reproductive dysfunction including hypogonadotropic hypogonadism and infer-  
230 tility [23]. Further, testosterone levels are known to be associated with iron  
231 levels, with testosterone playing a key role in erythropoiesis and thus iron  
232 incorporation and homeostasis [24]. Meanwhile, sicca syndrome is a condition  
233 affecting production of fluids including saliva and tears. It has previously been  
234 shown to have an association with oestrogen levels [25] and is most commonly  
235 diagnosed in postmenopausal women [26]. These associations could suggest  
236 a relationship between the structures producing reproductive hormones and  
237 their downstream effects across biological systems.



**Fig. 3 PheWAS of 18 significant GWAS associations** Scatter plot displaying the  $-\log_{10}(P\text{-values})$  from a PheWAS of 1839 phecodes extracted from UKB (Methods) and the 18 significant lead variants identified as associated with the volume of the hypothalamic grey matter. Statistically significant associations following Benjamini-Hochberg multiple test correction ( $P < 0.05$ ) are circled in black.

## 238 **Overlap of neuroendocrine genetics with neurological** 239 **phenotypes**

240 Ten of the 66 variants associated with the neuroendocrine IDPs have previ-  
241 ously been found to be significantly ( $P < 5 \times 10^{-8}$ ) associated with other brain  
242 phenotypes including substructure volumes, cortical thickness and brain con-  
243 nectivity (Table A3). To further explore the shared genetics underpinning brain  
244 structure phenotypes, we performed a PheWAS using Big40 [27], a database of  
245 GWAS summary statistics of  $\sim 4,000$  multi-modal MRI-derived phenotypes  
246 within the UKB. This identified significant associations ( $P < 1.93 \times 10^{-7}$ ,  
247 Bonferroni multiple testing correction applied accounting for 259,710 tests)  
248 with 823 of the neurological MRI IDPs (Fig. A6, Tables A10, A11, A12 &  
249 A13). Neuroendocrine IDP-associated genetic variants discovered in our study

250 were associated with 226 regional and tissue volumes and intensities. The  
251 most significant associations were between rs13107325 (a missense variant in  
252 *SLC39A8*) and the volume of the left ( $\beta = 0.43$ ,  $SE = 0.02$ ,  $P = 2.09 \times 10^{-175}$ )  
253 and right ventral striatum ( $\beta = 0.41$ ,  $SE = 0.02$ ,  $P = 2.51 \times 10^{-165}$ ), the vol-  
254 ume of the left ( $\beta = 0.39$ ,  $SE = 0.02$ ,  $P = 8.91 \times 10^{-147}$ ) and right ( $\beta = 0.40$ ,  
255  $SE = 0.02$ ,  $P = 5.25 \times 10^{-155}$ ) putamen, and the T1 intensity within the right  
256 hemisphere accumbens area ( $\beta = -0.40$ ,  $SE = 0.02$ ,  $P = 1.95 \times 10^{-146}$ ). The  
257 overlap in associations of genetic variants with the neuroendocrine IDPs and  
258 broader anatomical brain phenotypes is to be expected due to shared devel-  
259 opmental processes and common cell types, and has been seen between other  
260 brain phenotypes previously [6].

## 261 **Overlap of neuroendocrine genetics with testosterone** 262 **and other reproductive phenotypes**

### 263 **Testosterone related genetic variants**

264 Fourteen variants (21%) that we observed to be associated with the neuroen-  
265 docrine IDPs have been previously reported to be associated with reproductive  
266 traits (Table A3). Notably, there was a significant enrichment within the vari-  
267 ants associated with the IDPs for prior associations to testosterone ( $P$ -value  
268 =  $9.89 \times 10^{-12}$ ). Five loci have prior associations with testosterone: firstly,  
269 rs11941568 (hypothalamus volume  $\beta = 0.04$ ,  $SE = 0.01$ ,  $MAF = 0.40$ ), an  
270 upstream variant of *REST* which is a transcriptional repressor that represses  
271 neuronal genes in non-neuronal tissues [28]. Secondly, a downstream intronic  
272 variant of *CENPW* (rs2184968, pituitary gland volume  $\beta = 0.16$ ,  $SE = 0.007$ ,  
273  $MAF = 0.45$ ) which is also associated with male-pattern baldness [29], a phe-  
274 notype known to be largely controlled by testosterone [30]. Thirdly, rs76895963  
275 (pituitary gland volume  $\beta = 0.24$ ,  $SE = 0.029$ ,  $MAF = 0.02$ ) in an intronic  
276 region of *CCND2* has a prior association with testosterone [31]. rs76895963  
277 colocalised with an eQTL for *CCND2*, enriched in tissues including the pitu-  
278 itary gland, thyroid, tibial nerve and cerebellum [32]. *CCND2* deficient mice  
279 have hypoplastic testes in males and are sterile in females [33] with *CCND2*  
280 expression mirroring blood testosterone levels [34]. Further, it is thought that  
281 *CCND2* might help regulate spermatogenesis [34] and be involved in impotence  
282 [35]. Fourthly, a genetic variant upstream of *RNF32* (rs7808966, pituitary  
283 gland volume  $\beta = 0.03$ ,  $SE = 0.01$ ,  $MAF = 0.07$ ), a gene expressed in both  
284 the ovaries and testis, and expressed during spermatogenesis [36]. Fifthly,  
285 rs10220706 (pituitary gland volume  $\beta = -0.06$ ,  $SE = 0.01$ ,  $MAF = 0.30$ ),  
286 downstream of *PPM1A*, has prior associations to testosterone and also to age  
287 at menarche [29]. *PPM1A* has been shown to be involved in regulation of  
288 testosterone synthesis in Leydig cells within the testes [37].

## 289 Reproductive development and ageing-related genetic 290 variants

291 We also observed an overlap of variants found to be associated with the neuro-  
292 endocrine IDPs with the genetics of menstrual timing: both age at menarche  
293 and age at menopause (Table A3). For example, a missense variant on chromo-  
294 some X in *PJA1* (rs5937160, hypothalamus volume  $\beta = -0.04$ , SE = 0.01,  
295 MAF = 0.23) is associated with age at menarche in females, and the relative  
296 age of first facial hair and the relative age at which one's voice broke in males  
297 [29]. Additionally, a locus located in *STX6*, associated with hypothalamus vol-  
298 ume (rs1044595,  $\beta = -0.04$ , SE = 0.007, MAF = 0.40) and hypothalamic grey  
299 matter volume (rs35306826,  $\beta = -0.05$ , MAF = 0.41), has prior association to  
300 menopause status and the age at menopause, in addition to prior associations  
301 to other brain anatomy phenotypes [27, 38].

## 302 Genetic relationships with reproductive traits

303 We evaluated genetic correlations ( $r_g$ ) between the neuroendocrine IDPs and  
304 reproductive hormone (testosterone, FSH, LH, oestradiol and progesterone  
305 [39]) levels (Fig. A14). We found nominally significant  $r_g$  ( $P < 0.05$ ) between  
306 testosterone level and pituitary gland volume in a sex-combined analysis ( $P =$   
307  $0.03$ ,  $r_g = 0.07$ ), and between testosterone levels and both hypothalamus ( $P =$   
308  $9.9 \times 10^{-3}$ ,  $r_g = 0.11$ ) and olfactory bulb volume ( $P = 0.01$ ,  $r_g = 0.13$ )  
309 in females. We also assessed  $r_g$  between the neuroendocrine IDPs and six  
310 infertility phenotypes - five female infertility phenotypes (all cause, anatomical  
311 cause, anovulatory, unknown cause excluding idiopathic causes, and unknown  
312 cause including idiopathic fertility) and a male infertility phenotype [39]. None  
313 of the phenotype pairs displayed significant genetic correlation ( $P < 0.05$ ).

314 We carried out bidirectional Mendelian randomisation analysis to investi-  
315 gate causality between the neuroendocrine IDPs, and the reproductive  
316 hormone and infertility phenotypes outlined above. Phenotypes with at least  
317 five genetic instruments were used as exposures. We found no evidence of a  
318 causal relationship between the neuroendocrine IDPs and any of the repro-  
319 ductive phenotypes following multiple test correction ( $P < 4.17 \times 10^{-4}$  for all  
320 tests, Table A15).

321 We did not find any statistically significant evidence of colocalisation  
322 between genetic variants found to be associated with one of the neuroendocrine  
323 IDPs and reproductive hormones or fertility phenotypes (all  $P(\frac{H_4}{H_3}) < 5$  &  
324  $P(H_4) < 0.5$ ).

## 325 Tissue and pathway enrichment

326 The neuroendocrine structures studied here are involved in the production of  
327 reproductive hormones that act across many tissues. We therefore explored the  
328 colocalisation between the 66 neuroendocrine IDP genetic association signals  
329 with the tissue-specific expression quantitative trait loci (eQTLs) within the  
330 GTEx dataset [32]. We found that three genetic variants colocalised ( $P(\frac{H_4}{H_3}) >$

331 5 &  $P(H_4) > 0.5$ ) with eQTLs for genes known to be involved in sperm-  
332 oocyte fusion: rs4806665 with *TMEM190* enriched in the testes and lung [40],  
333 rs7550273 with *CD46* enriched in the oesophagus mucosa [41], and rs10800397  
334 with *SPATA46* enriched in the brain cerebellum [42]. Six genetic variants also  
335 colocalised with eQTLs for genes enriched for expression in regions of the brain  
336 including the amygdala, cerebellum and cortex (Table A16).

337 Furthermore, we performed gene-set and tissue enrichment analyses using  
338 DEPICT [43, 44]. We found an association between the genetic variants affect-  
339 ing hypothalamus volume in females and the cardiovascular system ( $P =$   
340  $5.54 \times 10^{-4}$ , Fig. A7, Tables A17, A18 & A19).

## 341 **Enrichment for gene expression in ovarian cell-types**

342 To test for enrichment of gene expression at the ovarian cell type level, we  
343 used an in-house single-cell ovary atlas consisting of a combination of pub-  
344 licly available and in-house single-cell RNA-seq datasets (paper in preparation,  
345 Ferreira *et al.*, see Methods for further details). Using CELLECT [45], the  
346 in-house annotated single-cell RNA-seq count data was combined with our neu-  
347 roendocrine IDP GWAS summary statistics to identify associated candidate  
348 aetiological cell types.

349 We identified statistically significant ( $P < 0.05$  after BH FDR correc-  
350 tion) enrichment of signal in each of the four neuroendocrine IDPs in gene  
351 regions specifically expressed in five ovarian cell types (Fig. 4A, Table A20).  
352 For example, the pituitary gland associated genetic variants discovered in the  
353 sex-combined sample were enriched for expression in genes specific to the  
354 early theca cells, theca interna, theca externa and the cumulus-oocyte com-  
355 plex. Variants associated with hypothalamus, hypothalamic grey matter and  
356 olfactory bulb volume, all in the sex-combined sample, and with hypothala-  
357 mus volume in the female only sample, were enriched for expression in theca  
358 externa cell genes. Additionally variants associated with hypothalamic grey  
359 matter and olfactory bulb volume in the sex-combined sample were enriched  
360 for early theca cell expressed genes. Theca cells are key to the endocrine  
361 function of the ovaries, responding to LH released by the pituitary gland to  
362 generate testosterone and precursors of oestradiol [46]. We also found sig-  
363 nificant enrichment of genes expressed in the surface epithelium for genetic  
364 variants associated with hypothalamic grey matter and olfactory bulb volume  
365 in the sex-combined sample, and hypothalamus volume in the sex-combined  
366 and female only sample. The ovarian surface epithelium has both FSH and  
367 LH receptors and gonadotropins have been shown to stimulate ovarian surface  
368 epithelium cell proliferation across several species. Additionally, GnRH acts as  
369 an autocrine growth inhibitor for the ovarian surface epithelium [47].



## 370 **Enrichment for gene expression in hypothalamic** 371 **cell-types**

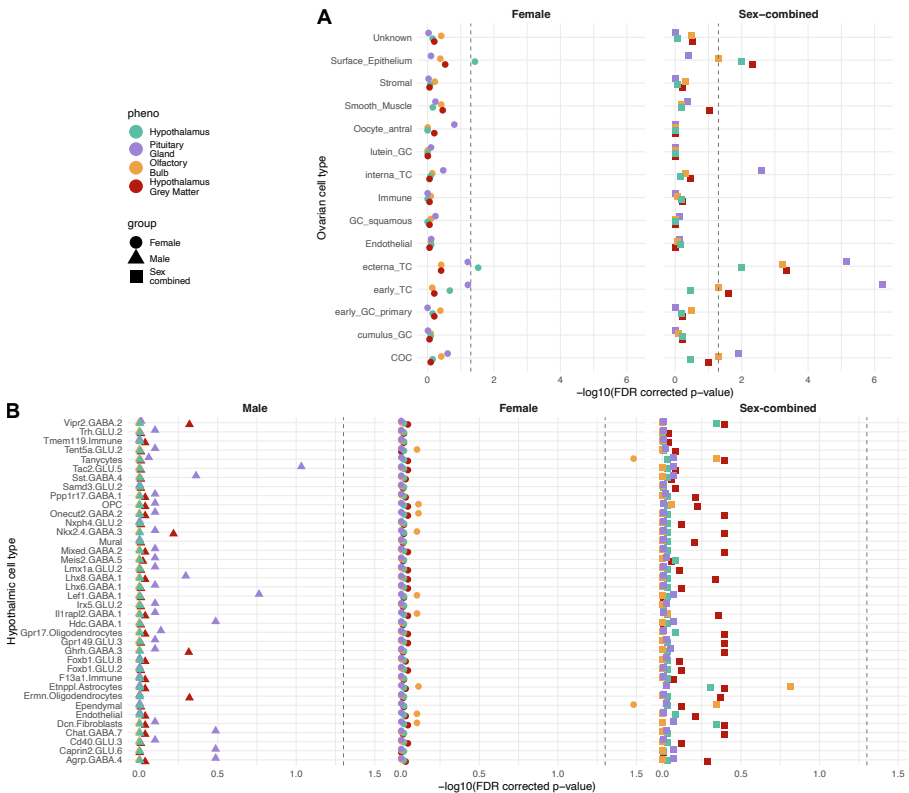
372 Using CELLECT [45] applied to published single-nucleus RNA-sequencing  
373 data from adult hypothalamus [48], we tested whether the genomic location  
374 of genes specifically expressed in any of 38 hypothalamic cell populations were  
375 enriched for association signals for any of the neuroendocrine IDPs (Table  
376 A21). Stratifying analysis based on variants identified in female, male and  
377 sex-combined GWASs, we identified two cell types exhibiting significant enrich-  
378 ment ( $P < 0.05$  after BH FDR correction) for olfactory bulb volume in the  
379 female sample GWAS signal, namely tanycytes and ependymal cells (Fig.4B),  
380 two related cell types residing along the ventricles in the mediobasal hypothala-  
381 mus. Tanycytes are a subtype of ependymal cells found in the third ventricle  
382 that project into the hypothalamus. There is evidence that tanycytes are  
383 involved in the release of GnRH due to the overlap in spatial distribution of the  
384 tanycytes and GnRH-releasing neurons [49]. Studies have shown that removal  
385 of tanycytes in rats impairs the release pulse of GnRH into the portal blood,  
386 and generation of the LH peak required for ovulation [50].

## 387 **Voxel-level genetic associations**

388 We ran linear models at voxel-level to explore the effect of each of the genetic  
389 variants discovered during the four sex-combined GWAS in the rest of the  
390 brain (i.e., outside of the original neuroendocrine structures). The Jacobian  
391 determinant maps estimated from the nonlinear warp, aligning each individual  
392 to the standard template brain, were used as a measure of relative volume dif-  
393 ferences (expansion or contraction) at each voxel. The allele dosage for each of  
394 the 66 identified loci was regressed onto the Jacobian determinant maps, with  
395 the linear models including the same confound regressors used in the GWAS.  
396 Summary maps were calculated across the output t-statistic maps (one for  
397 each SNP) associated with each of the four IDPs, to represent for each voxel  
398 the proportion of discovered loci strongly associated ( $t > 3.5$ ) with each vol-  
399 ume. These models allowed for visualisation of the effect of genetic variation  
400 across the whole brain (Fig. 5, Fig. A8). The strongest effects on volume for  
401 each genetic variant appear to correspond with voxels within the neuroen-  
402 docrine structures associated with that variant's discovery. While this is not  
403 surprising, due to the circular nature of this analysis using these structures,  
404 this result provides reassuring validation of the method used.

405 Further, for each of the sets of SNPs associated with the four neuro-  
406 endocrine IDPs, the visualisation indicates genetic effect on brain regions  
407 additional to those already found via the IDPs. For example, the summary  
408 map of the pituitary gland volume associated loci (Fig. 5) shows associations  
409 with the cavernous and petrous sinuses and the basal vein of Rosenthal. These  
410 structures drain venous blood and thereby hormones from the pituitary gland.  
411 Associations can also be seen with the medial part of the H-shaped orbital

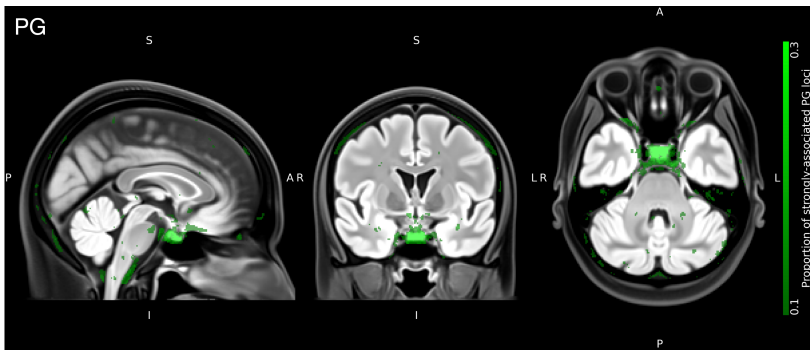




**Fig. 4 Enrichment for gene expression in ovarian and hypothalamic cell-types.** Forest plots of CELLECT-based BH FDR corrected  $P$ -values signifying the genomic colocalisation between the discovered genetic loci and genes specifically expressed across ovarian (A) and hypothalamic (B) cells. The results are presented for analyses performed with the genetic variants with each of the four neuroendocrine IDPs across three population samples: volume of the hypothalamus (turquoise), pituitary gland (purple), mean of left and right olfactory bulbs (gold), and hypothalamic grey matter (red) across a female (circle), male (triangle) and sex-combined (square) sample. Significance thresholds ( $P < 0.05$  after BH FDR correction) are indicated with a dashed grey line. A) Gene expression data from 15 ovarian cell types from five harmonised publicly available and four in-house expression datasets; GC, granulosa cell; TC, theca cell; COC, cumulus-oocyte complex. B) Gene expression data from 38 cell populations derived from a published human postmortem hypothalamus dataset [48].

412 sulcus, the hypothalamus, the amygdala and entorhinal cortex, as well as lob-  
 413 ule IX of the cerebellum. Associations in the nucleus accumbens (colliculus  
 414 nuclei caudati) and the midbrain structures can be seen in the hypothalamus  
 415 and hypothalamic grey matter maps (Fig. A8). Surprisingly, large parts of  
 416 the white matter, the fourth ventricle, and lateral ventricles around the for-  
 417 men of Monro are associated too. Similar to the regions highlighted in the  
 418 voxel-wise association of pituitary gland volume associated SNPs, we observe  
 419 associations with venous structures, which drain (among other areas) circum-  
 420 ventricular organs that lack the normal blood-brain barrier. Similarly to our

421 finding of associations between SNPs from the neuroendocrine IDPs GWAS  
422 and additional brain structure phenotypes within the Big40 dataset, it is per-  
423 haps unsurprising to observe associations across structures of the brain given  
424 shared developmental and cellular compositions. The associations to the addi-  
425 tional specified anatomical areas might rather point to areas of functionally  
426 related pathways.



**Fig. 5 Voxel-wise associations of discovered genetic variants from pituitary gland volume GWAS.** The summary map derived from the t-statistic maps overlaid on the OMM-1 T2-FLAIR reference template shows the proportion of the 16 variants identified in the GWAS of the pituitary gland volume in a sex-combined sample, that are also strongly-associated ( $t > 3.5$ ) with the volume at a voxel location. The intensity range reflecting the strongly-associated proportion of SNPs in the summary map was set to 0.1 - 0.3 (as shown in the colour bar) to highlight only voxels that are common in at least 10% of the identified PG SNPs.

## 427 Methods

### 428 Data

429 We used genetic information and multimodal brain imaging data from UKB  
430 [51]. UKB is one of the largest prospective epidemiological studies to date,  
431 containing genetic information as well as lifestyle and health measures from a  
432 cohort of ~500,000 individuals. Amongst other organ imaging efforts, UKB  
433 is acquiring brain MRI from 100,000 participants, with collection ongoing.

434 Here, we used MRI data from 37,330 UKB individuals (45-84 year age  
435 range) with usable T1-weighted (T1), T2-weighted fluid-attenuated inversion  
436 recovery (T2-FLAIR) and diffusion-weighted MRI (dMRI). T1 is one of the  
437 most commonly used sequences due to its strong contrast between grey mat-  
438 ter, white matter, and cerebrospinal fluid, allowing a high-resolution depiction  
439 of basic brain anatomy. T2-FLAIR as a second structural modality provides  
440 strong contrast in some subcortical grey matter structures and the olfactory  
441 bulbs, as well as between white matter hyperintensities and normal-appearing  
442 white matter. dMRI provides information about local microstructural tissue

443 properties and allows the estimation of a diffusion tensor, which adds informa-  
444 tion about axonal organisation. Scans had been acquired at one of three sites  
445 using identical 3T Siemens Skyra scanners and a standard 32-channel receive  
446 head coil. A detailed description of the acquisition protocol can be found in  
447 [51].

448 The complementary information provided by these spatially correspond-  
449 ing imaging modalities helps localisation of anatomical regions of interest  
450 (ROIs) and interpretation of results, and improves the accuracy when spatially  
451 normalising individuals to a template. Here, we used the recently released  
452 Oxford-MultiModal-1 (OMM-1) template [52] as the common space for align-  
453 ment. The OMM-1 is a fully-unbiased and internally-consistent multimodal  
454 template that represents the average brain anatomy and appearance of 240  
455 UKB subjects (randomly and uniformly sampled from the 50-55 year age  
456 group, 50% female). It comprises anatomically-corresponding volumes from  
457 the same modalities (T1, T2-FLAIR and DTI), which were also used to inform  
458 the construction process. Consequently, the OMM-1 provides optimal compat-  
459 ibility with respect to brain shape and appearance, and consistency for the  
460 current study population and task.

461 UKB has approval from the North West Multi-centre Research Ethics Com-  
462 mittee (<http://www.ukbiobank.ac.uk/ethics/>), which also covers the work in  
463 this study.

## 464 **Data preprocessing**

465 The UKB brain imaging data had undergone extensive manual and auto-  
466 mated quality control and preprocessing steps as described in [53]. In brief,  
467 the pipeline includes manual and semi-automated quality checks, gradient-  
468 distortion correction, cropping, brain extraction (removal of non-brain tissue  
469 such as the skull, scalp, and eyes), and intensity inhomogeneity correction. For  
470 each individual, the T2FLAIR and dMRI modalities are rigidly co-registered  
471 to the the space of the corresponding T1 image with FSL FLIRT [54]. Partial  
472 volume tissue segmentations are estimated for each of three tissue types (grey  
473 matter, white matter, and cerebrospinal fluid) with FSL FAST [55]. Tensors  
474 are fit to the dMRI data with FSL's DTIFIT tool [56] after correcting for  
475 susceptibility and eddy current-induced distortions, and head motion, using  
476 FSL's topup [57] and eddy [58, 59] tools.

## 477 **Image-derived phenotype extraction**

478 We extracted Image-derived phenotypes (IDPs), i.e., volume measurements  
479 for our ROIs, in individuals' native spaces through label propagation. This  
480 approach involved creating masks by segmenting anatomical ROIs in a com-  
481 mon reference template, spatial normalisation (estimation of transformations)  
482 of all individuals to this reference template, and application of the inverse  
483 transformations to these masks to propagate them from the reference space to  
484 the original spaces of the individuals.

485 We first created initial binary masks for the three primary anatomical ROIs  
486 (pituitary gland, hypothalamus and olfactory bulbs) in the space of the OMM-  
487 1. We manually delineated the olfactory bulbs (left and right separately) and  
488 pituitary gland in the T2-FLAIR and T1 volumes of the OMM-1 population  
489 average template, respectively. These binary masks were modulated by the  
490 corresponding image intensities to account for partial volume effects, which  
491 provided spatial segmentation maps with values in the 0-1 range. Note that,  
492 in contrast to the posterior pituitary gland (neurohypophysis, formerly called  
493 ‘pars cerebialis’ of the pituitary), the anterior pituitary gland (adenohypoph-  
494 ysis) and the pars intermedia (both formerly called ‘pars pharyngealis’ of the  
495 pituitary) are phylo- and ontogenetically not part of the brain but deriva-  
496 tives of Rathke’s pouch. Thus, the anterior pituitary is essentially an epithelial  
497 gland. The discrimination between posterior and anterior pituitary (including  
498 the pars intermedia) was not possible based on the available MRI dataset and  
499 would require more specific acquisition sequences. This has to be taken into  
500 consideration when interpreting the results.

501 A hypothalamus mask was obtained by transforming two existing atlases  
502 to OMM-1 space and combining them. The first hypothalamic atlas used here  
503 is the one constructed by Neudorfer *et al.* [60]. For the construction, two  
504 raters performed manual delineation in a high-resolution (initial resolution of  
505 0.7mm upsampled to 0.25mm isotropic), high-contrast template constructed  
506 from 990 individuals in the Human Connectome Project [61] after registering  
507 and transforming them to MNI space. We resampled this binary hypothala-  
508 mic mask into OMM-1 space using the transformations estimated during affine  
509 and nonlinear registrations between the MNI template and the T1 volume of  
510 the OMM-1 template. The second hypothalamus mask used was created by  
511 Pauli *et al.* [62]. They performed manual delineation in each of eight valida-  
512 tion templates before combining them into a probabilistic segmentation map.  
513 Each of the validation templates was constructed from 81 individuals sam-  
514 pled from the same random set of 168 HCP individuals [61]. We resampled  
515 this probabilistic hypothalamic map into OMM-1 space using the transforma-  
516 tions estimated during affine and nonlinear registrations of their study-specific  
517 group template to the T1 volume of the OMM-1 template. Both transformed  
518 atlases were merged in OMM-1 space by masking the probabilistic mask with  
519 the binary mask.

520 Next, we spatially normalised all individual brains to the reference space  
521 of the OMM-1 template. The brain-extracted T1 image of each individual was  
522 affinely (12 degrees of freedom) registered to the T1 volume of the OMM-1  
523 with FLIRT [54]. This affine transformation was used to initialise the nonlinear  
524 registration with FSL’s MultiModal Registration Framework (MMORF) [63].  
525 MMORF estimates one deformation field for all modalities and provides, from  
526 this, a derived Jacobian determinant map between an individual brain and the  
527 reference template. The deformation field provides a spatial mapping between  
528 anatomically corresponding ROIs and the Jacobian map provides a measure  
529 of expansion/contraction for each voxel. All three modalities, T1, T2-FLAIR

530 and DTI, from both individuals and the template were used to drive the regis-  
531 trations for optimal registration accuracy. Hypothalamus, pituitary gland, and  
532 olfactory bulb masks were transformed from OMM-1 space to each individ-  
533 ual's native space by applying the corresponding inverse transformations with  
534 trilinear interpolation.

535 Finally, due to the intensity modulation and effects of interpolation, the  
536 transformed masks in subject space had values in the 0-1 range, where a mask  
537 value of 0 indicates that the voxel definitely is not part of the anatomical  
538 structure and a value of 1 indicates that the voxel definitely is part of the  
539 structure. Values in-between indicate partial volumes, e.g., a value of 0.5 indi-  
540 cates that half of the voxel contains the structure of interest. The transformed  
541 masks were thresholded at 0.3 and binarised as this provided a good trade-off  
542 between over- and undersegmentation (assessed visually). Using these bina-  
543 rised masks we calculated the volume ( $\text{mm}^3$ ) for each structure, producing  
544 three IDPs (pituitary gland volume, hypothalamus volume and combined olfac-  
545 tory bulbs volume). In addition, we extracted the grey matter (GM) volume  
546 of the hypothalamus from the estimated GM partial volume maps, producing  
547 our fourth IDP (hypothalamic grey matter volume).

## 548 **Image derived phenotype characterisation**

549 The population was filtered using genotype principal component analysis  
550 to select for a population that represents the most densely populated well-  
551 mixed population, which here largely aligns with those who self-identify as  
552 White-European. To do so, in brief, we trained a random forest classifier to  
553 assign sample population labels using the 1000 Genomes [64] super-population  
554 labels (AFR=Africans, AMR=Admixed Americans, EAS=East Asians, Euro-  
555 peans=EUR, and South Asians=SAS) and genotype projected into principal  
556 component space. The resulting dataset contained 34,834 individuals. The  
557 mean and standard deviation of each IDP was calculated within this sample  
558 and in groups stratified by genetic sex, and age in bins of 10 years.

## 559 **Genome-wide association studies**

560 We performed GWASs using an additive linear model implemented using  
561 REGENIE [65], a high-speed methodology that accounts for relatedness within  
562 the sample. The analysis was done in both sex-combined and sex-stratified  
563 samples. We regressed confound covariates from each phenotype using a gen-  
564 eralised additive model in accordance with previous analysis: scan date and  
565 scan time, both with spline transformation applied; age; sex; head size; MRI  
566 table position in the  $z$  direction; brain centre of gravity in  $x$ ,  $y$  and  $z$  dimen-  
567 sions; mean head motion. Categorical covariates were numerically encoded  
568 and all covariates were mean-centred prior to regression. Residuals were then  
569 inverse-rank normalised. The first 21 genetic principal components (PCs)  
570 were also provided as covariates to REGENIE. Resultant summary statistics  
571 files were filtered to remove genetic variants with a minor allele frequency

572 <0.01, an imputation information score <0.3, a Hardy-Weinberg equilibrium  
573  $P$ -value of  $<1 \times 10^{-15}$  and any multi-allelic variants. We applied a genome-  
574 wide significance threshold of  $5 \times 10^{-8}$ . We performed LDSC [10] to check for  
575 uncontrolled stratification among the resultant test statistics and to estimate  
576 SNP-based heritability. Considering the general regression models applied  
577 within REGENIE that account for relatedness, this heritability can be consid-  
578 ered a lower-bound estimate. We identified significantly associated independent  
579 loci that are more than 10Mb apart using approximate conditional analysis  
580 implemented in GCTA-COJO [11]. Loci were manually annotated with the  
581 nearest coding gene and prior associations using ENSEMBL [66] and Open  
582 Targets Genetics [67].

## 583 Sexual dimorphism

584 For each phenotype, we tested each significant ( $P < 5 \times 10^{-8}$ ) lead variant  
585 identified across the different strata (sex-combined; and male- and female-  
586 specific;) for a sex-dimorphic effect by estimating the  $t$ -statistic:

$$t = \frac{\beta_{\text{female}} - \beta_{\text{male}}}{\sqrt{\text{SE}_{\text{female}}^2 + \text{SE}_{\text{male}}^2 - 2r \text{SE}_{\text{female}} \text{SE}_{\text{male}}}},$$

587 where  $\beta_{\text{sex}}$  and  $\text{SE}_{\text{sex}}$  ( $\text{sex} \in \{\text{female}, \text{male}\}$ ) are sex-specific effect sizes and  
588 SEs, respectively, and  $r$  is the genome-wide Spearman rank correlation between  
589 SNP effects in females and males. The  $t$ -statistic was estimated and the  
590 resulting  $P$ -value calculated [68] as implemented in the EasyStrata software  
591 [69].

## 592 Exome analysis

### 593 Sequencing Quality Control

594 To evaluate an initial set of high quality variants we ran sample and vari-  
595 ant level pre-filtering before calculating sample-level QC metrics. Following  
596 Karzcewski *et al.* [18], we then removed sample outliers using median abso-  
597 lute deviation (MAD) thresholds ( $> 4$  MADs from the median). Finally, we  
598 applied a genotype-level filter using genotype quality ( $\text{GQ} \geq 20$ ), depth ( $\text{DP}$   
599  $\geq 10$ ), and heterozygote allele balance ( $\text{AB} > 0.2$ ). After these filtering steps,  
600 the high-quality European call set consisted of 402,375 samples and 25,229,669  
601 variants.

## 602 Variant Annotation

603 To annotate variants we used Variant Effect Predictor (VEP) v105, GEN-  
604 CODE v39 [70] with the LOFTEE v1.04\_GRCh38 [71] and dbNSFP [72] plug-  
605 ins, annotating variants with CADD v1.6 [73], and REVEL [74] scores, and loss  
606 of function (LoF) confidence using LOFTEE. Full instructions to reproduce  
607 the annotation pipeline are provided within the VEP\_150\_LOFTEE repository  
608 ([https://github.com/BRaVa-#genetics/vep105\\_loftee](https://github.com/BRaVa-#genetics/vep105_loftee)). We then ran SpliceAI

609 v1.3 [75] with the GENCODE v39 annotation file to ensure alignment between  
610 VEP and SpliceAI transcript annotations. For variant-specific annotations we  
611 defined ‘canonical’ transcripts as follows: set MANE Select [76] as the canonical  
612 where available, otherwise set canonical and restrict to protein-coding genes:

- 613 1. High confidence predicted LoF (pLoF): high-confidence LoF variants, as  
614 defined by LOFTEE (LOFTEE HC).
- 615 2. Damaging missense/protein-altering, at least one of:
  - 616 • Variant annotated as missense/start-loss/stop-loss/in-frame indel and  
617 (REVEL  $\geq 0.773$  or CADD  $\geq 28.1$  (or both)).
  - 618 • Any variant with SpliceAI delta score (DS)  $\geq 0.2$  where SpliceAI DS  
619 is the maximum of the set {DS\_AG, DS\_AL, DS\_DG, DS\_DL} for each  
620 annotated variant (where DS\_AG, DS\_AL, DS\_DG, and DS\_DL are delta  
621 score (acceptor gain), delta score (acceptor loss), delta score (donor gain),  
622 and delta score (donor loss), respectively).
- 623 3. Low-confidence LoF variants, as defined by LOFTEE (LOFTEE LC)
- 624 4. Other missense/protein-altering:
  - 625 • Missense/start-loss/stop-loss/in-frame indel not categorised in (2) (Dam-  
626 aging missense/protein-altering).
- 627 5. Synonymous: synonymous variants with SpliceAI DS  $< 0.2$  in the gene (our  
628 ‘control’ set).

629 We selected the REVEL and CADD score thresholds based upon the American  
630 College of Medical Genetics and Genomics and the Association for Molecular  
631 Pathology’s (ACMG/AMP) PP3-moderate evidence criteria [77].

## 632 **Rare variant association testing**

633 Rare variant association testing was conducted using Scalable and Accurate  
634 Implementation of GEneralized mixed model (SAIGE, 1.1.9) [78], a mixed  
635 model framework that uses the saddlepoint approximation to calibrate the  
636 distribution of score test statistics in the presence of sample relatedness and  
637 case-control imbalance.

638 Beginning with our European-ancestry quality controlled subset of  
639 UKB we calculated a genetic relatedness matrix (GRM) using genotyp-  
640 ing array data. We then LD pruned variants using the PLINK option  
641 `--indep-pairwise 50 5 0.05` [79] and subset to 5,000 randomly selected  
642 markers to create a sparse GRM using SAIGE’s `createSparseGRM.R`.

643 For each trait we ran SAIGE step 1 to fit a null model with no  
644 genetic contribution using the sparse GRM with the first 20 genetic  
645 PCs and sex used as covariates. All default parameters were used  
646 except for `--relatednessCutoff 0.05`, `--useSparseGRMtoFitNULL TRUE`  
647 and `--isCateVarianceRatio TRUE`.

648 Finally, after fitting the null model we ran variant and gene-burden  
649 testing with SAIGE (step 2). All default parameters were used except



650 for `--maxMAF_in_groupTest=0.0001,0.001,0.01, --is_Firth_beta TRUE,`  
651 `--pCutoffforFirth=0.1` and `--is_fastTest TRUE`. For gene burden testing  
652 we used the annotations as previously defined, testing the following anno-  
653 tations groups: High confidence pLoF; Damaging missense/protein-altering;  
654 Other missense/protein-altering; Synonymous; High confidence pLoF or Dam-  
655 aging missense/protein-altering; and High confidence pLoF or Damaging  
656 missense/protein-altering or Other missense/protein-altering or Synonymous.  
657 The annotation results were also combined using a Cauchy combina-  
658 tion test [13] combining the tests for “High confidence pLoF”, “Damag-  
659 ing missense/protein-altering” and “High confidence pLoF AND Damaging  
660 missense/protein-altering” across all MAF limits. A conservative significance  
661 threshold of  $P < 6.25 \times 10^{-7}$  was chosen accounting for multiple testing in  
662 number of genes and phenotypes considered.

## 663 Phenome-wide association studies

664 We carried out PheWAS using ICD-10 codes from UKB, converted to phecodes  
665 as implemented in the PheWAS package in R [21]. Mean-centred age and sex  
666 and the first 21 genotype PCs were included as covariates and a significance  
667 threshold of  $P < 0.05$  following BH FDR correction was used. Further associa-  
668 tions were explored using the Big40 Oxford brain imaging genetics server. We  
669 extracted the marginal effect of the 66 variants significantly associated with  
670 the neuroendocrine IDPs on  $\sim 4,000$  image derived phenotypes from multi-  
671 modal brain imaging [27]. The  $P$ -value significance threshold was adjusted for  
672 multiple testing using the Bonferroni method ( $P < 1.93 \times 10^{-7}$ ).

## 673 Shared genetics underlying neuroendocrine anatomical 674 variation and reproductive traits

675 Here, we detail the analyses performed to investigate whether there is shared  
676 genetic effect on neuroendocrine IDPs, and reproductive traits. Reproductive  
677 traits examined included levels of testosterone, follicle stimulating hormone  
678 (FSH), luteinising hormone (LH), oestradiol and progesterone, alongside five  
679 phenotypes describing female infertility subtypes (all cause, anatomical cause,  
680 anovulatory, unknown cause excluding idiopathic infertility, and unknown  
681 cause including idiopathic infertility), and one male infertility phenotype.  
682 GWAS summary statistics of these traits were taken from Venkatesh *et al.*  
683 [39], in which further description of the phenotypes can be found. We used the  
684 summary statistics for the GWAS meta analysis of a European only sample.  
685 For all analyses apart from the enrichment for testosterone variants, we used  
686 the summary statistics from the meta analysis that excluded the UK Biobank  
687 data.

## 688 Enrichment for testosterone variants

689 We used a one-sided fishers exact test, implemented in R, to calculate enrich-  
690 ment for testosterone variants (5) within our 66 neuroendocrine IDP associated



691 variants. We obtained 254 lead variants associated with testosterone levels from  
692 the most-recent and largest meta-analysed GWAS of testosterone to our knowl-  
693 edge [39], among 1 million independent common variants across the genome  
694 [80].

## 695 Genetic correlation

696 We estimated genetic correlations between the four neuroendocrine IDPs and  
697 hormone levels and infertility phenotypes using LDSC [10, 81]. Summary  
698 statistics were intersected with HapMap3 SNPs. The remaining set of vari-  
699 ants were then used to estimate  $r_g$ . Traits with a heritability Z-score  $<4$  were  
700 excluded from  $r_g$  estimation.

## 701 Mendelian randomisation

702 Two-sample Mendelian randomisation was applied using the TwoSampleMR  
703 package in R [82, 83]. Bidirectional Mendelian randomisation was performed  
704 using the neuroendocrine IDP associated genetic variants as both exposure and  
705 outcome variables. Lead SNPs resulting from approximate conditional anal-  
706 ysis applied in GCTA-COJO [11] (see methods above) were used as genetic  
707 instruments when using the neuroendocrine IDPs as the exposure variable.  
708 When using the reproductive traits as the exposure variable, genetic instru-  
709 ments were selected as the lead SNPs, defined using distance-based pruning  
710 as the variant with the lowest  $P$ -value within a 1Mb window of each genome-  
711 wide significant ( $P < 5 \times 10^{-8}$ ) locus [39]. In both cases the internal clumping  
712 method was further applied using default parameters.

713 We applied five methods of meta-analyses: MR Egger, weighted median,  
714 inverse variance weighting, simple mode and weighted mode. Bonferroni mul-  
715 tiple test correction was applied resulting in a corrected  $P$ -value significance  
716 threshold of  $4.17 \times 10^{-3}$ . A secondary threshold was applied requiring the  
717 meta-analyses to include a minimum of five genetic instruments. Phenotype  
718 pairs were also tested for heterogeneity, horizontal pleiotropy, leave-one-out  
719 analysis and directionality [84–86].

## 720 Colocalisation with reproductive phenotypes

721 Genomic position of the neuroendocrine IDP summary statistics was lifted  
722 from build hg19 to hg38 using UCSC’s liftOver [87]. Colocalisation was per-  
723 formed using the coloc package in R [88] applying a Bayesian approach under  
724 a single causal variant assumption. Genetic variants within a 100kb window  
725 around each lead variant associated with the neuroendocrine IDPs were used  
726 in the analysis. In line with recommendation of the coloc developers [89] we  
727 set the prior probabilities of genetic variants within a locus being associated  
728 with each trait trait to  $1 \times 10^{-4}$ , and to both traits to  $1 \times 10^{-6}$ . We assessed  
729 five hypotheses:  $H_0$  - no association with either trait;  $H_1$  - association with  
730 trait A but not trait B;  $H_2$  - association with trait B but not trait A;  $H_3$  -  
731 association with both traits but different causal variants; and  $H_4$  - association

732 with both traits and the same causal variant. Associations were considered to  
733 be colocalised if  $P(\frac{H_4}{H_3}) > 5$  and  $P(H_4) > 0.5$ .

## 734 Colocalisation with tissue specific eQTLs

735 Colocalisation analysis with tissue specific eQTLs was performed using the  
736 coloc package in R [88] under a single causal variant assumption. The same pri-  
737 ors and genomic window were used as above. Colocalisation was assessed with  
738 publicly available eQTL data from GTEx [32]. All genes with a transcriptional  
739 start site within a 1Mb region about the lead SNPs associated with the neu-  
740 roendocrine IDPs were tested. Associations were considered to be colocalised  
741 if  $P(\frac{H_4}{H_3}) > 5$  and  $P(H_4) > 0.5$  [90].

## 742 Gene and Tissue Enrichment

743 To prioritise likely causal genes at the loci associated with our IDPs, and high-  
744 light enriched pathways and tissues where these genes are highly expressed, we  
745 performed gene-set and tissue enrichment analyses using DEPICT [43]. Briefly,  
746 DEPICT implements an integrative method that uses 14,461 gene sets recon-  
747 stituted on the basis of expression data from 77,840 samples. Reconstituted  
748 gene sets that are enriched for genes in the associated loci are identified, with  
749 genes in different associated loci that have similar predicted functions priori-  
750 tised. For tissue enrichment in DEPICT, microarray data from 37,427 human  
751 tissue samples are used to identify tissue and cell types in which genes from  
752 associated loci are highly expressed. A  $P$ -value threshold of 0.05 following BH  
753 FDR correction was used.

## 754 Enrichment for gene expression in ovarian and 755 hypothalamic cell-types

756 CELLEX (<https://github.com/perslab/CELLEX>) and CELLECT (<https://github.com/perslab/CELLECT>) were used to identify candidate aetiological  
757 cell types. In brief, CELLEX is used to compute robust cell-type expression  
758 specificity profiles by combining multiple expression specificity measures as  
759 described in Timshel *et al.* [45]. These estimates are used by CELLECT to  
760 test for association between genetic signal (heritability) taken from GWAS  
761 summary statistics, and cell-type expression specificity.

762 CELLEX was first applied to two datasets. Firstly gene expression  
763 measurements from 15 ovarian cell types from an in-house map compris-  
764 ing five harmonised publicly available ovarian gene expression datasets  
765 and four ovary samples processed in-house (paper in preparation, Ferreira  
766 *et al.*). Secondly from a dataset containing 38 hypothalamic cell popula-  
767 tions derived from a human postmortem single-nucleus RNA-seq dataset  
768 [48], downloaded from [https://storage.googleapis.com/linnarsson-lab-human/  
769 adult\\_human\\_20221007.loom](https://storage.googleapis.com/linnarsson-lab-human/adult_human_20221007loom).

770 CELLECT (version 1.1) was applied to GWAS summary statistics for each  
771 of the four neuroendocrine IDPs in sex-stratified and sex-combined analyses to  
772

773 find associations to the ovarian and hypothalamic cell-type expression speci-  
774 ficity. MAGMA covariate analysis was used as the genetic prioritisation model  
775 [91]. CELLECT was run with default parameters, and significantly associated  
776 cell types were defined using a by-trait BH FDR corrected  $P$ -value threshold  
777 of  $P < 0.05$ .

## 778 **Voxel-level genetic associations**

779 Similar to the approach taken in [6], we ran linear models to investigate the  
780 association between local variations in anatomical volume and allele dosage,  
781 of the 66 SNPs significantly associated with the neuroendocrine IDPs, at the  
782 voxel-level across the whole brain. The Jacobian determinant maps, which  
783 were obtained from the image registrations during the spatial normalisation  
784 processing step, were scaled by the determinant of the corresponding affine  
785 transformations to account for differences in brain size and used as a mea-  
786 sure of volumetric variation. The intensity value of each voxel in a Jacobian  
787 determinant map indicates the ratio of the volume change in the registered  
788 subject at that voxel relative to the reference template, i.e., no change ( $=1$ ),  
789 compression ( $<1$ ) or expansion ( $>1$ ). Allele dosages of SNPs that were previ-  
790 ously identified as significant and the same covariates that were used during  
791 the GWAS were used as explanatory variables. Genetic dosage was extracted  
792 using qctools (<https://www.chg.ox.ac.uk/~gav/qctool.v2/>).

793 The output t-statistic maps derived for each of the 66 SNPs were aggregated  
794 across their associated IDPs into four summary maps to highlight common  
795 brain regions strongly associated with an IDP. For example, for the PG we  
796 aggregated the t-statistic maps from 16 linear models, one for each variant  
797 identified as significantly associated with the PG IDP during the GWAS. These  
798 summary maps were created by taking the absolute values of the t-statistic  
799 maps followed by thresholding ( $t > 3.5$ ), binarizing and averaging, to highlight  
800 strongly associated brain regions. Two expert neuroanatomists investigated  
801 these summary maps qualitatively. When interpreting the maps the potential  
802 statistical risks due to the circular nature of these analyses should be taken  
803 into consideration, i.e. no claims of significance were made.

## 804 **Discussion**

805 We performed the first study to examine the genetic underpinnings of the  
806 anatomy of three neuroendocrine structures: the hypothalamus, pituitary  
807 gland and olfactory bulbs. Due to the functional nature of these structures  
808 within the hypothalamic-pituitary-gonadal axis, we hypothesised that the  
809 exploration of the genetics underlying the three neuroendocrine structures  
810 would lead to further understanding not only of their role in brain function,  
811 but also in broader endocrine and reproductive health. We extracted four  
812 quantitative image-derived phenotypes describing the volume of these largely

813 unexplored regions across almost 35,000 individuals. In our analyses, we identi-  
814 fied 66 independent loci significantly associated ( $P < 5 \times 10^{-8}$ ) with one of  
815 these IDPs.

816 We found initial evidence of a shared genetic effect on both neuroendocrine  
817 anatomy and testosterone, one of the structures' downstream products. There  
818 was significant enrichment ( $P$ -value =  $9.89 \times 10^{-12}$ ) of the genetic variants  
819 associated with the neuroendocrine IDPs for variants associated with testos-  
820 terone. Five of the 66 lead genetic variants identified in the GWASs had a  
821 prior association to testosterone levels in or near to *REST*, *CENPW*, *CCND2*,  
822 *RNF2* and *PPM1A* [31]. In particular, the colocalisation of rs76895963 with  
823 an eQTL for *CCND2* across numerous tissues strongly suggests this is the  
824 causal gene. *CCND2* has been robustly associated with testosterone across ani-  
825 mal models and is known to have a role in fertility and reproductive processes  
826 [33–35]. Rare variant exome-wide analyses identified an association between  
827 hypothalamus volume and *STAB1*, a gene involved in iron homeostasis. Iron  
828 metabolism was also significantly associated with hypothalamic grey matter  
829 volume associated variants. Both iron metabolism and the related process of  
830 erythropoiesis are regulated by testosterone [24], with diseases affecting iron  
831 levels linked to reproductive dysfunction and infertility [23]. Taken together,  
832 these results provide initial evidence for an overlap in genetic effect between the  
833 volume of the four neuroendocrine regions and their function in production and  
834 regulation of testosterone. This is further supported by nominally significant  
835 genetic correlations between testosterone and pituitary gland, hypothalamus  
836 and olfactory bulb volumes.

837 We identified a possible link between genetic effect on neuroendocrine  
838 anatomy and the gonadal tissues in which their downstream products are  
839 active. We found enrichment of ovarian theca cells for genetic variants asso-  
840 ciated with pituitary gland volume. Theca cells, in response to LH produced  
841 by the pituitary gland as part of the hypothalamic-pituitary-gonadal axis,  
842 produce testosterone and precursors of oestradiol [46]. A genetic variant was  
843 also found downstream of *PPM1A*, a gene known to be involved in regula-  
844 tion of testosterone synthesis in Leydig cells in the testis [37]. In both cases  
845 this suggests a commonality to the genetics affecting hormone production  
846 across neuroanatomical and gonadal tissues. These two findings could repre-  
847 sent shared pathways underlying the production and regulation of reproductive  
848 hormones, specifically testosterone, at different sites of the body. It could also  
849 be considered that the genetic links we are observing are caused by an effect of  
850 hormones produced in gonadal locations on the anatomy of the neuroendocrine  
851 structures. It would be ideal to analyse expression of genes across both gonadal  
852 and neuroanatomical tissues in the same individuals to aid the disentangle-  
853 ment of the causality underlying our observations, however no such datasets  
854 are currently available. Statistical analysis such as Mendelian randomisation  
855 could help further elucidate the relationship between the genetic variants effect  
856 on tissues that regulate hormones, and the hormone levels themselves, however  
857 we have thus far been underpowered to discover such causal relationships.

858 Given the androgen cascade controlled by the hypothalamic-pituitary-  
859 gonadal axis, there could still be underdiscovered links between anatomy of  
860 the neuroendocrine structures and additional reproductive hormones beyond  
861 testosterone. The pituitary gland is more directly involved in production of  
862 FSH and LH, yet we did not observe evidence for a genetic link between our  
863 phenotypes and the levels of these hormones beyond the ovarian cell enrich-  
864 ment. Currently, the discovery power of this analysis has been restricted by the  
865 limited sample size and resulting low SNP heritability of the hormone associ-  
866 ation studies. Particularly there is reduced statistical discovery power within  
867 the GWAS of FSH and LH, for which sample sizes in the GWAS which we  
868 used to look for genetic overlap were less than half that of testosterone. Fur-  
869 thermore, the genetic basis of testosterone has been more intensively studied  
870 than FSH and LH. Additional datasets with more accurate and comprehensive  
871 measurement of hormones in a healthy population are required to be able to  
872 further elucidate the link between neuroanatomical structures of the endocrine  
873 system and reproductive hormones. The addition of temporal collection of hor-  
874 monal data would also allow an understanding of intrapersonal variation that  
875 could be affecting statistical power of genetic discovery, and hence discovery  
876 of genetic overlap between the phenotypes.

877 We identified a possible shared role of the genetic variants associated with  
878 neuroendocrine anatomy on not just production of reproductive hormones, but  
879 also on mechanistic function of their release. Olfactory bulb volume associated  
880 variants were enriched in two cell types across hypothalamic cells tested, the  
881 ependymal cells, and their subtype, the tanycytes. Notably tanycytes have pre-  
882 viously been shown to influence the release of GnRH [49]. Further, employment  
883 of voxel-wise association studies identified association between the identified  
884 genetic variants and areas around the cavernous and petrous sinuses and the  
885 basal vein of Rosenthal, which are responsible for draining venous blood and  
886 hence hormones from the pituitary gland. This suggests a possible role of the  
887 discovered genetic variants not only in hormone production but also in the  
888 more mechanistic role of their release into the blood. The voxel-wise analy-  
889 sis allows us to identify regions of the brain affected by our genetic discovery  
890 set not captured by the quantitative image-derived phenotype summary mea-  
891 sures that are more commonly used within genetics or clinical practice. It must  
892 be acknowledged that associations with the original neuroendocrine IDPs are  
893 inherently circular and one must be cautious when interpreting results, restrict-  
894 ing only to structures assumed to be independent of the original phenotype  
895 measures. In this way it is still valuable as a method for visualising and explor-  
896 ing potential pleiotropy, and identifying regions of interest with potential for  
897 further investigation through the extraction of novel descriptive image-derived  
898 phenotypes for use in statistical genetics analyses.

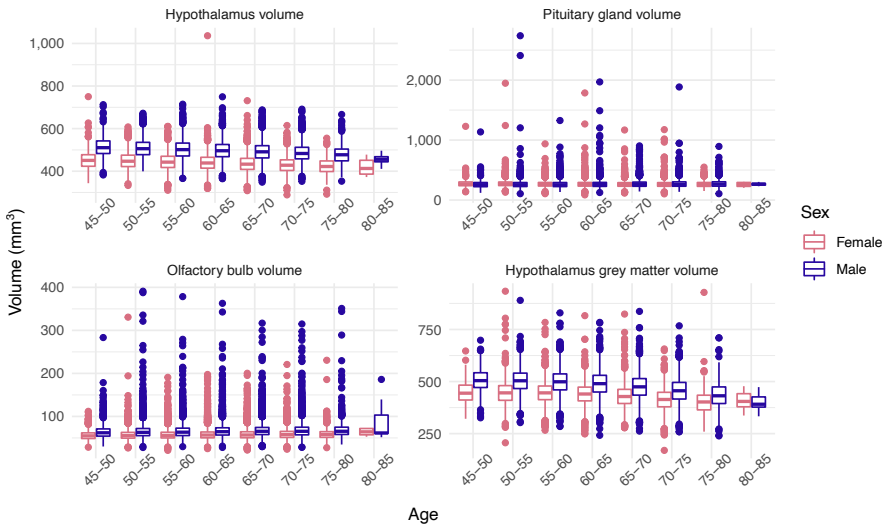
899 The GWAS further identified several genetic variants with prior associa-  
900 tion to menstrual timing, particularly menarche and menopause. Particularly  
901 two genetic variants within the *STX6* gene were identified to have both had  
902 prior associations to the age at menopause. We also identified sicca syndrome

903 to be associated with genetic variants associated with hypothalamic grey mat-  
904 ter volume. Sicca syndrome is a condition that is largely diagnosed at onset  
905 of menopause and causes dryness in numerous glands [26]. It could be consid-  
906 ered that the brain imaging that underlies the image-derived phenotypes was  
907 taken for many of the female participants at an age of substantial reproduc-  
908 tive disruption - the perimenopausal and menopausal age. There is preliminary  
909 evidence that this process causes changes to neurological structures [92]. This  
910 could be leading us to discover some genetic effects that are specific to this  
911 time point. This could also account for some of the sexual dimorphism in  
912 genetic effects that we observe, which might be in part driven by the repro-  
913 ductive ageing process affecting the sexes differently. Work targeting the topic  
914 of sex-specific ageing more directly is needed in datasets that have additional  
915 phenotypic collection about menopause and associated hormonal changes. In  
916 parallel, further work using an imaging dataset of individuals around their  
917 reproductive peak could reveal more about the link between brain structure  
918 and reproductive function.

919 The extraordinary opportunity presented in the UKB dataset, to study  
920 the biology underlying common variation in anatomy through the examina-  
921 tion of healthy population imaging coupled with genetics at a large scale, is  
922 unique. However, its uniqueness also presents a challenge in replication. The  
923 UKB is commonly acknowledged to represent a specific sociodemographic pop-  
924 ulation, here particularly pertinent is the cohort's older age. This might have  
925 affected the results as it is well documented that brain structure changes with  
926 advancing age. Ideally, replication would be conducted to confirm our results  
927 and assess the transferability of the results to cohorts of different population  
928 composition. However, there is currently a lack of accessible population-scale  
929 imaging datasets of non-disease specific individuals, especially those of ances-  
930 try outside of those that align with individuals who self-identify as "White  
931 European". The creation of new datasets with equally rich phenotyping, but  
932 additional increased diversity, would allow for replication and expansion of  
933 our knowledge beyond a single population. Additional datasets would further  
934 enable the proposal of novel questions that expand our understanding of the  
935 connection between neuroanatomical and endocrine systems.

936 **Appendix A**

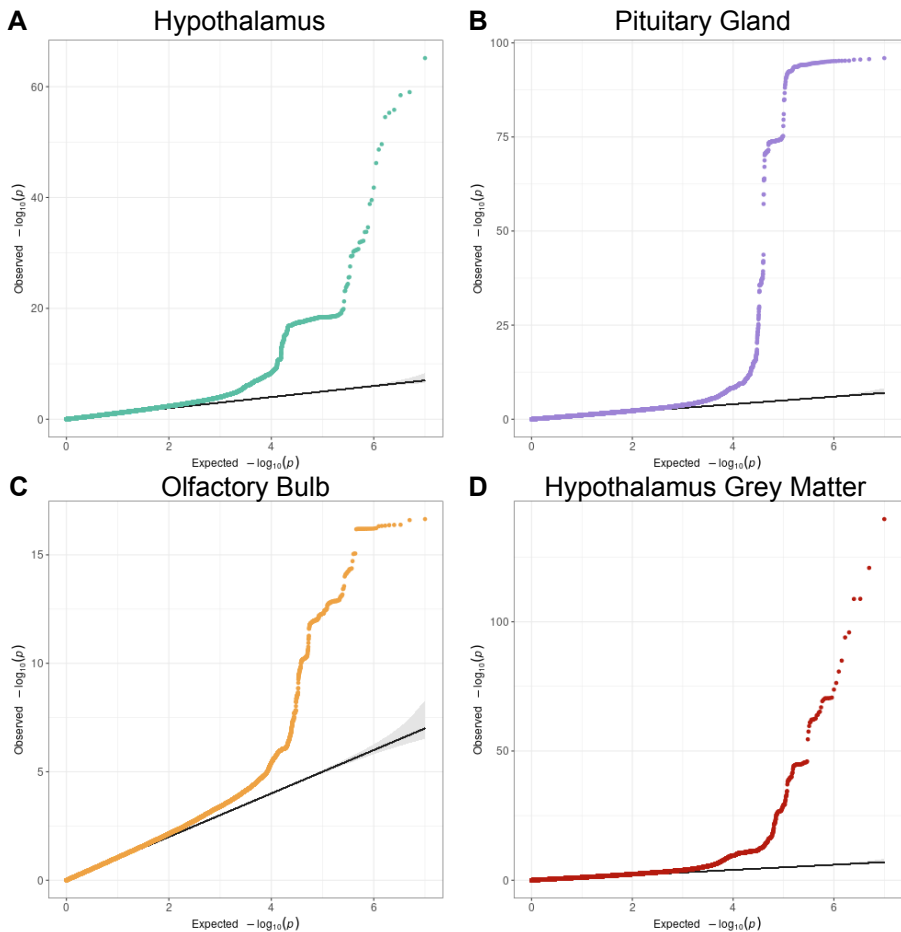
**Table A1** Mean volume ( $\pm$ standard deviation) of each neuroendocrine structure in the cohort stratified by age and sex.



**Fig. A1 Distribution of neuroendocrine volumes across sex and age.** The volumes (mm<sup>3</sup>) of the hypothalamus, pituitary gland, mean across left and right olfactory bulbs, and hypothalamic grey matter across age groups and sexes (as defined by genetic sex) within the study sample.

**Table A2** All statistically significant associations ( $P < 5 \times 10^{-8}$ ) from the four genome-wide association studies of the volume of the hypothalamus (HT), hypothalamic grey matter (HT-GM), pituitary gland (PG) and olfactory bulbs (OB). For each the following data is presented: nearest coding gene assigned using OpenTargets Genetics; rsid; chromosome (CHROM); genomic position (GENPOS) in GRCh37 coordinates; reference allele (A0); effect allele (A1); effect allele frequency (A1FREQ); info score (INFO); sample size (N); Effect size; SE; and p-value.

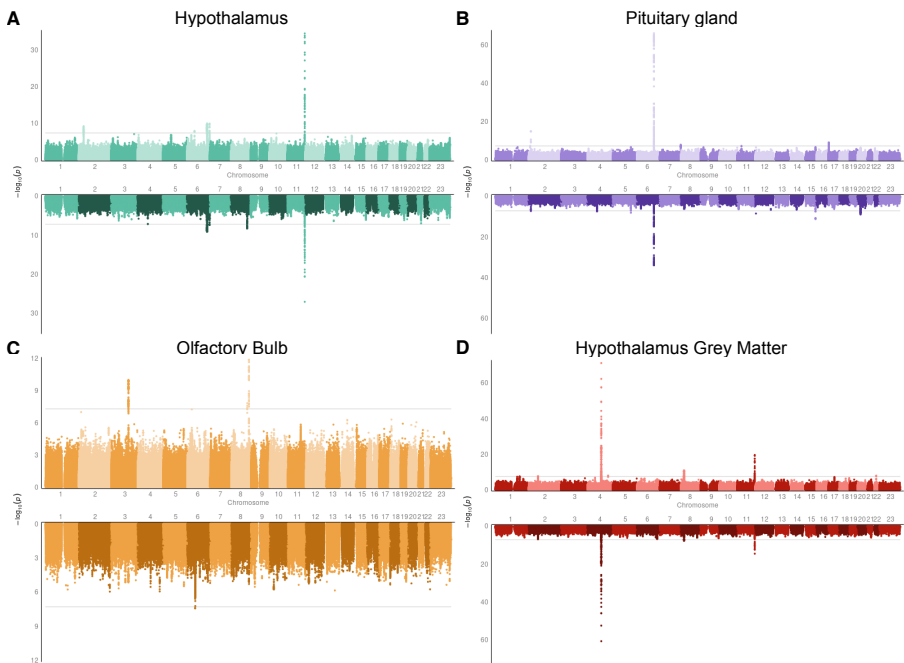
**Table A3** All statistically significant associations ( $P < 5 \times 10^{-8}$ ) from the four genome-wide association studies of the volume of the hypothalamus (HT), hypothalamic grey matter (HT-GM), pituitary gland (PG) and olfactory bulbs (OB). For each the following data is presented: nearest coding gene assigned using OpenTargets Genetics; rsid (ID); chromosome (CHR); phenotype;  $P$ -value; prior associations to reproductive traits; and prior associations to brain-related phenotypes, all assigned using OpenTargets Genetics [67] and ENSEMBL [66].



**Fig. A2** Quantile-quantile plots for GWAS of neuroendocrine IDPs. Plot of expected (x-axis) vs observed (y-axis)  $-\log_{10}(P\text{-values})$  ranked for each SNP tested across the four genome wide association studies: A) hypothalamus; B) pituitary gland; C) olfactory bulb; and D) hypothalamic grey matter.

**Table A4** Linkage score disequilibrium metrics for the four GWAS of the neuroendocrine IDPs: the hypothalamus (HT), the pituitary gland (PG), the mean of the left and right olfactory bulbs (OB) and the hypothalamic grey matter (HTGM). Metrics include lambda genomic control (GC), linkage disequilibrium score (LDSC) intercept, LDSC intercept standard error (SE), LDSC ratio, LDSC ratio SE and LDSC intercept  $P$ -value.





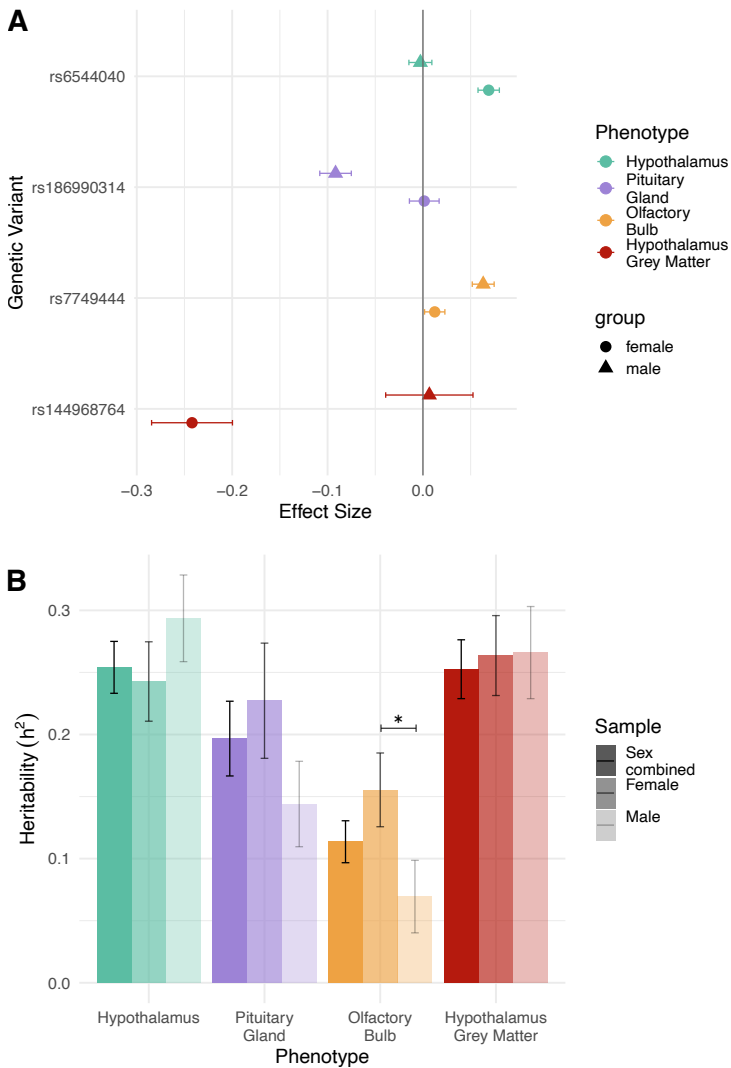
**Fig. A3 Genome-wide association studies of the four neuroendocrine IDPs in female and male samples.** Miami plots presenting the results of genome-wide association studies of the volume of the hypothalamus (A), pituitary gland (B), olfactory bulb (C) and hypothalamic grey matter (D) in a female sample ( $n=18,487$ , upper, light) and male sample ( $n=16,347$ , lower, dark).

**Table A5** SNPs with statistically significant sexual dimorphism in genetic effect on the neuroendocrine IDPs. For each SNP the following information is listed: neuroendocrine IDP with which the genetic variant is associated, SNP rsid, chromosome (CHR), genomic position (POS), other allele (OA), effect allele (EA), effect allele frequency (EAF), effect size (BETA), SE, and  $P$ -value ( $P$ ) in a male and female sample and the  $P$ -value from the sexual dimorphism test ( $P_{DIFF}$ ).

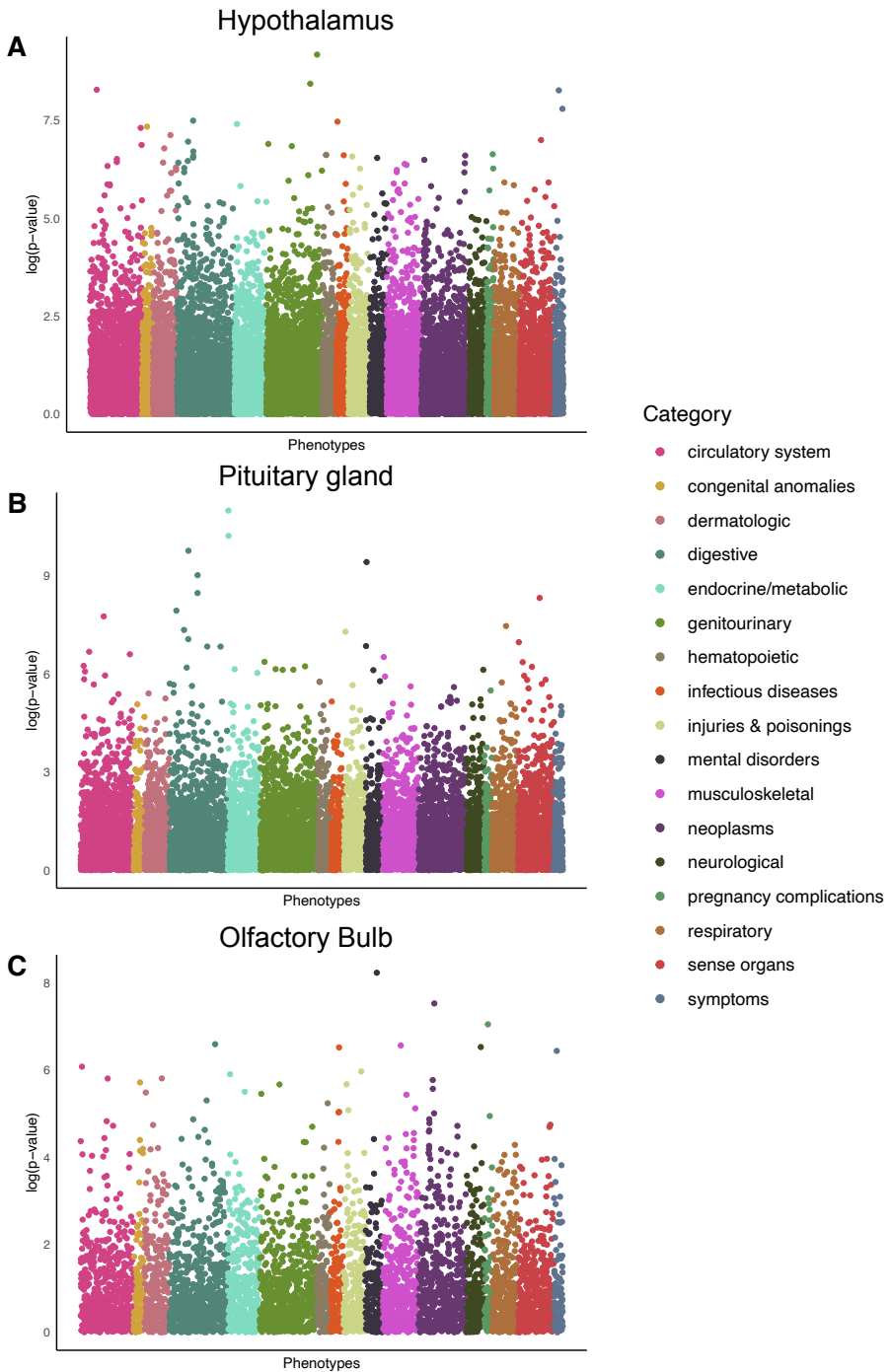
**Table A6** Results of phenome-wide association studies using genetic variants associated with hypothalamus volume and ICD10 codes from the UKB converted to phecodes. For all associations with phenotypes where there is a case sample size  $> 5$ , the following results are listed: rsid; phecode (phenotype), phecode description, phecode group, effect size (beta), SE, OR,  $P$ -value, BH FDR adjusted  $P$ -value (fdr\_adj\_p), case count, and control count.

**Table A7** Results of phenome-wide association studies using genetic variants associated with pituitary gland volume and ICD10 codes from the UKB converted to phecodes. For all associations with phenotypes where there is a case sample size  $> 5$ , the following results are listed: rsid; phecode (phenotype), phecode description, phecode group, effect size (beta), SE, OR,  $P$ -value, BH FDR adjusted  $P$ -value (fdr\_adj\_p), case count, and control count.

**Table A8** Results of phenome-wide association studies using genetic variants associated with olfactory bulb volume and ICD10 codes from the UKB converted to phecodes. For all associations with phenotypes where there is a case sample size  $> 5$ , the following results are listed: rsid; phecode (phenotype), phecode description, phecode group, effect size (beta), SE, OR,  $P$ -value, BH FDR adjusted  $P$ -value (fdr\_adj\_p), case count, and control count.

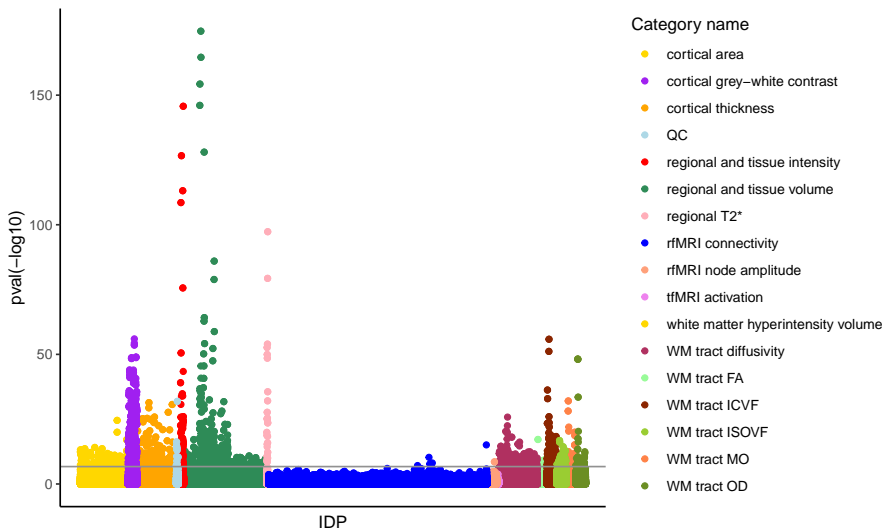


**Fig. A4 Sexual dimorphism in genetic effect on neuroendocrine IDPs.** A) The genetic effect of SNPs with statistically significant sexual dimorphism in genetic effect tested across all SNPs significantly associated in the female, male or sex-combined samples across the GWAS of the four neuroendocrine IDPs: hypothalamus (turquoise), pituitary gland (purple), olfactory bulb (gold) and hypothalamic grey matter (red). Genetic effect on females is represented as a circle, genetic effect on males is represented by a triangle. SE is presented as a horizontal line about the point. B) The heritability ( $h^2_{\text{SNP}}$ ) of the volume of the four neuroendocrine IDPs in the sex-combined (darkest), female (medium opacity) and male (lightest) samples as calculated in LDSC [10]. Statistically significant differences between males and females are highlighted with an asterisk.



**Fig. A5 Phenome-wide association studies using broader clinical phenotypes.** Manhattan plot presenting the  $-\log_{10}(P\text{-values})$  from a PheWAS between the UKB ICD10-codes converted to phecodes and the genetic variants identified as associated with the volume of A) hypothalamus; B) pituitary gland; and C) olfactory bulbs. Phecodes are categorised into broader categories represented by different colours.

**Table A9** Results of phenome-wide association studies using genetic variants associated with hypothalamic grey matter volume and and ICD10 codes from the UKB converted to phecodes. For all associations with phenotypes where there is a case sample size  $> 5$ , the following results are listed: rsid; phecode (phenotype), phecode description, phecode group, effect size (beta), SE, OR,  $P$ -value, BH FDR adjusted  $P$ -value (fdr\_adj\_p), case count, and control count.



**Fig. A6 Phenome-wide association study of image-derived phenotypes from Big40 database.** Manhattan plot presenting the  $-\log_{10}(P\text{-values})$  from a phenome-wide association study utilising image-derived phenotypes from multi-modal magnetic resonance imaging within the UKB stored within Big40 - a database of genetic effect on  $\sim 4000$  image-derived phenotypes from multimodal brain imaging [27], and the 66 genetic variants discovered across the four GWAS of the hypothalamus, pituitary gland, olfactory bulb and hypothalamic grey matter volume in a sex-combined sample.

**Table A10** All 1,140 statistically significant associations ( $P < 1.93 \times 10^{-7}$ ) between the genetic variants associated with hypothalamus volume in a sex-combined sample and the Big40 database IDPs. The IDP ID (IDP), chromosome (chr), rsid, genomic position (pos), effect allele (a1), reference allele (a2), effect size, standard error (se), pvalue, IDP short name and IDP category name are listed.

**Table A11** All 82 statistically significant associations ( $P < 1.93 \times 10^{-7}$ ) between the genetic variants associated with pituitary gland volume in a sex-combined sample and the Big40 database IDPs. The IDP ID (IDP), chromosome (chr), rsid, genomic position (pos), effect allele (a1), reference allele (a2), effect size, standard error (se), pvalue, IDP short name and IDP category name are listed.

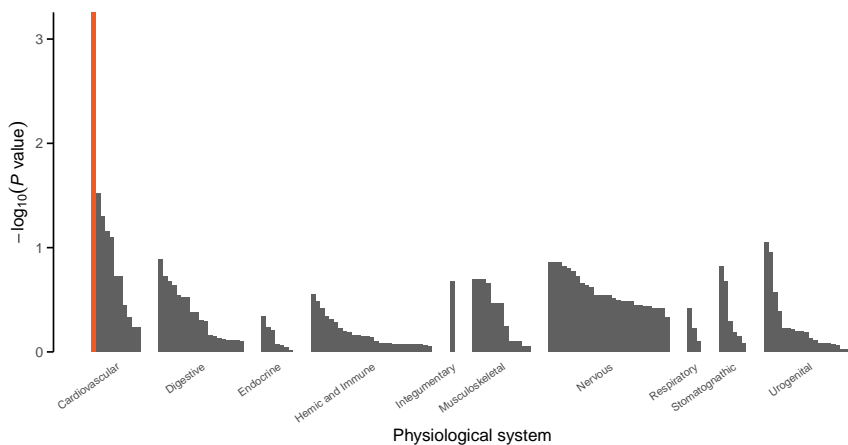
**Table A12** All 22 statistically significant associations ( $P < 1.93 \times 10^{-7}$ ) between the genetic variants associated with olfactory bulb volume in a sex-combined sample and the Big40 database IDPs. The IDP ID (IDP), chromosome (chr), rsid, genomic position (pos), effect allele (a1), reference allele (a2), effect size, standard error (se), pvalue, IDP short name and IDP category name are listed.

**Table A13** All 1,228 statistically significant associations ( $P < 1.93 \times 10^{-7}$ ) between the genetic variants associated with hypothalamic grey matter volume in a sex-combined sample and the Big40 database IDPs. The IDP ID (IDP), chromosome (chr), rsid, genomic position (pos), effect allele (a1), reference allele (a2), effect size, standard error (se), pvalue, IDP short name and IDP category name are listed.

**Table A14** Genetic correlation between the four neuroendocrine IDPs and reproductive traits. The reproductive traits include testosterone (test), follicle stimulating hormone (FSH), luteinising hormone (LH), oestradiol (oest) and progesterone (prog) levels, female all cause infertility (infertility1), female anatomical cause infertility (infertility2), female anovulatory infertility (infertility3), female unknown cause infertility excluding idiopathic infertility (infertility4), female unknown cause infertility including idiopathic infertility (infertility5), and male infertility (male\_infert).

**Table A15** Results from Mendelian randomisation analyses of relationships between the volume of the hypothalamus, pituitary gland, olfactory bulb and hypothalamic grey matter volume, and reproductive phenotypes including: testosterone (test), follicle stimulating hormone (FSH), luteinising hormone (LH), oestradiol (oest) and progesterone (prog) levels, female all cause infertility (infertility1), female anatomical cause infertility (infertility2), female anovulatory infertility (infertility3), female unknown cause infertility excluding idiopathic infertility (infertility4), female unknown cause infertility including idiopathic infertility (infertility5), and male infertility (male\_infert).

**Table A16** Genetic variants significantly colocalised ( $P(\frac{H_4}{H_3}) > 5$  and  $P(H_4) > 0.5$ ) with tissue-specific eQTLs within the GTEx dataset [32].



**Fig. A7 Pathway enrichment of hypothalamus associated genetic variants in a female sample.** Significance of enrichment of the genetic variants associated with hypothalamus volume in females across different pathways within physiological systems, assessed using DEPICT [43]. Significant associations ( $P < 0.05$  after BH FDR correction) are highlighted in red.

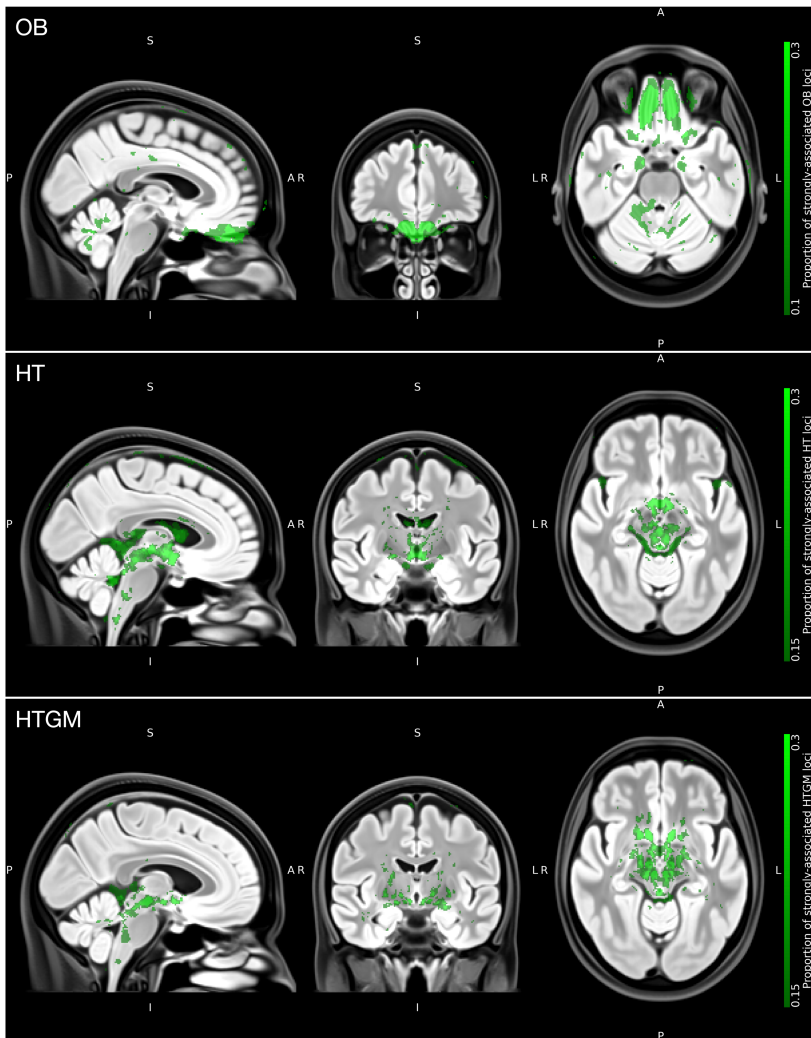
**Table A17** Gene prioritisation of genetic variants associated with the four neuroendocrine IDPs. Prioritisation was assessed using DEPICT [43] and BH FDR correction used with a significance threshold of 0.05.

**Table A18** Pathway enrichment of genetic variants associated with the four neuroendocrine IDPs. Enrichment was assessed using DEPICT [43] and BH FDR correction used with a significance threshold of 0.05.

**Table A19** Tissue enrichment of genetic variants associated with the four neuroendocrine IDPs. Enrichment was assessed using DEPICT [43] and BH FDR correction used with a significance threshold of 0.05.

**Table A20** All results from enrichment analysis for gene expression in ovarian cell-types using CELLEX and CELLECT [45]. Gene expression data represents 15 cell populations derived from five harmonised publicly available ovarian datasets and four ovary samples processed in-house; GC, granulosa cell; TC, theca cell; COC, cumulus-oocyte complex.

**Table A21** All results from enrichment analysis for gene expression in hypothalamic cell-types using CELLEX and CELLECT [45]. Gene expression data represents 38 cell populations derived from a published human postmortem hypothalamus dataset [48].



**Fig. A8** Voxel-wise associations of neuroendocrine IDP associated genetic variants. Summary maps overlaid on the OMM-1 reference template showing the proportion of strongly-associated loci for each voxel, from each of the three sets of loci identified in GWAS (olfactory bulbs (OB), hypothalamus (HT), and hypothalamic grey matter (HTGM) volume in a sex-combined sample).

## References

937

938 [1] Herbison, A.E.: The Gonadotropin-Releasing hormone pulse generator.  
939 *Endocrinology* **159**(11), 3723–3736 (2018)

940 [2] Cho, H.-J., Shan, Y., Whittington, N.C., Wray, S.: Nasal placode devel-  
941 opment, GnRH neuronal migration and kallmann syndrome. *Front Cell*  
942 *Dev Biol* **7**, 121 (2019)

943 [3] Sheng, J.A., Bales, N.J., Myers, S.A., Bautista, A.I., Roueifar, M., Hale,  
944 T.M., Handa, R.J.: The Hypothalamic-Pituitary-Adrenal axis: Develop-  
945 ment, programming actions of hormones, and Maternal-Fetal interactions.  
946 *Front. Behav. Neurosci.* **14**, 601939 (2020)

947 [4] Tsunoda, A., Okuda, O., Sato, K.: MR height of the pituitary gland as  
948 a function of age and sex: especially physiological hypertrophy in ado-  
949 lescence and in climacterium. *AJNR Am. J. Neuroradiol.* **18**(3), 551–554  
950 (1997)

951 [5] Koenigkam-Santos, M., Santos, A.C., Versiani, B.R., Diniz, P.R.B.,  
952 Junior, J.E., de Castro, M.: Quantitative magnetic resonance imaging  
953 evaluation of the olfactory system in kallmann syndrome: correlation with  
954 a clinical smell test. *Neuroendocrinology* **94**(3), 209–217 (2011)

955 [6] Elliott, L.T., Sharp, K., Alfaro-Almagro, F., Shi, S., Miller, K.L., Douaud,  
956 G., Marchini, J., Smith, S.M.: Genome-wide association studies of brain  
957 imaging phenotypes in UK biobank. *Nature* **562**(7726), 210–216 (2018)

958 [7] Pirruccello, J.P., Bick, A., Wang, M., Chaffin, M., Friedman, S., Yao,  
959 J., Guo, X., Venkatesh, B.A., Taylor, K.D., Post, W.S., Rich, S., Lima,  
960 J.A.C., Rotter, J.I., Philippakis, A., Lubitz, S.A., Ellinor, P.T., Khera,  
961 A.V., Kathiresan, S., Aragam, K.G.: Analysis of cardiac magnetic reso-  
962 nance imaging in 36,000 individuals yields genetic insights into dilated  
963 cardiomyopathy. *Nat. Commun.* **11**(1), 2254 (2020)

964 [8] Currant, H., Hysi, P., Fitzgerald, T.W., Gharahkhani, P., Bonnemaier,  
965 P.W.M., Senabouth, A., Hewitt, A.W., UK Biobank Eye and Vision Con-  
966 sortium, International Glaucoma Genetics Consortium, Atan, D., Aung,  
967 T., Charng, J., Choquet, H., Craig, J., Khaw, P.T., Klaver, C.C.W.,  
968 Kubo, M., Ong, J.-S., Pasquale, L.R., Reisman, C.A., Daniszewski, M.,  
969 Powell, J.E., Pébay, A., Simcoe, M.J., Thiadens, A.A.H.J., van Duijn,  
970 C.M., Yazar, S., Jorgenson, E., MacGregor, S., Hammond, C.J., Mackey,  
971 D.A., Wiggs, J.L., Foster, P.J., Patel, P.J., Birney, E., Khawaja, A.P.:  
972 Genetic variation affects morphological retinal phenotypes extracted from  
973 UK biobank optical coherence tomography images. *PLoS Genet.* **17**(5),  
974 1009497 (2021)



975 [9] Bethlehem, R.A.I., Seidlitz, J., White, S.R., Vogel, J.W., Anderson, K.M.,  
976 Adamson, C., Adler, S., Alexopoulos, G.S., Anagnostou, E., Arces-  
977 Gonzalez, A., Astle, D.E., Auyeung, B., Ayub, M., Bae, J., Ball, G.,  
978 Baron-Cohen, S., Beare, R., Bedford, S.A., Benegal, V., Beyer, F.,  
979 Blangero, J., Blesa Cábez, M., Boardman, J.P., Borzage, M., Bosch-  
980 Bayard, J.F., Bourke, N., Calhoun, V.D., Chakravarty, M.M., Chen,  
981 C., Chertavian, C., Chetelat, G., Chong, Y.S., Cole, J.H., Corvin, A.,  
982 Costantino, M., Courchesne, E., Crivello, F., Croypley, V.L., Crosbie, J.,  
983 Crossley, N., Delarue, M., Delorme, R., Desrivieres, S., Devenyi, G.A.,  
984 Di Biase, M.A., Dolan, R., Donald, K.A., Donohoe, G., Dunlop, K.,  
985 Edwards, A.D., Elison, J.T., Ellis, C.T., Elman, J.A., Eyster, L., Fair,  
986 D.A., Feczko, E., Fletcher, P.C., Fonagy, P., Franz, C.E., Galan-Garcia,  
987 L., Gholipour, A., Giedd, J., Gilmore, J.H., Glahn, D.C., Goodyer, I.M.,  
988 Grant, P.E., Groenewold, N.A., Gunning, F.M., Gur, R.E., Gur, R.C.,  
989 Hammill, C.F., Hansson, O., Hedden, T., Heinz, A., Henson, R.N., Heuer,  
990 K., Hoare, J., Holla, B., Holmes, A.J., Holt, R., Huang, H., Im, K.,  
991 Ipser, J., Jack, C.R. Jr, Jackowski, A.P., Jia, T., Johnson, K.A., Jones,  
992 P.B., Jones, D.T., Kahn, R.S., Karlsson, H., Karlsson, L., Kawashima,  
993 R., Kelley, E.A., Kern, S., Kim, K.W., Kitzbichler, M.G., Kremen, W.S.,  
994 Lalonde, F., Landeau, B., Lee, S., Lerch, J., Lewis, J.D., Li, J., Liao, W.,  
995 Liston, C., Lombardo, M.V., Lv, J., Lynch, C., Mallard, T.T., Marcelis,  
996 M., Markello, R.D., Mathias, S.R., Mazoyer, B., McGuire, P., Meaney,  
997 M.J., Mechelli, A., Medic, N., Mistic, B., Morgan, S.E., Mothersill, D.,  
998 Nigg, J., Ong, M.Q.W., Ortinau, C., Ossenkoppele, R., Ouyang, M.,  
999 Palaniyappan, L., Paly, L., Pan, P.M., Pantelis, C., Park, M.M., Paus,  
1000 T., Pausova, Z., Paz-Linares, D., Pichet Binette, A., Pierce, K., Qian,  
1001 X., Qiu, J., Qiu, A., Raznahan, A., Rittman, T., Rodrigue, A., Rollins,  
1002 C.K., Romero-Garcia, R., Ronan, L., Rosenberg, M.D., Rowitch, D.H.,  
1003 Salum, G.A., Satterthwaite, T.D., Schaare, H.L., Schachar, R.J., Schultz,  
1004 A.P., Schumann, G., Schöll, M., Sharp, D., Shinohara, R.T., Skoog, I.,  
1005 Smyser, C.D., Sperling, R.A., Stein, D.J., Stolicyn, A., Suckling, J.,  
1006 Sullivan, G., Taki, Y., Thyreau, B., Toro, R., Traut, N., Tsvetanov,  
1007 K.A., Turk-Browne, N.B., Tuulari, J.J., Tzourio, C., Vachon-Preseau,  
1008 É., Valdes-Sosa, M.J., Valdes-Sosa, P.A., Valk, S.L., van Amelsvoort, T.,  
1009 Vandekar, S.N., Vasung, L., Victoria, L.W., Villeneuve, S., Villringer, A.,  
1010 Vértes, P.E., Wagstyl, K., Wang, Y.S., Warfield, S.K., Warriar, V., West-  
1011 man, E., Westwater, M.L., Whalley, H.C., Witte, A.V., Yang, N., Yeo,  
1012 B., Yun, H., Zalesky, A., Zar, H.J., Zettergren, A., Zhou, J.H., Ziaud-  
1013 deen, H., Zugman, A., Zuo, X.N., 3R-BRAIN, AIBL, Alzheimer's Disease  
1014 Neuroimaging Initiative, Alzheimer's Disease Repository Without Bor-  
1015 ders Investigators, CALM Team, Cam-CAN, CCNP, COBRE, cVEDA,  
1016 ENIGMA Developmental Brain Age Working Group, Developing Human  
1017 Connectome Project, FinnBrain, Harvard Aging Brain Study, IMAGEN,  
1018 KNE96, Mayo Clinic Study of Aging, NSPN, POND, PREVENT-AD  
1019 Research Group, VETSA, Bullmore, E.T., Alexander-Bloch, A.F.: Brain

- 1020 charts for the human lifespan. *Nature* **604**(7906), 525–533 (2022)
- 1021 [10] Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J.,  
1022 Schizophrenia Working Group of the Psychiatric Genomics Consortium,  
1023 Patterson, N., Daly, M.J., Price, A.L., Neale, B.M.: LD score regression  
1024 distinguishes confounding from polygenicity in genome-wide association  
1025 studies. *Nat. Genet.* **47**(3), 291–295 (2015)
- 1026 [11] Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Genetic Investiga-  
1027 tion of ANthropometric Traits (GIANT) Consortium, DIABetes Genet-  
1028 ics Replication And Meta-analysis (DIAGRAM) Consortium, Madden,  
1029 P.A.F., Heath, A.C., Martin, N.G., Montgomery, G.W., Weedon, M.N.,  
1030 Loos, R.J., Frayling, T.M., McCarthy, M.I., Hirschhorn, J.N., God-  
1031 dard, M.E., Visscher, P.M.: Conditional and joint multiple-SNP analysis  
1032 of GWAS summary statistics identifies additional variants influencing  
1033 complex traits. *Nat. Genet.* **44**(4), 369–7513 (2012)
- 1034 [12] Tadayon, S.H., Vaziri-Pashkam, M., Kahali, P., Ansari Dezfouli, M.,  
1035 Abbassian, A.: Common genetic variant in VIT is associated with human  
1036 brain asymmetry. *Front. Hum. Neurosci.* **10**, 236 (2016)
- 1037 [13] Liu, Y., Xie, J.: Cauchy combination test: a powerful test with analytic  
1038 p-value calculation under arbitrary dependency structures. *J. Am. Stat.*  
1039 *Assoc.* **115**(529), 393–402 (2020)
- 1040 [14] van der Meer, D., Kaufmann, T., Shadrin, A.A., Makowski, C., Frei, O.,  
1041 Roelfs, D., Monereo-Sánchez, J., Linden, D.E.J., Rokicki, J., Alnæs, D.,  
1042 de Leeuw, C., Thompson, W.K., Loughnan, R., Fan, C.C., Westlye, L.T.,  
1043 Andreassen, O.A., Dale, A.M.: The genetic architecture of human cortical  
1044 folding. *Sci Adv* **7**(51), 9446 (2021)
- 1045 [15] Shadrin, A.A., Kaufmann, T., van der Meer, D., Palmer, C.E., Makowski,  
1046 C., Loughnan, R., Jernigan, T.L., Seibert, T.M., Hagler, D.J., Smeland,  
1047 O.B., Motazed, E., Chu, Y., Lin, A., Cheng, W., Hindley, G., Thomp-  
1048 son, W.K., Fan, C.C., Holland, D., Westlye, L.T., Frei, O., Andreassen,  
1049 O.A., Dale, A.M.: Vertex-wise multivariate genome-wide association study  
1050 identifies 780 unique genetic loci associated with cortical morphology.  
1051 *Neuroimage* **244**, 118603 (2021)
- 1052 [16] Kzhyshkowska, J., Mamidi, S., Gratchev, A., Kremmer, E., Schmutter-  
1053 maier, C., Krusell, L., Haus, G., Utikal, J., Schledzewski, K., Scholtze,  
1054 J., Goerdt, S.: Novel stabilin-1 interacting chitinase-like protein (SI-CLP)  
1055 is up-regulated in alternatively activated macrophages and secreted via  
1056 lysosomal pathway. *Blood* **107**(8), 3221–3228 (2006)

- 1057 [17] Gong, W., Fu, Y., Wu, B.-S., Du, J., Yang, L., Zhang, Y.-R., Chen, S.-  
1058 D., Kang, J., Mao, Y., Dong, Q., Tan, L., Feng, J., Cheng, W., Yu, J.-  
1059 T.: Whole-exome sequencing identifies protein-coding variants associated  
1060 with brain iron in 29,828 individuals. *Nat. Commun.* **15**(1), 5540 (2024)
- 1061 [18] Karczewski, K.J., Solomonson, M., Chao, K.R., Goodrich, J.K., Tiao, G.,  
1062 Lu, W., Riley-Gillis, B.M., Tsai, E.A., Kim, H.I., Zheng, X., Rahimov,  
1063 F., Esmaeeli, S., Grundstad, A.J., Reppell, M., Waring, J., Jacob, H.,  
1064 Sexton, D., Bronson, P.G., Chen, X., Hu, X., Goldstein, J.I., King, D.,  
1065 Vittal, C., Poterba, T., Palmer, D.S., Churchhouse, C., Howrigan, D.P.,  
1066 Zhou, W., Watts, N.A., Nguyen, K., Nguyen, H., Mason, C., Farnham, C.,  
1067 Tolonen, C., Gauthier, L.D., Gupta, N., MacArthur, D.G., Rehm, H.L.,  
1068 Seed, C., Philippakis, A.A., Daly, M.J., Davis, J.W., Runz, H., Miller,  
1069 M.R., Neale, B.M.: Systematic single-variant and gene-based association  
1070 testing of thousands of phenotypes in 394,841 UK biobank exomes. *Cell*  
1071 *Genom* **2**(9), 100168 (2022)
- 1072 [19] Thomas, G.E.C., Leyland, L.A., Schrag, A.-E., Lees, A.J., Acosta-  
1073 Cabronero, J., Weil, R.S.: Brain iron deposition is linked with cognitive  
1074 severity in parkinson’s disease. *J. Neurol. Neurosurg. Psychiatry* **91**(4),  
1075 418–425 (2020)
- 1076 [20] Wang, Q., Zhang, Y., Wang, M., Song, W.-M., Shen, Q., McKenzie, A.,  
1077 Choi, I., Zhou, X., Pan, P.-Y., Yue, Z., Zhang, B.: The landscape of mul-  
1078 tiscale transcriptomic networks and key regulators in parkinson’s disease.  
1079 *Nat. Commun.* **10**(1), 5234 (2019)
- 1080 [21] Carroll, R.J., Bastarache, L., Denny, J.C.: R PheWAS: data analysis and  
1081 plotting tools for phenome-wide association studies in the R environment.  
1082 *Bioinformatics* **30**(16), 2375–2376 (2014)
- 1083 [22] Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: A prac-  
1084 tical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B*  
1085 *Stat. Methodol.* **57**(1), 289–300 (1995)
- 1086 [23] Gabrielsen, J.S., Lamb, D.J., Lipshultz, L.I.: Iron and a man’s reproduc-  
1087 tive health: the good, the bad, and the ugly. *Curr. Urol. Rep.* **19**(8), 60  
1088 (2018)
- 1089 [24] Beggs, L.A., Yarrow, J.F., Conover, C.F., Meuleman, J.R., Beck, D.T.,  
1090 Morrow, M., Zou, B., Shuster, J.J., Borst, S.E.: Testosterone alters iron  
1091 metabolism and stimulates red blood cell production independently of  
1092 dihydrotestosterone. *Am. J. Physiol. Endocrinol. Metab.* **307**(5), 456–61  
1093 (2014)
- 1094 [25] McCoy, S.S., Sampene, E., Baer, A.N.: Association of sjögren’s syn-  
1095 drome with reduced lifetime sex hormone exposure: A Case-Control study.

- 1096 Arthritis Care Res. **72**(9), 1315–1322 (2020)
- 1097 [26] Baer, A.N., Walitt, B.: Sjögren syndrome and other causes of sicca in  
1098 older adults. *Clin. Geriatr. Med.* **33**(1), 87–103 (2017)
- 1099 [27] Smith, S.M., Douaud, G., Chen, W., Hanayik, T., Alfaro-Almagro, F.,  
1100 Sharp, K., Elliott, L.T.: An expanded set of genome-wide association stud-  
1101 ies of brain imaging phenotypes in UK biobank. *Nat. Neurosci.* **24**(5),  
1102 737–745 (2021)
- 1103 [28] Hwang, J.-Y., Zukin, R.S.: REST, a master transcriptional regulator in  
1104 neurodegenerative disease. *Curr. Opin. Neurobiol.* **48**, 193–200 (2018)
- 1105 [29] NealeLab: Neale Lab summary statistics. [http://www.nealelab.is/  
1106 uk-biobank/](http://www.nealelab.is/uk-biobank/)
- 1107 [30] Ong, J.-S., Seviiri, M., Dusingize, J.C., Wu, Y., Han, X., Shi, J., Olsen,  
1108 C.M., Neale, R.E., Thompson, J.F., Saw, R.P.M., Shannon, K.F., Mann,  
1109 G.J., Martin, N.G., Medland, S.E., Gordon, S.D., Scolyer, R.A., Long,  
1110 G.V., Iles, M.M., Landi, M.T., Whiteman, D.C., MacGregor, S., Law,  
1111 M.H.: Uncovering the complex relationship between balding, testosterone  
1112 and skin cancers in men. *Nat. Commun.* **14**(1), 5962 (2023)
- 1113 [31] Ruth, K.S., Day, F.R., Tyrrell, J., Thompson, D.J., Wood, A.R., Mahajan,  
1114 A., Beaumont, R.N., Wittemans, L., Martin, S., Busch, A.S., Erzurum-  
1115 luoglu, A.M., Hollis, B., O’Mara, T.A., Endometrial Cancer Association  
1116 Consortium, McCarthy, M.I., Langenberg, C., Easton, D.F., Wareham,  
1117 N.J., Burgess, S., Murray, A., Ong, K.K., Frayling, T.M., Perry, J.R.B.:  
1118 Using human genetics to understand the disease impacts of testosterone  
1119 in men and women. *Nat. Med.* **26**(2), 252–258 (2020)
- 1120 [32] GTEx Consortium: The Genotype-Tissue expression (GTEx) project.  
1121 *Nat. Genet.* **45**(6), 580–585 (2013)
- 1122 [33] Sicinski, P., Donaher, J.L., Geng, Y., Parker, S.B., Gardner, H., Park,  
1123 M.Y., Robker, R.L., Richards, J.S., McGinnis, L.K., Biggers, J.D., Eppig,  
1124 J.J., Bronson, R.T., Elledge, S.J., Weinberg, R.A.: Cyclin D2 is an FSH-  
1125 responsive gene involved in gonadal cell proliferation and oncogenesis.  
1126 *Nature* **384**(6608), 470–474 (1996)
- 1127 [34] Tan, K.A.L., Turner, K.J., Saunders, P.T.K., Verhoeven, G., De Gendt,  
1128 K., Atanassova, N., Sharpe, R.M.: Androgen regulation of stage-  
1129 dependent cyclin D2 expression in sertoli cells suggests a role in modulat-  
1130 ing androgen action on spermatogenesis. *Biol. Reprod.* **72**(5), 1151–1160  
1131 (2005)
- 1132 [35] Zhao, L., Han, S., Su, H., Li, J., Zhi, E., Li, P., Yao, C., Tian, R., Chen,

- 1133 H., Chen, H., Luo, J., Shi, C., Ji, Z., Hu, J., Wu, G., Zhou, W., Tang,  
1134 Y., Chen, Y., Lin, G., Lue, T.F., Wu, D., Li, Z.: Single-cell transcriptome  
1135 atlas of the human corpus cavernosum. *Nat. Commun.* **13**(1), 4302 (2022)
- 1136 [36] van Baren, M.J., van der Linde, H.C., Breedveld, G.J., Baarends, W.M.,  
1137 Rizzu, P., de Graaff, E., Oostra, B.A., Heutink, P.: A double RING-H2  
1138 domain in RNF32, a gene expressed during sperm formation. *Biochem.*  
1139 *Biophys. Res. Commun.* **292**(1), 58–65 (2002)
- 1140 [37] Chen, Y., Wang, J., Xu, D., Xiang, Z., Ding, J., Yang, X., Li, D., Han,  
1141 X.: m6a mRNA methylation regulates testosterone synthesis through  
1142 modulating autophagy in leydig cells. *Autophagy* **17**(2), 457–475 (2021)
- 1143 [38] Zhao, B., Luo, T., Li, T., Li, Y., Zhang, J., Shan, Y., Wang, X., Yang, L.,  
1144 Zhou, F., Zhu, Z., Alzheimer’s Disease Neuroimaging Initiative, Pediatric  
1145 Imaging, Neurocognition and Genetics, Zhu, H.: Genome-wide associa-  
1146 tion analysis of 19,629 individuals identifies variants influencing regional  
1147 brain volumes and refines their genetic co-architecture with cognitive and  
1148 mental health traits. *Nat. Genet.* **51**(11), 1637–1644 (2019)
- 1149 [39] Venkatesh, S.S., Wittemans, L.B.L., Palmer, D.S., Baya, N.A., Ferreira,  
1150 T., Hill, B., Lassen, F.H., Parker, M.J., Reibe, S., Elhakeem, A., Banasik,  
1151 K., Bruun, M.T., Erikstrup, C., Jensen, B.A., Juul, A., Mikkelsen, C.,  
1152 Nielsen, H.S., Ostrowski, S.R., Pedersen, O.B., Rohde, P.D., Sorensen,  
1153 E., Ullum, H., Westergaard, D., Haraldsson, A., Holm, H., Jonsdottir, I.,  
1154 Olafsson, I., Steingrimsdottir, T., Steinthorsdottir, V., Thorleifsson, G.,  
1155 Figueredo, J., Karjalainen, M.K., Pasanen, A., Jacobs, B.M., Hubers, N.,  
1156 DBDS Genomic Consortium, FinnGen, Lippincott, M., Fraser, A., Lawlor,  
1157 D.A., Timpson, N.J., Nyegaard, M., Stefansson, K., Magi, R., Laivuori,  
1158 H., van Heel, D.A., Boomsma, D.I., Balasubramanian, R., Seminara, S.B.,  
1159 Chan, Y.-M., Laisk, T., Lindgren, C.M.: Genome-wide analyses identify  
1160 21 infertility loci and over 400 reproductive hormone loci across the allele  
1161 frequency spectrum (2024)
- 1162 [40] Nishimura, H., Gupta, S., Myles, D.G., Primakoff, P.: Characterization of  
1163 mouse sperm TMEM190, a small transmembrane protein with the trefoil  
1164 domain: evidence for co-localization with IZUMO1 and complex formation  
1165 with other sperm proteins. *Reproduction* **141**(4), 437–451 (2011)
- 1166 [41] Frolikova, M., Sebkova, N., Ded, L., Dvorakova-Hortova, K.: Characteri-  
1167 zation of CD46 and  $\beta$ 1 integrin dynamics during sperm acrosome reaction.  
1168 *Sci. Rep.* **6**, 33714 (2016)
- 1169 [42] Chen, J., Gu, Y., Zhang, Z., Zheng, W., Yang, L., Huang, W., Lin, S., Li,  
1170 Y., Guo, H., Luo, M., Ma, Q., Jiang, Z., Tang, A., Gui, Y.: Deficiency of  
1171 SPATA46, a novel nuclear membrane protein, causes subfertility in male  
1172 mice. *Biol. Reprod.* **95**(3), 58 (2016)

- 1173 [43] Pers, T.H., Karjalainen, J.M., Chan, Y., Westra, H.-J., Wood, A.R.,  
1174 Yang, J., Lui, J.C., Vedantam, S., Gustafsson, S., Esko, T., Frayling,  
1175 T., Speliotes, E.K., Genetic Investigation of ANthropometric Traits  
1176 (GIANT) Consortium, Boehnke, M., Raychaudhuri, S., Fehrmann,  
1177 R.S.N., Hirschhorn, J.N., Franke, L.: Biological interpretation of genome-  
1178 wide association studies using predicted gene functions. *Nat. Commun.*  
1179 **6**, 5890 (2015)
- 1180 [44] 1000 Genomes Project Consortium, Abecasis, G.R., Altshuler, D., Auton,  
1181 A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., McVean,  
1182 G.A.: A map of human genome variation from population-scale sequenc-  
1183 ing. *Nature* **467**(7319), 1061–1073 (2010)
- 1184 [45] Timshel, P.N., Thompson, J.J., Pers, T.H.: Genetic mapping of etiologic  
1185 brain cell types for obesity. *Elife* **9** (2020)
- 1186 [46] Richards, J.S., Ren, Y.A., Candelaria, N., Adams, J.E., Rajkovic, A.:  
1187 Ovarian follicular theca cell recruitment, differentiation, and impact on  
1188 fertility: 2017 update. *Endocr. Rev.* **39**(1), 1–20 (2018)
- 1189 [47] Auersperg, N., Wong, A.S., Choi, K.C., Kang, S.K., Leung, P.C.: Ovarian  
1190 surface epithelium: biology, endocrinology, and pathology. *Endocr. Rev.*  
1191 **22**(2), 255–288 (2001)
- 1192 [48] Siletti, K., Hodge, R., Mossi Albiach, A., Lee, K.W., Ding, S.-L., Hu,  
1193 L., Lönnberg, P., Bakken, T., Casper, T., Clark, M., Dee, N., Gloe,  
1194 J., Hirschstein, D., Shapovalova, N.V., Keene, C.D., Nyhus, J., Tung,  
1195 H., Yanny, A.M., Arenas, E., Lein, E.S., Linnarsson, S.: Transcriptomic  
1196 diversity of cell types across the adult human brain. *Science* **382**(6667),  
1197 7046 (2023)
- 1198 [49] Prevot, V., Bellefontaine, N., Baroncini, M., Sharif, A., Hanchate, N.K.,  
1199 Parkash, J., Campagne, C., de Seranno, S.: Gonadotrophin-releasing hor-  
1200 mone nerve terminals, tanycytes and neurohaemal junction remodelling  
1201 in the adult median eminence: functional consequences for reproduction  
1202 and dynamic role of vascular endothelial cells. *J. Neuroendocrinol.* **22**(7),  
1203 639–649 (2010)
- 1204 [50] Rodríguez, E.M., Blázquez, J.L., Pastor, F.E., Peláez, B., Peña, P.,  
1205 Peruzzo, B., Amat, P.: Hypothalamic tanycytes: a key component of  
1206 brain-endocrine interaction. *Int. Rev. Cytol.* **247**, 89–164 (2005)
- 1207 [51] Miller, K.L., Alfaro-Almagro, F., Bangerter, N.K., Thomas, D.L., Yacoub,  
1208 E., Xu, J., Bartsch, A.J., Jbabdi, S., Sotiropoulos, S.N., Andersson,  
1209 J.L.R., Griffanti, L., Douaud, G., Okell, T.W., Weale, P., Dragonu, I.,  
1210 Garratt, S., Hudson, S., Collins, R., Jenkinson, M., Matthews, P.M.,  
1211 Smith, S.M.: Multimodal population brain imaging in the UK biobank

- 1212 prospective epidemiological study. *Nat. Neurosci.* **19**(11), 1523–1536  
1213 (2016)
- 1214 [52] Arthofer, C., Smith, S.M., Douaud, G., Bartsch, A.J., Andersson, J.,  
1215 Lange, F.J.: Multimodal MRI brain templates from UK Biobank: Oxford-  
1216 MM. *Open Science Framework* (2022)
- 1217 [53] Alfaro-Almagro, F., Jenkinson, M., Bangerter, N.K., Andersson, J.L.R.,  
1218 Griffanti, L., Douaud, G., Sotiropoulos, S.N., Jbabdi, S., Hernandez-  
1219 Fernandez, M., Vallee, E., Vidaurre, D., Webster, M., McCarthy, P.,  
1220 Rorden, C., Daducci, A., Alexander, D.C., Zhang, H., Dragonu, I.,  
1221 Matthews, P.M., Miller, K.L., Smith, S.M.: Image processing and qual-  
1222 ity control for the first 10,000 brain imaging datasets from UK biobank.  
1223 *Neuroimage* **166**, 400–424 (2018)
- 1224 [54] Jenkinson, M., Smith, S.: A global optimisation method for robust affine  
1225 registration of brain images. *Med. Image Anal.* **5**(2), 143–156 (2001)
- 1226 [55] Zhang, Y., Brady, M., Smith, S.: Segmentation of brain MR images  
1227 through a hidden markov random field model and the expectation-  
1228 maximization algorithm. *IEEE Trans. Med. Imaging* **20**(1), 45–57 (2001)
- 1229 [56] Basser, P.J., Mattiello, J., LeBihan, D.: Estimation of the effective self-  
1230 diffusion tensor from the NMR spin echo. *J. Magn. Reson. B* **103**(3),  
1231 247–254 (1994)
- 1232 [57] Andersson, J.L.R., Skare, S., Ashburner, J.: How to correct susceptibility  
1233 distortions in spin-echo echo-planar images: application to diffusion tensor  
1234 imaging. *Neuroimage* **20**(2), 870–888 (2003)
- 1235 [58] Andersson, J.L.R., Sotiropoulos, S.N.: An integrated approach to cor-  
1236 rection for off-resonance effects and subject movement in diffusion MR  
1237 imaging. *Neuroimage* **125**, 1063–1078 (2016)
- 1238 [59] Andersson, J.L.R., Graham, M.S., Zsoldos, E., Sotiropoulos, S.N.: Incor-  
1239 porating outlier detection and replacement into a non-parametric frame-  
1240 work for movement and distortion correction of diffusion MR images.  
1241 *Neuroimage* **141**, 556–572 (2016)
- 1242 [60] Neudorfer, C., Germann, J., Elias, G.J.B., Gramer, R., Boutet, A.,  
1243 Lozano, A.M.: A high-resolution in vivo magnetic resonance imaging atlas  
1244 of the human hypothalamic region. *Sci Data* **7**(1), 305 (2020)
- 1245 [61] Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E.J., Yacoub,  
1246 E., Ugurbil, K., WU-Minn HCP Consortium: The WU-Minn human  
1247 connectome project: an overview. *Neuroimage* **80**, 62–79 (2013)



- 1248 [62] Pauli, W.M., Nili, A.N., Tyszka, J.M.: A high-resolution probabilistic in  
1249 vivo atlas of human subcortical brain nuclei. *Sci Data* **5**, 180063 (2018)
- 1250 [63] Lange, F.J., Arthofer, C., Bartsch, A., Douaud, G., McCarthy, P.,  
1251 Smith, S.M., Andersson, J.L.R.: MMORF—FSL’s MultiMODal registra-  
1252 tion framework. *Imaging Neuroscience* **2**, 1–30 (2024)
- 1253 [64] 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin,  
1254 R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L.,  
1255 McCarthy, S., McVean, G.A., Abecasis, G.R.: A global reference for  
1256 human genetic variation. *Nature* **526**(7571), 68–74 (2015)
- 1257 [65] Mbatchou, J., Barnard, L., Backman, J., Marcketta, A., Kosmicki, J.A.,  
1258 Ziyatdinov, A., Benner, C., O’Dushlaine, C., Barber, M., Boutkov, B.,  
1259 Habegger, L., Ferreira, M., Baras, A., Reid, J., Abecasis, G., Maxwell,  
1260 E., Marchini, J.: Computationally efficient whole-genome regression for  
1261 quantitative and binary traits. *Nat. Genet.* **53**(7), 1097–1103 (2021)
- 1262 [66] Martin, F.J., Amode, M.R., Aneja, A., Austine-Orimoloye, O., Azov,  
1263 A.G., Barnes, I., Becker, A., Bennett, R., Berry, A., Bhai, J., Bhurji, S.K.,  
1264 Bignell, A., Boddu, S., Branco Lins, P.R., Brooks, L., Ramaraju, S.B.,  
1265 Charkhchi, M., Cockburn, A., Da Rin Fiorretto, L., Davidson, C., Dodiya,  
1266 K., Donaldson, S., El Houdaigui, B., El Naboulsi, T., Fatima, R., Giron,  
1267 C.G., Genez, T., Ghattaoraya, G.S., Martinez, J.G., Guijarro, C., Hardy,  
1268 M., Hollis, Z., Hourlier, T., Hunt, T., Kay, M., Kaykala, V., Le, T., Lemos,  
1269 D., Marques-Coelho, D., Marugán, J.C., Merino, G.A., Mirabueno, L.P.,  
1270 Mushtaq, A., Hossain, S.N., Ogeh, D.N., Sakthivel, M.P., Parker, A.,  
1271 Perry, M., Piližota, I., Prosovetskaia, I., Pérez-Silva, J.G., Salam, A.I.A.,  
1272 Saraiva-Agostinho, N., Schuilenburg, H., Sheppard, D., Sinha, S., Sipos,  
1273 B., Stark, W., Steed, E., Sukumaran, R., Sumathipala, D., Suner, M.-M.,  
1274 Surapaneni, L., Sutinen, K., Szpak, M., Tricomi, F.F., Urbina-Gómez, D.,  
1275 Veidenberg, A., Walsh, T.A., Walts, B., Wass, E., Willhoft, N., Allen, J.,  
1276 Alvarez-Jarreta, J., Chakiachvili, M., Flint, B., Giorgetti, S., Haggerty,  
1277 L., Ilesley, G.R., Loveland, J.E., Moore, B., Mudge, J.M., Tate, J., Thy-  
1278 bert, D., Trevanion, S.J., Winterbottom, A., Frankish, A., Hunt, S.E.,  
1279 Ruffier, M., Cunningham, F., Dyer, S., Finn, R.D., Howe, K.L., Harrison,  
1280 P.W., Yates, A.D., Flicek, P.: Ensembl 2023. *Nucleic Acids Res.* **51**(D1),  
1281 933–941 (2023)
- 1282 [67] Ghossaini, M., Mountjoy, E., Carmona, M., Peat, G., Schmidt, E.M.,  
1283 Hercules, A., Fumis, L., Miranda, A., Carvalho-Silva, D., Buniello, A.,  
1284 Burdett, T., Hayhurst, J., Baker, J., Ferrer, J., Gonzalez-Uriarte, A.,  
1285 Jupp, S., Karim, M.A., Koscielny, G., Machlitt-Northen, S., Malangone,  
1286 C., Pendlington, Z.M., Roncaglia, P., Suveges, D., Wright, D., Vrousou,  
1287 O., Papa, E., Parkinson, H., MacArthur, J.A.L., Todd, J.A., Barrett, J.C.,

1288 Schwartzentruber, J., Hulcoop, D.G., Ochoa, D., McDonagh, E.M., Dun-  
1289 ham, I.: Open targets genetics: systematic identification of trait-associated  
1290 genes using large-scale genetics and functional genomics. *Nucleic Acids*  
1291 *Res.* **49**(D1), 1311–1320 (2021)

1292 [68] Randall, J.C., Winkler, T.W., Kutalik, Z., Berndt, S.I., Jackson, A.U.,  
1293 Monda, K.L., Kilpeläinen, T.O., Esko, T., Mägi, R., Li, S., Workalemahu,  
1294 T., Feitosa, M.F., Croteau-Chonka, D.C., Day, F.R., Fall, T., Ferreira,  
1295 T., Gustafsson, S., Locke, A.E., Mathieson, I., Scherag, A., Vedantam,  
1296 S., Wood, A.R., Liang, L., Steinthorsdottir, V., Thorleifsson, G., Derm-  
1297 mitzakis, E.T., Dimas, A.S., Karpe, F., Min, J.L., Nicholson, G., Clegg,  
1298 D.J., Person, T., Krohn, J.P., Bauer, S., Buechler, C., Eisinger, K., DIA-  
1299 GRAM Consortium, Bonnefond, A., Froguel, P., MAGIC Investigators,  
1300 Hottenga, J.-J., Prokopenko, I., Waite, L.L., Harris, T.B., Smith, A.V.,  
1301 Shuldiner, A.R., McArdle, W.L., Caulfield, M.J., Munroe, P.B., Grönberg,  
1302 H., Chen, Y.-D.I., Li, G., Beckmann, J.S., Johnson, T., Thorsteinsdottir,  
1303 U., Teder-Laving, M., Khaw, K.-T., Wareham, N.J., Zhao, J.H., Amin,  
1304 N., Oostra, B.A., Kraja, A.T., Province, M.A., Cupples, L.A., Heard-  
1305 Costa, N.L., Kaprio, J., Ripatti, S., Surakka, I., Collins, F.S., Saramies,  
1306 J., Tuomilehto, J., Jula, A., Salomaa, V., Erdmann, J., Hengstenberg,  
1307 C., Loley, C., Schunkert, H., Lamina, C., Wichmann, H.E., Albrecht, E.,  
1308 Gieger, C., Hicks, A.A., Johansson, A., Pramstaller, P.P., Kathiresan, S.,  
1309 Speliotes, E.K., Penninx, B., Hartikainen, A.-L., Jarvelin, M.-R., Gyl-  
1310 lensten, U., Boomsma, D.I., Campbell, H., Wilson, J.F., Chanoock, S.J.,  
1311 Farrall, M., Goel, A., Medina-Gomez, C., Rivadeneira, F., Estrada, K.,  
1312 Uitterlinden, A.G., Hofman, A., Zillikens, M.C., den Heijer, M., Kiemeny,  
1313 L.A., Maschio, A., Hall, P., Tyrer, J., Teumer, A., Völzke, H., Kovacs, P.,  
1314 Tönjes, A., Mangino, M., Spector, T.D., Hayward, C., Rudan, I., Hall,  
1315 A.S., Samani, N.J., Attwood, A.P., Sambrook, J.G., Hung, J., Palmer,  
1316 L.J., Lokki, M.-L., Sinisalo, J., Boucher, G., Huikuri, H., Lorentzon,  
1317 M., Ohlsson, C., Eklund, N., Eriksson, J.G., Barlassina, C., Rivolta, C.,  
1318 Nolte, I.M., Snieder, H., Van der Klauw, M.M., Van Vliet-Ostaptchouk,  
1319 J.V., Gejman, P.V., Shi, J., Jacobs, K.B., Wang, Z., Bakker, S.J.L.,  
1320 Mateo Leach, I., Navis, G., van der Harst, P., Martin, N.G., Medland,  
1321 S.E., Montgomery, G.W., Yang, J., Chasman, D.I., Ridker, P.M., Rose,  
1322 L.M., Lehtimäki, T., Raitakari, O., Absher, D., Iribarren, C., Basart,  
1323 H., Hovingh, K.G., Hyppönen, E., Power, C., Anderson, D., Beilby, J.P.,  
1324 Hui, J., Jolley, J., Sager, H., Bornstein, S.R., Schwarz, P.E.H., Kristians-  
1325 son, K., Perola, M., Lindström, J., Swift, A.J., Uusitupa, M., Atalay,  
1326 M., Lakka, T.A., Rauramaa, R., Bolton, J.L., Fowkes, G., Fraser, R.M.,  
1327 Price, J.F., Fischer, K., Krjutå Kov, K., Metspalu, A., Mihailov, E., Lan-  
1328 genberg, C., Luan, J., Ong, K.K., Chines, P.S., Keinanen-Kiukaanniemi,  
1329 S.M., Saaristo, T.E., Edkins, S., Franks, P.W., Hallmans, G., Shungin,  
1330 D., Morris, A.D., Palmer, C.N.A., Erbel, R., Moebus, S., Nöthen, M.M.,  
1331 Pechlivanis, S., Hveem, K., Narisu, N., Hamsten, A., Humphries, S.E.,  
1332 Strawbridge, R.J., Tremoli, E., Grallert, H., Thorand, B., Illig, T., Koenig,

- 1333 W., Müller-Nurasyid, M., Peters, A., Boehm, B.O., Kleber, M.E., März,  
1334 W., Winkelmann, B.R., Kuusisto, J., Laakso, M., Arveiler, D., Cesana, G.,  
1335 Kuulasmaa, K., Virtamo, J., Yarnell, J.W.G., Kuh, D., Wong, A., Lind,  
1336 L., de Faire, U., Gigante, B., Magnusson, P.K.E., Pedersen, N.L., Dedous-  
1337 sis, G., Dimitriou, M., Kolovou, G., Kanoni, S., Stirrups, K., Bonnycastle,  
1338 L.L., Njølstad, I., Wilsgaard, T., Ganna, A., Rehnberg, E., Hingorani,  
1339 A., Kivimaki, M., Kumari, M., Assimes, T.L., Barroso, I., Boehnke, M.,  
1340 Borecki, I.B., Deloukas, P., Fox, C.S., Frayling, T., Groop, L.C., Haritu-  
1341 nians, T., Hunter, D., Ingelsson, E., Kaplan, R., Mohlke, K.L., O'Connell,  
1342 J.R., Schlessinger, D., Strachan, D.P., Stefansson, K., van Duijn, C.M.,  
1343 Abecasis, G.R., McCarthy, M.I., Hirschhorn, J.N., Qi, L., Loos, R.J.F.,  
1344 Lindgren, C.M., North, K.E., Heid, I.M.: Sex-stratified genome-wide asso-  
1345 ciation studies including 270,000 individuals show sexual dimorphism in  
1346 genetic loci for anthropometric traits. *PLoS Genet.* **9**(6), 1003500 (2013)
- 1347 [69] Winkler, T.W., Kutalik, Z., Gorski, M., Lottaz, C., Kronenberg, F., Heid,  
1348 I.M.: EasyStrata: evaluation and visualization of stratified genome-wide  
1349 association meta-analysis data. *Bioinformatics* **31**(2), 259–261 (2015)
- 1350 [70] McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann,  
1351 A., Flicek, P., Cunningham, F.: The ensembl variant effect predictor.  
1352 *Genome Biol.* **17**(1), 122 (2016)
- 1353 [71] Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J.,  
1354 Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P.,  
1355 Gauthier, L.D., Brand, H., Solomonson, M., Watts, N.A., Rhodes, D.,  
1356 Singer-Berk, M., England, E.M., Seaby, E.G., Kosmicki, J.A., Walters,  
1357 R.K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A.,  
1358 Seed, C., Whiffin, N., Chong, J.X., Samocha, K.E., Pierce-Hoffman, E.,  
1359 Zappala, Z., O'Donnell-Luria, A.H., Minikel, E.V., Weisburd, B., Lek,  
1360 M., Ware, J.S., Vittal, C., Armean, I.M., Bergelson, L., Cibulskis, K.,  
1361 Connolly, K.M., Covarrubias, M., Donnelly, S., Ferriera, S., Gabriel, S.,  
1362 Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Llanwarne, C., Munshi,  
1363 R., Novod, S., Petrillo, N., Roazen, D., Ruano-Rubio, V., Saltzman, A.,  
1364 Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G., Talkowski,  
1365 M.E., Genome Aggregation Database Consortium, Neale, B.M., Daly,  
1366 M.J., MacArthur, D.G.: The mutational constraint spectrum quantified  
1367 from variation in 141,456 humans. *Nature* **581**(7809), 434–443 (2020)
- 1368 [72] Liu, X., Li, C., Mou, C., Dong, Y., Tu, Y.: dbNSFP v4: a comprehensive  
1369 database of transcript-specific functional predictions and annotations for  
1370 human nonsynonymous and splice-site SNVs. *Genome Med.* **12**(1), 103  
1371 (2020)
- 1372 [73] Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., Kircher, M.:  
1373 CADD: predicting the deleteriousness of variants throughout the human

- 1374 genome. *Nucleic Acids Res.* **47**(D1), 886–894 (2019)
- 1375 [74] Ioannidis, N.M., Rothstein, J.H., Pejaver, V., Middha, S., McDonnell,  
1376 S.K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D., Cannon-  
1377 Albright, L.A., Teerlink, C.C., Stanford, J.L., Isaacs, W.B., Xu, J.,  
1378 Cooney, K.A., Lange, E.M., Schleutker, J., Carpten, J.D., Powell, I.J.,  
1379 Cussenot, O., Cancel-Tassin, G., Giles, G.G., MacInnis, R.J., Maier, C.,  
1380 Hsieh, C.-L., Wiklund, F., Catalona, W.J., Foulkes, W.D., Mandal, D.,  
1381 Eeles, R.A., Kote-Jarai, Z., Bustamante, C.D., Schaid, D.J., Hastie, T.,  
1382 Ostrander, E.A., Bailey-Wilson, J.E., Radivojac, P., Thibodeau, S.N.,  
1383 Whittemore, A.S., Sieh, W.: REVEL: An ensemble method for predicting  
1384 the pathogenicity of rare missense variants. *Am. J. Hum. Genet.* **99**(4),  
1385 877–885 (2016)
- 1386 [75] Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Dar-  
1387 bandi, S.F., Knowles, D., Li, Y.I., Kosmicki, J.A., Arbelaez, J., Cui,  
1388 W., Schwartz, G.B., Chow, E.D., Kanterakis, E., Gao, H., Kia, A., Bat-  
1389 zoglou, S., Sanders, S.J., Farh, K.K.-H.: Predicting splicing from primary  
1390 sequence with deep learning. *Cell* **176**(3), 535–54824 (2019)
- 1391 [76] Morales, J., Pujar, S., Loveland, J.E., Astashyn, A., Bennett, R., Berry,  
1392 A., Cox, E., Davidson, C., Ermolaeva, O., Farrell, C.M., Fatima, R., Gil,  
1393 L., Goldfarb, T., Gonzalez, J.M., Haddad, D., Hardy, M., Hunt, T., Jack-  
1394 son, J., Joardar, V.S., Kay, M., Kodali, V.K., McGarvey, K.M., McMahon,  
1395 A., Mudge, J.M., Murphy, D.N., Murphy, M.R., Rajput, B., Rangwala,  
1396 S.H., Riddick, L.D., Thibaud-Nissen, F., Threadgold, G., Vatsan, A.R.,  
1397 Wallin, C., Webb, D., Flicek, P., Birney, E., Pruitt, K.D., Frankish, A.,  
1398 Cunningham, F., Murphy, T.D.: A joint NCBI and EMBL-EBI transcript  
1399 set for clinical genomics and research. *Nature* **604**(7905), 310–315 (2022)
- 1400 [77] Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J.,  
1401 Grody, W.W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm,  
1402 H.L., ACMG Laboratory Quality Assurance Committee: Standards and  
1403 guidelines for the interpretation of sequence variants: a joint consensus  
1404 recommendation of the american college of medical genetics and genomics  
1405 and the association for molecular pathology. *Genet. Med.* **17**(5), 405–424  
1406 (2015)
- 1407 [78] Zhou, W., Bi, W., Zhao, Z., Dey, K.K., Jagadeesh, K.A., Karczewski,  
1408 K.J., Daly, M.J., Neale, B.M., Lee, S.: SAIGE-GENE+ improves the effi-  
1409 ciency and accuracy of set-based rare variant association tests. *Nat. Genet.*  
1410 **54**(10), 1466–1469 (2022)
- 1411 [79] Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R.,  
1412 Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham,

- 1413 P.C.: PLINK: a tool set for whole-genome association and population-  
1414 based linkage analyses. *Am. J. Hum. Genet.* **81**(3), 559–575 (2007)
- 1415 [80] Pe'er, I., Yelensky, R., Altshuler, D., Daly, M.J.: Estimation of the mul-  
1416 tiple testing burden for genomewide association studies of nearly all  
1417 common variants. *Genet. Epidemiol.* **32**(4), 381–385 (2008)
- 1418 [81] Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh,  
1419 P.-R., ReproGen Consortium, Psychiatric Genomics Consortium, Genetic  
1420 Consortium for Anorexia Nervosa of the Wellcome Trust Case Control  
1421 Consortium 3, Duncan, L., Perry, J.R.B., Patterson, N., Robinson, E.B.,  
1422 Daly, M.J., Price, A.L., Neale, B.M.: An atlas of genetic correlations across  
1423 human diseases and traits. *Nat. Genet.* **47**(11), 1236–1241 (2015)
- 1424 [82] Hemani, G., Tilling, K., Davey Smith, G.: Orienting the causal relation-  
1425 ship between imprecisely measured traits using GWAS summary data.  
1426 *PLoS Genet.* **13**(11), 1007081 (2017)
- 1427 [83] Hemani, G., Zheng, J., Elsworth, B., Wade, K.H., Haberland, V., Baird,  
1428 D., Laurin, C., Burgess, S., Bowden, J., Langdon, R., Tan, V.Y.,  
1429 Yarmolinsky, J., Shihab, H.A., Timpson, N.J., Evans, D.M., Relton, C.,  
1430 Martin, R.M., Davey Smith, G., Gaunt, T.R., Haycock, P.C.: The MR-  
1431 Base platform supports systematic causal inference across the human  
1432 phenome. *Elife* **7** (2018)
- 1433 [84] Davies, N.M., Holmes, M.V., Davey Smith, G.: Reading mendelian ran-  
1434 domisation studies: a guide, glossary, and checklist for clinicians. *BMJ*  
1435 **362**, 601 (2018)
- 1436 [85] Burgess, S., Thompson, S.G., CRP CHD Genetics Collaboration: Avoid-  
1437 ing bias from weak instruments in mendelian randomization studies. *Int.*  
1438 *J. Epidemiol.* **40**(3), 755–764 (2011)
- 1439 [86] Bowden, J., Davey Smith, G., Burgess, S.: Mendelian randomization with  
1440 invalid instruments: effect estimation and bias detection through egger  
1441 regression. *Int. J. Epidemiol.* **44**(2), 512–525 (2015)
- 1442 [87] Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H.,  
1443 Zahler, A.M., Haussler, D.: The human genome browser at UCSC.  
1444 *Genome Res.* **12**(6), 996–1006 (2002)
- 1445 [88] Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani,  
1446 A.D., Wallace, C., Plagnol, V.: Bayesian test for colocalisation between  
1447 pairs of genetic association studies using summary statistics. *PLoS Genet.*  
1448 **10**(5), 1004383 (2014)
- 1449 [89] Wallace, C.: Eliciting priors and relaxing the single causal variant

- 1450 assumption in colocalisation analyses. *PLoS Genet.* **16**(4), 1008720 (2020)
- 1451 [90] Schmiedel, B.J., Rocha, J., Gonzalez-Colin, C., Bhattacharyya, S., Madri-  
1452 gal, A., Ottensmeier, C.H., Ay, F., Chandra, V., Vijayanand, P.: COVID-  
1453 19 genetic risk variants are associated with expression of multiple genes  
1454 in diverse immune cell types. *Nat. Commun.* **12**(1), 6760 (2021)
- 1455 [91] Skene, N.G., Bryois, J., Bakken, T.E., Breen, G., Crowley, J.J., Gaspar,  
1456 H.A., Giusti-Rodriguez, P., Hodge, R.D., Miller, J.A., Muñoz-Manchado,  
1457 A.B., O'Donovan, M.C., Owen, M.J., Pardiñas, A.F., Ryge, J., Walters,  
1458 J.T.R., Linnarsson, S., Lein, E.S., Major Depressive Disorder Work-  
1459 ing Group of the Psychiatric Genomics Consortium, Sullivan, P.F.,  
1460 Hjerling-Leffler, J.: Genetic identification of brain cell types underlying  
1461 schizophrenia. *Nat. Genet.* **50**(6), 825–833 (2018)
- 1462 [92] Mosconi, L., Berti, V., Dyke, J., Schelbaum, E., Jett, S., Loughlin,  
1463 L., Jang, G., Rahman, A., Hristov, H., Pahlajani, S., Andrews, R.,  
1464 Matthews, D., Etingin, O., Ganzer, C., de Leon, M., Isaacson, R., Brinton,  
1465 R.D.: Menopause impacts human brain structure, connectivity, energy  
1466 metabolism, and amyloid-beta deposition. *Sci. Rep.* **11**(1), 10867 (2021)