

1 Identification of rare loss of function variation regulating body fat distribution

2 **Authors:** Mine Koprulu^{a#}, Yajie Zhao^{a#}, Eleanor Wheeler^a, Liang Dong^b, Nuno Rocha^b, Satish Patel ^b,
3 Marcel Van de Streek ^c , Craig A. Glastonbury^d, Isobel D. Stewart ^a , Felix R. Day^a, Jian'an Luan^a,
4 Nicholas Bowker^a, Laura B. L. Wittemans^a, Nicola D. Kerrison^a, Debora M. E. Lucarelli^a, Inês Barroso^c,
5 Mark I. McCarthy ^f, Robert A. Scott ^a, Vladimir Saudek ^b, Kerrin S. Small ^c, Nicholas J. Wareham^a,
6 Robert K. Semple ^g, John R. B. Perry^a, Stephen O'Rahilly^b, Luca A. Lotta^a, Claudia Langenberg^{a,h*},
7 David B. Savage^{b*}

- 8 a. MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Box 285
9 Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge, United Kingdom,
10 CB2 0QQ
11 b. University of Cambridge Metabolic Research Laboratories, Wellcome Trust-MRC Institute of
12 Metabolic Science, Box 289, Cambridge Biomedical Campus, Cambridge, United Kingdom, CB2
13 0 QQ
14 c. Department of Twin Research and Genetic Epidemiology, King's College London, St Thomas'
15 Campus, Lambeth Palace Road, London, SE1 7EH, United Kingdom
16 d. BenevolentAI Limited, 4-8 Maple St, London W1T 5HD
17 e. Wellcome Sanger Institute, Hinxton, Cambridge, CB10 1SA, United Kingdom
18 f. Wellcome Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, OX3
19 7BN, United Kingdom
20 g. Centre for Cardiovascular Science, University of Edinburgh, 47 Little France Crescent,
21 Edinburgh, EH16 4TJ, UK
22 h. Computational Medicine, Berlin Institute of Health at Charité – Universitätsmedizin Berlin,
23 Germany

24
25 # authors contributed equally

26 * authors contributed equally

27

28 Correspondence to:

29 Claudia Langenberg (claudia.langenberg@mrc-epid.cam.ac.uk)

30 David B. Savage (dbs23@medschl.cam.ac.uk)

31

32 **ABSTRACT**

33 Biological and translational insights from large-scale, array-based genetic studies of fat distribution, a key
34 determinant of metabolic health, have been limited by the difficulty in linking identified predominantly
35 non-coding variants to specific gene targets. Rare coding variant analyses provide greater confidence that
36 a specific gene is involved, but do not necessarily indicate whether gain or loss of function would be of
37 most therapeutic benefit. Here we use a dual approach that combines the power of genome-wide analysis
38 of array-based rare, non-synonymous variants in 184,246 individuals of UK Biobank with exome-sequence-
39 based rare loss of function gene burden testing. The data indicates that loss-of-function (LoF) of four genes
40 (*PLIN1*, *INSR*, *ACVR1C* and *PDE3B*) is associated with a beneficial impact on WHR_{adjBMI} and increased
41 gluteofemoral fat mass, whereas *PLIN4* LoF adversely affects these parameters. This study robustly
42 implicates these genes in the regulation of fat distribution, providing new and in some cases somewhat
43 counter-intuitive insight into the potential consequences of targeting these molecules therapeutically.

44

45 Word count: 161

46 INTRODUCTION

47 Fat distribution is a heritable trait, commonly estimated by the relative amounts of waist and hip fat (waist-
48 to-hip ratio, WHR) for a given body size. Genetic mechanisms linked to either relatively *lower*
49 gluteofemoral or *higher* abdominal fat or both, have been shown to contribute to a greater WHR and its
50 consistently adverse cardiometabolic consequences (1). Genome-wide array-based association studies have
51 robustly identified many loci linked to WHR but thus far provided relatively limited biological and
52 translational insights due to poor coverage of rare protein-coding variants and uncertainties connecting
53 associated non-coding variants to functional genes (2, 3). Consequently, very few genes have been
54 definitively linked to WHR and it is generally unknown whether a gain or loss of gene function is likely to
55 drive observed associations.

56 The low frequency of rare (minor allele frequency [MAF] <0.5%, as defined by the 1000 Genomes Project
57 (4)) functional variants which may have sizable effects on the encoded protein, may be a consequence of
58 selective pressure acting against them, and previous studies have shown inverse relationships between
59 allele frequency and effect size for complex traits (5, 6). Rare variants that occurred recently are also likely
60 to be in low linkage disequilibrium (LD) with nearby common variants, facilitating fine-mapping and
61 identification of causal variants and genes (7). However, rare variants are difficult to impute (8) so their
62 study requires large, homogeneous samples and direct genotyping. To date, the vast majority of studies
63 have explored the contribution of common variants in relation to WHR including the largest meta-analysis
64 of imputed genome-wide association studies which included up to 694,649 individuals but only identified
65 two variants at MAF 0.1-0.5% (3). The only other study which investigated the role of rare variants for
66 WHR was a subsequent trans-ethnic Exomechip effort that identified 9 low frequency or rare variants with
67 a lowest MAF of 0.1% (9).

68 The contribution of the full spectrum of rare variants to WHR using sequence data has not been studied,
69 yet has the potential to provide a more direct link between gene and phenotype, and to facilitate translation
70 from gene identification to drug development. Whilst the identification of coding variants in a specific gene
71 clearly increases confidence in linking that particular gene to a trait, the impact of individual coding variants
72 can still be very, or at least relatively, subtle. Individual variant testing, even using exome sequencing data
73 in large populations, therefore still provides limited power and leaves residual uncertainty about the benefits
74 of gain or loss of function of a particular gene. Exome wide scans of gene-based burden of rare loss-of-
75 function variants have the potential to address this limitation (10-14). In this study, we use a dual approach
76 that combines the power of large-scale genome-wide analysis of array-based rare, non-synonymous variants
77 with exome-sequence-based rare gene burden testing to identify the putative function of variants, genes and

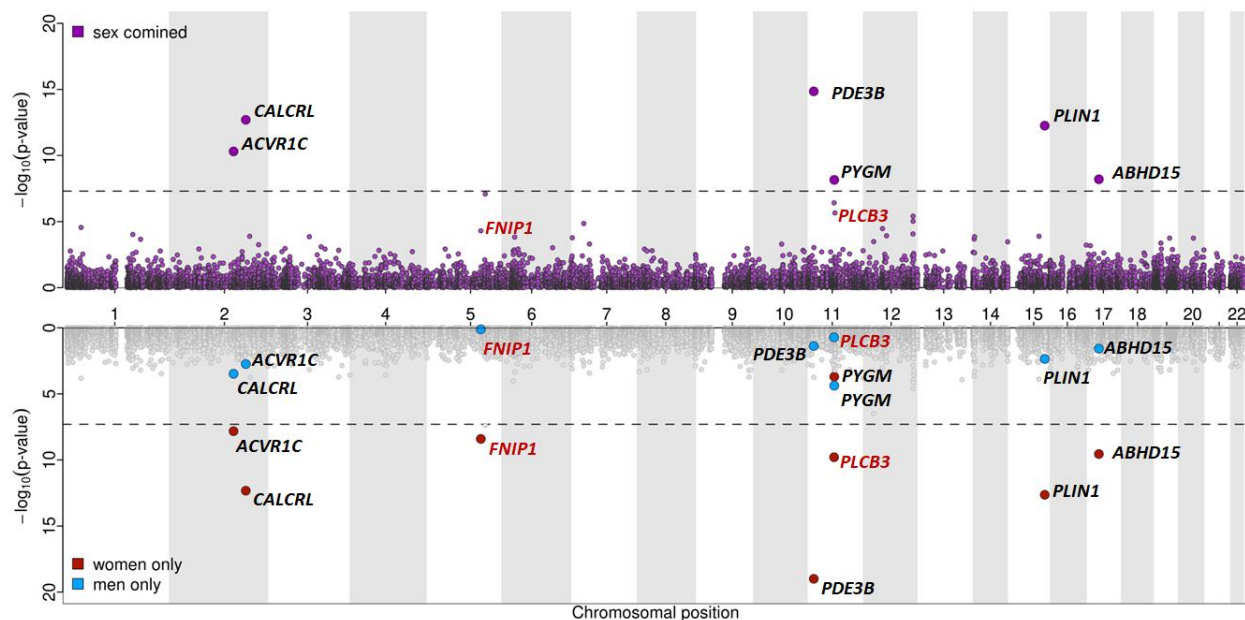
78 pathways regulating body shape and fat distribution (assessed by BMI-adjusted WHR [WHR_{adjBMI}]) and to
79 determine their effects on body composition and metabolic health.

80 RESULTS

81 A genome-wide analysis of directly genotyped, rare ($0.1\% \leq \text{MAF} \leq 0.5\%$) non-synonymous variants
82 associated with WHR_{adjBMI} at $p < 5 \times 10^{-8}$ in 450,562 European ancestry individuals from UK Biobank
83 identified lead variants in *PLIN1* p.L90P (rs139271800, EAF=0.1%), *PDE3B* p.R783X (rs150090666,
84 EAF=0.1%), *ACVR1C* p.I195T (rs56188432, EAF=0.2%), *CALCRL* p.L87P (rs61739909; EAF=0.3%),
85 *ABHD15* p.G147D (rs141385558; EAF=0.2%) and *PYGM* p.R50X (rs116987552, EAF=0.4%) (Figure 1,
86 Supplementary Table 1). We observed a correlation of 0.99 and minor allele concordance of 0.99 comparing
87 genotyped to whole-exome sequenced rare ($0.1\% \leq \text{MAF} \leq 0.5\%$) non-synonymous variants when testing the
88 validity of rare, genotyped variants using exome-sequencing data from the overlapping samples
89 (Supplementary Table 2).

90 *Sex-differences in genetic effects of rare variants on WHR_{adjBMI}*

91 Common variant analyses have provided evidence of differences in genetic associations with fat distribution
92 between men and women (3). In line with this, we found evidence of significant sex interactions, with
93 stronger genetic effects in women, compared to men for all lead variants, except for *ACVR1C* p.I195T and
94 *PYGM* p.R50X (Table 1). We therefore conducted sex-specific analyses which revealed two additional
95 variants, *PLCB3* p.V806I (rs145502455, EAF=0.4%) and *FNIP1* p.R518Q (rs115209326, EAF=0.3%) to
96 be genome-significant in ($p < 5 \times 10^{-8}$) in women, with no effect in men (Table 1, Supplementary Table 1).
97 No variants reached genome-wide significance in men only.



98

99 **Figure 1. Miami plot for sex-combined and sex-specific single marker association results for WHR_{adjBMI} .** *Top:*
 100 Manhattan plot representing results from the main, sex-combined GWAS for WHR_{adjBMI} for genotyped, rare non-
 101 synonymous variants ($0.1\% \leq MAF \leq 0.5\%$, correlation and rare allele concordance >0.9 when compared to the exome
 102 sequencing data). Gene annotations for the genome-wide significant variants from the main, sex-combined analyses
 103 are shown in black; gene annotations and significance from the main, sex-combined analyses for variants that were
 104 genome-wide significant in sex-specific analyses (women) only are shown in red. *Bottom:* Sex-specific significance
 105 of the variants highlighted above.

106

Variant	rsID	Sex interaction p value	p value Women	p value Men	Beta Women	Beta Men	SE Women	SE Men
<i>PLIN1 p.L90P</i>	rs139271800	4.07×10^{-3}	1.90×10^{-13}	4.50×10^{-3}	-0.273	-0.126	0.039	0.042
<i>PDE3B p.R783X</i>	rs150090666	5.66×10^{-5}	1.00×10^{-19}	4.10×10^{-2}	-0.392	-0.102	0.042	0.048
<i>ACVR1C p.I195T</i>	rs56188432	3.61×10^{-1}	1.10×10^{-8}	3.40×10^{-4}	-0.16	-0.109	0.028	0.032
<i>CALCRL p.L87P</i>	rs61739909	4.30×10^{-3}	4.30×10^{-13}	2.50×10^{-3}	-0.171	-0.082	0.024	0.026
<i>ABHD15 p.G147D</i>	rs141385558	4.53×10^{-3}	3.40×10^{-10}	3.80×10^{-2}	0.168	0.057	0.026	0.029
<i>PYGM p.R50X</i>	rs116987552	3.56×10^{-1}	2.90×10^{-4}	5.40×10^{-5}	0.079	0.095	0.021	0.024
<i>PLCB3 p.V806I</i>	rs145502455	2.34×10^{-2}	1.60×10^{-10}	2.00×10^{-1}	0.126	0.029	0.021	0.024
<i>FNIP1 p.R518Q</i>	rs115209326	6.09×10^{-4}	4.80×10^{-9}	8.40×10^{-1}	-0.128	-0.003	0.023	0.026

107 **Table 1: Sex-stratified results for variants identified in joint and sex specific analyses of genotyped rare variants**
 108 **in UK Biobank.** Unshaded rows present results discovered from the sex-combined analysis and grey shaded rows
 109 represent results from variants identified only in the sex-stratified (women) analysis. Abbreviations: p value Women,

110 BOLT LMM p-value in women; p value Men, BOLT LMM p-value in men; Beta Women, effect size in women; Beta
111 Men, effect size in men; SE Women, standard error in women; SE Men, standard error in men.

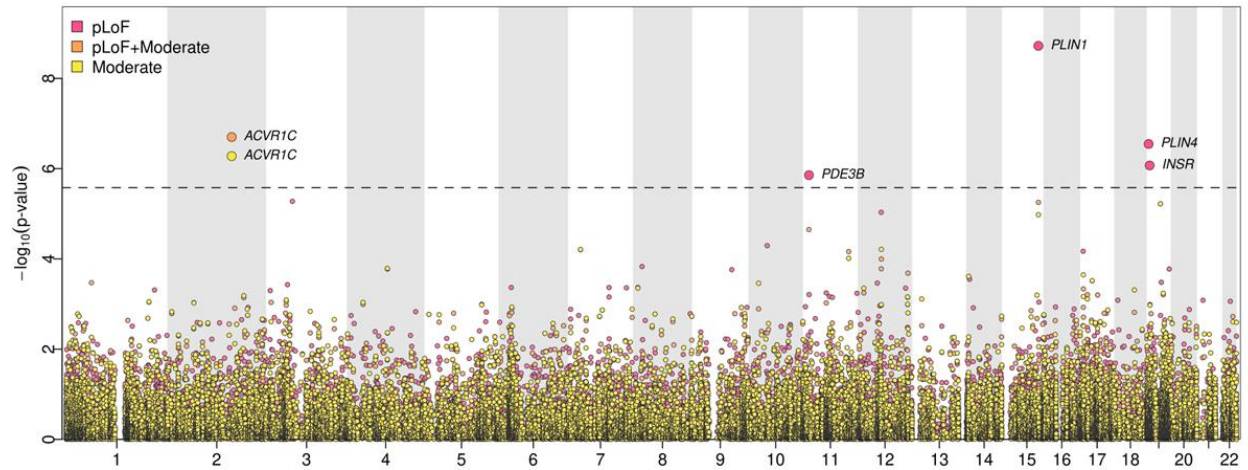
112 ***Genomic context and fine-mapping analyses***

113 We found strong statistical evidence for causal associations of rare non-synonymous variants in *PLIN1*,
114 *PDE3B*, *ACVR1C* and *CALCRL* (Supplementary Note 1, Supplementary Table 3, Supplementary Figure 1)
115 through conditional analysis and fine-mapping, whereas genomic context analyses did not support the
116 causality of the identified rare lead variants in *ABHD15* or *PYGM* from the joint (sex-combined) analysis,
117 and of *PLCB3* and *FNIP1* in the women-only analysis (Supplementary Note 2, Supplementary Table 4).
118 Bioinformatic analysis of these variants strongly predicted that the *PDE3B* variant p.R783X would truncate
119 *PDE3B* within the catalytic site, impairing *PDE3B* catalytic activity if expressed (Supplementary Note 3),
120 whereas predictions of the functional impact of the *PLIN1*, *ACVR1C* and *CALCRL* variants were less
121 conclusive (Supplementary Note 3).

122 ***Exome-sequenced based burden testing of rare, loss-of function variants***

123 Next, we considered the genes identified in the single variant analysis for exome-sequence-based gene rare
124 LoF and missense burden testing in 184,246 individuals in UK Biobank (see Methods, Gene-based
125 association testing) and found that *PLIN1*, *PDE3B*, *ACVR1C*, and *CALCRL* were all significantly associated
126 with lower WHR_{adjBMI} at a Bonferroni corrected threshold ($p < 0.0125$). Predicted loss of function (pLoF)
127 variants showed the most significant association for *PLIN1* and *PDE3B*, moderate impact variants for
128 *CALCRL*, and the combination of pLoF with moderate impact variants for *ACVR1C* (Supplementary Table
129 5).

130 In order to identify additional genes where loss of function may regulate fat distribution, we extended this
131 approach to a hypothesis free, exome-wide analysis ($p < 2.53 \times 10^{-6}$) for WHR_{adjBMI} using more stringent
132 quality control (QC) parameters (see Methods). This identified *PLIN4* and *INSR* in at least one of the variant
133 categories (see Methods), in addition to *PLIN1*, *ACVR1C* and *PDE3B* (Figure 2, Supplementary Table 6).
134 *PLIN4*, *INSR* and *PDE3B* all showed significantly larger standardized effect sizes for women compared to
135 men ($p < 0.05$) in gene-based analyses, in line with the single marker results (Supplementary Table 7).



136

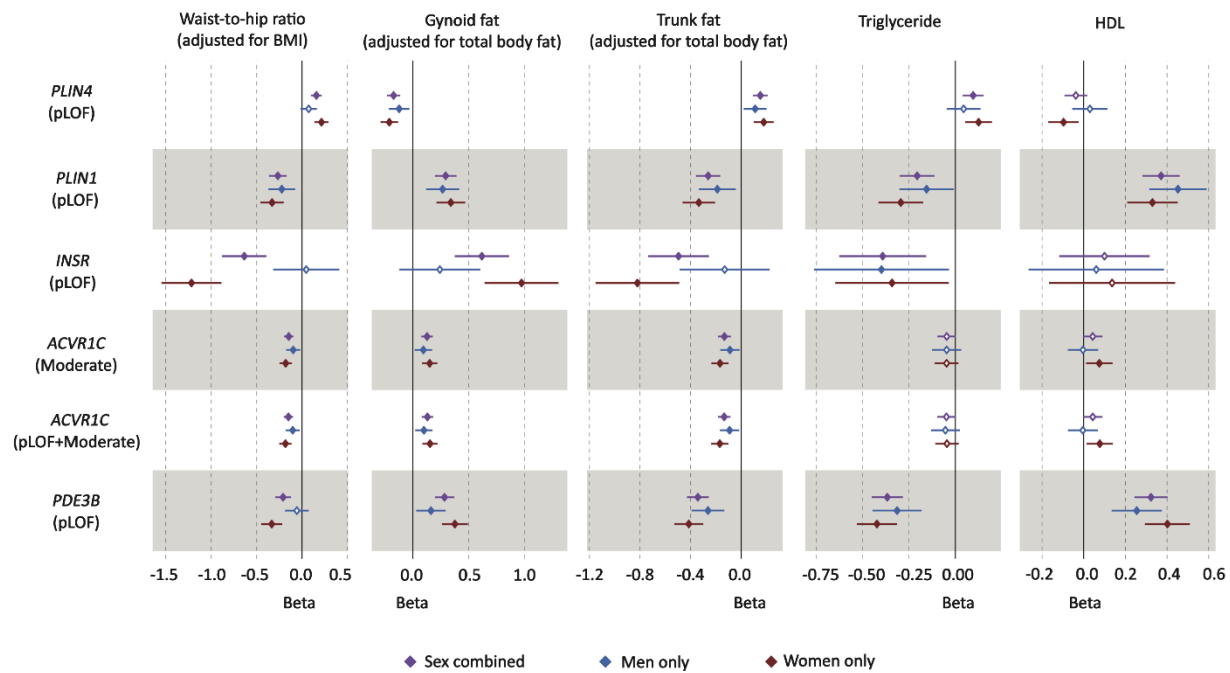
137 **Figure 2. Gene-based association results.** Gene-based exome-wide discovery results for WHR_{adjBMI} . The horizontal
138 dashed line represents the exome-wide significance threshold ($p=2.53 \times 10^{-6}$).

139

140 While the joint effect of rare LoF variants in *PLIN4* (65 variants, 1,065 carriers) was associated with a
141 higher WHR_{adjBMI} (Beta = 0.16 [0.10 – 0.22], $p=5.86 \times 10^{-7}$), the combination of rare LoF variants in *PLIN1*
142 (31 variants, 393 carriers) was associated with a lower WHR_{adjBMI} , (Beta = -0.27 [-0.17 – -0.36], $p=9.82 \times 10^{-9}$) (Supplementary Table 6). The lead *PLIN1* LoF variant (*PLIN1* p.T338DfsX51, rs750619494) is predicted
143 to result in a frameshift from amino acid 338 with a premature stop at amino acid 388, though it may well
144 be subject to nonsense mediated RNA decay. Several additional *PLIN1* variants are similarly expected to
145 result in early truncations or nonsense mediated RNA decay. Several additional *PLIN1* variants are similarly expected to
146 result in early truncations or nonsense mediated RNA decay (Supplementary Table 8). In either instance,
147 these variants are expected to impair Plin1 interaction with *ABHD5* and thus its regulation of adipose
148 triglyceride lipase (*ATGL*) (15). In the case of *PLIN4* (p.Q372X, rs201581703), the variant list also included
149 early frameshift/premature stop variants predicted to result in nonsense mediated RNA decay.

150 We next assessed phenotypic associations with refined measures of fat distribution and cardiometabolic
151 parameters and diseases. Bioelectrical Impedance Analysis (BIA) derived body fat compartment
152 measurements (16) showed that *PLIN4* (pLoF) was associated with higher android and trunk fat, and lower
153 gynoid and leg fat (Figure 3, Supplementary Figure 2, Supplementary Table 9) whereas *PLIN1* (pLoF)
154 acted in the opposite direction. Fat distribution is strongly linked to insulin resistance, but as direct
155 indicators of insulin resistance are not currently available in UK Biobank, we evaluated the impact of these
156 genes on metabolic indicators typically associated with insulin resistance (17, 18) (Figure 3, Supplementary
157 Figure 2, Supplementary Table 9). *PLIN4* LoF was associated with higher triglycerides (TGs), TG/HDL
158 (triglyceride/high-density lipoprotein cholesterol) ratio and higher HbA1c levels. The associations for
159 *PLIN1* consistently contrasted with those of *PLIN4* with lower TGs, TG/HDL ratio and additionally higher

160 HDL cholesterol levels, in keeping with a beneficial impact on insulin sensitivity. In keeping with these
 161 findings, *PLIN4* LoF was nominally associated with an increased risk for type 2 diabetes (T2D) (OR=1.36
 162 [1.06-1.66], $p=0.04$) in the Type 2 Diabetes Knowledge Portal (T2DKP; <https://t2d.hugeamp.org/>), though
 163 none of the genes showed a significant association with T2D in UK Biobank through our analysis or in the
 164 AstraZeneca PheWAS Portal (<https://azphewas.com/>) (Supplementary Figure 2; Supplementary Table 9;
 165 Supplementary Table 10). *PLINI* LoF showed nominal significance for a lower risk of cardiovascular heart
 166 disease (CHD) ($p=0.03$, OR=0.55[0.31-0.91]) in our analysis of UK Biobank, a finding supported by
 167 association between *PLINI* LoF and reduced susceptibility to chronic ischemic heart disease (OR=0.40
 168 [0.32-0.75], $p=4.49 \times 10^{-4}$) in the AstraZeneca PheWAS Portal (14).

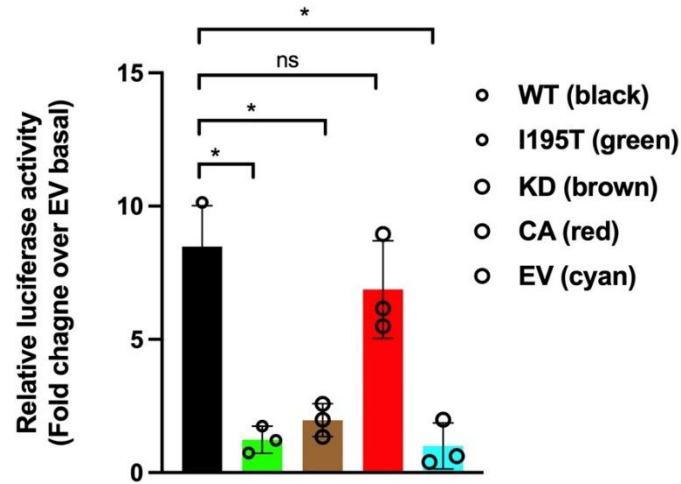


169
 170 **Figure 3. Forest plot of phenotypic associations for significant variant-gene categories.** Black represents the sex-
 171 combined, blue represents the men-only and red represents the women-only analysis. Horizontal lines represent 95%
 172 confidence intervals. Waist-to-hip ratio adjusted for BMI (n=184,246), gynoid fat adjusted for total body fat (n=
 173 178,143), trunk fat adjusted for total body fat (178,143), triglyceride levels (n=175,271) and HDL cholesterol
 174 (n=161,239) were all driven from UK Biobank (See Supplementary Table 12 for details).

175 Similarly to *PLINI*, the combined effect of LoF variants in the *INSR* (27 variants, 61 carriers) was
 176 associated with lower WHR_{adjBMI} (Beta=-0.64[-0.39 – -0.88], $p=6.21 \times 10^{-7}$; Supplementary Table 6).
 177 Although a few common intronic variants (rs1035942, rs1035940, rs62124511, rs34194998) and a low-

178 frequency synonymous variant (rs1799815) in the *INSR* have previously been associated with WHR_{adjBMI}
179 (3, 19), the causal mechanism underlying these associations remains unknown. Our gene-based findings
180 indicate that the *INSR* can alter body fat distribution through loss of function. Given the fact that the *INSR*
181 gene encodes the insulin receptor itself and that both biallelic and heterozygous loss of function variants in
182 this gene have long been linked with monogenic severe insulin resistance syndromes (20), this evidence for
183 loss of function of the *INSR* having a seemingly beneficial effect on fat distribution i.e. lower WHR_{adjBMI} is
184 surprising. Importantly, none of the *INSR* mutations previously linked to monogenic disease were present
185 in our UK Biobank analysis. The lead *INSR* variant (p.R525X) is predicted to result in truncation of the
186 protein within the extracellular domain preventing interaction of the extra- and intra-cellular domains, and
187 thus formation of a functional receptor. In the homozygous state, this would be expected to lead to
188 monogenic severe insulin resistance. In terms of body fat distribution, heterozygous *INSR* LoF was also
189 associated with higher gynoid and leg fat, and lower android and trunk fat mass (Figure 3, Supplementary
190 Figure 2, Supplementary Table 9). Similarly to the cardio-metabolic associations of *PLIN1* indicating a
191 beneficial effect, heterozygous loss of *INSR* was associated with lower TGs and a lower TG/HDL ratio
192 (Figure 3, Supplementary Figure 2, Supplementary Table 9). It was also associated with lower LDL (low-
193 density lipoprotein) cholesterol levels but was not associated with altered HDL (Figure 3, Supplementary
194 Figure 2, Supplementary Table 9). Despite these seemingly favourable changes in fat distribution and
195 plasma lipids, *INSR* LoF showed a nominal association for increased susceptibility to T2D in the T2DKP
196 (OR=3.67 [2.50-4.83], $p=0.02$) (Supplementary Table 10).

197 For *ACVR1C*, the genetic architecture of gene-based results was slightly different, gene-based association
198 results were significant for (i) the combined burden of pLoF and moderate impact variants and (ii) for
199 moderate impact variants only. There were 130 rare moderate impact variants and 9 rare high impact
200 variants included in this analysis [1414 and 16 carriers, respectively]. The combined effect of pLoF and
201 moderate impact variants and moderate impact only variants were both associated with lower WHR_{adjBMI}
202 (Beta=-0.15 [-0.10 - -0.20] and -0.15[-0.10 - -0.20], $p=1.68\times 10^{-7}$ and 4.57×10^{-7} , respectively;
203 Supplementary Table 6). In this instance, the highest-ranking variant was the previously reported p.I195T
204 variant (19, 21). In silico predictions including M-CAP (22), REVEL (23), SIFT (24), PolyPhen-2 (25) and
205 PROVEAN (26) all assess this variant to be damaging to the protein and structural modelling also suggests
206 that it is likely to have a sizable impact (Supplementary Note 3). CADD (27) also estimates this variant to
207 be among the top 1% of the deleterious variants ranked by CADD (score = 27.1). To test this prediction,
208 we performed a luciferase reporter assay in HEK293 cells which strongly suggested that the mutation
209 impairs *ACVR1C* signalling (Figure 4).



210
211
212
213
214
215
216
217
218
219
220

Figure 4. Functional impact of ACVR1C I195T variant on Smad signalling. HEK293 cells were transiently transfected with ACVR1C expression constructs and their receptor components, along with firefly and *Renilla* luciferase expression plasmids. Firefly luciferase activity was normalised to *Renilla* activity and the luciferase activity in non-stimulated cells transfected with empty vector (EV) was set to 1. A constitutively active (CA) ACVR1C variant T194D and a kinase dead (KD) variant K222R were included for comparison. Results from three independent experiments are presented as mean ± SD. Statistical significance was evaluated by one-way ANOVA with Tukey's post hoc test for multiple comparisons between pairs. WT, wild type ACVR1C. * $P < 0.001$, ns not significant.

221 The phenotypic associations for ACVR1C LoF for fat categories were similar to *PLIN1* LoF, with significant
222 associations with higher gynoid and leg fat, and lower android and trunk fat. However, the cardio-metabolic
223 associations were less clear for ACVR1C with an association with lower TG and but not HDL or the
224 TG/HDL ratio (Figure 3, Supplementary Figure 2; Supplementary Table 9).

225
226 Finally, the combined effect of LoF variants in *PDE3B* was also significantly associated with lower
227 WHR_{adjBMI} . However, leave-one-out analysis suggested that this association was mainly driven by the
228 premature stop variant (p.R783X, rs150090666; p -value after dropping the variant=0.49; Supplementary
229 Table 11). All other candidate genes remained at least nominally significant after dropping the most
230 significant variant (Supplementary Table 11). Our analysis of the *PDE3B* p.R783X variant was in line with
231 previous reports associating it with lower triglyceride levels and higher HDL (28, 29). *PDE3B* p.R783X
232 has also been reported to be associated with higher apolipoprotein B, lower apolipoprotein A1 levels and
233 other haematological traits (30). This variant was reported to be significantly associated with cardiovascular
234 disease when meta-analysed in UK Biobank and other cohorts (29, 31).

235
236
237

238 DISCUSSION

239 Central adiposity has long been linked to insulin resistance and metabolic disease (32-37) but exactly why
240 this is the case and, other than sex hormones, what exactly determines fat distribution remains incompletely
241 understood. So, what have we learnt from human genetics thus far? Firstly, that inheritance contributes to
242 WHR (3, 38). Secondly, monogenic partial lipodystrophies indicate that single gene variants can be
243 sufficient to mediate substantial changes in fat distribution, classic examples being mutations in *LMNA* and
244 *PPARG*. Interestingly, both proteins are expressed in all white adipocytes and yet specific loss of function
245 variants are consistently associated with loss of hip and leg fat whereas visceral fat is preserved (39).
246 Thirdly, whilst the beneficial impact of thiazolidinediones, one of very few drugs that clearly improve
247 insulin sensitivity, was recognised before the discovery of *PPARG* mutations in patients with partial
248 lipodystrophy, this link attests to the potential for human genetics to inform drug discovery. Fourthly,
249 GWAS studies have identified many loci associated with WHR (2, 3, 9) though these have yet to be
250 translated into therapeutic targets. Finally, Mendelian Randomization has been used to establish that genetic
251 mechanisms linked to greater WHR_{adjBMI} can be causally linked to the risk of cardiovascular disease and
252 type 2 diabetes through either relatively *lower* gluteofemoral or *higher* abdominal fat or both (40).
253 Furthermore, these associations are very likely to be underpinned by insulin resistance as the genetic risk
254 score for WHR_{adjBMI} was also shown to be strongly associated with elevated fasting insulin, higher
255 triglycerides and lower HDL cholesterol (40).

256 Our WHR_{adjBMI} single variant analysis in samples from 450,562 UK Biobank participants revealed missense
257 variants in 3 genes (*CALCRL*, *PLINI* and *ACVR1C*) and a nonsense variant in *PDE3B*. All these genes are
258 highly expressed in adipose tissue in keeping with emerging evidence that adiposity itself is largely
259 centrally mediated whereas where excess energy is stored is regulated within adipose tissue itself (2, 41).
260 *CALCRL* is also expressed in a host of other tissues and its role in adipose tissue remains to be established
261 (42, 43). *PLINI* is a lipid droplet surface protein almost exclusively expressed in adipocytes and has a well-
262 established role in regulating both triglyceride and diacylglycerol hydrolysis (44). *PDE3B* is expressed in
263 many tissues but has long been linked to adipocyte lipolysis, and specifically to insulin mediated inhibition
264 of lipolysis (45, 46). Several lines of evidence have recently implicated *ACVR1C* in the regulation of
265 lipolysis, but it is expressed in many tissues in addition to adipose tissue and further work is required to
266 convincingly establish exactly what it does in adipocytes (47-49).

267 In terms of the impact of the specific mutations present in each gene, the *PDE3B* p.R783X is clearly
268 expected to impair *PDE3B* catalytic activity, and thus potentially to increase cAMP levels and lipolysis,
269 but the impact of the other three variants is far less certain (see Supplementary Note 3). Interestingly, gene
270 based LoF burden analyses for all four genes were at least nominally significant, suggesting that the single

271 variants were most likely to impair function of the encoded proteins. At least when transiently transfected
272 into cultured cells, our functional data is also consistent with the LoF predictions for the *ACVR1C* p.I195T
273 variant.

274 Our exome wide analyses confirmed significant effects of LoF variants in *PLIN1*, *ACVR1C* and *PDE3B*. In
275 the cases of *ACVR1C* and *PDE3B*, the lead LoF variants were the same as single variants reported above,
276 namely the *ACVR1C* p.I195T and *PDE3B* p.R783X variants. The fact that the phenotypic associations of
277 the *PLIN1* p.L90P variant are directionally consistent with the *PLIN1* LoF gene burden data suggests that
278 this variant is likely to be a loss of function variant too. Both the lead *PLIN1* variant and several additional
279 *PLIN1* variants are expected to result in early truncations or nonsense mediated RNA decay. These data are
280 consistent with Laver *et al.*'s assertion that *PLIN1* haploinsufficiency is not associated with lipodystrophy
281 (50). Interestingly, several heterozygous *PLIN1* frameshift variants had previously been linked to partial
282 lipodystrophy (51). None of these variants overlap with those identified in UK Biobank to date and none
283 are predicted to result in nonsense mediated RNA decay. Instead, in several cases, immunoblotting of
284 adipose tissue lysates confirmed expression of an elongated form of Perilipin 1 in addition to the wildtype
285 copy, so perhaps expression of these mutant forms with an altered carboxy-terminus, accounts for the
286 seemingly 'opposite' phenotypes (51, 52).

287 Gene burden testing also highlighted the role of *PLIN4* LoF variants in fat distribution, though in this case,
288 these had an adverse impact on WHR_{adjBMI} . Whilst a higher WHR_{adjBMI} would conventionally be deemed to
289 be metabolically adverse, it is possible that this need not be the case for all genetic perturbations, however,
290 our phenotypic analyses were consistent with the predicted outcomes for all the above genes, in fact, the
291 phenotypic associations for *PLIN1* and *PLIN4* were consistently opposite. Similar to *PLIN1*, *PLIN4* is
292 highly expressed in adipose tissue, but it is also expressed in heart and skeletal muscle, and the *PLIN4*
293 knockout mouse has not been reported to have an adipose tissue phenotype to date (53).

294 The last gene identified in the exome wide gene burden analysis was the *INSR*. In this instance, the lead
295 variant (p.R525X) is expected to truncate the protein in the alpha subunit shortly before the disulphide bond
296 normally connecting alpha and beta subunits. This is expected to abrogate synthesis of functional receptor.
297 Even if truncated protein were synthesised this would not be able to dimerise with or exert dominant
298 negative activity over co-expressed wild type receptor, and so heterozygosity for the truncating variant
299 would be expected to reduce functional receptor protein by ~50%. In keeping with this, biallelic mutations
300 in this domain usually cause extreme IR classified as Donohue or Rabson-Mendenhall syndrome. Parents
301 of affected children have not been systematically studied and are generally held to be metabolically normal.
302 In contrast, heterozygous *INSR* variants in the intracellular beta subunit, which are synthesised and interfere

303 with wild type receptor function, cause type A insulin resistance (20). Whilst fat mass is often reduced in
304 Donohue's syndrome, heterozygous variants associated with type A insulin resistance are not reported to
305 be associated with fat redistribution and interestingly do not typically lead to fatty liver or dyslipidaemia
306 (20). Our data suggest, surprisingly, that the *INSR* LoF variants favourably impact WHR_{adjBMI} and LDL
307 cholesterol. Whilst the *INSR* LoF association with T2D was relatively weak statistically and not seen in all
308 the cohorts assessed, it is conceivable that *INSR* LoF might adversely affect pancreatic beta cell function
309 and/or insulin sensitivity despite the apparently beneficial impact on fat distribution. The change in
310 triglycerides is somewhat reminiscent of the well described absence of dyslipidaemia in patients with
311 monogenic severe insulin resistance due to bi- or mono-allelic *INSR* mutations, so again this does not
312 preclude the *INSR* LOF variants being associated with reduced insulin sensitivity.

313 Our sex-specific analyses consistently revealed stronger effects in women than in men. These data are
314 consistent with the fact that WHR is more strongly associated with insulin resistance in women than in men
315 (54). Fat mass in women is consistently significantly higher than men of a similar BMI, who typically have
316 higher lean/muscle mass. The adverse impact of a lack of lower limb/gluteofemoral fat on metabolism is
317 strikingly apparent in patients with familial partial lipodystrophy, particularly types 2 and 3, due to specific
318 mutations in *LMNA* and *PPARG* respectively (39). In these and in fact in all forms of partial lipodystrophy,
319 metabolic disease manifests considerably earlier and is typically more severe in women than in men (39,
320 55, 56).

321 Our analyses have several limitations which future work should help to resolve. Firstly, the statistical power
322 to detect associations, particularly when examining rare variants, depends on the sample size. Hence, there
323 is the opportunity to discover additional findings when the WES data is released in the full UK Biobank
324 cohort or other large-scale studies. Secondly, the phenotypic follow up of cardiometabolic diseases for the
325 candidate genes was primarily conducted in UK Biobank, a population cohort with a limited number of
326 cases of specific diseases. Our follow-up in T2DKP revealed the potential of datasets enriched for cases to
327 increase statistical power in phenotypic follow ups. Thirdly, fat distribution is strongly associated with
328 insulin resistance, but the UK Biobank cohort did not provide fasting samples so direct measures of insulin
329 and inferred indices of insulin sensitivity are not available.

330 In conclusion, our analysis strongly implicates at least four genes in the regulation of fat distribution.
331 Furthermore, the data suggests that inhibitors of *PLIN1*, *PDE3B* and *ACVR1C* might favourably impact fat
332 distribution and associated metabolic phenotypes whereas *PLIN4* inhibition is likely to have adverse health
333 consequences. The data in *PLIN1* needs to be tempered by the earlier reports linking some specific *PLIN1*
334 LoF variants with partial lipodystrophy. Finally, the data on the *INSR* seemed to suggest a potential
335 disconnect between an apparently favourable impact of LoF variants on WHR and an apparently adverse

336 impact on T2D risk. These findings provide valuable insight into the potential of these genes as therapeutic
337 targets.

338

339 **METHODS**

340

341 **The UK Biobank Resource**

342 The UK Biobank is a large-scale prospective population-based study of approximately 500,000 participants
343 aged 40 – 69 at the time of recruitment (57). Recruitment took place between 2006 – 2010 in centres across
344 the United Kingdom and participants have deep phenotypic information collected from initial and repeat
345 assessment visits, health records, self-reported survey information, linkage to death and cancer registries,
346 urine and blood biomarkers and other phenotypic endpoints. A Seca 200 cm tape was used to measure
347 waist and hip circumference at the baseline visit, and BMI was calculated from height and weight
348 measurements. WHR_{adjBMI} was constructed as the ratio of waist and hip circumferences adjusted for age,
349 age^2 and BMI (measured at the baseline assessment visit). Residuals were calculated for men and women
350 separately and then transformed using the rank-based inverse normal transformation. All additional
351 phenotypes are described in Supplementary Table 12.

352

353 **Genome-wide association scan of genotyped rare nonsynonymous genetic variants**

354 Genetic variants were genotyped in UK Biobank using the Affymetrix UK BiLEVE or the Affymetrix UK
355 Biobank Axiom arrays (57). Genotyping underwent quality control procedures including (a) routine quality
356 checks carried out during the process of sample retrieval, DNA extraction, and genotype calling; (b) checks
357 and filters for genotype batch effects, plate effects, departures from Hardy Weinberg equilibrium, sex
358 effects, array effects, and discordance across control replicates; and (c) individual and genetic variant call
359 rate filters as previously described (57). We further excluded genetic variants with a genotype call rate
360 below 95% and variants that were (i) not rare ($0.1\% \leq MAF \leq 0.5\%$) or (ii) not non-synonymous or (iii) had
361 poor correlation ($r < 0.9$) or rare allele concordance (< 0.9) when compared to the whole exome sequence
362 data (Supplementary Table 2). Genomic annotations were performed using the ANNOVAR software (58).
363 The coordinates of genotyped rare variants were lifted over from GRCh37 to GRCh38 using liftOver and
364 all reported positions in this study are in GRCh38. A total of 13,181 genetic variants in 7,481 genes were
365 available for analysis. Genome-wide association analyses were performed using the BOLT-LMM software
366 (59) in 450,562 participants of European Ancestry defined using a K-means clustering approach applied to
367 the first four principal components calculated from genome-wide SNP genotypes.

368 Sex-specific genome-wide association analyses were performed using the BOLT-LMM software (59) in
369 206,082 men and 244,478 women of European ancestry from UK Biobank. Evidence for sex differences at
370 the variants identified in the sex combined analysis were formally tested in unrelated individuals using a

371 linear regression model including an interaction term between the genetic variant and sex using the same
372 covariates used in the discovery analysis.

373

374 **Conditional analyses and fine-mapping**

375 At each associated genomic region, we conducted systematic analyses of the genomic context of
376 associations. Our goal was to establish whether or not the identified rare nonsynonymous variants are likely
377 to be the causal variants for the association with WHR_{adjBMI} . At each region 1 Mb either side of the
378 nonsynonymous genetic variants associated with WHR_{adjBMI} , we conducted both approximate and formal
379 conditional analyses. We considered the association of all genetic variants in the regions regardless of
380 functional annotation or allele frequency using directly-genotyped and imputed data (imputed using the
381 Haplotype Reference Consortium (HRC) and UK10K haplotype resource). First, approximate conditional
382 analyses were conducted on summary-level estimates using GCTA (60, 61) to identify sets of conditionally-
383 independent index genetic variants ($p < 5 \times 10^{-8}$). Individual-level genotypes for the conditionally-
384 independent variants identified in this first step were then extracted in 350,721 unrelated European ancestry
385 participants of UK Biobank and their independent association was confirmed in multivariable linear
386 regression models including all variants put forward from approximate analyses. Then, at each region, we
387 statistically decomposed the identified index signals by conditioning on the other conditionally-independent
388 index variants. We then performed Bayesian fine-mapping (62) to estimate the posterior probability of
389 association for each variant (PPA, where 0% indicates that the variant is not causal and 100% indicates the
390 highest possible posterior probability that the variant is causal) and define the 99% credible set at that signal
391 (i.e. a set of variants in a genomic window that accounts for 99% of the PPA at that association signal). To
392 generate credible sets, the association results at each locus were converted to Bayes factors (BF) for each
393 variant within the locus boundary. The posterior probability that a variant- j was causal was defined by:

$$394 \quad \Phi_j = \frac{BF_j}{\sum_k BF_k}$$

395

396 where, BF_j denotes the BF for the j^{th} variant, and the denominator is the sum of BF_s for all included variants
397 at that signal. A 99% credible set of variants was created by ranking the posterior probabilities from highest
398 to lowest and summing them until the cumulative posterior probability exceeded 0.99 (i.e. 99%).

399

400 **UK Biobank exome-sequence data processing and QC**

401 Whole exome sequencing (WES) data of 200,643 UK Biobank participants made available in October 2020
402 were downloaded in VCF and PLINK formats. The details of the UK Biobank WES data processing are
403 provided in detail elsewhere (63, 64). Further data processing and quality control has been described
404 previously (65). In brief, we did not apply additional QC based on QUAL (variant site-level quality score,
405 Phred scale) or AQ measures (variant site-level allele quality score reflecting evidence for each alternate
406 allele, Phred scale). Site-level filtering was applied for targeted biallelic calls if the AB ratio (no. of alternate
407 allele reads/total depth) was ≤ 0.25 or ≥ 0.8 . Variant-level QC filters were applied if any of the variants had
408 (i) genotype missingness $> 5\%$, (ii) maximum read depth (DP) of less than 10 across samples or (iii) had
409 GQ less than 20 for over 20% of the calls. After applying these filters, 7.3% of the variants were flagged as
410 poor quality and not taken forward for further analysis.

411

412 **Variant annotation and definition of gene burden sets**

413 We annotated variants released in UK Biobank 200K whole exome sequencing VCF files in build hg38,
414 using the Variant Effect Predictor (VEP) tool release 99 provided by Ensembl (66). In addition to default
415 VEP features such as the consequence and impact of the variant, overlapping gene, position at cDNA and
416 protein level and codon and amino acid change, if applicable, we have used the following plugins for
417 annotation: (i) SIFT (67), which predicts whether an amino acid substitution affects protein function based
418 on sequence homology and the physical properties of amino acid, (ii) Polyphen-2 (25), which predicts
419 possible impact of an amino acid substitution on the structure and function of a human protein, (iii) CADD
420 (27) which provides deleteriousness prediction scores for all variants based on diverse genomic features
421 and (iv) LOFTEE (68) which provides loss of function prediction for variants. We annotated each variant
422 using the most severe consequence across overlapping transcripts in Ensembl. We defined loss of function
423 variants as those with ‘high’ impact predict by VEP. This includes frameshift variants, transcript ablating
424 or transcript amplifying variants, splice acceptor or splice donor variants, stop lost, start gained or stop
425 gained variants. ‘Moderate impact’ variants include missense variants, in-frame deletion or insertions,
426 missense variants and protein altering variants.

427

428

429 **Gene-based association testing**

430 In our discovery stage, we used the method STAAR (variant-Set Test for Association using Annotation
431 infoRmation), which is a computationally scalable method for very large whole-exome sequence (WES)
432 and whole-exome sequence (WGS) studies and large-scale biobanks. STAAR uses a Generalized Linear
433 Mixed Model (GLMM) framework that includes linear and logistic mixed models and can also account for
434 both relatedness and population structure for both quantitative and dichotomous traits (69). In our analysis,
435 we used the genotype dosage matrix as the genotype input and covariates including age at first check (age),
436 age², sex, genotyping array, top ten genetically derived principal components (PC1-PC10) generated from
437 the SNP array data, exome sequencing batch and the sparse GRM. For rank-based inverse normal
438 transformed WHR_{adjBMI} , we also added BMI as a covariate. We excluded the samples from our analysis if
439 they did not pass UK Biobank quality control parameters, were non-European ancestry or if they withdrew
440 consent from the study (n=184,246) (65).

441 We ran STAAR with its default options without additional functional annotations. For each gene, with at
442 least two variants with $MAF \leq 0.5\%$, we conducted gene-based association analysis for the following three
443 variant categories: rare variants predicted by VEP to be a) loss of function (pLoF; i.e. high impact), b)
444 missense (Moderate; i.e. moderate impact) or c) both (pLoF+Moderate). For each variant clustering of a
445 gene, STAAR will provide p-values for several collapsing burden tests including SKAT (sequence kernel
446 association test), Burden test, and ACAT-V (set-based aggregated Cauchy association test). In addition, the
447 output of STAAR also includes the omnibus p-value (STAAR-O) by using the combined Cauchy
448 association test to aggregate the association across the different tests.

449 After identifying the genes with STAAR-O p-value over the threshold for exome-wide significance
450 ($p < 2.53 \times 10^{-6}$), we applied more stringent QC filters on the genotype calls of the included variants. We set
451 to missing genotype calls which did not meet the following QC criteria: 1. Genotype Quality (GQ) ≥ 20 for
452 heterozygous variants; 2. Depth (DP) ≥ 7 for SNVs and $DP \geq 10$ for InDels; 3. A binomial test on allelic
453 balance using the Allelic Depth (AD) FORMAT field for heterozygous variants with $p \geq 1 \times 10^{-3}$. We then
454 repeated the STAAR analysis using the filtered genotype dosage matrix.

455 To examine the extent to which the gene-based association is driven by single variants, we conducted a
456 sensitivity leave-one-out analysis for each significant gene ($p < 2.53 \times 10^{-6}$), testing the significance of the
457 gene-based association after excluding each variant.

458

459 **Secondary association testing**

460 We created dichotomous dummy variables using the filtered genotype dosage matrix for each identified
461 gene, where samples with one or more rare alleles were set as “1” and the samples without rare alleles were

462 set as “0” for different variant clustering settings of each gene. Then we combined these dummy variables
463 into a single file and transformed it to BGEN format, which was used as the genotype input for association
464 testing using a linear mixed model implemented in BOLT-LMM to account for cryptic population structure
465 and relatedness (59). The GRM in BOLT-LMM was generated from the autosomal genetic variants that
466 were common (MAF > 1%), passed quality control in all 106 batches, and were present on both genotyping
467 arrays (65). Covariates included age, sex, and PC1-PC10, genotyping chip and exome sequencing batch.
468 For rank-based inverse normal transformed WHR_{adjBMI} , covariates also included BMI. We excluded the
469 same group of samples as we did for STAAR analyses.

470 To test for heterogeneity of effect sizes between men and women for significant genes identified in the
471 gene-based analyses, we used a Z-test to compare effect size estimates for each gene calculated in the sex-
472 specific analyses.

473

474 **Phenotypic associations**

475 The gene-based phenotypic associations using the same STAAR and BOLT-LMM pipelines for the
476 following continuous phenotypes: BMI, BIA-derived gynoid fat, BIA-derived leg fat, BIA-derived android
477 fat, BIA-derived trunk fat, BIA-derived arm fat, triglyceride levels, HDL cholesterol, LDL cholesterol,
478 HbA1c levels (see Supplementary Table 12 for phenotype details). Body fat compartments were predicted
479 using bioimpedance measurements in UK Biobank. The details for the prediction of body fat compartments
480 in UK Biobank are described elsewhere (16).

481 We have also investigated gene-based phenotypic associations for binary disease outcomes: type 2 diabetes
482 and cardiovascular heart disease. As BOLT-LMM is based on the linear mixed model which cannot give
483 an accurate effect estimate for binary variables, we have also applied a generalized linear model (GLM) to
484 estimate the Odd Ratio (OR) for binary phenotypes. We also looked up these binary outcomes in other
485 resources such as the AstraZeneca PheWAS Portal (<https://azphewas.com/>, accessed on 02/09/2021) (14)
486 and the Type 2 Diabetes Knowledge Portal (T2DKP; <https://t2d.hugeamp.org/>, accessed on 02/09/2021).
487 The AstraZeneca PheWAS Portal also uses UK Biobank as their primary resource, but have access to a
488 larger dataset of 281,104 exomes. We looked up results for T2D (N cases = 1,671; N controls = 160,949)
489 and chronic ischaemic heart disease (defined by ICD-10 code I25; N cases =24,147, N controls = 176,170).
490 In the T2DKP, we also looked up results for T2D (N=43,125).

491

492

493 **ACVR1C dual-luciferase assay**

494 HEK293 cells were seeded at a density of 150,000 cells per well in 24-well tissue culture plates pre-treated
495 with poly-D-lysine. On the following day, medium was replaced with Opti-MEM I Reduced Serum medium
496 and a total of 550 ng of plasmid DNA; this included different pcDNA3.1 based ACVR1C constructs listed
497 in the table below, as well as constructs encoding receptor components (ACVR-IIB and CRIPTO) along
498 with firefly (consisting of the SMAD binding elements) and *Renilla* (control) luciferase reporter plasmids.
499 Lipofectamine 3000 Reagent was used for the transfection according to the manufacturer's protocol. Opti-
500 MEM I Reduced Serum medium was then replaced with DMEM growth medium 6 hours post transfection.

501

pcDNA3.1 construct	Receptor components
WT	ACVR-IIB, CRIPTO, pGL4.48[<i>luc2P</i> /SBE/Hygro], pRL-SV40
I195T	ACVR-IIB, CRIPTO, pGL4.48[<i>luc2P</i> /SBE/Hygro], pRL-SV40
K222R	ACVR-IIB, CRIPTO, pGL4.48[<i>luc2P</i> /SBE/Hygro], pRL-SV40
T194D	ACVR-IIB, CRIPTO, pGL4.48[<i>luc2P</i> /SBE/Hygro], pRL-SV40
EV (empty vector)	ACVR-IIB, CRIPTO, pGL4.48[<i>luc2P</i> /SBE/Hygro], pRL-SV40

502

503 The dual-luciferase reporter assay was performed according to the manufacturer's protocol (Promega,
504 USA). Cells were washed once with DPBS, followed by an active lysis procedure. Briefly, 125 μ l of passive
505 lysis buffer was added in each well and the cells were subjected to one cycle of a freeze-thaw process. Cell
506 lysates were cleared of cell debris by centrifugation at 21,130 x g for one minute. The assay was conducted
507 in a 96-well plate format. In each assay, 20 μ l of cleared supernatant was pre-dispensed, followed by
508 sequential measurement of firefly and *Renilla* luciferase using a Tecan Spark 10M plate reader (Tecan,
509 Switzerland). Firefly luciferase activity was normalised for *Renilla* luciferase activity, and then further
510 normalised with values from non-stimulated cells transfected with empty pcDNA3.1 vector (EV). We also
511 studied a constitutively active (ACVR1C p.T194D) mutant and a kinase dead (ACVR1C p.K222R) mutant
512 for comparison (70, 71). The experiment was repeated with fresh transfections on three separate occasions.

513

514

515 **REFERENCES**

- 516 1. Lotta LA, Wittemans LBL, Zuber V, Stewart ID, Sharp SJ, Luan J, et al. Association of Genetic
517 Variants Related to Gluteofemoral vs Abdominal Fat Distribution With Type 2 Diabetes, Coronary Disease,
518 and Cardiovascular Risk Factors. *JAMA*. 2018;320(24):2553-63.
- 519 2. Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Mägi R, et al. New genetic
520 loci link adipose and insulin biology to body fat distribution. *Nature*. 2015;518(7538):187-96.
- 521 3. Pulit SL, Stoneman C, Morris AP, Wood AR, Glastonbury CA, Tyrrell J, et al. Meta-analysis of
522 genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry.
523 *Hum Mol Genet*. 2019;28(1):166-74.
- 524 4. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated
525 map of genetic variation from 1,092 human genomes. *Nature*. 2012;491(7422):56-65.
- 526 5. Claussnitzer M, Cho JH, Collins R, Cox NJ, Dermitzakis ET, Hurles ME, et al. A brief history of
527 human disease genetics. *Nature*. 2020;577(7789):179-89.
- 528 6. Wang Q, Dhindsa RS, Carss K, Harper A, Nag A, Tachmazidou I, et al. Surveying the contribution
529 of rare variants to the genetic architecture of human disease through exome sequencing of 177,882 UK
530 Biobank participants. *bioRxiv*. 2020:2020.12.13.422582.
- 531 7. Raychaudhuri S. Mapping rare and common causal alleles for complex human diseases. *Cell*.
532 2011;147(1):57-69.
- 533 8. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel
534 of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48(10):1279-83.
- 535 9. Justice AE, Karaderi T, Highland HM, Young KL, Graff M, Lu Y, et al. Protein-coding variants
536 implicate novel genes related to lipid homeostasis contributing to body-fat distribution. *Nat Genet*.
537 2019;51(3):452-69.
- 538 10. Akbari P, Gilani A, Sosina O, Kosmicki JA, Khrimian L, Fang YY, et al. Sequencing of 640,000
539 exomes identifies *GPR75* variants associated with protection from obesity. *Science*. 2021;373(6550).
- 540 11. Dewey FE, Murray MF, Overton JD, Habegger L, Leader JB, Fetterolf SN, et al. Distribution and
541 clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study.
542 *Science*. 2016;354(6319).
- 543 12. Van Hout CV, Tachmazidou I, Backman JD, Hoffman JD, Liu D, Pandey AK, et al. Exome
544 sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature*. 2020;586(7831):749-56.
- 545 13. Park J, Lucas AM, Zhang X, Chaudhary K, Cho JH, Nadkarni G, et al. Exome-wide evaluation of
546 rare coding variants using electronic health records identifies new gene-phenotype associations. *Nat Med*.
547 2021;27(1):66-72.

- 548 14. Wang Q, Dhindsa RS, Carss K, Harper AR, Nag A, Tachmazidou I, et al. Rare variant contribution
549 to human disease in 281,104 UK Biobank exomes. *Nature*. 2021.
- 550 15. Granneman JG, Moore HP, Krishnamoorthy R, Rathod M. Perilipin controls lipolysis by regulating
551 the interactions of AB-hydrolase containing 5 (Abhd5) and adipose triglyceride lipase (Atgl). *J Biol Chem*.
552 2009;284(50):34538-44.
- 553 16. Powell R, De Lucia Rolfe E, Day FR, Perry JRB, Griffin SJ, Forouhi NG, et al. Development and
554 validation of total and regional body composition prediction equations from anthropometry and single
555 frequency segmental bioelectrical impedance with DEXA. *medRxiv*. 2020:2020.12.16.20248330.
- 556 17. McLaughlin T, Abbasi F, Cheal K, Chu J, Lamendola C, Reaven G. Use of metabolic markers to
557 identify overweight individuals who are insulin resistant. *Ann Intern Med*. 2003;139(10):802-9.
- 558 18. Freeman AM, Pennings N. *Insulin Resistance*. StatPearls. Treasure Island (FL): StatPearls
559 Publishing
- 560 Copyright © 2021, StatPearls Publishing LLC.; 2021.
- 561 19. Zhu Z, Guo Y, Shi H, Liu CL, Panganiban RA, Chung W, et al. Shared genetic and experimental
562 links between obesity-related traits and asthma subtypes in UK Biobank. *J Allergy Clin Immunol*.
563 2020;145(2):537-49.
- 564 20. Semple RK, Savage DB, Cochran EK, Gorden P, O'Rahilly S. Genetic syndromes of severe insulin
565 resistance. *Endocr Rev*. 2011;32(4):498-514.
- 566 21. Emdin CA, Khera AV, Aragam K, Haas M, Chaffin M, Klarin D, et al. DNA Sequence Variation
567 in *ACVR1C* Encoding the Activin Receptor-Like Kinase 7 Influences Body Fat Distribution and Protects
568 Against Type 2 Diabetes. *Diabetes*. 2019;68(1):226-34.
- 569 22. Jagadeesh KA, Wenger AM, Berger MJ, Guturu H, Stenson PD, Cooper DN, et al. M-CAP
570 eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat Genet*.
571 2016;48(12):1581-6.
- 572 23. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: An
573 Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet*.
574 2016;99(4):877-85.
- 575 24. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids*
576 *Res*. 2003;31(13):3812-4.
- 577 25. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations
578 using PolyPhen-2. *Curr Protoc Hum Genet*. 2013;Chapter 7:Unit7.20.
- 579 26. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid
580 substitutions and indels. *Bioinformatics*. 2015;31(16):2745-7.

- 581 27. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness
582 of variants throughout the human genome. *Nucleic Acids Res.* 2019;47(D1):D886-D94.
- 583 28. Sinnott-Armstrong N, Tanigawa Y, Amar D, Mars N, Benner C, Aguirre M, et al. Genetics of 35
584 blood and urine biomarkers in the UK Biobank. *Nat Genet.* 2021;53(2):185-94.
- 585 29. Klarin D, Damrauer SM, Cho K, Sun YV, Teslovich TM, Honerlaw J, et al. Genetics of blood
586 lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. *Nat Genet.*
587 2018;50(11):1514-23.
- 588 30. Vuckovic D, Bao EL, Akbari P, Lareau CA, Mousas A, Jiang T, et al. The Polygenic and
589 Monogenic Basis of Blood Traits and Diseases. *Cell.* 2020;182(5):1214-31.e11.
- 590 31. Emdin CA, Khera AV, Chaffin M, Klarin D, Natarajan P, Aragam K, et al. Analysis of predicted
591 loss-of-function variants in UK Biobank identifies variants protective for disease. *Nat Commun.*
592 2018;9(1):1613.
- 593 32. Pischon T, Boeing H, Hoffmann K, Bergmann M, Schulze MB, Overvad K, et al. General and
594 abdominal adiposity and risk of death in Europe. *N Engl J Med.* 2008;359(20):2105-20.
- 595 33. Wang Y, Rimm EB, Stampfer MJ, Willett WC, Hu FB. Comparison of abdominal adiposity and
596 overall obesity in predicting risk of type 2 diabetes among men. *Am J Clin Nutr.* 2005;81(3):555-63.
- 597 34. Canoy D. Distribution of body fat and risk of coronary heart disease in men and women. *Curr Opin*
598 *Cardiol.* 2008;23(6):591-8.
- 599 35. Mason C, Craig CL, Katzmarzyk PT. Influence of central and extremity circumferences on all-
600 cause mortality in men and women. *Obesity (Silver Spring).* 2008;16(12):2690-5.
- 601 36. Bogardus C, Lillioja S, Mott DM, Hollenbeck C, Reaven G. Relationship between degree of obesity
602 and in vivo insulin action in man. *Am J Physiol.* 1985;248(3 Pt 1):E286-91.
- 603 37. Hocking S, Samocha-Bonet D, Milner KL, Greenfield JR, Chisholm DJ. Adiposity and insulin
604 resistance in humans: the role of the different tissue and cellular lipid depots. *Endocr Rev.* 2013;34(4):463-
605 500.
- 606 38. Rose KM, Newman B, Mayer-Davis EJ, Selby JV. Genetic and behavioral determinants of waist-
607 hip ratio and waist circumference in women twins. *Obes Res.* 1998;6(6):383-92.
- 608 39. Lim K, Haider A, Adams C, Sleight A, Savage DB. Lipodistrophy: a paradigm for understanding
609 the consequences of "overloading" adipose tissue. *Physiol Rev.* 2021;101(3):907-93.
- 610 40. Emdin CA, Khera AV, Natarajan P, Klarin D, Zekavat SM, Hsiao AJ, et al. Genetic Association of
611 Waist-to-Hip Ratio With Cardiometabolic Traits, Type 2 Diabetes, and Coronary Heart Disease. *JAMA.*
612 2017;317(6):626-34.
- 613 41. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass
614 index yield new insights for obesity biology. *Nature.* 2015;518(7538):197-206.

- 615 42. Fischer JP, Els-Heindl S, Beck-Sickinger AG. Adrenomedullin - Current perspective on a peptide
616 hormone with significant therapeutic potential. *Peptides*. 2020;131:170347.
- 617 43. Erratum to: "Altered Expression of Adrenomedullin 2 and its Receptor in the Adipose Tissue of
618 Obese Patients". *J Clin Endocrinol Metab*. 2020;105(12).
- 619 44. Sztalryd C, Brasaemle DL. The perilipin family of lipid droplet proteins: Gatekeepers of
620 intracellular lipolysis. *Biochim Biophys Acta Mol Cell Biol Lipids*. 2017;1862(10 Pt B):1221-32.
- 621 45. Degerman E, Ahmad F, Chung YW, Guirguis E, Omar B, Stenson L, et al. From PDE3B to the
622 regulation of energy homeostasis. *Curr Opin Pharmacol*. 2011;11(6):676-82.
- 623 46. DiPilato LM, Ahmad F, Harms M, Seale P, Manganiello V, Birnbaum MJ. The Role of PDE3B
624 Phosphorylation in the Inhibition of Lipolysis by Insulin. *Mol Cell Biol*. 2015;35(16):2752-60.
- 625 47. Yogosawa S, Mizutani S, Ogawa Y, Izumi T. Activin receptor-like kinase 7 suppresses lipolysis to
626 accumulate fat in obesity through downregulation of peroxisome proliferator-activated receptor γ and
627 C/EBP α . *Diabetes*. 2013;62(1):115-23.
- 628 48. Ibáñez CF. Regulation of metabolic homeostasis by the TGF- β superfamily receptor ALK7. *FEBS*
629 *J*. 2021.
- 630 49. Guo T, Marmol P, Moliner A, Björnholm M, Zhang C, Shokat KM, et al. Adipocyte ALK7 links
631 nutrient overload to catecholamine resistance in obesity. *Elife*. 2014;3:e03245.
- 632 50. Laver TW, Patel KA, Colclough K, Curran J, Dale J, Davis N, et al. PLIN1 Haploinsufficiency Is
633 Not Associated With Lipodystrophy. *J Clin Endocrinol Metab*. 2018;103(9):3225-30.
- 634 51. Jéru I, Vantyghem MC, Bismuth E, Cervera P, Barraud S, Auclair M, et al. Diagnostic Challenge
635 in PLIN1-Associated Familial Partial Lipodystrophy. *J Clin Endocrinol Metab*. 2019;104(12):6025-32.
- 636 52. Gandotra S, Le Dour C, Bottomley W, Cervera P, Giral P, Reznik Y, et al. Perilipin deficiency and
637 autosomal dominant partial lipodystrophy. *N Engl J Med*. 2011;364(8):740-8.
- 638 53. Chen W, Chang B, Wu X, Li L, Sleeman M, Chan L. Inactivation of Plin4 downregulates Plin5
639 and reduces cardiac lipid accumulation in mice. *Am J Physiol Endocrinol Metab*. 2013;304(7):E770-9.
- 640 54. Lagou V, Mägi R, Hottenga JJ, Grallert H, Perry JRB, Bouatia-Naji N, et al. Sex-dimorphic genetic
641 effects and novel loci for fasting glucose and insulin variability. *Nat Commun*. 2021;12(1):24.
- 642 55. Garg A. Gender differences in the prevalence of metabolic complications in familial partial
643 lipodystrophy (Dunnigan variety). *J Clin Endocrinol Metab*. 2000;85(5):1776-82.
- 644 56. Mann JP, Savage DB. What lipodystrophies teach us about the metabolic syndrome. *J Clin Invest*.
645 2019;129(10):4009-21.
- 646 57. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource
647 with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-9.

- 648 58. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-
649 throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164.
- 650 59. Loh PR, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, et al. Efficient
651 Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet.* 2015;47(3):284-
652 90.
- 653 60. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis.
654 *Am J Hum Genet.* 2011;88(1):76-82.
- 655 61. Yang J, Ferreira T, Morris AP, Medland SE, Madden PA, Heath AC, et al. Conditional and joint
656 multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex
657 traits. *Nat Genet.* 2012;44(4):369-75, S1-3.
- 658 62. Maller JB, McVean G, Byrnes J, Vukcevic D, Palin K, Su Z, et al. Bayesian refinement of
659 association signals for 14 loci in 3 common diseases. *Nat Genet.* 2012;44(12):1294-301.
- 660 63. Szustakowski JD, Balasubramanian S, Kvikstad E, Khalid S, Bronson PG, Sasson A, et al.
661 Advancing human genetics research and drug discovery through exome sequencing of the UK Biobank.
662 *Nat Genet.* 2021;53(7):942-8.
- 663 64. Yun T, Li H, Chang PC, Lin MF, Carroll A, McLean CY. Accurate, scalable cohort variant calls
664 using DeepVariant and GLnexus. *Bioinformatics.* 2021.
- 665 65. Zhao Y, Stankovic S, Koprulu M, Wheeler E, Day FR, Lango Allen H, et al. GIGYF1 loss of
666 function is associated with clonal mosaicism and adverse metabolic health. *Nat Commun.* 2021;12(1):4178.
- 667 66. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect
668 Predictor. *Genome Biol.* 2016;17(1):122.
- 669 67. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein
670 function using the SIFT algorithm. *Nat Protoc.* 2009;4(7):1073-81.
- 671 68. Karczewski K. Loftee (Loss-of-Function Transcript Effect Estimator) 2015.
- 672 69. Li X, Li Z, Zhou H, Gaynor SM, Liu Y, Chen H, et al. Dynamic incorporation of multiple in silico
673 functional annotations empowers rare variant association analysis of large whole-genome sequencing
674 studies at scale. *Nat Genet.* 2020;52(9):969-83.
- 675 70. Bondestam J, Huotari MA, Morén A, Ustinov J, Kaivo-Oja N, Kallio J, et al. cDNA cloning,
676 expression studies and chromosome mapping of human type I serine/threonine kinase receptor ALK7
677 (ACVR1C). *Cytogenet Cell Genet.* 2001;95(3-4):157-62.
- 678 71. Wieser R, Wrana JL, Massagué J. GS domain mutations that constitutively activate T beta R-I, the
679 downstream signaling component in the TGF-beta receptor complex. *EMBO J.* 1995;14(10):2199-208.

680

681

682 **Acknowledgements**

683
684 This research has been conducted using the UK Biobank resource. Access to the UK Biobank genotype and
685 phenotype data is open to all approved health researchers (<http://www.ukbiobank.ac.uk/>). This study was
686 funded by the United Kingdom's Medical Research Council through grants MC_UU_12015/1,
687 MC_PC_13046, MC_PC_13048 and MR/L00002/1. This work was supported by the MRC Metabolic
688 Diseases Unit (MC_UU_12012/5) and the Cambridge NIHR Biomedical Research Centre and EU/EFPIA
689 Innovative Medicines Initiative Joint Undertaking (EMIF grant: 115372). R.K.S, D.B.S. and S.O'R. are
690 supported by the Wellcome Trust (WT 210752, WT 219417 and WT 214274 respectively) the MRC
691 Metabolic Disease Unit, the National Institute for Health Research (NIHR) Cambridge Biomedical
692 Research Centre and the NIHR Rare Disease Translational Research Collaboration. K.S.S. is supported by
693 MRC Project Grant L01999X/1. Some computation was enabled through access granted to K.S.S. to the
694 MRC eMedLab Medical Bioinformatics infrastructure, supported by the Medical Research Council (grant
695 number MR/L016311/1). M.McC. is a Wellcome Senior Investigator supported by Wellcome grants
696 098381, 090532, 106130, 203141. M.McC. declares that the views expressed in this article are those of the
697 authors and not necessarily those of the NHS, the NIHR, or the Department of Health; he has served on
698 advisory panels for Pfizer, Novo Nordisk and Zoe Global, has received honoraria from Merck, Pfizer, Novo
699 Nordisk and Eli Lilly, and research funding from Abbvie, Astra Zeneca, Boehringer Ingelheim, Eli Lilly,
700 Janssen, Merck, Novo Nordisk, Pfizer, Roche, Sanofi Aventis, Servier, and Takeda. M.K. is supported by
701 the Gates Cambridge Trust. The authors gratefully acknowledge the help of the MRC Epidemiology Unit
702 Support Teams, including Field, Laboratory and Data Management Teams.

703

704 **Competing interests**

705 M.McM. is an employee of Genentech, and a holder of Roche stock. D.M.E.L. is currently an employee of
706 Enhanc3D Genomics Ltd. N.B. is an employee of GlaxoSmithKline Plc. (GSK). R.A.S. is an employee
707 and shareholder of GlaxoSmithKline Plc. (GSK). C.A.G. is an employee of Benevolent AI. L.A.L. is an
708 employee of Regeneron Genetics Center and receives salary, stocks and stock options from Regeneron
709 Pharmaceuticals Inc.

710

711 **Author contributions**

712 Data collection and analysis: M.K., Y.Z., E.W., L.D., N.R., S.P., M.V.S., C.G., I.D.S, F.R.D, J.L., N.B.,
713 L.B.L.W., N.D.K., V.S.,

714 Study supervision: D.M.E.L., I.B., M.I.McM., R.A.S., K.S.S., N.J.W., R.K.S., L.A.L, J.R.B.P., S.O'R.,
715 C.L., D.B.S.

716

717 **Data availability**

718 This research has been conducted using the UK Biobank resource (application no. 44448 and 9905).

719 Access to the UK Biobank genotype and phenotype data is open to all approved health researchers

720 (<http://www.ukbiobank.ac.uk/>).